Clustering Educational Digital Library Usage Data: A Comparison of Latent Class Analysis and K-Means Algorithms

BEIJIE XU,
Department of Instructional Technology & Learning Sciences
Utah State University
beijie.xu@aggiemail.usu.edu

MIMI RECKER,
Department of Instructional Technology & Learning Sciences
Utah State University
mimi.recker@usu.edu

XIAOJUN QI,
Department of Computer Science
Utah State University
xiaojun.qi@usu.edu

NICHOLAS FLANN,
Utah State University
Department of Computer Science
nick.flann@usu.edu

LEI YE,
Department of Instructional Technology & Learning Sciences
Utah State University
lei.ye@aggiemail.usu.edu

This article examines clustering as an educational data mining method. In particular, two clustering algorithms, the widely used K-means and the model-based Latent Class Analysis, are compared, using usage data from an educational digital library service, the Instructional Architect (IA.usu.edu). Using a multi-faceted approach and multiple data sources, three types of comparisons of resulting clusters are presented: 1) Davies-Bouldin indices, 2) clustering results validated with user profile data, and 3) cluster evolution. Latent Class Analysis is superior to K-means on all three comparisons. In particular, LCA is more immune to the variance of feature variables, and clustering results turn out well with minimal data transformation. Our research results also show that LCA perform better than K-means in terms of providing the most useful educational interpretation for this dataset.

Keywords: Educational Data Mining, Educational Web Mining, Clustering, Latent Class Analysis, K-means, Digital Libraries, Teacher Users.

---

1. INTRODUCTION

A growing interest in applying Data Mining (DM) to evaluate web-based educational systems makes Educational Data Mining (EDM) a rising and promising research field (Romero and Ventura, 2007). With access to massive user logs and DM strategies, researchers can analyze fine-grained usage data to better understand a web-based educational system's users and their behaviors. One particular DM approach, clustering, can be used to group similar users, a set of pages with similar contents, or similar navigation patterns (Antonenko et al., 2012). Despite its increased popularity in the field of EDM, work remains in demonstrating its utility. For example, the basis for choosing a particular clustering algorithm should be justified. In addition, clustering results from the selected algorithm and a competing algorithm should be rigorously compared, evaluated, and validated.

This research employs user data from an educational digital library service, the Instructional Architect (IA.usu.edu), as a test bed for applying clustering approaches to help identify different user groups and, more importantly, to compare approaches. As will be described below, the IA supports teachers in authoring and sharing instructional activities using online learning resources (Recker et al., 2006; 2007). It has been in use for approximately eight years, yet little is known about its over 7,000 teacher users.

A particular focus of this work is comparing results from the more commonly used K-means clustering algorithm with Latent Class Analysis (LCA) (Collins and Lanza 2010; Magidson and Vermunt 2004). K-means is a process of partitioning n-dimensional data into *k* sets to minimize the mean distance within each set. The most commonly used distance measures are the squared Euclidean distance and the sum of the squared differences across variables.

In contrast, LCA (Magidson and Vermunt, 2004) is a model-based clustering analysis technique in that a statistical model (a mixture of probability distributions) is postulated for the population based on a set of sample data. The most common applications of LCA have been in health, clinical research (Campbell and Morgan-Lopez 2009; Pence et al. 2009), social, and psychological studies (Klonsky and Olino, 2008; Nylund, Bellmore, Nishina, and Graham, 2007). Despite its purported advantages, it has been applied less commonly in educational research (Roussos, 2007).

The purpose of this article is to first apply LCA to cluster an educational dataset, and then to compare its results with a benchmark and widely used clustering algorithm, K-means. Three types of criteria are used: 1) Davies-Bouldin indices, 2) clustering results validated with user profile data, and 3) cluster evolution. Results show that Latent Class Analysis is superior to K-means on all three comparisons for this educational dataset. In particular, LCA's performance is more stable and produces results supporting the most useful educational interpretation for this dataset.

The next sections review clustering studies in the context of educational research, provide an overview of the target user environment (the Instructional Architect) and previous research, and describe data sources and analyses. The results section presents results from three comparisons of clustering results obtained by applying LCA and K-means on three kinds of preprocessed data.

2. LITERATURE REVIEW

As an emerging discipline, Educational Data Mining (EDM) is concerned with applying DM methods for exploring datasets automatically collected in educational settings (Baker and Yacef, 2009) in order to address questions about a system's impact on users and usage patterns. The increasing availability of large educational datasets and the evolution of DM algorithms have made EDM a growing interdisciplinary area, lying between the fields of education and information/computer sciences. Many DM methods have been applied to investigate educational problems and phenomena. However, this study focuses exclusively on a particular type of DM method, clustering, to address how well clustering algorithms can be used to better understand an online educational system's usage patterns.

In general, clustering is an unsupervised method to group physical or abstract objects into classes, most often based on two measures: the similarity between the data objects within the same cluster (minimal intra-cluster distance), and the dissimilarity between the data objects of different clusters (maximal inter-cluster distance) (Romero and Ventura, 2007). As a common DM method, clustering has been applied to seek patterns in educational datasets.

However, we identify four common problems with clustering studies in EDM. First, a particular clustering algorithm can be applied to an educational dataset without the choice being justified either in theory or in practice. Second, a comparison between a selected algorithm and a competing algorithm is seldom presented. Third, as is a common problem for unsupervised machine learning algorithms, there is no standard method for comparing and evaluating the clustering results; therefore, researchers often present results without validating their findings. Lastly, a powerful educational system can store many aspects of user information. However, few studies employ users' performance or profile data to complement and validate the clustering results. Table I summarizes the EDM clustering studies included in our literature review, and the extent they addressed these problems. These studies are briefly described next in four categories.

Table I. Summary of the Clustering Studies in the Field of EDM

Study	Clustering methods	Compared with	Choice justified	Comparison measure	Validation dataset
Durfee et al. (2007)	SOM	-	no	-	students' performance
Anaya & Boticario (2009)	EM	-	no	-	expert opinion
Wang et al. (2004)	ISODATA	-	yes	-	-
Shih et al. (2010)	Step-wise HMM	-	yes	-	students' learning outcome
Hübscher et al. (2007)	Hierarchical clusters; K-means	-	yes	-	-
Maull et al. (2010)	K-means; EM	-	yes	-	-
Lee (2007)	PCA over SOM K-means	PCA only; hierarchical agglomerative clustering	yes	within-cluster variance	-
Dogan & Camurcu (2008)	K-means	Fuzzy c-means	yes	within-cluster variation	-
Perera et al. (2009)	K-means	EM clustering; hierarchical agglomerative clustering	yes	students' membership	group performance

Category 1: Studies applying only one clustering algorithm

Using factor analysis (Gorsuch, 1983; Hair et al., 2006) and Self-Organizing Map (SOM) (Harp et al. 1995; Vesanto and Alhoniemi 2000) techniques, Durfee et al. (2007) analyzed the relationship between student characteristics and their adoption and use of computer-based educational technology; subsequently a t-test on performance scores showed significant group differences, and thereby supported the clustering decisions. Anaya and Boticario (2009) classified users of a learning forum based on their level of interaction, such as the number of threads started and messages sent. Three clusters of learners were identified through the Expectation Maximization (EM) algorithm, and the results were validated by asking an expert to manually label students according to their collaboration levels. Wang et al. (2004) combined sequential pattern mining with a clustering algorithm to study students' learning portfolios. The authors first found each student's set of frequent sequences of learning activities. Then, a clustering algorithm called ISODATA was used to group learners into four clusters according to their learning features (Hall and Ball 1965). In Shih et al.'s study (2010), students' problem-solving behaviors while using a geometry cognitive tutor were broken down into actions. A series of actions represented a learning tactic. A step-wise Hidden Markov Model clustering (Baum and Petrie 1966) was developed to discover interpretable tactics, which were then related to learning outcomes.

Although some of the above studies validated the clustering results with other measures of students' performance, a common limitation with those studies is that they did not compare their choice of algorithm with another benchmark algorithm regardless of how well the decisions were justified. The next category of studies, however, did compare algorithm performance.

Category 2: Studies applying two clustering algorithms, but without comparisons to other (non-clustering) algorithms or validation with other data

Hübscher et al. (2007) used K-means and hierarchical clustering techniques (Hastie et al. 2009) respectively to group students who have used an educational hypermedia system that helps students understand relationships between science concepts and principles. Maull et al. (2010) applied a clustering approach to discover patterns among

teachers using an online curriculum planner. This study used K-means and expectation-maximum likelihood to cluster the user sessions based on 27 selected features. The two algorithms identified very similar patterns in the largest clusters. However, there was not complete agreement on top cluster features or cluster sizes for different algorithms. Finally, although both studies used two clustering algorithms, they did not validate results by triangulating with other data to help determine educational significance.

Category 3: Studies applying one algorithm compared with other algorithms, but without validation

Lee (2007) proposed to assess student knowledge and infer important knowledge states (mastery levels) in an integrated online environment using a SOM (Harp et al. 1995; Vesanto and Alhoniemi, 2000), K-means and principal component analysis (PCA) (Smith 2002). K-means was used to cluster the SOM generated from the student data set (SOM K-means). Comparisons with other algorithms, such as PCA only, showed that applying PCA over SOM K-means could reveal more significant student knowledge states than PCA itself. Dogan and Camurcu (2008) compared two clustering algorithms, K-means and fuzzy c-means (Bezdek, 1981; Dunn, 1973), to cluster students' exam results on six different concepts when using an intelligent tutoring system. K-means produced smaller squared error values when the number of clusters was four, five, and seven; fuzzy c-means gave a better result when there were six clusters. Similar to studies by Hübscher et al. (2007) and Maull et al. (2010), these two studies did not validate the clustering results using students' performance data. However, they compared the clustering results obtained from different algorithms.

Category 4: A study addressing all concerns

To the best of our knowledge, there is only one educational clustering study that addressed our concerns in a comprehensive way. Perera et al. (2009) explored group dynamics in a software development project by extracting patterns distinguishing the better from the weaker teams to gain insights on factors leading to success. K-means, EM algorithm and hierarchical agglomerative clustering were used to cluster project teams based on 11 numeric attributes, which captured the salient factors of using an online collaboration tool. All methods converged on the same results. K-means was also used to

cluster individual students based on their online activities to reveal information that was missing from team-wise clustering. The results revealed interesting patterns characterizing the work of stronger and weaker students, with a high number of active events associated with positive outcomes. However, this was a very small-scale study, involving only 43 students from 7 teams.

3. THE INSTRUCTIONAL ARCHITECT

This research is set within the context of the Instructional Architect (IA.usu.edu), a lightweight, web-based tool developed for supporting teachers in authoring simple instructional activities using online learning resources from the National Science Digital Library (NSDL.org) and the Web (Recker et al., 2006). With the IA, teachers are able to search for, select, sequence, annotate, and reuse online learning resources to create instructional web pages, called IA projects. These IA projects (or, *projects*, for short) can be kept private (private-view), made available to students only (student-view), or to the wider Web (public-view). Anyone can visit a public-view IA project, students can access their teachers' student-view IA projects through their student accounts, and private IA projects are viewable only by the author. Any registered teacher can make a copy of any public IA project by clicking the copy button at the bottom of the project. In this way, the IA supports a teacher community around creating and sharing instructional resources and activities.

To use the IA, a teacher must first register by creating a free IA account, which provides exclusive access to his/her saved resources and projects. As part of the registration process, the new user may optionally provide profile data, some of which is used as part of our study and discussed below.

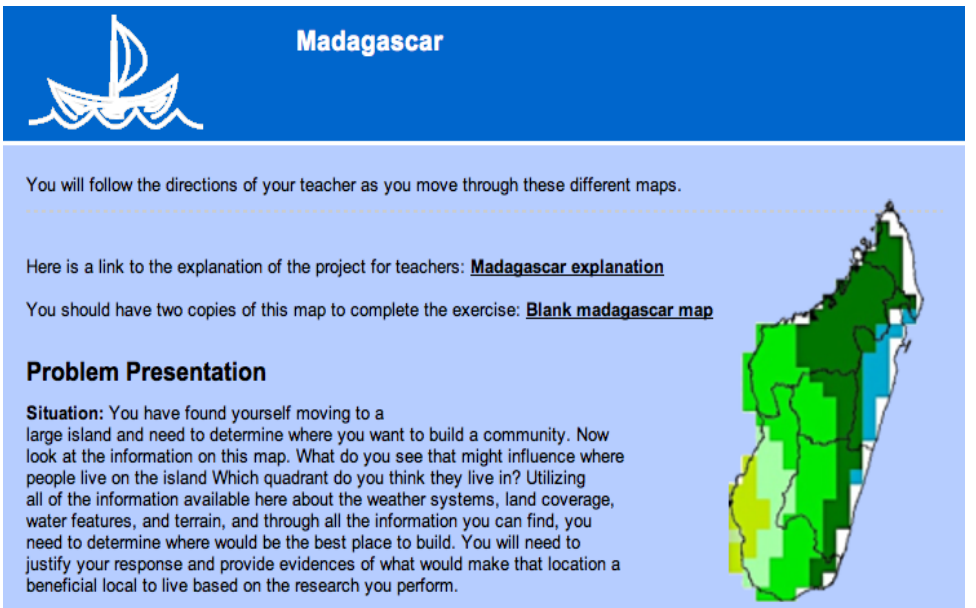

Fig. 1. Screenshot of a teacher-created IA project, containing links and a graphic.

Figure 1 shows an example of a teacher-created IA project. Since 2004, over 7,200 teachers have registered with the IA, who have created more than 17,000 IA projects using over 71,000 online resources. Public IA projects have been viewed over 1.7 million times.

4. PRIOR WORK

In a previous related work (Xu and Recker, 2012), we used Latent Class Analysis (LCA) to cluster IA users, and also provide interpretations of the kinds of teachers in each cluster. Based on the recurring cluster patterns, the three resulting clusters were labeled as: 1) *inactive islanders*, 2) *insular classroom practitioners*, and 3) *key brokers*, respectively. The first group tends to be isolated from other IA users and does not fully exploit the full range of features available in the IA. The second group is generally more interested in creating IA projects for students but produces lower quality IA projects. The last group comprises focused users, who are willing to observe and learn from others in the IA community while also giving back to it by publishing valued IA projects. This

group of users is also comprised those reporting the highest comfort with technology and the most teaching experience.

One of the limitations of our prior work is that no benchmark clustering method is used to compare with the performance of LCA. As pointed out in the literature review, an educational data mining study should justify the choice of clustering method, compare the selected method with a benchmark method, and more importantly, use a different data source to triangulate/validate clustering results. The study described in this paper is designed to fill gaps identified above by 1) applying LCA to an educational dataset, 2) comparing clustering algorithm results to a benchmark algorithm, namely K-means, using three comparison measures, and 3) evaluate the clustering results using teacher profile data

5. DATA, DATA PREPROCESSING, AND LCA

In this section, we first introduce the dataset and the data preprocessing procedures used in this research. We then provide an introduction to LCA (Goodman, 1974; Lazarsfeld and Henry, 1968; Magidson and Vermunt 2004; McCutcheon, 1987), a clustering algorithm that has been well researched but not given much attention in EDM. Next, we introduce the K-means algorithm as the benchmark algorithm for comparison purposes, and finally, present results from the comparison measures.

5.1 Data Set and Preprocessing Procedures

The IA is fueled by a relational database that stores not only users' profile data, but also their transaction data, such as their created IA projects, collected resources, and browsing activities. The data for this study is extracted from the relational database and aggregated to serve as the user features and input for the clustering study.

IA teachers are the focus of this study. Within the IA environment, a teacher can assume three general roles: project authoring, project usage, and navigation. Data from

these three roles are included in the feature space for representing a teacher's online behavior, as summarized below.

Role I – Project authoring. This includes five metrics: the number of public or student projects created, the number of copied projects, and three project quality indicators, namely, the mean number of resources used per project, the mean number of words in the project body, and the mean number of words in the project overview.

Role II – Project usage. This is measured by counting visits, excluding authors' visits to their own projects and visits referred from other websites. The latter is excluded because links to some (but not all) IA projects are automatically harvested into other digital libraries, for example the NSDL, and may result in an inflated count. To remove these potentially confounding factors, we only include the number of student visits and the number of visits from other IA users (called peer visits).

Role III – Navigation. This includes navigating through the IA website, as well as viewing and copying other teachers' projects. Note that the latter action also belongs to Role I.

Table II. User Feature Space

Role	Raw data	Transformed data	Min	Max	Mean
Project authoring	# of projects	Number of projects	1	10	2.46
	Project content	Mean number of resources per project	0	44	4.49
	Project content	Mean number of words in project body	0	2843	174.03
	Project overview	Mean number of words in project overview	0	293	22.50
	Project originality	* Number of copied projects	0	18	0.55

Project usage	Project visits	Maximum number of peer visits	0	164	1.84
	Project visits	Maximum number of student visits	0	1022	10.96
Navigation	Transaction data	Number of visits to the IA	1	57	7.85
	Transaction data	Number of project browses	0	134	8.58
	Project originality	*Number of copied projects	0	293	22.50

*Number of copied projects belongs to both the project authoring and navigation roles.

The data is collected from IA teachers who registered in 2009. One-time visitors and those who have never created any public IA projects are excluded in the collection process. As a result, the data from 661 teachers (out of a total of 1164 registered teachers during that period) are included to collect necessary information for constructing the feature space. Table II lists all the features extracted from the valid 661 teachers together with the minimum, maximum, and mean values for each feature. Since all the information discussed above may not be explicitly logged, extensive data transformation and aggregation are conducted to convert raw data into the desired features.

5.2 Latent Class Analysis (LCA)

This research uses LCA (Magidson and Vermunt, 2004) to classify registered teachers into groups. LCA is a model-based cluster analysis technique in that a statistical model (a mixture of probability distributions) is postulated for the population based on a set of

sample data. LCA offers several advantages over traditional clustering approaches such as K-means: 1) for each data point, LCA assigns a probability to the cluster membership, instead of relying on the distances to cluster means; 2) LCA provides various diagnostics such as common statistics, Log-Likelihood (LL), Bayesian information criterion (BIC) and *p*-value to determine the number of clusters and the significance of variables' effects; 3) LCA accepts variables of mixed types without the need to standardize or normalize them; and 4) LCA allows for the inclusion of demographic and other exogenous variables either as active or inactive factors (Magidson and Vermunt 2004; Vermunt and Magidson 2002). Inactive covariates do not affect initial parameter estimation; they are only taken into consideration at a later stage when the model without such covariates has been estimated.

The basic structure of an LCA model for continuous y variables is:

$$f(y_i) = \sum_{x=1}^{K} P(x) f(y_i|x), \quad (1)$$

where $f(y_i)$ is the distribution of a random manifest variable y_i, and $P(x)$ is the probability of latent class x regardless of any other information, and $f(y_i|x)$ is the distribution of y_i within latent class x. Then, the least restrictive model is obtained by assuming that all y's follow class-specific multivariate normal distributions, that is:

$$f(y_i|x) = (2\pi)^{-K/2} |\Sigma_x|^{-1/2} exp\{-\frac{1}{2}(y_i - \mu_x)' \Sigma_x^{-1} (y_i - \mu_x)\}. \quad (2)$$

In this model, each latent class has its own means μ_x and variance-covariance matrix Σ_x, which leaves too many parameters to be estimated.

In recent years, LCA has been further developed to include the mixed scale type (nominal, ordinal, continuous, and count), and to allow for both complete and partial local dependence in order to accommodate more research situations (Collins and Lanza, 2010; Magidson and Vermunt, 2004; Nylund, Asparouhov, and Muthén, 2007; Vermunt and Magidson, 2002). To reduce the number of parameters and to restrict an LCA model, one can either set cluster-independent error variances and covariances to zero, or set some off-diagonal elements of the covariance matrix to zero.

Finally, after an LCA model is constructed, cases are assigned to the latent class that can help achieve the highest □(□□) (Magidson and Vermunt, 2004).

LCA uses the maximum likelihood method for parameter estimation. It starts with an EM algorithm and then switches to the Newton-Raphson algorithm (Minka, 2002; Ypma, 1995) when it begins to converge on a final solution. In this way, the advantages of both algorithms, that is, the stability of EM and the speed of Newton-Raphson when it is close to the optimum solution (Vermunt and Magidson, 2005), are exploited.

5.3 Comparing LCA and K-means Algorithms

Because of its widespread use, the K-means algorithm is used as a benchmark algorithm for comparison with the LCA algorithm, which is purported to have better performance. As shown in Table II, the variables (or features) fall into different ranges. Unlike LCA, K-means relies on distance as a measure of cluster variance, which means scaling needs to be applied to the variables. In this study, all variables were adjusted to the range between 0 ~ 1, and called a Type 1 data transformation.

Since all variables in the feature space are positively skewed, the values on the right tails are much larger than the means. In order to alleviate the effect of outliers, 5% of the largest values are first winsorized to the 95th percentile of the entire set of data points, and then all values are converted to the range between 0 ~ 1. This is called a Type 2 data transformation. To ensure a fair comparison, both LCA and K-means algorithms are applied to the adjusted values obtained by Type 1 and Type 2 data transformation processes, respectively.

Table III. Data Segmentation (Type 3) Results

Transformed data	Equal interval	Range of original values
Number of projects	yes	1 2 ~ 10
Mean # of resources per project	yes	0~2 3~4 5 ~ 44
Mean # of words in project body	yes	0~32 33-167 168 ~ 2843
Mean # of words in project overview	yes	0~11 12~21 22 ~ 293
Number of copied projects	no	0 1 2 ~ 18
Maximum number of peer visits	no	0 1 2 ~ 164
Maximum number of student visits	no	0 1~5 6 ~ 1022
Number of visits to the IA	yes	1~4 5~8 9 ~ 57
Number of projects viewed	yes	0 1~4 5 ~ 134

Since the original features are either in continuous or count format, the outliers cannot be completely eliminated through data transformation. Outliers lead to an overly long tail of a sample dataset, inflated variance and error rate, and distorted estimation of

parameters in statistical models (Zimmerman, 1994). Therefore, data segmentation is used to further reduce the effect of outliers. Most variables are segmented into three equal intervals. However, some features are extremely skewed, leaving a huge number of data points on the far left and only a few cases on the tail side. In that case, it is impossible to segment the data into equal intervals. To this end, the authors made the data segmentation decisions based on their first-hand observation of IA teachers and IA usage statistics. The equal interval-based or data segmentation-based data transformation procedure is called Type 3 data transformation. Table III shows these segmentation decisions. Among the features, *mean # of words in project overview*, *mean # of words in project body*, and *number of visits to IA* are segmented into equal intervals.

In summary, our preprocessing procedures produce three types of transformed data. Type 1 is data from a linear transformation to make every variable fall into the range 0 ~ 1. Type 2 is data where the top 5% values of each variable are trimmed before making the linear transformation. Type 3 is data that have undergone the defined data segmentation-based preprocessing procedure.

LCA is applied to all three types of data (called LCA1, LCA2, and LCA3). K-means is only applied to Type 1 and Type 2 (called K-means1 and K-means2). Here, we do not apply K-means on the third type of preprocessed data because K-means is a distance-based algorithm. When every feature only differs by two levels at most, K-means is unable to distinguish subtle differences and appropriately separate the dissimilar data points. Finally, since clustering performance varies as the number of clusters k varies, we set k to be 3 to 15 to evaluate the average performance of each method in a fair setting. In theory, we can set the number of clusters k as large as it allows. However, LCA involves manual adjustment of parameters, which makes unlimited sets of clustering results unwieldy. Furthermore, too many clusters may lead to unnecessary splits of functional classes. Similarly, too few clusters are not very useful due to the lack of precision.

6. RESULTS

Comparisons between K-means and LCA are presented using three measures: 1) Davies-Bouldin indices, 2) clustering results validated with user profile data, and 4) cluster evolution. The first one relies on the internal criteria of the dataset, the second one relies on an interpretation of the clustering results, and the last one examines the evolution of the clusters when changing the value of *k*.

6.1 Davies-Bouldin Index

Clustering results can be evaluated using two criteria: minimal intra-cluster similarity and maximal inter-cluster dissimilarity. The Davies-Bouldin index is a cluster separation measure that strikes a balance between the two by taking both intra-cluster closeness and inter-cluster dispersion into consideration (Davies and Bouldin, 1979). The index is computed as follows:

$$DBI = \frac{1}{k}\sum_{i=1,j=1,i\neq j}^{k} \max\left(\frac{S_i+S_j}{M_{i,j}}\right), \qquad (3)$$

where k is the number of clusters, and S_i and S_j are the dispersions of cluster i and cluster j respectively. The dispersion is calculated as follows:

$$s_i = \sqrt[q]{\frac{1}{T_i}\sum_{l=1}^{T_i} |X_l - A_i|^q} \qquad (4)$$

where T_i is the number of data points in cluster i, and A_i is cluster i's centroid. M_{ij} is the distance between the centroids of the two clusters i and j, and is calculated by:

$$M_{ij} = \sqrt[p]{\sum_{l=1}^{N} |a_{li} - a_{lj}|^p} \qquad (5)$$

where a_{li} and a_{lj} is the lth component of the N-dimensional vectors a_i and a_j, respectively. Here, a_i and a_j are the centroid of clusters i and j, respectively; q in equation

(4) and p in equation (5) are usually set to 2, meaning all distances are in the Euclidean distance measure.

Since the Davies-Bouldin index considers both within-cluster and between-cluster dispersion, an algorithm that produces a smaller index is preferred over one producing a larger index. As can be seen in Table IV and Figure 2, LCA consistently produces smaller Davies-Bouldin indices than K-means for all cluster numbers ranging from 3 to 15. In addition, when using LCA, the Davies-Bouldin Indices always fall into a small range regardless of k; however, K-means' Davies-Bouldin Indices differ widely under different k. This means K-means is not as stable as LCA. As a result, LCA outperforms K-means for each clustering result under investigation.

A Friedman analysis of variance with repeated measures (a non-parametric repeated measures comparisons treating indices as ranks) was conducted (Friedman 1940; Howell 2002) using *Stata 11* (stata.com). The dependent variable was DB index (the means for the two methods were calculated for each *k)*, and the independent variable was cluster method with two levels (K-Means vs. LCA). The results showed that LCA significantly outperforms K-means, $Friedman = 13.00$, $p = .0003$. Finally, the DB index values for LCA are more stable regardless of the choice of k compared to K-means, which shows several troughs and spikes. For K-means, the DB index gets smaller as k increases.

Table IV. Davies-Bouldin Index on Results from different Clustering methods.

K	LCA1	LCA2	LCA3	K-means1	K-means2
3	2.21	2.16	**2.00**	11.87	16.19
4	**2.45**	2.62	2.49	26.57	18.95
5	**2.16**	2.37	2.42	10.02	9.78
6	**2.31**	3.60	2.93	11.38	8.94
7	2.88	**2.67**	2.71	18.33	13.07
8	3.01	3.03	**2.84**	6.87	9.13
9	2.98	3.02	**2.92**	8.12	9.94
10	3.14	**2.56**	3.70	13.65	12.03
11	2.97	**2.40**	3.57	5.88	6.92
12	3.06	**2.80**	3.29	8.37	9.28
13	3.26	**2.72**	3.13	8.08	8.16
14	2.91	**2.77**	3.60	10.18	8.44
15	**2.67**	3.04	3.35	7.79	7.69

Note: Bold values indicates the best result for a certain *k*.

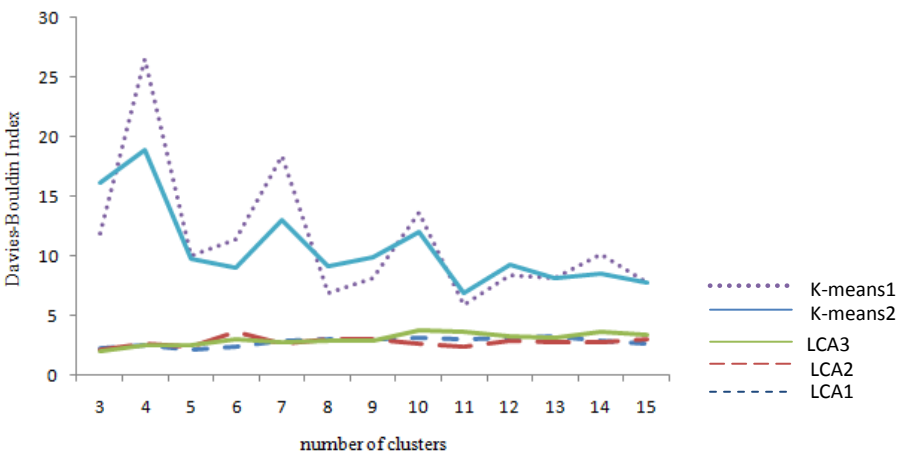

Fig. 2. Davies-Bouldin indices for different *k*.

6.2 Clustering Results Validated with User Profile Data

The Davies-Bouldin index only examines internal cluster criteria. However, to contribute to educational data mining research, a clustering study should also address educational questions. As such, we believe it is important to validate cluster results by triangulating with other sources of data.

Our previous research (Xu and Recker 2012) discovered strong relationships between teacher characteristics and their IA activities. Specifically, results showed that teachers with more teaching experience were more likely to be key brokers in the online IA community, whereas those with less teaching experience were more likely to be inactive users.

In order to examine the association between different clustering results and users' profile data, we looked for relationships through a common theme – teachers' effectiveness in using the IA.

In particular, this section describes a type of cluster validation by associating the clustering results with users' profile data collected as part of the registration process. When teachers first create their IA account, they are asked to optionally state their years of teaching experience. This dataset contains values for 233 teachers, and they are divided into two groups: Novice (N = 133, 1 – 3 years) and Veteran (N = 100, more than 3 years).

In order to triangulate teacher clusters with teaching experience, each cluster is 1) defined by its usage pattern, and then 2) related to self-reported teaching experience.

6.2.1 Extract Usage Patterns. Each cluster's characteristics are defined by behaviors on the nine features presented in Table III. Since the distribution of each feature is skewed, the mean value of each feature cannot be used to define a user. In this study, all variables are segmented into three levels. Since LCA3 already uses segmented data, a different segmentation method from that of LCA3 is used. Here, equal interval is used for all variables. In cases having a variable that is highly skewed and cannot be segmented

into three equal parts, the level for smaller values is assigned more users than the one with larger values.

After all variables are segmented, every cluster is next converted to a piece of usage pattern, which is a conjunction of the themes of individual features within a cluster in the form of $f_1 = t_1 \wedge f_2 = t_2 \wedge \ldots \wedge f_n = t_n$, where $<f_1, f_2, \ldots, f_n>$ denotes the user feature space, and $<t_1, t_2, \ldots, t_i>$ denotes the themes for each feature (Xu 2011; Xu and Recker 2012).

The theme of a user feature for a certain cluster is defined based on the following heuristic rules:

1. If one of its levels has 75% or more users, it is considered the dominant level, and the value for that level is the dominant theme.

2. If two neighboring levels consist of more than 75% of the total users, and the upper level and lower level differ by at least 10%, the combination of the two neighboring levels with more than 75% of total users is considered the dominant theme.

3. If this feature does not have a dominant theme, it is not considered for the user pattern *k* at all.

Since a feature's values rarely fall completely into one level, we develop the 75% heuristic rule. A too lenient threshold cannot justify the representativeness of a case, while a too stringent threshold will filter out too many dominant themes. Five themes might exist for a three-level indicator: the lowest level is dominant, the lower two levels are dominant, the middle level is dominant, the higher two levels are dominant, and the highest level is dominant. After all of the themes are defined, dominant themes are combined to represent cluster-wise usage patterns. A cluster's usage pattern is a more descriptive summary of a group of teachers' online behaviors.

As an example, Figure 3 shows the distributions of every feature into its respective three levels generated by LCA2 (when $k = 4$). Each set of stacked bars represents three tiers of a feature, with the leftmost stack being the lowest level and the rightmost stack being the highest level. For example, for the *number of peer visits* feature in cluster 1, 62.2% of the users fall into the lowest level (a long dotted bar on the left), 33.3% fall into

the middle level (a small square-pattern bar in the middle), and only 4.4% fall into the highest level (a tiny semi-solid bar on the right). We also use arrows to mark all dominant patterns of individual features of each cluster. Then, based on observed patterns, clusters are labeled in terms of how well they represent effective use of the IA. Due to space limitations, other cases are not presented in this article and details can be found at http://edm.usu.edu/publications/appendix.pdf.

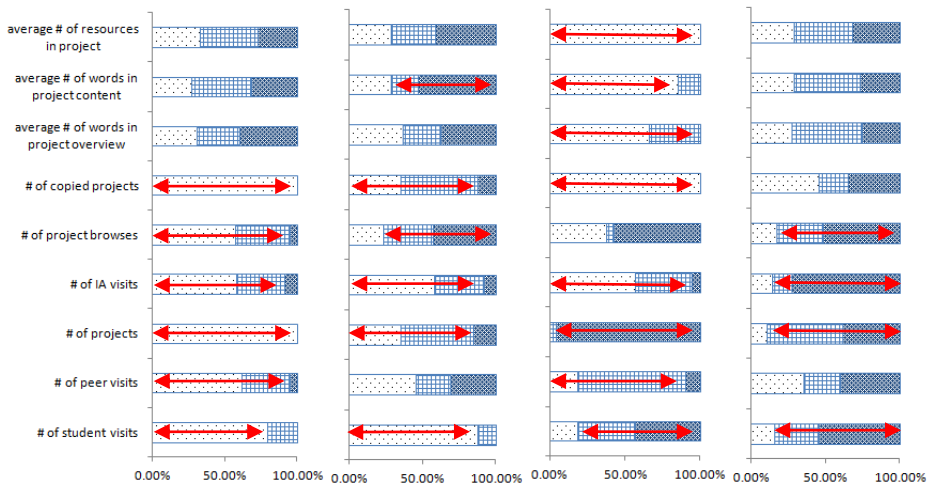

Fig. 3. Visualization of the dominant patterns of each feature when LCA2 and k = 4, where each column represents 1 cluster

6.2.2 Associate Usage Patterns with User Profile Data. Prior research has shown that successful use of a digital library is related to teachers' teaching experience and information literacy (Chen & Doty, 2005; Perrault, 2007). Multinomial logistic regression models were fitted to model potential associations between users' effectiveness in using the IA and their self-reported teaching experience. Because K-means performance is not stable as shown via the DB index, we sample clustering results from different k's. Note that since teaching experience is a nominal variable and it is impossible to tell the subtle group-wise differences when there are too many clusters, we

only examine clusters with k = 3, 4, and 5. Fifteen (3 x 5) models in total were fitted in order to compare those five clustering methods.

Multinomial logistic regression, also known as polychotomous logistic regression, is used when the dependent variable has more than two categories, and the explanatory variable is numerical or categorical (Chatterjee and Hadi, 2006). When there is no natural ordering of the dependent variable, one category of the dependent variable is considered as the base level (reference group), and multinomial logistic regression can be applied to estimate the relative risk ratios that a particular outcome is present in other categories (comparison groups) instead of in the reference group under the influence of the explanatory variables (Hosmer and Lemeshow. 2000; Kwak and Clayton-Matthews. 2002).

In each model, teachers' cluster labeling is set as the dependent variable, and teaching experience is set as the binary predictor variable (comparing veteran teachers with novice teachers). By examining the relative-risk ratios, we model the relative risk of being clustered in one category rather than in the reference category for a unit change in the predictor variable.

Taking the four cluster (k = 4) case generated by LCA2 as an example, the multinomial logistic regression result shows that compared with novice teachers, veteran teachers are more likely to be clustered in cluster 2 and cluster 4 than in cluster 1 (relative risk ratio= 2.54, $p < .01$, relative risk ratio= 4.78, $p < .01$, for clusters 2 and 4 respectively); the probability of being categorized in cluster 2 or in cluster 4 do not have a statistically significant difference ($p = .10$).

In examining the usage patterns derived from LCA2, we find a significant relationship between teaching experience and clustering results. Cluster 4 achieves high levels on four features (*number of project browses*, *number of IA visits*, *number of projects*, and *number of student visits*). Its four dominant themes indicate this group represents the most effective IA users. The regression model shows that veteran teachers are more likely to be in this cluster. Conversely, cluster 1 achieves low levels on six features (*number of copied projects*, *number of project browses*, *number of IA visits*, *number of projects*, *number of peer visits*, and *number of student visits*). Its six dominant

themes suggest that this group represents the most ineffective IA users. The regression model supports this interpretation, as novice teachers are more likely to be in this cluster than any other cluster. Cluster 2 and cluster 3 show mixed patterns, in that users achieve high levels on two features and low levels on four and six features respectively. These two groups should be comprised of more effective users than cluster 1, but less effective than cluster 4. The former is confirmed by the regression model, while the latter contradicts it because cluster 2 and cluster 4 should have similar proportion of veteran teachers.

One of the values of educational data mining is that its results should enrich our knowledge of the subject of interest. In our case, we investigate whether clustering results show a relationship with teachers' profiles, specifically to separate novice from veteran teachers. Based on the above analysis, cluster 1, cluster 3 and cluster 4 produced by LCA2 are able to correctly profile teachers. But there is a discrepancy between the expected and actual teaching experience in cluster 2; when k = 4, LCA2 profiler's accuracy is 75%.

Table V. Clustering methods' accuracy as a teacher profiler

		Kmeans1	Kmeans2	LCA1	LCA2	LCA3
k = 3	correct	-	-	3	2	2
	incorrect	-	-	0	1	1
	No distinctive profile	3	3	0	0	0
	accuracy	-	-	100%	67%	67%
k = 4	correct	-	-	4	3	3
	incorrect	-	-	0	1	1
	No distinctive profile	3	3	0	0	0
	accuracy	-	-	100%	75%	75%
k = 5	correct	-	-	4	3	4
	incorrect	-	-	1	0	1
	No distinctive profile	3	3	0	2	0
	accuracy	-	-	75%	-	75%

Note: "-" indicates no distinctive teacher profile, and thus no way to calculate accuracy rate.

In this fashion, we compare the five clustering methods in terms of their profiler accuracies. Table V shows the accuracy rate of each clustering methods for different *k*'s. None of the clusters produced by K-means methods shows a distinctive teacher profile that can distinguish a group of teachers as more experienced than others. In this case, it is impossible to use the clustering results as a teacher profile, let alone compare it with the benchmark self-report profile; therefore, we are not able to calculate the accuracy rate. When $k = 5$, the 2nd cluster produced by LCA2 has too few cases to produce a valid multinomial logistic regression result. This is because users who haven't reported their teaching experience have been excluded from this analysis. Therefore, cluster 2 is excluded from multinomial logistic regression analysis. In addition, cluster 5 doesn't show significant difference in teaching experience than any other cluster. Since we cannot get a full picture from this particular case, we cannot calculate accuracy for LCA2 when $k = 5$.

In sum, this analysis shows that K-means fails in finding user groups whose online usage behaviors could be accounted for by their teaching experience. For our dataset, K-means does not appear to create the clusters we would expect based on teacher characteristics. Conversely, in this context, LCA methods are able to produce educationally meaningful clusters. In particular, all three LCA methods showed a strong association between users' teaching experience and how well they can use the IA. Among them, LCA1 appears the best at clustering users whose online behaviors could be explained by their teaching experiences. LCA3 appears to have the lowest performance, probably because it has lost the subtle difference between users after data segmentation.

6.3 Cluster evolution

As illustrated above, the generated IA patterns vary with the algorithms and data transformations used, as well as the number of clusters *k*. In addition, with LCA, even as *k* increases, the patterns demonstrated by each user group overlap with their counterparts for smaller *k*. Based on observations of cluster formation, we conjecture that LCA does

not produce a completely different set of clusters; instead, cases are taken from the existing clusters to form each new cluster.

To examine this more closely, for each method and for k = 3-5, we group similar clusters together based on the themes of the nine features to examine the evolving process of cluster formation. Figure 4 shows a visualization of how LCA1 groups users into different clusters. Initially, 329 users are assigned to cluster 3-1, from which LCA1 assigns 244 users to cluster 4-1, and the remainder to cluster 4-2; next, cluster 5-1 keeps almost everyone from 4-1. Similarly, cluster 5-2 evolves from cluster 3-2 through cluster 4-3. Cluster 5-5 can be traced back to cluster 3-3. Clusters 3-1 and 3-2 have users of similar characteristics (low peer visits, very few projects, and no copied projects), and those users are grouped together to form cluster 4-2. But when k = 5, it is split up again to become 5-3 and 5-4. In this way, the visualization confirms our conjecture about how new clusters evolve from old ones, thereby partitioning users into newer, smaller clusters.

In contrast, Figure 5 shows how K-means2 groups users. As can be seen, there appears to be few similarities between clusters. Unlike a probability model, K-means' greedy approach leads to a series of solutions that bear little resemblance to the previous ones. The fact that clusters produced by K-means do not evolve gradually explains why a consistent clustering performance is not observed when measured by the Davies-Bouldin index, and the indices vary widely under different k's.

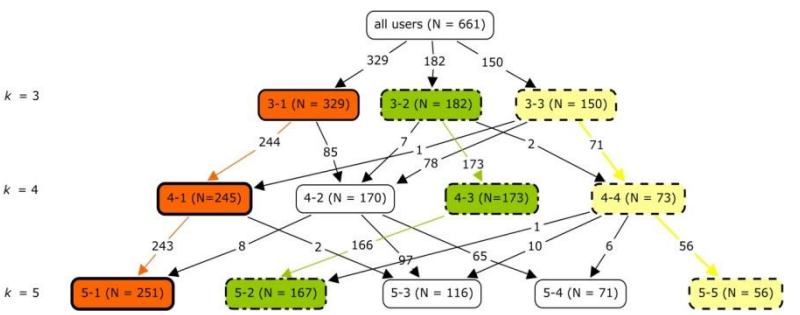

Fig. 4. Evolution of LCA1 clusters.

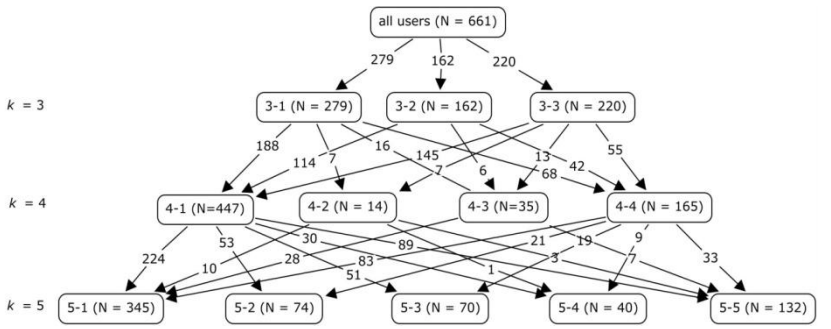

Fig. 5. Evolution of K-means2 clusters.

7. IMPLICATIONS FOR THE INSTRUCTIONAL ARCHITECT

The LCA study shows that teachers' teaching experience is correlated with their effectiveness in using the Instructional Architect. In particular, veteran teachers are more likely to demonstrate active engagement when browsing and viewing others IA projects, and in return create more projects and encourage their students to use them. On the other hand, novice teachers appear less likely to transfer their teaching strategies to an online tool.

Several activities could be done to lower teachers' barriers in adopting the IA. Firstly, a small number of teachers have participated in face-to-face professional development workshops on using the IA. These teachers showed large, significant gains in terms of their knowledge of and attitudes toward using the IA and online resources in their teaching (Walker et al., 2011). These workshops could benefit more people if moved to a self-paced, online environment. Secondly, teachers are not able to easily collaborate within the IA. Future versions of the IA could mine IA project content or self-reported demographic data in order to form interest groups based on shared subject areas and grade levels; and teachers could even pair up to create IA projects. In this way, the IA becomes a service for online collaboration to support teachers in sharing their pedagogies with each other, and in designing high quality IA-based learning activities.

8. CONCLUSIONS

In light of the increasing interest in employing clustering techniques in EDM research, the work reported in this article used data from an educational digital library service, the Instructional Architect, as a test bed for conducting a comprehensive clustering study.

Our literature review found few educational data mining studies that applied different clustering methods to the same dataset, and also found it rare to use different sources of data to complement one and another within one study. In our study, two clustering algorithms - LCA and K-means, a widely used algorithm in EDM studies - are compared using one year's worth of Instructional Architect usage data. Our contribution to EDM lies beyond simply comparing LCA with a well-known benchmark algorithm. Instead, we compared the two clustering results from different angles: the Davies-Bouldin index, accuracy as a teacher profiler, cluster evolution, and using different data sets: teachers' online behaviors vs. their self-reported teaching experience. This makes our contributions unique and multi-faceted.

In the first comparison, the Davies-Bouldin index is used to compare the internal quality of the identified clusters for each method. This index integrates intra-cluster similarity and inter-cluster dissimilarity. On this measure, LCA performs much better than K-means algorithm regardless of the type of data preparation used. In addition, unlike K-means, LCA's performance in terms of its Davies-Bouldin index grows slowly as the number of clusters *k* grows.

Secondly, in order to examine each method's utility in addressing educational questions, we model the relationship between data from users' profiles (specifically, teaching experience) and each resulting cluster. In general, K-means fails to find any association between teaching experience and usage patterns as defined by each cluster. LCA methods, in particular LCA1, perform better than K-means in terms of associating clusters with teaching experience, supporting findings from previous studies (Chen & Doty, 2005; Perrault, 2007).

Thirdly, a visualization of how K-means and LCA generate new clusters when *k* increases shows that K-means produces a very different set of clusters, whereas

LCA models with higher *k*'s seem to results in new clusters that are partitions of those from lower *k* models. Overall, the analysis shows that LCA is less sensitive to the variance with feature variables, and clustering results turn out well with minimal data transformation.

There are several limitations to this work. More generally, clustering results are sensitive to the kinds of algorithms used. They can also sometimes find structure in a dataset when none exists. More specifically, this study only shows the superiority of LCA over K-means, using one particular educational dataset. Moreover, although the dataset is characterized using a theoretically motivated feature set, other data and feature sets could produce different results. Finally, only two kinds of clustering algorithms are compared, and others exist.

In conclusion, although LCA has widespread applications in health, marketing, survey, sociology, psychology, and education research, it has not been extensively utilized in EDM research. Through this study, LCA's utility as an EDM method has been demonstrated and discussed. Compared with the more widely used K-means, it appears more useful in clustering IA users in educationally meaningful ways and, as a statistical model, its performance is stable across different numbers of clusters for this dataset. As such, it is worth considering by researchers who are interested in studying usage patterns in educational contexts.

ACKNOWLEDGMENTS

This material is partially based upon work supported by the National Science Foundation under grant #0840745. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the many teacher users of the Instructional Architect, Bart Palmer, and members of the IA research group.

REFERENCES

ANTONENKO, P., SERKAN TOY, S., AND NIEDERHAUSER, D. 2012. Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 6*(3), 383-398.

BAKER, R. S. J. D. AND YACEF, K. 2009. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining 1,* 3-17.

BAUM, L. E. AND PETRIE, T. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics* 37, 1554-1563.

BEZDEK, J. C. 1981. *Pattern recognition with fuzzy objective function algorithms*, New York, Plenum Press.

CAMPBELL, S. B. AND MORGAN-LOPEZ, A. A. 2009. A Latent Class Analysis of maternal depressive symptoms over 12 Years and offspring adjustment in adolescence. *Journal of Abnormal Psychology 118*, 479-493.

CHATTERJEE, S. AND HADI, A. S. 2006. *Regression analysis by example*, 4th ed. John Wiley and Sons, Inc.

CHEN, H., & DOTY, P. 2005. A conceptual framework for digital libraries for k–12 mathematics education: part 1, information organization, information literacy, and integrated learning. *The Library Quarterly, 75*, 231-261.

COLLINS, L.M., AND LANZA, S.T. 2010. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences.* New York: Wiley.

DAVIES, D. L. AND BOULDIN, D. W. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence 1*, 224-227.

DOGAN, B. AND CAMURCU, A. Y. 2008. Visual Clustering of multidimensional Educational data from an intelligent tutoring system. *Computer Applications in Engineering Education 18*, 375-382.

DUNN, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *J Cybernetics 3*, 32-57.

DURFEE, A., SCHNEBERGER, S., AND D. L. AMOROSO. 2007. Evaluating students computer-based learning using a visual data mining approach. Journal of Informatics Education Research 9, 1-28.

FRIEDMAN, M. 1940. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11, 86-92.

GOODMAN, L. A. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. I, 2, 215-231.

GORSUCH, R. L. 1983. *Factor Analysis.* Lawrence Erlbaum, Hillsdale, NJ.

HAIR, J., BLACK, B., BABIN, B., ANDERSON, R. E., AND TATHAM, R. L. 2006. *Multivariate Data Analysis*, 6th edition, Pearson Prentice Hall, New Jersey.

HALL, D. J. AND BALL, G.B. 1965. ISODATA: A novel method of data analysis and pattern classification. Technical report, Stanford Research Institute, Menlo park CA.

HARP, S. A., SAMAD, T., AND VILLANO, M. 1995. Modeling student knowledge with self-organizing feature maps. *IEEE Transactions on Systems, Man and Cybernetics 25*, 727-737.

HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2009.*The Elements of Statistical Learning* (pp. 520-529), 2nd edition. Springer, New York.

HOSMER, D. W. AND LEMESHOW, S. 2000. *Applied Logistic Regression*, 2nd ed. John Wiley and Sons, Inc.

HOWELL, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury Thomson Learning, Inc.

HÜBSCHER, R., PUNTAMBEKAR, S. AND NYE, A. H. 2007. Domain specific interactive data mining. In *Proceedings of Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*.

KLONSKY, E. D. AND OLINO, T. 2008. M. Identifying clinically distinct subgroups of self-injurers among young adults: A Latent Class Analysis. *Journal of Counseling and Clinical Psychology 76*, 22-27.

KWAK, C. AND CLAYTON-MATTHEWS, A. 2002. Multinomial logistic regression. *Nursing Research 51*, 404-410.

LAZARSFELD, P. F. AND HENRY, N. W. 1968. *Latent Structure Analysis*. Houghton Mifflin.

LEE, C. 2007. Diagnostic, predictive and compositional modeling with data mining in integrated learning environments. *Computers & Education 49*, 562-580.

MAULL, K. E., SALDIVAR, M. G., AND SUMNER, T. 2010. Online curriculum planning behavior of teachers. In *Proceedings of the 3rd International Conference on Educational Data Mining*.

MAGIDSON, J. AND VERMUNT. J. 2004. Latent class models. In *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, Ed. Sage Publications, Thousand Oaks, CA, 175-198.

MCCUTCHEON, A. L. 1987. Latent class analysis. *Quantitative Applications in the Social Sciences Series 64. Sage Publication, Thousand Oaks, California.*

MINKA, T. P. 2002. Beyond Newton's method. http://research.microsoft.com/en-us/um/people/minka/papers/minka-newton.pdf.

NYLUND, K.L., ASPAROUHOV, T., & MUTHÉN, B. 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling 14*, 535-569.

NYLUND, K., BELLMORE, A., NISHINA, A., AND GRAHAM, S. 2007. Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development 78*, 1706-1722.

PENCE, B. W., MILLER, W. C., AND GAYNES, B. N. 2009. Prevalence estimation and validation of new instruments in psychiatric research: An application of latent class analysis and sensitivity analysis. *Psychology Assessment 21*, 235-219.

PERERA, D., KAY, J., KOPRINSKA, I., YACEF, K., AND ZAI¨ANE, O. R. 2009. Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering 21*, 759-772.

PERRAULT, A. M. 2007. An Exploratory Study of Biology Teachers' Online Information Seeking Practices. *School Library Media Research*, 10.

RECKER, M., DORWARD, J., DAWSON, D., MAO, X., YE, L., PALMER, B, HALIORIS, S., AND PARK, J. 2006. *The Annual Meeting of the American Education Research Association*.

RECKER, M., WALKER, A., GIERSCH, S., MAO, X., PALMER, B., JOHNSON, D. LEARY, H., AND ROBERTSHAW, B. 2007. A study of teachers' use of online learning resources to design classroom activities. *New Review of Hypermedia and Multimedia 13*, 117 - 134

ROMERO, C. AND VENTURA, S. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33, 135-146.

ROUSSOS, L. A., TEMPLIN, J. L., AND HENSON, R. A. 2007. Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement 44*, 293–311.

SHIH, B., KOEDINGER, K. R., AND SCHEINES, R. 2010. Unsupervised discovery of student learning tactics. In *Proceedings of the 3rd International Conference on Educational Data Mining*.

SMITH, L. I. 2002. *A tutorial on principal component analysis*. Cornell University

VERMUNT, J., K. AND MAGIDSON, J. 2002. Latent class cluster analysis. In *Applied Latent Class Analysis*, J. Hagenaars and A. McCutcheon, Eds. Cambridge University Press, 89-106.

VERMUNT, J.K. AND MAGIDSON, J. 2005. *Technical guide for Latent GOLD 4.0: Basic and advanced*. Statistical Innovations Inc.

VESANTO, J., AND ALHONIEMI, E. 2000. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks 11*, 586–600.

WALKER, A., RECKER, M., ROBERTSHAW, B., OLSEN, J., LEARY, H., YE, L., & SELLERS, H., 2011. Integrating technology and problem-based learning: A mixed methods study of two teacher professional development approaches. *Interdisciplinary Journal of Problem-based Learning* 5, 70-94.

WANG, W., WENG, J., SU, J., AND TSENG, S. 2004. Learning portfolio analysis and mining in SCORM compliant environment. *The 34th ASEE/IEEE Frontiers in Education Conference*.

XU, B. 2011. Clustering educational digital library usage data: Comparisons of latent class analysis and K-Means algorithms. *All Graduate Theses and Dissertations. Paper 954.* http://digitalcommons.usu.edu/etd/954

XU, B., AND RECKER, M. 2012. Teaching analytics: A clustering and triangulation study of digital library user data. *Journal of Educational Technology & Society 15*, 3, 103-115.

YPMA. T. J. 1995. Historical development of the Newton-Raphson Method. *SIAM Review 37*, 5, 531-551.

ZIMMERMAN, D. W. 1994. A note on the influence of outliers on parametric and nonparametric tests. *Journal of General Psychology 121*, 4, 391-401.