The 8th International Scientific Conference
eLearning and software for Education
Bucharest, April 26-27, 2012
10.5682/2066-026X-12-119

ASSESSING STUDENT RETENTION AND PROGRESSION: A MULTI-MODAL APPROACH

Phil ICE
American Public University System, 111 West Congress Street, Charles Town, West Virginia, USA
E-mail: pice@apus.edu

***Abstract**: Student retention and progression are important measures of success for postsecondary education. They are key factors by which online programs, in particular, are currently scrutinized. This presentation reviews a successfully implemented, multi-modal approach that leverages data mining and quantitative analysis, supported by text analytics. American Public University System, began assessing student retention with an exploratory model utilizing regression analysis with 32 variables from the student information system. Quickly, the initiative expanded include the use of data mining across all campus systems touched by students and the integration of criterion nodes into neural network models. Despite an 87% degree of accuracy in predicting retention within a 125 hour window across 187 variables, the issue of causality remained opaque. Incorporation of text analytics to student input provided a means of ontological ordering of qualitative data that could then be converged back onto relevant data points across high probability nodes of disenrollment. The merging of these techniques has provided APUS with both a means of creating actionable business intelligence to assist in retaining students, as well as a causal understanding of systemic issues. As previously noted, the merging of data mining, neural network analysis, conventional regression analysis and text analytics has provided a robust framework for intervention at both the short term and long term horizons. Through actionable intelligence, provided by the explanatory data derived from text analytics and semantic analysis, the APUS data team has been able to provide insight to the instructional design team, faculty members and administrative stakeholders. This has translated into a richer basis for continuous quality improvement of course materials, pedagogical strategies and student services. The impact on retention and student satisfaction has been considerable with 31% and 19% increases respectively since implementation. Participants will be introduced to the data collection, federation and modeling techniques utilized at APUS. This will include exploration of methodology and required technical infrastructure. The presentation will be in case study format with numerous examples and resource links. Participants will be encourage to raise questions at any point and to consider how similar techniques might be used at their institution. Significant coverage will be given to exploration of perceived problems associated with both technical infrastructure and stakeholder buy-in.*

***Keywords:** Data mining, Exploratory statistics, Community of Inquiry Framework, Text analytics, Online learning*

I. INTRODUCTION

The evolution of student retention and progression over the years has yielded considerable methods and approaches toward the goal of coming closer to understanding causal factors—both student and institutional, that may trigger early student disenrollment from postsecondary study. Further, accountability guidelines established by the U.S. Department of Education, as well as organizational accreditation agencies [1,2] have also contributed to the collective mission for postsecondary student success. Although well-intentioned, researchers, agencies, organizations, institutions and all stakeholders involved in this mission have fallen short of this goal for a number of reasons.

Historically, education has maintained a lasting reputation of failing to proactively devise strategies that accommodate and that are designed to quickly respond to future economic and societal trends [3]. If history is an indicator of the current disruptive landscape, perhaps former CEO of General Electric, Jack Welch, said it best in his observation, “If the rate of change on the outside exceeds the rate of change on the inside, the end is near” [4].

Additionally, online universities have the ability to collect large quantities of qualitative data. Unfortunately qualitative data has been underutilized due to the inefficiency of traditional methods of analysis.

The purpose of this study is to introduce a new multi-modal approach to assess nontraditional student retention and progression. Framed by the Community of Inquiry, student retention and progression is examined by the American Public University System (APUS) through a tripartite methodological lens using (1) descriptive, (2) inferential, and (3) exploratory data. Although very informative, APUS collects volumes of explanatory data that cannot be efficiently analyzed using traditional qualitative methodologies. Therefore, recognizing this issue, APUS presents an innovative and efficient multi-modal approach to analyze text.

1.1. Institutional Assessment: The APUS Retention and Progression Model

As a foundational guide, mission statements should be central to higher education programs and institutions when developing assessment strategies [5,6]. The APUS Retention and Progression Model illustrates the continuous interconnectedness among analyses and stakeholders within a data-driven decision-making culture. Both internal and external benchmarking provides a wealth of explanatory, exploratory, inferential, and descriptive data toward understanding who is likely to disenroll and the reasons why.

1.2. Levels of Analysis

Analysis was conducted at the individual record level. Each students qualitative replies were disaggregated from the overall data and entered into the analytics engine. As described below, these were correlated with the Teaching and Cognitive Presence dimension of the Community of Inquiry Framework

1.3. Descriptive Statistics at APUS

Descriptive statistics are used to present rudimentary data characteristics and additionally provide basic information regarding the population and the measures. Descriptive statistics are commonly included in quantitative analyses and furthered displayed in a chart, graph, or table. Descriptive statistics can (1) be analyzed by a single system; (2) are subjective in interpretation; and (3) represent nearly 90% of solutions. APUS collects student demographic data that include gender, ethnicity, race, military classification and military branch. Pie charts are used below in Figure 1 to display APUS student demographics and enrollment percentages.

1.4. Inferential Statistics at APUS

Inferential statistics provide extended conclusions whereby inferences can be made on observed differences between groups and then can be further applied to general conditions. APUS uses inferential statistics such as regression, factor analysis, and decision trees to examine areas regarding retention, learning effectiveness, instructional design, and beta-level technology integration. Figure 3 illustrates a federation of multiple demographic and transactional data sets using predictive modeling.

Figure 1. APUS Student Demographics Using Descriptive Statistics

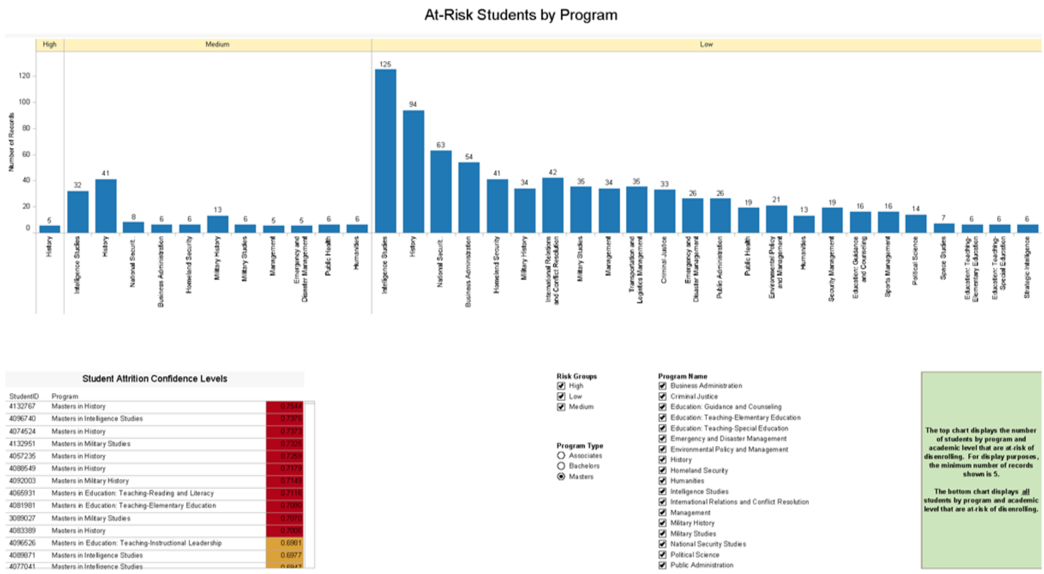

Figure 2. APUS At-Risk Students by Program Using Predictive Modeling

1.5. Retention and Causality

Utilizing only quantitative measures is not adequate for implementing systematic change across the institution, nor is using only qualitative measures. Therefore, APUS collects both qualitative and quantitative data and analyzes this data by using a combination of descriptive, regression, and factor analysis, thus providing a more comprehensive examination of retention and progression.

Although this mixed methods analysis appears conceptually practical, the large volume of qualitative data, in particular, that is collected at APUS makes traditional qualitative data analysis immensely impractical. Moreover, exploration of the data should extend beyond branch/node analysis toward the discovery of participant perceptions, opinions, and personal accounts. Based upon these observations, consensus was established to develop an efficient way to analyze large volumes of qualitative data.

1.6. Using the Community of Inquiry Framework to Analyze Text

APUS employs an end-of-course survey based upon in the Community of Inquiry Framework (CoI) [7]. The CoI is a process model of learning in online and blended educational environments. The model is grounded in a collaborative constructivist view of higher education [8] that assumes effective online learning requires the development of a community of learners to support meaningful inquiry and deep learning (Figure 3 below).

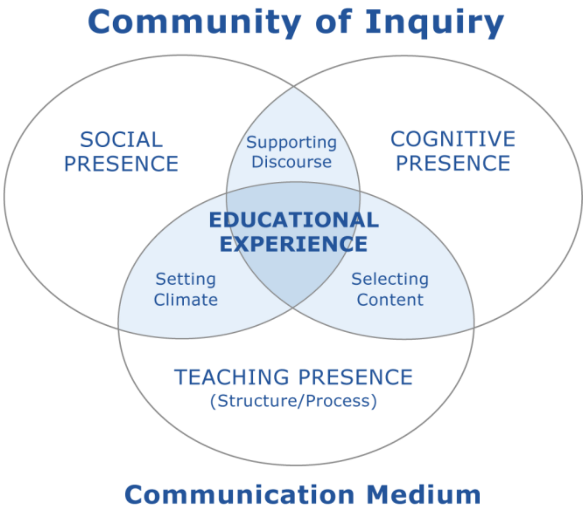

Figure 3. The Community of Inquiry Framework Model (Garrison, Anderson, Archer, 2000).

The developers of this model assert that interdependence among three distinct constructs--social, cognitive and teaching presence, ultimately yields an educational experience.

Social Presence. Social presence is described as "the ability of participants to identify with the community (e.g., course of study), communicate purposefully in a trusting environment, and develop inter-personal relationships by way of projecting their individual personalities." [9]. Subcategories used to describe characteristics of social presence include affective expression, open communication, and group cohesion.

Teaching Presence. Teaching Presence supports and encourages the realization of objectives and goals using the design, facilitation, and direction of cognitive and social processes [7]. Examples of activities that evidence teaching presence include designing the curriculum and associated activities, shaping and guiding constructive discourse, and focusing and resolving arising issues.

Cognitive Presence. Cognitive presence is explained by the sustained reflection and discourse necessary for learners to construct and confirm meaning [10] Specific indicators of cognitive presence includes triggering events (sense of puzzlement), exploration (sharing information & ideas), integration (connecting ideas), and resolution (synthesizing & applying new ideas).

1.7. Community of Inquiry Survey

Based on the Community of Inquiry Framework, the CoI Survey was developed toward measuring the existence of the three presences [11]. The CoI survey includes 9 social presence items (3 affective expression, 3 open communication, 3 group cohesion); 12 cognitive presence items (3 triggering, 3 exploration, 3 integration, 3 resolution); and 13 teaching presence items (4 design & facilitation, 6 facilitation of discourse, 3 direct instruction). Table 1 shows the Based on research conducted around Social, Cognitive, and Teaching Presence, this instrument was further validated using principal component factor analysis whereby a three factor model was confirmed. More than 500,000 learners have used this instrument, thus creating a strong baseline for further research.

Table 1.. Community of Inquiry Presence, Category, and Associated Indicator(s)

CoI Presence	Category	Indicator
Teaching	Design and Organization Facilitating Discourse Direct Instruction	Learning climate/risk- free expression Group identity/collaboration Self-projection/expressing emotions
Cognitive	Triggering Event Exploration Integration Resolution	Sense of puzzlement Information exchange Connecting ideas Applying new ideas

1.8. Considerations

Although libraries exist for text analytics, they are not specific to the Community of Inquiry Framework or higher education. Conversely, however, existing Opinions library helped identify positive and negative comments. Therefore, the determination was made to define CoI categories and identify the terms that were related to each category and then further test the text analysis model to establish accuracy.

Although the CoI Framework comprises three presences, the researchers of this study decided to focus on only two of the three; (1) Teaching Presence; and (2) Cognitive Presence. From the CoI Survey, questions were developed to begin defining the categories. Further, a positive (+), or negative (-) grouping was given to each Teaching Presence and Cognitive Presence indicators to show whether the response was a positive descriptor, or a negative descriptor (Example: Direct Instruction, Direct Instruction-Positive, Direct Instruction-Negative). Finally, for each Teaching Presence and Cognitive Presence indicator, coding guidelines and an example set of responses were developed to provide coders with keywords and phrases that directly link to each indicator, Table 2.

Table 2. Example of Participant Responses, CoI Indicators, and Coding Technique

Response		**Design & Organization**		**Facilitation**	
Positive Comment	Negative Comment	Design & Organization (Negative)	Design & Organization (Positive)	Facilitation (Negative)	Facilitation (Positive)
Professor Smith's discussion board questions and her contribution to the DB made the course enjoyable. (Facilitation-positive)	One thing that she changed during our course was giving more paper topic options at the beginning of the week it was due. This threw me off a bit because I was prepared for one topic and then in mid-stream changed topics.	1 (Design & Organization-negative)	0	0	1

Data from a two month sample were collected and entered into the IBM/SPSS Text Analytics for Surveys model. Accuracy of 80% was established with a sample of 100 records by comparing results from the text analysis model to the results of data that were analyzed using traditional qualitative coding methods. It is important to note that the model will never reach 100% accuracy due to inconsistent or sarcastic phrasing. For example, the response "I don't think anything was exceptional during this class" may be flagged as positive, or "That professor was so dynamic – as dynamic as an old shoe." may be flagged as positive as the terms "exceptional" and "dynamic" would be considered positive terms and thus coded incorrectly.

Results of the pilot study not only informed further refinement of the model prior to the current study, it provided baseline criteria for data selection using the Text Analytics for Surveys Model.

II. DATA AND METHODS

2.1. Inclusion Criteria

Archival end-of-course survey records were purposively selected for text analysis inclusion based upon the following criteria: (1) the course must be identified as 3 or 4 factor course; (2) the survey must contain responses with comments; (3) at least one of the categories (such as Exploration) must have a mean average score that fell below 3 (on a scale of 1-5). The data included qualitative responses from the CoI-based end-of-curse survey from 134 undergraduate and graduate level students who were enrolled in a course during the (put season and year here). The American Public University System is a fully online university located in the northeastern United States with a total student population of over 100,000.

2.2. Procedures

Data were loaded into the IBM/SPSS Text Analytics for Surveys model and further exported into an Excel spreadsheet. Keyword text matching was used on a first pass analysis. This yielded numerous matches and mismatches. Human coders sorted through 200 random matches and noted areas for inclusion of additional keywords. 200 random non-matches were also reviewed and additional keywords and keyword strings were added to the Text Analytics thesaurus. This procedure was repeated for two more iterative cycles. After the second iteration, coders reviewed 428 samples and agreed with the Text Analytics matches 80.1% of the time. A second round of validation was conducted with 134 random samples and an 82% match was found.

III. DATA AND METHODS

Though much work remains to be completed, the use of Text Analytics to inform quantitative survey analyses appears to be feasible. Given APUS's extremely large volume of student surveys, examining qualitative data by hand is simply not feasible. However, such information is much needed as quantitative data only provides sentiment and correlational data; it does not help inform causation. For this researchers require the rich data that is included in qualitative comments. Having validated the text analytics procedure, the APUS data team is currently conducting further key word correlations. As an example, student scores may indicated that issues exist around exploration phase of cognitive presence, providing researchers an area to look for issues related to efficacy. From the text analytics data an extraction may be performed that reveals students are having issues with being able to access digital textbooks and related, reader-based resources. Further inspection of the text analytics results may reveal that this is related specifically to certain courses. This gives the data team to needed information to review those course structures and work with instructional designers to improve the user experience through design modifications. Using the same example it may be possible to trace the problems with digital assets to low integration scores. Such data thus helps inform causality and provide solutions for course developers. While still in the early stages APUS believes that this project has yielded a wealth of data and intends to expand the use of this technology to improve the online learning experience. Notably, this should not be construed as being a comprehensive solution, but part of the overall ecosystem described.

References

[1] Council for Higher Education Accreditation. (2010c). Recognition policy and procedures. Washington, DC: Council for Higher Education Accreditation.

[2] Eaton, J. (2010). Accreditation and the federal future of higher education. Academe, 96(5), 21-24. [Online] Retrieved from http://www.aaup.org/AAUP/pubsres/academe/2010/SO/feat/eato.htm

[3] Clark, (2011). Blog http://donaldclarkplanb.blogspot.com/2011/11/educations-slow-learner-lessons-from.html

[4] Lowe J. (1998). Jack Welch speaks. New York: Wiley; 1998.

[5] Meacham, J. A., & Gaff, Jerry, G. (2006). Learning goals in mission statements: implications for educational leadership. Liberal Education, 92(1), 6-13.

[6] Rozycki, E. G. (2004). Mission and vision in education. Educational Horizons, 82, pp. 94-98.

[7] Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing environment. Journal of Asynchronous Learning Networks, 5 (2).

[8] Dewey, J. How we think. Boston: D.C. Health, 1933.

[9] Garrison, D. R. (2009). Communities of inquiry in online learning: Social, teaching and cognitive presence. In C. Howard et al. (Eds.), Encyclopedia of distance and online learning (2nd ed., pp. 352-355). Hershey, PA: IGI Global.

[10] Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. The Internet and Higher Education, 2(2-3), 87-105

[11] Arbaugh, J.B., Cleveland-Innes, M., Diaz, S.R., Garrison, D.R., Ice, P., Richardson, & Swan, K.P. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. The Internet and higher Education, 11(3-4), 133-136.