Community College Journal of Research and Practice, 00: 1–16, 2014

ISSN: 1066-8926 print/1521-0413 online
DOI: 10.1080/10668926.2012.720865

Routledge
Taylor & Francis Group

In Their Own Words: A Text Analytics Investigation of College Course Attrition

Greg V. Michalski

Office of Student Analytics and Research, Florida State College at Jacksonville, Jacksonville, Florida, USA

Excessive course attrition is costly to both the student and the institution. While most institutions have systems to quantify and report the numbers, far less attention is typically paid to each student's reason(s) for withdrawal. In this case study, text analytics was used to analyze a large set of open-ended written comments in which students explained their reason(s) for course withdrawal in their own words. The text for all comments was extracted verbatim from the course withdrawals database of Florida State College at Jacksonville, a large, diverse, multicampus institution located in northeast Florida. An initial set of 616 comment records from the beginning of the fall 2010 term was used to develop a preliminary text analytics model. This model revealed 11 major category nodes and successfully classified 96.1% of all withdrawal records into one or more categories. The model was retained and further tested using a second set of 679 records from the spring 2011 term and found to successfully classify 98.7% of the spring records. At the broadest level, withdrawal explanations were found to include both academic and nonacademic student rationales. Leading academic rationales involve course scheduling adjustments, delivery method preference changes (e.g., classroom vs. online), and faculty related reasons. Leading nonacademic rationales include personal issues especially involving job/work, family, financial, and health matters. The limitations of the study along with implications for practice, administrative decision making, and future directions for the expanded use of text analytics in institutional research are discussed.

The purpose of this case study is to examine the question of why students withdraw from college courses using text analytics data consisting of verbatim student course withdrawal explanations. Reflecting the larger United States college graduation crisis (Complete College America, 2011), the problem of students withdrawing from college courses in which they originally enrolled with the intent of successfully completing is often described in dollar terms for both the student and the institution. Specific costs include time to degree completion, physical resource allocation and planning (e.g., available classroom/facilities space), and redundancies associated with the entire reenrollment process—assuming it occurs at all. According to the Florida Department of Education (2011), in the state college system alone between 2007–2008 and 2009–2010, there were a total of 668,854 course withdrawals, representing 11.3% of the total enrollments. That report summarized the problem as follows:

The author would like to acknowledge and thank Dr. Kathleen Ciez-Volz, PhD, for her manuscript reading and suggestions, and Ms. Karen Stearns, MPS, for her technical and practical assistance.

Address correspondence to Greg V. Michalski, Office of Student Analytics and Research, Florida State College at Jacksonville, 501 W. State Street, Jacksonville, FL 32202. E-mail: gmichals@fscj.edu

> When students enroll in, but fail to complete a course, it costs the student and the state money, reduces available classroom space, and increases the amount of time for the student to complete their degree. Clearly, many withdrawals are necessary for personal and academic reasons, but when withdrawals become excessive they pose a significant burden on the student, the college, and the state. (p. 1)

Renewed efforts to tie state funding to various performance indicators, including course and degree completion, have also raised the stakes across the nation. In reviewing current literature and policy on the impacts of state performance funding systems on higher education, Dougherty & Reddy (2011) focused on Florida, Tennessee, and Washington; but they also noted Indiana, Arkansas, Colorado, Illinois, Massachusetts, Missouri, North Carolina, Ohio, and South Carolina. At the state level, so called performance funding (PF) programs emphasize both ultimate outcome indicators (e.g., graduations); through bonuses (PF 1.0); and intermediate achievement indicators (e.g., course completions) through directly embedded funding (PF 2.0) to colleges. Summarizing this work, Hermes (2012) further noted that 17 states currently implement (or are considering) performance-based funding measures.

While most postsecondary institutions track and report aggregate course withdrawal numbers, far fewer have processes in place to efficiently track and analyze individual student course withdrawal reasons. This gap may be related to the lack of published research on the use of standardized methods and applications, such as the use of text analytics, available for this purpose. Although traditional quantitative descriptive reports may answer questions regarding the who, what, when, where, and how of withdrawal based upon student demographics and common academic system data, the use of text analytics illuminates the actual reason(s) behind each withdrawal—effectively answering the question of why a student withdrew. Such information is valuable for administrative decision making affecting actions to decrease withdrawal rates or realize other quality enhancements associated with student learning and program completion. The following review focuses on empirical research in the area of college course withdrawal with emphasis on the problem in community colleges and four-year (nongraduate) institutions. Particular studies were selected on the basis of their explicit description of student course withdrawal reasons and the methodological means used to obtain such.

COLLEGE COURSE ATTRITION

The literature on college course attrition (course withdrawal) is related to, but distinct from, that concerning the complete withdrawal of the student from college. Some authors distinguish the two by referring to the former as *course attrition* and the latter simply as *attrition* (Conklin, 1997). While perhaps constituting a more positive outcome, compared to complete withdrawal from college, individual course withdrawal is problematic in its own right. Compared to the broader area of student retention in higher education, which has been widely studied and now comprises an extensive body of emerging theory and research spanning at least four decades (see Tinto, 2006), the set of empirical studies focusing on selective or discretionary student withdrawal from individual courses is relatively less developed although it has received some recent attention (Buglear, 2009; Fincher, 2010; Hagedorn, Maxwell, Cypers, Hye, & Lester, 2007).

Prior empirical work suggests that student rationales for withdrawing from college courses can be considered along a continuum based on reasons ranging from the purely logical and/or clearly necessary at one end (e.g., wrong course), to the creatively justified and/or tenuously fanciful at the other (e.g., didn't like the classroom). Much of the research done in the area is based upon student surveys and questionnaires. These often provide a fixed set of withdrawal reason choices and in some cases limited open-ended comment sections to accommodate more precise explanations. In any case, data consist of what students say and as such there are possible validity concerns regarding respondent truthfulness regarding their genuine course withdrawal reasons. While such concerns tend to be recognized, highlighted, and even formally addressed in studies involving more personally or socially sensitive areas (e.g., mental health and substance abuse), this is not the case in the area of student course withdrawal, perhaps due to relative lower-stakes perceptions from the student's perspective regarding the topic area.

Taken as a whole, most withdrawal reasons can be classified into either of the following two broad categories: (a) largely academic reasons, related to areas such as grades, instructors, and the course, or (b) nonacademic reasons related to areas such as job/work, family, military commitment, and illness (Astin, 1997; Dunwoody & Frank, 1995; Hope, 2009; Scoggin & Styron, 2006; Thompson, 1969). Although published work focused specifically on course attrition can also be traced back at least four decades, the body of peer reviewed work in the area is especially scarce; and peer-reviewed work performed in the context of the two-year postsecondary institution is even scarcer. Nevertheless, various researchers have looked at the problem and documented their findings.

For example, in a study conducted at Macomb Community College to determine reasons that its students withdrew from classes, Thompson (1969) mailed 3,568 questionnaires to students who had dropped a total of 6,081 courses. Based on a response rate of just over 40%, (N = 1,434 responding students), he found the primary academically related course withdrawal reasons to include wrong program, academic difficulty, and conflict with the teacher. Nonacademic reasons included job conflict, lack of interest, financial problems, and military responsibilities. Since then, other researchers have documented similar academic and nonacademic student rationales.

In researching why students withdrew from classes at the University of Washington, Lunneborg, Lunneborg, & de Wolf (1974) found main academic reasons to include reducing course load, disappointment with the class, instructor, or probable grade. Nonacademic rationales identified by the researchers included personal reasons and conflict with job or other activities. In another university-based case study, Reed (1981) performed survey research and identified three factors associated with course withdrawal. These factors involve student satisfaction with course performance, student motivation, and student impressions of the instructor. The probability of students persisting in a course was viewed as a function of their confidence that they could handle the course material, perceptions of the relevance of the course, and impressions of how likable/helpful the instructor was.

Focusing primarily on course attrition at community colleges, Friedlander (1980, 1981) reviewed a range of empirical work and found support for both academic and nonacademic course withdrawal reasons. He listed the seven most frequently cited reasons (in descending order) as follows: (a) job conflict, (b) inadequate preparation for the course, (c) dislike of the class, (d) assignments too heavy, (e) indefinite motivation, (f) illness, and (g) dislike of instruction. He further discussed these reasons in terms of the degree of control the institution has to correct or improve course attrition attributable to specific causes. For example, "the major reason for withdrawing from a course (job conflict) is one over which the college does not have much control"

(Friedlander, 1981, p. 3). As discussed later, however, the institution may still have the ability to ameliorate even withdrawals that may ostensibly appear outside of its initial control.

Further acknowledging the relative dearth of work in the area of course withdrawal, Dunwoody and Frank (1995) found empirical support for course attrition in terms of academic and nonacademic reasons. They surveyed both faculty ($n = 30$) and students ($n = 151$) at a four-year state college and performed factor analysis to identify two main reasons why students withdrew from courses. They found the reasons as involving either personal considerations independent of the course and instructor, or course considerations primarily concerned with understanding the material and liking the course. Factor 1 examples include "I ran out of money and needed to work" and "family problems" (p. 555). Factor 2 examples specifically referenced (a) grade happiness, (b) understanding material, (c) liking the course, (d) liking the professor, and (e) uninteresting subject matter (p. 556).

Other researchers have also noted the lack of empirical work in the area while contributing results of their own. Studies include Miller's (1997) logistic regression study of variables affecting student decisions to withdraw from liberal arts and science courses at community colleges; Conklin's (1997) five-year perspective on student course withdrawal in which she identified the top three reasons as being nonacademically based involving work-schedule conflicts, inconvenient time, and personal problems; and Sumner's (2000, 2001) survey research identifying work schedule and personal problems as the top two course withdrawal reasons cited by students. Aldridge and Rowley (2001) also conducted a college withdrawal survey using semistructured interviews and found both academic and nonacademic reasons including course not as expected, travel difficulties, institution not as expected, domestic difficulties, and financial difficulties.

From a research perspective, there is evidence of a general broadening of the methodological spectrum used to describe, understand, and even predict course withdrawals (see Bambara, Harbour, Davies, & Athey, 2009; Buglear, 2009; Charlton, Barrow, & Hornby-Atkinson, 2006). However, most studies, especially those involving large institutions, continue to rely heavily (or exclusively) upon traditional quantitative/statistical measures, survey research, and more recently, the implementation of business intelligence and data mining processes (Hope, 2009). Such efforts generally involve counting and comparing course withdrawal frequencies in relation to various continuous or categorical dimensions such as time (e.g., term/semester); course/credit type; student demographics; student major; and others (see Hagedorn, Maxwell, Cypers, Moon, & Lester, 2003; Hall, Smith, Boeckman, Ramachandra, & Jasin, 2003; Mery, 2001). While enumerative comparisons are certainly useful, the use of text analytics represents an opportunity to further contextualize and/or complement such results.

TEXT ANALYTICS/MINING

With the steady rise and continued growth of predictive analytics and data mining in the field of education, most emphasis has been on classical mining techniques and procedures involving structured sources (Romero, Ventura, Pechenizkiy, & Baker, 2011). Yet credible estimates indicate that well over half of the information in most organizations is stored in various unstructured forms, especially those involving text-based sources and documents (Nisbet, Elder, Elder, & Miner, 2009). Text analytics refers to automated processes for analyzing unstructured, text-based data from a wide range of sources including document transcriptions/repositories, open-ended surveys, websites, blogs, wikis, databases, and other electronic sources. While technical distinctions can be made between processes involving text analytics and text mining (e.g., based on qualitative vs. quantitative analytical or postprocessing approaches), the two concepts are used interchangeably here.

Text mining is also considered a form of qualitative analysis involving the discovery of new, previously unknown information extracted and organized from different written sources. In brief, text mining involves the discovery of useful and previously unknown gems of information from textual document repositories based upon patterns extracted from natural language (Zhang & Segall, 2010). Currently a topic of considerable importance in academia and industry, text mining theory and practice have also benefited from increased multidisciplinary interest spanning the public and private sectors involving multinational government, business and industry, and university research (Berry & Kogan, 2010). The field of text analytics has also been recently proposed as a standard graduate business school topic (Edgington, 2011).

Text mining is generally considered to be part of the broader field of data mining, and it involves the discovery of novel information such as associations, hypotheses, or trends that are not explicitly present in the text sources being analyzed (Nisbet et al., 2009). The field of text mining, and the many applications now available to engage in such, has evolved rapidly over the past two decades and is tied closely to the concurrent growth of foundational technologies in areas that include computer science, artificial intelligence, and machine learning among others. Two broad text categorization approaches include (a) the knowledge engineering approach in which expert knowledge about categories is encoded directly and (b) machine learning in which a category is constructed from a set of existing examples according to a general inductive process (Feldman & Sanger, 2007). While distinctions are made between purely statistical approaches and those based upon artificial intelligence, a particularly promising direction of development is based upon Natural Language Processing (NLP), which is rooted in the realm of machine learning traceable back to the work of Turing (1950).

The current study employs a NLP-based application that references techniques used to develop a text analysis and knowledge mining system described by Nasukawa and Nagano (2001). Major process steps involve the extraction of key terms, patterns, and concepts and the organization of these into a number of labeled categories. Extraction and categorization are done according to defined analysis objectives using default and/or customized linguistic resource libraries consisting of candidate terms, synonyms, types, patterns, and concepts. Although a detailed description of the entire process is beyond the present scope, the main steps in the overall text analytics process include textual data input, cleansing, and standardization; candidate term and equivalence class identification; synonym definition/integration; type assignment and indexing; relationship pattern matching/extraction; category building; and finally, individual record categorization within the model. A major advantage of this approach is model reusability—specifically, the ability to replicate results with different data sets (e.g., course withdrawal comments from different academic terms) using exactly the same rules of analysis and categorization.

CONTEXT AND METHODOLOGY

The context for the study was Florida State College at Jacksonville, a large multicampus institution with an annual (2009–2010) unduplicated student enrollment of over 84,000. According to the Florida Department of Education (2011), the 28-college system as a whole had an average course withdrawal rate (withdrawals as a percentage of total course enrollments 2007–2008 through 2009–2010) of 11.3% with a range of 6.3% to 15.5%. With a withdrawal rate of 8.1% based on 364,179 enrollments and 29,655 withdrawals, Florida State College at Jacksonville had a better (i.e., lower) than average course withdrawal rate compared to the system average.

Beginning in the fall 2009 term, students withdrawing from courses at Florida State College were provided an opportunity to explain the reason(s) for withdrawal. As part of the course withdrawal process, the student encounters the following request: Please provide your reason for requesting a withdrawal. The student is then provided with a text entry area to enter a response. These open-ended (textual) explanations are collected in a Course Withdrawal Database (CWDB) along with associated information (such as a unique withdrawal identification number, course reference number, student identification number, and withdrawal submission date). Including all recorded data back to inception, the CWDB now contains well over 10,000 records and was used to provide the raw textual data used in the current study.

The current research project design involved two phases that included (a) model construction and (b) model validation. Student course withdrawal comments from the beginning of the fall 2010 academic term were used to build the text analytics model. The model was further tested and validated using an equivalent set of comments taken from the beginning of the spring 2011 term. Student comments for course withdrawals that occurred between September 1 and September 26, 2010 were used for model construction. This period corresponds to the first three full weeks of the fall 2010 term. Model validation was accomplished using an equivalent set of comments from the spring 2011 term for withdrawals that occurred between January 19 and February 6, 2011.

All comments were taken directly from the CWDB using Microsoft SQL Server 2008 Management Studio® and then imported to Microsoft Excel® for cleansing and organization. The cleansing process consisted of performing a descending alphabetical sort to identify and delete noncomment entries. Deleted entry examples include those in which a student simply typed a random character or entered other nonresponse character combinations such as n/a, no, none, or other random (nonsensical) key combinations. Common terms and abbreviations used at the college were standardized. These included commonly used abbreviations such as FSCJ to refer to the college name, bb or BB to refer to Blackboard (an academic learning content and management system), and others. Finally, a spell check was performed using Microsoft Excel®. The organized and cleansed set of comments in the fall term data set consisted of 616 comments and included several associated reference variables (i.e., course, gender, etc.) for each record. This data set was then imported into IBM/PASW Text Analytics for Surveys® (v. 3.0.1) for mining and analysis using default library resources for initial extraction and analysis.

To enable comparisons among final model categories, a preliminary set of seven reference variables was also defined and extracted in phase one of the project. The reference variables included (a) campus/location, (b) class time block, (c) course credit type, (d) course identifier, (e) student gender, (f) student race, and (g) instructor name. Of these, only results based on the course identifier are discussed in the present paper. The course identifier was used to check and compare the proportionality of courses in the withdrawal comment data set against that of all courses withdrawn from in the entire fall 2010 term (i.e., all academic history W grades for the term). The remaining reference variables were initially explored for representativeness across demographic groups but not used for analysis in the current study. Finally, although a discussion of the results of such are beyond the scope of the current paper, coded results from both terms were also exported and further analyzed using a range of quantitative methods including correlation matrix analysis, cluster analysis, principal components analysis, and multiple correspondence analysis.

RESULTS

Initial extraction of terms and categorization was performed using default text analytics application library resources for course withdrawal explanations provided by students in the fall 2010 academic term. Following the iterative cycle of model building typical of text analytics projects, several initial models were produced, examined, and refined. The refinement process encompassed a range of activities including the definition of synonymous terms (e.g., faculty, instructor, teacher, professor); identification of context specific terms (e.g., campus location identifiers); and exclusion from extraction of terms not relevant to the study. A final model consisting of 11 distinct withdrawal explanation categories (also referred to as model nodes) emerged. A working label for each model category was produced based on commonalities among the explanations coded into the category and with reference to the empirical withdrawal literature.

Final model categories were labeled; (a) time-schedule, (b) job-work, (c) family, (d) health, (e) financial, (f) personal-other, (g) information technology, (h) faculty negative, (i) course negative, (j) online course, and (k) federal service. Figure 1 is a view of the category web model produced in project phase one. As an efficient way to represent complex coding relationships

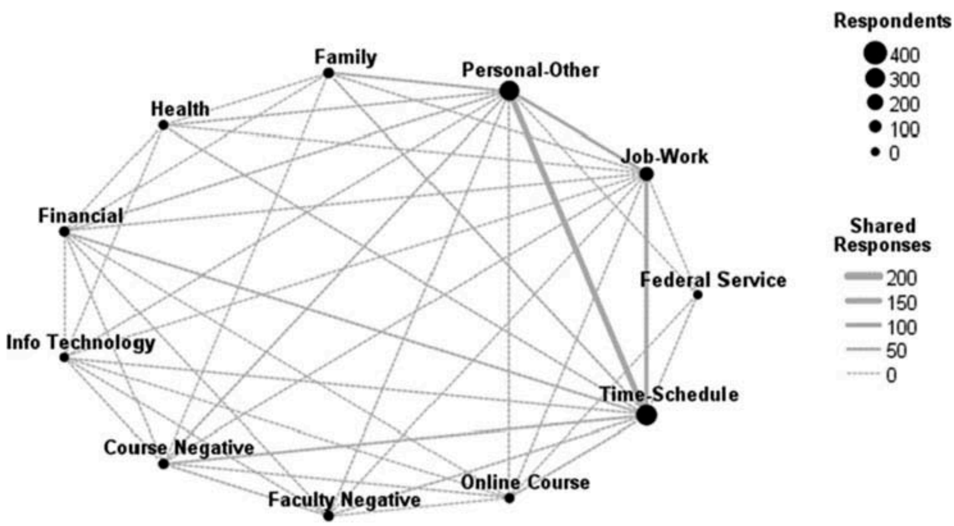

FIGURE 1 Category web model showing eleven labeled nodes. The model uses relative circle diameter to represent the number of student withdrawal comments coded into each node, and relative line thickness to represent the number of comments shared between nodes. The phase one model categorized 96.1% of student course withdrawal explanations from the fall 2010 academic term.

in the text data, this model visually depicts the categorization of 592 of the 616 (96.1%) phase one student responses. The model uses relative circle diameter to represent the number of cases (responses) coded into each node and relative line thickness to represent the number of responses shared between nodes. As shown, the top three categories accounting for the largest percentage of explanations are labeled time-schedule, personal-other, and job-work. These three categories also contain the largest percentages of shared responses in which more than one reason is mentioned in a single withdrawal explanation (see Figure 1).

The exact number of withdrawal comments coded into each category and shared between categories is shown in Table 1. The top row of the table shows that 331 withdrawal comments were coded into the Time-Schedule category (node), and 301 records were coded into the Personal-Other category. A total of 151 student comment (records) contained both schedule and personal course withdrawal reasons and were coded into (shared between) both model nodes. To enable additional and future comparisons, seven reference variables were also defined. These included both academic/institutional and student demographic descriptors with specific codes for (a) campus/location, (b) class time block, (c) course credit type, (d) course identifier, (e) student gender, (f) student race/ethnicity, and (g) instructor name (see Table 1). While several preliminary comparisons were made using the reference variables, a discussion of those results is beyond the scope of the present study, and only the course identifier is discussed here. The course identifier was used to validate the phase one model by comparing the course withdrawal proportions in the comment data set against that of all course withdrawals from in the fall 2010 term.

Course Withdrawal Proportionality Comparisons

Because the phase one model was developed using a sample of comments taken from the beginning of the fall 2010 term, the course proportions represented in the sample were compared to those of all course withdrawals for the entire academic term. The course identifier was used to make these comparisons. The results were favorable with the same subset of high withdrawal courses represented in both the sample and the full term. As shown in Figure 2, the results are positively and significantly correlated (Pearson's $r = 0.84$, $p < .01$), and the same set of high withdrawal courses was represented in both the text analytics sample and the full term academic term (see Figure 2).

Model Testing and Validation

Using the same text analytics library resources and extraction settings, the phase one model was further tested and validated using an equivalent set of withdrawal comments taken from the beginning of the spring 2011 academic term. Two key objectives of this validation phase involved (a) comparing the total proportion of records categorized by the initial model (using fall term data) with data from an equivalent period in a different academic term and (b) comparing the relative proportions of comments coded into each category by academic term as well as overall for both terms combined.

To perform the validation, a second sample of 679 comments, taken from the beginning of the spring term, was independently introduced to the phase one model. Of the 679 spring term withdrawal comments, 670 records (98.7%) were successfully coded by the model. The number and relative proportion of withdrawal explanations coded into each of the 11 model categories

TABLE 1
Withdrawal Comment Record Counts Within and Shared Between Category Nodes

Category 1		*Category 2*		
Node title	*Record count*	*Node title*	*Record count*	*Shared records (both categories)*
Personal-Other	301	Time-Schedule	331	151
Job-Work	146	Time-Schedule	331	79
Job-Work	146	Personal-Other	301	38
Course Negative	43	Time-Schedule	331	32
Financial	43	Time-Schedule	331	28
Time-Schedule	331	Family	54	21
Online Course	34	Time-Schedule	331	20
Course Negative	43	Personal-Other	301	16
Personal-Other	301	Family	54	16
Time-Schedule	331	Faculty Negative	48	16
Time-Schedule	331	Health	30	14
Financial	43	Personal-Other	301	13
Info Technology	14	Time-Schedule	331	11
Online Course	34	Personal-Other	301	11
Personal-Other	301	Health	30	10
Faculty Negative	48	Course Negative	43	8
Federal Service	11	Time-Schedule	331	7
Job-Work	146	Course Negative	43	7
Course Negative	43	Financial	43	6
Faculty Negative	48	Info Technology	14	6
Faculty Negative	48	Financial	43	6
Info Technology	14	Personal-Other	301	6
Online Course	34	Course Negative	43	6
Info Technology	14	Course Negative	43	5
Job-Work	146	Financial	43	5
Online Course	34	Job-Work	146	5
Faculty Negative	48	Personal-Other	301	4
Info Technology	14	Financial	43	4
Job-Work	146	Family	54	4
Job-Work	146	Info Technology	14	4
Job-Work	146	Faculty Negative	48	4
Online Course	34	Info Technology	14	3
Online Course	34	Financial	43	3
Family	54	Health	30	2
Financial	43	Family	54	2
Job-Work	146	Health	30	2
Family	54	Course Negative	43	1
Financial	43	Health	30	1
Info Technology	14	Health	30	1
Job-Work	146	Federal Service	11	1
Online Course	34	Federal Service	11	1

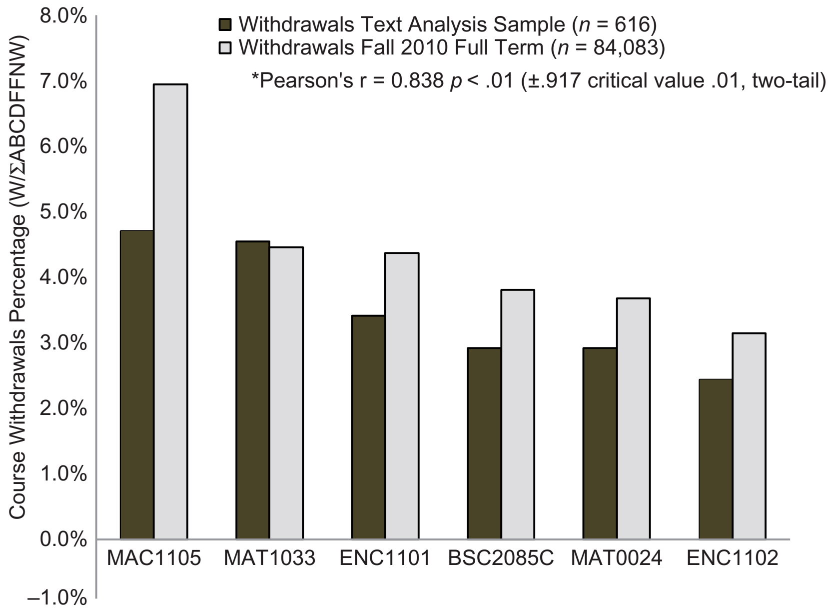

FIGURE 2 Withdrawal percentage comparisons between phase one sample ($n = 616$) and full fall 2010 term ($n = 84{,}083$). The same subset of six courses was present in both the phase one sample and full fall term data.

was very similar for both academic terms with the same order and relative proportions being represented in each model category.

For both terms combined, the exact number of comments coded into each category, as well as the number of comments shared between categories, is displayed in Table 2. Using all data from both terms ($N = 1{,}295$), this table displays the total number of comments in each model node as the numbers printed diagonally, and the number of comments shared between categories as numbers at the row-column intersections. As shown, the time-schedule category contains the highest number of comments (680), followed by personal-other (665), and job-work (301). There were 322 comments coded into both the time-schedule and personal-other category, while 178 comments were coded into both the time-schedule and job-work category. Similarly, the number of comments contained within any single node, as well as the number of comments shared between any two model nodes, can be easily viewed using the frequency counts displayed in the table (see Table 2). Next, several student withdrawal comments are presented to illustrate how comments were coded in the text analytics model.

Comment Coding Examples

Taken as a whole, the complete set of student comments includes a wide breadth and diversity of withdrawal explanations ranging from single word explanations (e.g., work) to complex, multi-theme, detailed descriptions of paragraph length (or greater). As such, the text analytics model

TABLE 2
Withdrawal Explanation Counts by Model Node Category for Fall 2010 and Spring 2011 Combined

Category node	*Time-schedule*	*Personal-other*	*Job-work*	*Family*	*Course negative*	*Faculty negative*	*Financial*	*Online course*	*Health*	*Info tech*	*Federal service*
Time-Schedule	680	–	–	–	–	–	–	–	–	–	–
Personal-Other	322	665	–	–	–	–	–	–	–	–	–
Job-Work	178	123	301	–	–	–	–	–	–	–	–
Family	44	55	29	118	–	–	–	–	–	–	–
Course Negative	74	48	15	4	101	–	–	–	–	–	–
Faculty Negative	32	28	9	0	29	95	–	–	–	–	–
Financial	49	40	20	6	9	8	95	–	–	–	–
Online Course	48	40	16	2	22	7	6	82	–	–	–
Health	24	24	6	10	1	2	2	1	59	–	–
Info Technology	18	13	7	2	8	6	4	7	1	24	–
Federal Service	7	2	1	0	0	0	1	1	0	0	19

Note. Comment frequency is for both terms combined ($n = 1{,}295$).

may code a comment into one or more categories based on both the coding parameters of the model and content of the comment. In this section, several examples of verbatim comments are provided to illustrate how comments were coded in the text analytics model. In the following examples, the student comment is presented followed in brackets by the category/node name(s) into which it was coded within the model:

- Don't have the time needed to complete this class with my current job. [time-schedule, job-work]
- No time. Work. [time-schedule, job-work]
- I just can't find the time due to the fact that so busy at work. I cannot apply myself as needed. I will try to work it out before next semester. [time-schedule, job-work, personal-other]
- The schedule times interfere with my daughters' daycare. The class ends at 6 and her daycare closes at 6. [family, time-schedule]
- We've recently been forced to deal with an estate issue on her father's behalf and have realized that a full schedule is too much while taking care of our special needs three year old. [family, time-schedule, personal-other]
- Due to newly received medical treatment on this day I was advised by my doctor to reschedule this course. I was advised that if I don't make changes to my schedule it may affect my treatment. I did not know at the beginning of class that I would have to receive treatment on this day. [health, time-schedule, personal-other]
- I'm requesting a withdrawal because although I'm enrolled in this course now, I don't have the money to pay for my books right now. So I was hoping to drop this course and re-register in a later dated course so that my financial aid will pay the expenses. [financial, time-schedule]
- I am withdrawing because my class wasn't paid for at the time I registered. [financial]
- (I) lost internet connection for approximately 2 weeks during the beginning of the semester (and) missed an important assignment that would not allow me to continue. [info technology, personal-other]
- I would rather take the class in a classroom. I don't like taking the class by computer. Too much information too fast and too many distractions at home. I will sign up for it again after I have taken some math classes. [info technology, personal-other, course negative, online course]
- Better instructor. [faculty negative]
- Don't like the class. [course negative]
- The class is boring and not engaging. [course negative]
- Not enough time, boring class. [course negative, time-schedule]
- I think I need to do this one in a classroom atmosphere. I am worried about it being online. [online course]
- Deploying to Iraq soon. [federal service]
- Going Active Army. Cannot Move forward in Class. [federal service]
- I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well. [federal service, time-schedule, online course, job-work]

The federal service category was included to account for all military as well as other federal service commitments such as Homeland Security. As illustrated in the last example, how any given comment is coded is a direct result of the word combinations, context, and phrases used in the response. The text analytics application allows for sufficient parameter adjustment, iterative development, and library refinement to construct models that can be used (and reused) to accurately code and categorize a broad range of possible textual responses.

DISCUSSION

The results are best understood in terms of the purpose and institutional context of the study. Of the 11 attrition categories identified in the text analytics model, the top four involve nonacademic reasons related to time-schedule, personal-other, job-work, and family. These are followed by a combination of both academically and nonacademically related withdrawal reasons involving negative views of the course and/or faculty, financial issues, online course issues, health, information technology (info technology), and federal service commitments.

Although the college now has a growing number of four-year degree programs, it has operated as a traditional, urban, two-year institution throughout most of its history. Key student demographics also play a part in explaining the results. As of the fall 2010 academic term, the median student age for all students was over 25 years, 73% of all students were part-time, and 75% worked at least part-time. For students who completed a program (summer 2010 through spring 2011), the median age is over 27 years (and the mean > 30 years). With older, part-time, working students comprising a majority at the college, nonacademic reasons for course withdrawals are not surprising and have been discussed in the literature. For example, in citing the work of Tinto (1993), Charlton et al. (2006) indicated that student levels of responsibility associated with age, maturity, marital status, and general family commitments can play a substantive role in student withdrawal:

> Older students are likely to differ from younger students in a number of respects. They are more likely to be married, have children and be based at home, and will therefore typically have more demands on their time resulting in lesser social integration with other students, greater problems in obtaining academic support, and less study time. If commitment to their studies is low, these external pressures can make them particularly prone to withdraw. (p. 35)

The comment coding examples (stated earlier) provide only a glimpse into the realities faced by such students. An examination of the complete set of comments reveals that such issues are often complex and interrelated, as illustrated by the number of comments coded into more than one model category.

While comments coded into the course negative and faculty negative categories as well as those in the financial and health categories are typical and representative of those discussed in the literature summarized earlier (see Aldridge & Rowley, 2001; Dunwoody & Frank, 1995; Friedlander, 1981; Lunneborg et al., 1974; Thompson, 1969), a particularly interesting outcome in the present study involves many of the comments coded into the online course category. As illustrated in the comment coding examples, many of these comments express the desire to withdraw from an online course and take the same course in a traditional classroom environment. This result may also reflect the older (adult) student demographic as students describe the need to focus on the course away from the distractions of home. The remaining category, federal service, includes a range of comments, many of which are related to military commitments from all branches.

LIMITATIONS

The lack of comparable, published work represents both a limitation and a challenge. Although it would be ideal to compare results from several demographically similar institutions using the same methodological approach, the iterative and adaptive nature of model building and optimization using text analytics limits such direct comparison. Fortunately, there are options to address this challenge. One is a common model approach in which the same model is used on textual course withdrawal data from different institutions. Another option is widened cooperation and collaboration across institutional boundaries. This option would require a concerted effort among collaborating institutions to define, standardize, share, and use a common set of text analytics linguistic resources such as common terms, concepts, types, rules, patterns, templates, and libraries. In any case, increased collaboration and ongoing research are recommended. While there may be differences between institutions based on a range of characteristics (e.g., public vs. private, large vs. small, and urban vs. rural), it seems reasonable to expect that there is enough commonality to allow for expanded collaboration. The idea of expanded partnerships and data sharing has received recent attention and interest as highlighted at the 50th annual meeting of the Association for Institutional Research (2010). Text analytics research in the area of college course attrition (as well as other areas) would benefit from an increased collaborative focus.

IMPLICATIONS FOR PRACTICE AND CONCLUSIONS

If raising college completion and graduation rates is a priority, then even small or incremental changes that serve to reduce course attrition are important. The expanded use of text analytics represents an efficient option to empirically complement purely quantitative procedures to further understand and effectively respond to the problem.

While it may be easy enough to concede many (or even most) course withdrawals as inevitable, upon closer look the institution may indeed be in a position to mitigate the impact of even ostensibly nonacademic withdrawals. Possibilities include increasing course scheduling flexibility, expanding resources available to adult students (e.g., 24-hour drop-in study facilities, increased child care services), and improving counseling/advising availability, access, and quality. Systems-and process-based solutions also exist. For example, Florida State College has implemented an early alert system in which course withdrawal requests are held for 48-hour period to provide the instructor an opportunity for review of the request. Such systems, especially combined with expanded use of the incomplete (I) grade, can help because many institutions allow up to 12 months for a student to make up missed work for the satisfactory completion of courses.

The expanded application of text analytics technology represents a promising new direction to understand and respond to the problem of excessive course attrition. At the same time, there are several important factors for those considering such practice. Methodologically, potential practitioners should recognize that—especially as a dynamically evolving field—even the best examples of text analytics technology do not represent perfect or fully automatic solutions to account for all the nuanced complexities in students' academic and personal lives that may impact course withdrawal decisions. There are also practical considerations of implementation.

And although a full discussion of the benefits and challenges of data mining in general, and text mining in particular (see, e.g., McDonald & Kelly, 2012), is beyond the present scope, some guidance is in order.

Institutions considering the adoption of text analytics should examine their purpose and goals, especially in the context of existing measures used to support decision making. Scale and text volume are main considerations. When the amount of text is relatively small, text analytics may be unnecessary in light of the institution's ability to address problems such as course attrition individually via counseling/advising, for example. However, this approach may be impractical at very large institutions, particularly given the reality of ever shrinking resources. Organizationally, the data/text mining function can be positioned in a variety of ways, for example, within institutional research. Colleges with limited resources (such as those with one person shops) can grow text mining capability in collaboration with similarly interested institutions. Such collaborative partnerships would also serve to promote desirable synergism, addressing the limitations discussed while enabling more institutions to benefit. Text analytics technology offers promise as a means to more completely understand and respond to the problem of course attrition as explained by students in their own words.

REFERENCES

Aldridge, S., & Rowley, J. (2001). Conducting a withdrawal survey. *Quality in Higher Education*, *7*(1), 55–63.

Association for Institutional Research. (2010). *Expanding our partnerships, consortia and data sharing.* Chicago, IL: Association for Institutional Research Annual Forum.

Astin, A. (1997). *What matters in college?* San Francisco, CA: Jossey-Bass.

Bambara, C., Harbour, C., Davies, T., & Athey, S. (2009). The lived experience of community college students enrolled in high-risk online courses. *Community College Review*, *36*(3), 219–238.

Berry, M., & Kogan, J. (Eds.). (2010). *Text mining: Applications and theory*. Chichester, UK: John Wiley & Sons.

Buglear, J. (2009). Logging in and dropping out: Exploring student non-completion in higher education using electronic footprint analysis. *Journal of Further & Higher Education*, *33*(4), 381–393.

Charlton, J., Barrow, C., & Hornby-Atkinson, P. (2006). Attempting to predict withdrawal from higher education using demographic, psychological and educational measures. *Research in Post-Compulsory Education*, *11*(1), 31–47.

Complete College America. (2011). *Time is the enemy*. Washington, DC: Author. Retrieved from http://www.completecollege.org/docs/Time_Is_the_Enemy.pdf

Conklin, K. (1997). Course attrition: A 5-year perspective on why students drop classes. *Community College Journal of Research & Practice*, *21*(8), 753–759.

Dougherty, K. J., & Reddy, V. (2011). *The impacts of state performance funding systems on higher education institutions: Research literature review and policy recommendations* (CCRC Working Paper 37). New York, NY: Community College Research Center, Teachers College, Columbia University.

Dunwoody, P., & Frank, M. L. (1995). Why students withdraw from classes. *The Journal of Psychology*, *129*(5), 553–558.

Edgington, T. M. (2011). Introducing text analytics as a graduate business school course. *Journal of Information Technology Education: Innovations in Practice*, *10*(2011), 207–234.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. New York, NY: Cambridge University Press.

Fincher, M. (2010). Adult student retention: A practical approach to retention improvement through learning enhancement. *Journal of Continuing Higher Education*, *58*(1), 12–18.

Florida Department of Education. (2011, March). *Commonly withdrawn courses*. Tallahassee, FL: Division of Florida Colleges, Florida Department of Education.

Friedlander, J. (1980). An ERIC review: The problem of withdrawal from college credit courses. *Community College Review*, *8*(1), 57–66.

Friedlander, J. (1981). *Why students drop courses*. Los Angeles, CA: Junior College Resource Review.

Hagedorn, L., Maxwell, B., Cypers, S., Hye, S. M., & Lester, J. (2007). Course shopping in urban community colleges: An analysis of student drop and add activities. *Journal of Higher Education*, *78*(4), 464–485.

Hagedorn, L., Maxwell, W., Cypers, S., Moon, H., & Lester, J. (2003). *Course-shopping in the urban community colleges: An analysis of student drop and add activities*. Los Angeles, CA: Center for Higher Education Policy Analysis, University of Southern California, Los Angeles.

Hall, M., Smith, K., Boeckman, D., Ramachandra, V., & Jasin, J. (2003, October). *Why do students withdraw from courses?* Paper presented at the 2003 Southern Association for Institutional Research Conference, San Antonio, TX.

Hermes, J. (2012). States trend toward performance funding. *Community College Journal*, *82*(4), 26–29.

Hope, J. (Ed.). (2009). Withdrawal data helps solve retention issues. *Successful Registrar*, *9*(2), 6–7.

Lunneborg, C., Lunneborg, P., & de Wolf, V. (1974). *Why do UW students withdraw from classes?* Seattle, WA: Educational Assessment Center, University of Washington.

McDonald, D., & Kelly, U. (2012). *The value and benefit of text mining to UK further and higher education*. London, United Kingdom: Joint Information Systems Committee (JISC), Digital Infrastructure Directions Report.

Mery, P. (2001). *Students leaving before census*. San Francisco, CA: Office of Institutional Development, Research, and Planning, City College of San Francisco.

Miller, J. C. (1997). Variables affecting the decision to withdraw from liberal arts and science courses. *Community College Review*, *25*(3), 39.

Nasukawa, T., & Nagano, T. (2001). Text analysis and knowledge mining system. *IBM Systems Journal*, *40*(4), 976–984.

Nisbet, R., Elder, J., Elder, J. F., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Amsterdam, Netherlands: Academic Press/Elsevier.

Reed, J. G. (1981). Dropping a college course: Factors influencing students' withdrawal. *Journal of Educational Psychology*, *73*(3), 376–385.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2011). *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Scoggin, D., & Styron, R. A. (2006). Factors associated with student withdrawal from community college. *Community College Enterprise*, *12*(1), 111–124.

Sumner, P. (2000). *JCC drop study/attrition rates, fall 1999*. Overland Park, KS: Office of Institutional Research, Johnson County Community College.

Sumner, P. (2001). *Drop study/attrition rates, fall 2000*. Overland Park, KS: Office of Institutional Research, Johnson County Community College.

Thompson, J. R. (1969). *Why students drop courses*. Retrieved from ERIC database. (ED026994)

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Tinto, V. (2006). Research and practice of student retention: What next? *Journal College Student Retention*, *8*(1), 1–19.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, *49*(236), 433–460.

Zhang, Q., & Segall, R. (2010). Review of data, text and web mining software. *Kybernetes*, *39*(4), 625–655.