Utilizing Business Intelligence to Enhance Online Education at For-Profit and Non-Profit Institutions

Ehi Aimiuwu, Department of Information Sciences & Systems, Morgan State University, Baltimore, MD, USA

Sanjay Bapna, Department of Information Sciences & Systems, Morgan State University, Baltimore, MD, USA

ABSTRACT

With the increase in the demand of online education, especially for working people, there is a need for institutions that offer online education to identify, target, and market their services to people who are in need of furthering their education or who would be interested in using online education to advance their careers. The aim of this paper is to explore how Business Intelligence (BI) can be utilized to enhance online education for profit and non-profit organizations. The goal is to identify, target, and market online education to an audience who may not have thought of online education, but will appreciate the opportunity.

Keywords: Business Intelligence, Business Management, Education, E-Learning, For-Profit, Information Systems, Non-Profit, Online Education

INTRODUCTION

According to Onlineeducation.org (2011), the Internet has given educators a new way to communicate with students and a study by Sloan Consortium shows that online education enrollment increased by 13% annually compared to traditional classrooms, which grew at 1.2% annually. Also, the website shows that people interested in online education are full-time workers, those hospitalized, and those unable to attend traditional classrooms. Shaffer (2011) says that over 5.6 million students study online, a 19% increase since 2005, compared to a 1.8% increase for traditional education. The author also states that the online education graduation rate is 71%, and for-profit universities, such as University of Phoenix, will take the lead in the continuous growth of online education. So there is a high demand for online education in today's world, which is growing faster than traditional education.

Onlineeducation.org mentions five types of online education, which are adult, hybrid, online continuing, online distance, and online higher educations. Adult education is where adults are taught basic math, reading, science, and problem solving skills to combat illiteracy. Hybrid education is a mixture of online and traditional classroom learning. Online continuing education attends to students who already have degrees but want to take a course or two to further their skills. Online distance education is for students taking traditional classes but may also want to take some online courses, while online higher education is for students who want to get their entire degree online. Nevertheless, studies show that students who take both online and classroom instruction (i.e., hybrid classes) perform the best (Jaschik, 2009). So there are various forms of online education that may fit the need of various individuals who need further education to better their lives in this new global economy.

Brantley (2006) gives ten reasons why individuals would want to enroll in online education. They include: the flexibility to do assignments at any time and from any location (given variable family situations, one's inability to travel, and scheduling conflicts); saving money on gas one would utilize to go to school; working a decent full-time job while going to school; saving time between traditional classes and on driving back and forth between home and school; the ability to simultaneously take multiple courses that may have had been offered at conflicting times in the traditional classroom; not needing resources and classroom supplies because these are now electronically accessible; and engaging with more classmates through venues like the online discussion board. You can still be a parent and access class from a single location (e.g., from the comfort of your home), and do house work and other tasks, while attending any school from any location with adequate internet access. Shaffer (2011) adds other plus points of this platform. They include referrals from colleagues who have had a great online education experience, and those who travel a lot for work, as well as people who

might want to graduate early by taking more classes, even during summer breaks. She also suggests that online education is more suitable for people who wish to learn at their own pace, including working ahead, staying up all night, and studying in the early part of the morning. Onlineeducation.org mentions the value of online education for those who have been suspended from school, drop-outs, and those who are home-schooled. These individuals are very likely to be interested in online education. So there are people in society (e.g., prisoners, delinquents, and those on probation) whose lifestyle and/or condition make online education the most viable and accessible mechanism for furthering their careers and becoming productive members of the society.

Lohr (2009) says that online students performed better than students in classroom instruction and over a 12-year span, online students ranked in the 59th percentile in test performance, while classroom students ranked 50th percentile. The author insist that web-video, instant messaging, and collaboration tools have transformed online education, provided learning experiences tailored to individuals, transformed traditional learning, and have students teach one another. Much of the benefit of online learning is not based on technology per se, reports Jaschik (2009), but rather, on the quality and quantity of time students spend on online tasks, as compared to classroom students. So online education may even be more productive to students in terms of active learning and tested performance than traditional classrooms. Onlineeducation.org states that tuition for for-profit online education is lower than for traditional education, because many states fund virtual school programs and students can apply for financial aid if the online school is accredited by an impartial agency recognized by the U.S. Department of Education. Many non-profit online education programs are becoming very cheap or even free.

Ali (2003) explains that the transition from lecturing to student-centered learning demands a teaching model that enhances interactive learning and critical thinking, so online learning

and education should embrace facilitated and engaged communication between students and faculty. Burdett (2003) observes that universities have introduced information and technology communication to enhance teaching and learning as well as to improve the performance of the institution, including its faculty and students. Most universities are now providing online education through web-based learning and teaching, to serve both classroom and online students (Samarawickrema & Stacey, 2007). Since the purpose of introducing online education was to enhance learning and teaching, it is imperative to ensure that the transition from lecturing to online student-centered learning is a successful alternative and supplement to classroom-based learning.

In this paper, the role of business intelligence (BI) in enhancing online education is discussed. This is followed by a model of business intelligence to acquire and retain students for non-profit and for-profit online education institutions are recommended. The role of certain components of business intelligence and how they affect both non-profit as well as for-profit online education institutions are then discussed. This is followed by how the tools of BI could be used to enhance student retention and recruitment. Business intelligence strategies are analyzed for both for-profit and non-profit towards acquiring and retaining students, and finally, the conclusion and limitations are addressed.

BI IN ONLINE EDUCATION

BI systems use operational data along with analytic tools to present complex information to planners and decision makers, and the objective of BI is to provide usable information at the right time, right location, and in the right form in order to assist decision makers (Negash, 2004). BI describes technologies, applications, and processes used to gather, store, access, and analyze data to help users make better decisions, and it has become necessary to compete in today's marketplace (Wixom & Watson, 2010). It is in the best interest of organizations or institutions offering online education to utilize BI in order to compete for customers or students through the acquisition, integration, and analysis of huge volumes of diversified data.

If online education institutions are looking to increase their student enrollment and attendance, they need to depend on both their structured and semi-structured data. The internal sources of structured data are Enterprise Resources Planning (ERP) and the external sources include Customer Relationship Management (CRM), while the sources for semi-structured data are business processes and news items (Negash, 2004). These institutions can utilize ERP to locate patterns and relationships from data of current students who may have potential student leads within their families. They could utilize CRM to find potential students, using the data of shopping centers in the locality where their current students reside.

Negash (2004) says the BI architecture of structured data is based on data warehouse. The author adds that data is extracted from operational systems and distributed through Internet browser technologies into the data mart used by planners and decision makers. Then the data is retrieved from the data mart through inquiries from the web user and OLAP analysts (Negash, 2004). Processes are important in BI because it does the extracting, loading, and storing of data as well as metadata for IT and users, and prioritize BI projects (Wixom & Watson, 2010). Data originates from a variety of sources, such as transactional and Enterprise Resource Planning (ERP), and it gets into the data warehouse through data integration, then dependable data marts are used by users to retrieve information from the data warehouse (Wixom & Watson, 2010). BI is necessary for online education institutions because a variety of data about potential students from ERP, CRM, or the news outlets needs to be retrieved into the data warehouse through data integration, and then into dependable data marts for analysis of the data to take place.

The semi-structured data's architecture is divided into 5 components. The function model that shows what the organization does; process model that shows how organization performs its functions; data model that shows what data describes the organization; application inventory shows where the architecture resides; and metadata repository supports the metadata capture and usage (Negash, 2004). According to Wixom & Watson (2010), the BI environment also includes metadata, data quality, and governance, where the metadata supports the IT people and end users who need the quality data and the governance includes the committee and processes that make BI meet its organizational goal. It also states that BI has three targets: it needs specific applications in place, it needs to clean up and define the data, and it needs to transform itself organizationally (Wixom & Watson, 2010). Online education institutions need to define how they collect their data to fit the purpose of their business of acquiring and retaining students in order to know how the data architecture is designed, as well as how the metadata should work.

According to Wixom & Watson (2010), an example was given to illustrate the use of BI. Harrah's Entertainment saw new market opportunities with the change in legislation, so they integrated the hotels and casinos so that customers would come to gamble at their properties through Customer Relationship Management (CRM) and Total Rewards programs. They created the Winners Information Network (WINet) from the casino, hotels, and special programs, and gave loyalty cards to customers to use at slot machines. This card gathered all the necessary information and habits of each customer, so they knew how to target them with deals and coupons. Data from shopping cards from local Kroger, Giant, Sam's Club, and Costco stores, along with data from local banks and cash centers can be utilized to detect patterns for potential students for online education.

According to Davenport, Harris, De Long, & Jacobson (2001), when turning transactional data into knowledge and then results, the three elements required are *context, transformation*, and *outcomes*. This study also states that the context includes the strategy, skills and experience, organization and culture, as well as technology and data. The transformation has to do with analysis and decision making, while outcomes has the financial, process and program, and behavioral. In organizational culture, we have sophistication of the analysis where work is done in centralized groups and the amount of local knowledge are needed where decentralized analysts are because of knowledge of local market and firm's cultural orientation (Davenport et al., 2001). Online education institutions need decentralized analysts who understand the local markets from which data is collected to analyze patterns and centralized groups can analyze potential students for both for-profit and non-profit institutions.

ROLES OF BI IN ONLINE EDUCATION

Some of the components of BI needed by online education institutions in order to acquire and retain students are data warehousing, data warehousing context and success, data quality and data cleaning, privacy and security, data mining and knowledge discovery databases, as well as performance measurement and strategy. The roles of each of these BI components in enhancing online education are discussed below.

Performance Measurement and Strategy

Balance scorecards are used for motivating and measuring business unit performance, which consists of four perspectives, namely, financial, customer, internal business processes, and learning and growth (Kaplan & Norton, 1996). According to Kaplan & Norton (1996), financial performance measure is the long-term objective of the business unit, which has rapid growth, sustain, and harvest as stages. Rapid growth is the initial stage of the life cycle where products and services are developed and expanded, the sustain stage has majority of

the business that attracts investments for great returns, and the harvest stage is where equipment and capabilities are maintained. Financial objective in growth stage deals with sales, in the sustain stage deals with operating income and gross margin, while the harvest stage deals with cash flow. The three financial themes are revenue growth, cost reduction, and asset utilization. So an online education institution needs to measure how well it is doing in terms of finance, students, internal business processes, as well as learning and growth to know how their BI implementation is doing.

Customer performance measure deals with customer and market segments and the generic outcome measures market share, customer retention, customer acquisition, customer satisfaction, and customer profitability (Kaplan & Norton, 1996). Internal business processes deals with traditional approach that measures improved business processes with time-based metrics as well as balanced scoreboard approach which deals with innovation processes (Kaplan & Norton, 1996). Learning and growth deals with the infrastructure needed to create long-term growth and improvements. Goals that lead to business improvement and performance must be linked to economic and financial results (Kaplan & Norton, 1996). Performance measures (PM) recommendation, either PM frameworks or PM systems, are the initial building blocks of all PM initiatives (Folan & Browne, 2005). A PM framework involves the utilization of particular sets of PM recommendations, which could suggest the development of a structural framework (balanced scorecards) or procedural framework (Folan & Browne, 2005).

Data Warehousing

Jukic (2006) states that data warehousing has become mandatory for most large companies globally and it is created within companies as a separate data store whose main purpose is to help with data analysis for supporting the decision making process. Also, Jukic observes that it is separate in order not to compete with operational queries, and the database structure for operational purposes is different from that used for analytical purposes. Chaudhuri & Dayal (1997) conclude that data warehousing and online analytical processing (OLAP) have become important in decision support. The authors also say that data warehousing is now the focus of the database industry, where it is used to make decision makers make better and faster decisions, and it is a subject-oriented, integrated, time varying, and non-volatile collection of data. They also say that data warehousing is separated from operational databases because it's functional and performance requirements need OLAP, while the transactional databases need online transaction processing (OLTP). Online education institutions need to separate their transactional database from the operational database needed for BI analysis and this requires OLAP processing for better decision-making.

Jukic (2006) explain that in any data warehouse, the facilities that retrieve the data from operational databases into the data warehouses are called "Extraction, Transformation, and Load" (ETL), which take 50%-70% of development time, while data warehouse combines the data in the operational database and the data marts focus on departments or subjects. Data marts have fact tables that contain one or more measures, while the dimensional tables contain various descriptive attributes that relate to the subjects in the fact table (Jukic, 2006). March & Hevner (2005) insist that profit maximization is the economic goal of a business, so decision and information technology must align with economic goals. They discuss that internal information includes foundation information (accounting and financial measurement), productive information (effectiveness and efficiency of organizational production), competence information (competence that enable business to succeed), and resource-allocation information (effective and efficient use of scarce resources). Online education institutions should build an effective bridge between their operational databases, data warehouses, ETL, and data marts, and then link them to the economic goal of the business, which is to attract and retain students.

Organizations should choose their database modeling methodology based on their current and future analytical needs, which requires high expertise and high time commitment (Jukic, 2006). OLAP operations include rollup, drill-down, as well as slice and dice. Data warehouses might use relational OLAP (ROLAP), which has SQL extensions to implement multi-dimensional data operations, but multi-dimensional OLAP (MOLAP) stores the multi-dimensional data and implements the OLAP itself (Chaudhuri & Dayal, 1997). Online education institutions have the option to utilize ROLAP or MOLAP to analyze their data.

Data quality in terms of accuracy, timeliness, completeness, and consistency remains a major concern for both operational systems and data warehouses where users are trained and managed by the organization (March & Hevner, 2005). Companies use data warehousing to support strategy and mission-critical applications, which transforms data into information and knowledge, and are disseminated to decision makers (March & Hevner, 2005). The main problems are how the information was disseminated to end users, the selection and implementation of models, analytic tools, and data mining tools as well as privacy, security of data, system performance, training and support (March & Hevner, 2005). There are many issues that need to be tackled by online education institutions before the benefits of BI can be appreciated by transforming information from various data into knowledge for decision makers. Figure 1 shows the business intelligence model for acquiring and retaining students for for-profit and non-profit online education.
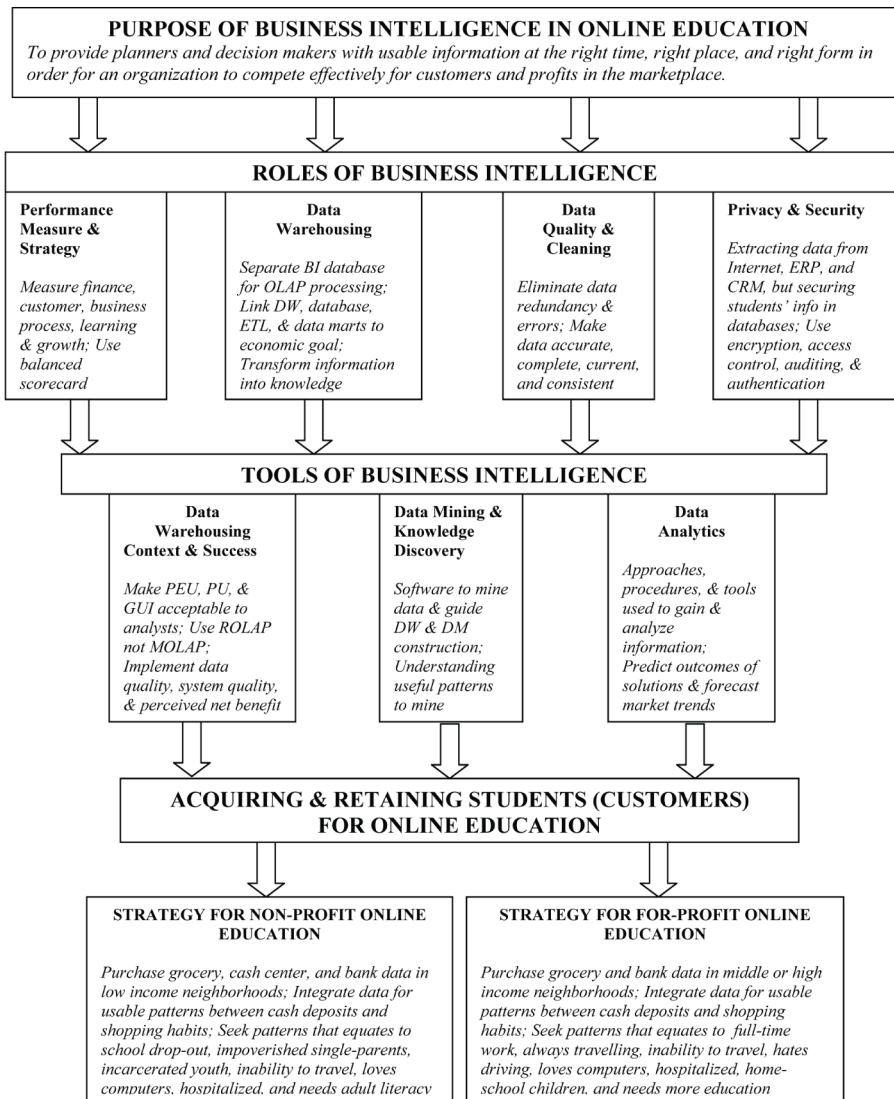
Data Quality and Data Cleaning

Data cleaning is also called data cleansing or scrubbing, which is needed in the case of misspellings, missing information, or invalid data (Rahm & Do, 2000). The need for it increases when multiple data sources need to be integrated because they may have redundant data in different representations (Rahm & Do, 2000). A major step to data reuse in order to attract new users and customers is to guarantee its quality through a data cleaning campaign and it is especially important due to the increasing need to integrate data from various sources (Scannapieco, Missier, & Batini, 2005). Some of the commonly referenced dimensions of data quality are accuracy, completeness, currency, and consistency. To enjoy the net benefit of BI, online education institutions must ensure that redundancy is eliminated in its integrated data from multiple sources.

Huge data for potential online students acquisition and retention must the cleaned for accuracy and completeness. Data cleaning approaches should be able to detect and remove all major errors and inconsistencies, minimize manual inspection and programming efforts, and should not be performed in isolation (Rahm & Do, 2000). Accuracy results in disparate levels of data models which range from single values to the entire database and it could be syntactic accuracy dealing with comparison function that evaluates the distance between "a" and "a!", and semantic accuracy capturing the various cases in which the correct value of "a" differs from "a!" (Scannapieco, Missier, & Batini, 2005). Data needs to be free from errors and accurate in order for adequate decisions to be made. Completeness is the sufficiency of data for the task at hand, which could be schema completeness when dealing with the degree of lack of missing data from the schema, column completeness when dealing with missing value in a column of a table, and population completeness when working with missing values for a particular population (Scannapieco, Missier, & Batini, 2005).

Numerous data for the acquisition and retention of potential online students must the cleaned for data currency and consistency. Stable data, like one's name, date of birth, and SSN, do not need updates, but time-variable data, like age, occupation, and address do. The three common time-related dimensions are currency, timeliness, and volatility. Currency is how promptly data is updated, timeliness is how current data is for the task at hand, and volatility is the frequency with which data vary

Figure 1. Business Intelligence strategy for non-profit & for-profit online educational institutions

PURPOSE OF BUSINESS INTELLIGENCE IN ONLINE EDUCATION
To provide planners and decision makers with usable information at the right time, right place, and right form in order for an organization to compete effectively for customers and profits in the marketplace.

ROLES OF BUSINESS INTELLIGENCE

Performance Measure & Strategy

Measure finance, customer, business process, learning & growth; Use balanced scorecard

Data Warehousing

Separate BI database for OLAP processing; Link DW, database, ETL, & data marts to economic goal; Transform information into knowledge

Data Quality & Cleaning

Eliminate data redundancy & errors; Make data accurate, complete, current, and consistent

Privacy & Security

Extracting data from Internet, ERP, and CRM, but securing students' info in databases; Use encryption, access control, auditing, & authentication

TOOLS OF BUSINESS INTELLIGENCE

Data Warehousing Context & Success

Make PEU, PU, & GUI acceptable to analysts; Use ROLAP not MOLAP; Implement data quality, system quality, & perceived net benefit

Data Mining & Knowledge Discovery

Software to mine data & guide DW & DM construction; Understanding useful patterns to mine

Data Analytics

Approaches, procedures, & tools used to gain & analyze information; Predict outcomes of solutions & forecast market trends

ACQUIRING & RETAINING STUDENTS (CUSTOMERS) FOR ONLINE EDUCATION

STRATEGY FOR NON-PROFIT ONLINE EDUCATION

Purchase grocery, cash center, and bank data in low income neighborhoods; Integrate data for usable patterns between cash deposits and shopping habits; Seek patterns that equates to school drop-out, impoverished single-parents, incarcerated youth, inability to travel, loves computers, hospitalized, and needs adult literacy

STRATEGY FOR FOR-PROFIT ONLINE EDUCATION

Purchase grocery and bank data in middle or high income neighborhoods; Integrate data for usable patterns between cash deposits and shopping habits; Seek patterns that equates to full-time work, always travelling, inability to travel, hates driving, loves computers, hospitalized, home-school children, and needs more education

in time (Scannapieco, Missier, & Batini, 2005). Data must be complete and current; otherwise erroneous decisions may be made that may be too costly and might be misconstrued as evidence of the ineffectiveness of BI. Consistency captures the violation of semantic rules in the data items, where all instances of data schema must satisfy the integrity constraints, which are intra-relation and inter-relation constraints

(Scannapieco, Missier, & Batini, 2005). Intra-relation constraint means that attributes or attributes of a relation must match, while inter-relation constraint means that specific relations and their attributes must match (Scannapieco, Missier, & Batini, 2005). Data must also be consistent in order for decision makers to arrive at the right decision. Data cleaning approaches include data analysis, definition of transforma-

tion workflow, verification, transformation, and backflow of cleaned data. Data analysis could be data profiling or data mining while defining data transformation could use structured query language (SQL) by utilizing application-specific language and particular user-defined function (UDF) (Rahm & Do, 2000).

Privacy and Security

Encryption, access control, auditing, and authentication are a priority for securing students' information in the online education databases. Informational privacy could be psychological or associative. Data collected without the knowledge of the people involved is considered psychological, while associative is where people are aware about data being collected about them but have no say about how the data is to be used (Tavani, 1999). Many believe that the unstructured web is more open to mining data than regular data mining (DM) from structured data warehouses and that the Internet is an emerging frontier for data mining because one major difference between data extracted data warehouse and the web is that the former is hidden from public view while the later is not (Tavani, 1999). Aside from relying on ERP and CRM, the internet could be an excellent place to extract information to find potential customers for online education, but institutions have to protect students' information in its DW.

Data warehouses (DW) are primarily built on open systems which may compromise the analytical discovery process, which helps to create a security conflict by making valuable and sensitive data easily accessible (Priebe & Pernul, 2000). During the data load process, highly sensitive data needs to be protected from eavesdropping and secrecy threats with encryption, while authentication may be sufficient for operating systems that allow sign-on strategy, and auditing will be best for the multidimensional level of OLAP (Priebe & Pernul, 2000). Access control on the back-end side controls access to data warehouse and source databases by the extract/transform/load processes, while the front-end has more novel access control issues where traditional DW are queried by high level users (Priebe & Pernul, 2000).

TOOLS OF BI IN ONLINE EDUCATION

Data Warehousing Context and Success

For a firm to have competitive advantage, lots of data needs to be managed, analyzed, and fed into the decision-making process with the use of OLAP tools, which gives decision makers valuable insights into customer and business behavior that can help maximize profits (Gorla, 2003). Perceived ease of use (PEU) and perceived usefulness (PU) are the two determinants of technology acceptance. PU is the degree to which a user believes the system will make the work more efficient, while PEU is the degree of how a user believes the system will be free of effort (Gorla, 2003). OLAP can organize data in multidimensional tables called data cubes, and provide access to the data warehouse through an interactive GUI (Gorla, 2003). For online education institutions to become profitable and have a competitive advantage in acquiring and retaining students efficiently, the PEU, PU, and GUI should be acceptable to the analyst utilizing the data warehouse to derive knowledge for the decision makers.

In MOLAP, data is cleaned, aggregated in multiple dimensions, and updated into the data cube periodically to be stored in multi-dimensional arrays, while in ROLAP, data is aggregated and stored along with relational databases (Gorla, 2003). PU was measured based on OLAP benefits such as improving decision making, providing accurate analysis, providing all required information, and increasing productivity, while PEU was measured based on

if OLAP was easy to use and was user friendly (Gorla, 2003). Cognos software (PowerPlay or Oracle) is an example of MOLAP while Business Objects (SQL Base and RDBMS) is a ROLAP. According to Gorla (2003), all features of ROLAP were useful, while only visualization and summarization were useful in MOLAP. So ROLAP will most likely appear to be more beneficial and useful to analysts of the online education institutions.

Gorla (2003) also gives a list of conclusions: In ROLAP, PEU was significantly related to PU and all features were easy to use and were user friendly. Visualization had positive effect in ROLAP but it was negative in MOLAP, but were both positive in usefulness. The summarization was both positive in ROLAP and MOLAP. In the query function, only PEU was positive for MOLAP, MOLAP was easy to use, but was not useful. In sophistication analysis, it was positive PU in ROLAP, but not for MOLAP. In dimensionality, PU was significant in ROLAP, but not in MOLAP. The positive correlation between performance and PU shows the importance of system performance in ROLAP. It is therefore in the best interest of online education institutions to utilize ROLAP instead of MOLAP because it has a better PEU and PU analysis.

Empirical research about data warehouse projects was necessary to know the critical roles of success in implementing an information system and the result shows that there is a relationship between the three dimensions of business success, which are system quality, data quality, and perceived net benefit (Wixom & Watson, 2001). Data quality is the quality of data available from the data warehouse; system quality is the performance measure of flexibility, integration, response time, and reliability, while net benefit is when a system displays high data quality and system quality (Wixom & Watson, 2001). For data warehouse implementation success, it must have three facets which are success with organizational issues, success with project issues, and success

with technical issues (Wixom & Watson, 2001). Online education institutions must make sure that their organizational plan of acquiring and retaining students, as well as the technical ability to utilize DW effectively and successfully is linked to the successful implementation of data quality, system quality, and perceived net benefit.

For organizational implementation success, implementation of a system is not successful unless its produce is accepted into the organization and integrated into the work processes, project implementation success must have complex array of tasks and roles that must be well managed by highly skilled teams, and technical implementation success must have the ability to understand, reconcile, and coordinate large volume of data that must be extracted, transformed, loaded, and maintained (Wixom & Watson, 2001). Understanding the successful adoption of information technology is based on the linkage of quality, satisfaction, and usage, which has four dormant views of quality that are the excellence that meets some standard, quality as value, conformance with specifications, and quality as meeting expectations (Nelson, Todd, & Wixom, 2005). Quality is the degree to which a service exceeds the customer's expectations (Nelson, Todd, & Wixom, 2005).

Data Mining and Knowledge Discovery

Feelders, Daniels, & Holsheimer (2000) state that data mining has become important due to the need for corporations to understand valuable information from the vast amount of data collected on their customers. They also say that it is this vital information from customers that can lead to improving business processes that enhance the construction of data warehouses (DW) and data marts. All data hold valuable information that could be used to improve business decisions and optimize success. Databases contain too much data so mining software is

needed to analyze them for valuable decision making information (Goebel & Gruenwald, 1999). Online education needs data mining software to mine customer or student data from databases and the information from the data can also guide the how the DW and DM are constructed.

According to Feelders, Daniels, & Holsheimer (2000), data mining process includes problem definitions, acquisition of background knowledge, selection data, pre-processing of data, analysis and interpretation, as well as reporting and use. The role of modeling is not pure prediction but rather to gain understanding of particular groups of customers in order to develop marketing strategies (Feelders, Daniels, & Holsheimer, 2000). Knowledge discovery and data mining are techniques needed to discover strategic hidden information in large databases. (Goebel & Gruenwald, 1999). Knowledge discovery in databases (KDD) is the process of identifying potentially useful and ultimately understandable pattern in data, while data mining (DM) is the extraction of patterns or models from observed data (Goebel & Gruenwald, 1999). Knowledge Discovery in databases (KDD) is a process model and a life cycle rather than a methodology (Marban, Mariscal, & Segovia, (2009). Useful and understood patterns in the students' data are essential in order to know what needs to be mined from the databases in order for the decision makers to make effective decisions.

Data Analytics

Business Analytics (BA) are a set of approaches, procedures, and tools used to gain and analyze information, as well as to predict the outcomes of solutions to problems during planning, sourcing, making, and delivering (Trkman, McCormack, Valadares de Oliveira, & Ladeira, 2010). BA is useful in providing valuable information and knowledge to decision makers towards minimizing cost, accurately forecasting market trends,

enhancing competitive advantage, as well as improving performance (Trkman et al., 2010). It is the ability to analyze the patterns in mined data and utilize the knowledge derived from it to predict outcomes that lead to reduced cost, enhanced competitive advantage, and improved profitability.

Buchner & Mulvenna (1998) state that the main reason of utilizing online retail models is that customers display behavioral patterns, marketing orientation with products, and descriptive varied products and services. The authors also say that there are three types of data online retail model, which are server data, marketing data, as well as web metadata. The server data include Server logs, which are stored in extended logfile format to provide meta information about the log file, Error logs, which stores failed requests of missing links and authentication failures, and Cookies, which tracks user clients access across hosted web pages. The marketing data stores information about the customers, products, transactions, as well as customer's socio-graphic information, distribution details, and advertising information. The web metadata is the data about the website where activity is taking place or being updated. Data has to be stored in specific formats before it can be analyzed effectively to provide the exact knowledge needed by the decision makers to attract new customers and retain old ones.

The two important parts of customer attraction is the selection of potential customers as well as the acquisition of the selected candidates, and a marketing strategy to utilize is to find already established information and behavioral patterns for profitable and non-profitable customers (Buchner & Mulvenna, 1998). Customer retention is the attempt to make customers as loyal as possible by understanding the behaviors of customers and this data can be mined through resource discovery, information extraction, and generalizations for marketing knowledge in order to perform commercial activities like

personalization, adaptation, customization, profiling, and recommendation (Buchner & Mulvenna, 1998). Analytics help to categorize customers, based on established behavioral patterns in order to maximize profit and the customer base through customer acquisition and retention.

BI STRATEGIES: FOR-PROFIT AND NON-PROFIT ONLINE EDUCATION

After various data from banks, grocery stores, and cash centers have been purchased by online education institutions, it is recommended that it is stored, integrated, and analyzed for useful patterns that can be used by these institutions to acquire and retain students (customers). These useful and needed patterns are then mined with data mining software based on the knowledge discovery results, which will be used by decision makers to increase their students and profits.

For the non-profit online education institution, where the main goal is to provide free (or small fee) educational service to those who need adult literacy, school drop-outs, the hospitalized, those with inadequate transportation, inability to travel, the incarcerated, and impoverished single-parents who want better education, data needs to be collected from banks, grocery stores, and cash centers in low income neighborhoods. Banks already have customer information while grocery stores, such as, Giant and Kroger, have customer information on their discount cards. Wixom & Watson (2010), states how Harrah Entertainment used Customer Relationship Management (CRM) and Total Rewards programs through their loyalty cards to acquire information and habits about their customers. These population and customer base would most likely benefit from educational services that would be funded by county, state, and federal government in order to increase the educational level of the citizens as long as the online education institution is accredited by a recognized and reputable agency.

The institution's analysts are to search for useful patterns that match cash deposits with shopping habits. The sources of structured data internally are Enterprise Resources Planning (ERP) and externally are the Customer Relationship Management (CRM), while the sources for semi-structured data are business processes and news items (Negash, 2004). The analysts can use the ERP to study the useful patterns of enrolled students they already have in order to see what kind of patterns they can derive from their cash deposits and grocery shopping habits. This information can then be used to retain students in the ERP, or used to seek similar patterns in CRM and the Internet in order to know the most willing students to acquire for their educational services.

The derived pattern could be how deposited cash relate to frequency of grocery shopping, amount of money spent regularly on food, quality of diet, and types of products purchased, or how groceries relate to those who use the cash centers to cash checks rather than to have a bank account. Those who shop irregularly may lack adequate transportation and thereby carry grocery by hand or may buy a lot at once because they do not have much for gas. This is why patterns must be analyzed and understood in order for decision makers to make the right decision in retaining and recruiting students.

For the for-profit online education institution, the analyst should be more interested in collecting, integrating, and analyzing data from middle or high income neighborhoods. Their target for students to recruit are those with full-time jobs, travel frequently, single-parents, inadequate transportation, home-schooling, loves computers, and need more education. These should be people that can afford to pay for their education or qualify for financial aid. The same process for the non-profit online education data in the ERP, CRM, banks, and groceries can be applied for the for-profit, except the population and student target are different, as well as the fact that these people usually have bank accounts rather than utilizing cash centers.

CONCLUSION AND LIMITATIONS

It is in the best interest of organizations or institutions of online education to utilize BI in order to compete for customers or students through the acquisition, integration, and analysis of a huge and diverse body of data. These institutions could utilize ERP to find patterns and relationships from data of current students who may have potential student leads within their families. They can utilize CRM to find potential students from the data of shopping centers in the locality where their current students reside.

BI is necessary to online education institutions because variety of data about potential students from ERP, CRM, or the news outlets needs to be retrieved into the data warehouse through data integration, and then into dependable data marts for analysis of the data to take place. Online education institutions need to define how they collect their data to fit the purpose of their business of acquiring and retaining students in order to know how the data architecture should be designed as well as how the metadata should work. Data from shopping cards from local Kroger, Giant, Sam's Club, and Costco along with data from data from local banks and cash centers can be utilized to detected patterns for potential students for online education.

Online education institution need decentralized analysts who understand the local markets from which data is collected to analyze patterns and centralized groups that may make potential students for both for-profit and non-profit institutions. Online education institution need to separate their transactional database from the database needed for BI analysis and it requires OLAP processing for better decision making. Online education institutions should build an effective bridge between its operational databases, data warehouses, ETL, and data marts, then link them to the economic goal of the business.

For online education institutions to become profitable and have a competitive advantage in acquiring students efficiently, the PEU, PU, and GUI must be acceptable to the analyst utilizing the data warehouse to derive knowledge for the decision makers. It is in the best interest of online education institutions to utilize ROLAP instead of MOLAP because it has a better PEU and PU analysis. Online education institutions must make sure that their organizational plan of acquiring and retaining students, as well as their technical ability to utilize DW effectively is linked to the successful implementation of data quality, system quality, and perceived net benefit.

To enjoy the net benefit of BI, Online education institutions must ensure that redundancy is eliminated from their integrated data from multiple sources. Data needs to be free from errors and accurate, in order for adequate decisions to be made. Data must be complete and current, otherwise erroneous decisions may be made that may be costly and be seen as evidence of the ineffectiveness of BI. Data must also be consistent in order for decision-makers to arrive at the right decision. Aside from relying on ERP and CRM, the internet could be an excellent place to extract information to find potential customers for online education, but institutions have to protect students' information in its DW.

Encryption, access control, auditing, and authentication are a priority, to secure students' information in online education databases. Online education needs data mining software to mine customer or student data from databases and the information from the data can also guide how the DW and data marts are constructed. Useful and understood patterns in the students' data are essential in order to know what needs to be mined from the databases in order for decision makers to make effective decisions. So an online education institution needs to measure how well it is doing in terms of finance, students, internal business processes, as well as learning and growth to know how their BI implementation is doing.

This paper is conceptual because no data was collected from any bank, grocery store, or cash center to know if any useful pattern can be derived in order to know what kind of students to retain or recruit for both non-profit

and for-profit online education. If a pattern was derived based on the knowledge discovery, there is no guarantee that these identified potential students will respond favorably to the online education opportunity presented to them. Also, data from banks, cash centers, and grocery stores may not be the best possible data sources to acquire students, but it would be a good start since people usually need their money and food to survive on a regular basis.

Despite the fact that we separated potential students for non-profit online education from the for-profit online education based on economic and residential status, there may be a fine line or certain members in each population could qualify for services provided by both kinds of online education. Regardless, an investment in business intelligence towards generating any kind of useful patterns is worthwhile, because it is one of the best ways of increasing and retaining one's customer base, as well as improving profits.

REFERENCES

Ali, A. (2003). Instructional design and online instruction: Practices and perception. *TechTrends*, *47*(5), 42–45. doi:10.1007/BF02763205.

Brantley, A. (2006). 10 reasons why you should take online classes, http://www.associatedcontent.com/article/54397/10_reasons_why_you_should_take_online.html?cat=4

Buchner, A. G., & Mulvenna, M. D. (1998). Discovering internet marketing intelligence through online analytical web usage mining, ACM SIGMOD proceedings, 1-8.

Burdett, J. (2003). A switch to online takes time: Academics' experiences of ICT innovation, Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE), 84-93.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, *26*(1), 1–10. doi:10.1145/248603.248616.

Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to result: Building an analytic capability, California Management Review, 4, 43 (2), 117-138.

Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodology and practical aspects of data mining. *Information & Management*, *37*, 271–281. doi:10.1016/S0378-7206(99)00051-8.

Folan, P., & Browne, J. (2005). A review of performance measurement: Towards performance management. *Computers in Industry*, *56*, 663–680. doi:10.1016/j.compind.2005.03.001.

Goebel, M., & Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. *ACM SIGKDD*, *1*(1), 20–33. doi:10.1145/846170.846172.

Gorla, N. (2003). Features to consider in a data warehousing system. *Communications of the ACM*, *46*(11), 111–115. doi:10.1145/948383.948389.

Jaschik, S. (2009). The evidence of online education, http://www.insidehighered.com/news/2009/06/29/online

Jukic, N. (2006). Modeling Strategies and alternatives for data warehousing projects. *Communications of the ACM*, *49*(4), 83–88. doi:10.1145/1121949.1121952.

Kaplan, R. S., & Norton, D. P. (1996). Linking the balanced scorecard to strategy. *California Management Review*, *39*(1), 53–79. doi:10.2307/41165876.

Lohr, S. (2009). Study finds that online education beats the classroom, http://bits.blogs.nytimes.com/2009/08/19/study-finds-that-online-education-beats-the-classroom/

Marban, O., Mariscal, G., Segovia, J. (2009). A data mining and knowledge discovery process model, Data Mining and Knowledge Discovery in Real Life Applications, 1-16.

March, S. T., & Hevner, A. R. (2005). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, *43*, 1031–1043. doi:10.1016/j.dss.2005.05.029.

Negash, S. (2004). Business Intelligence. *Communications of the AIS*, *13*, 177–195.

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, *21*(4), 199–235.

Onlineeducation.org. http://www.onlineeducation.org/introduction-online-education

Onlineeducation.org. http://www.onlineeducation.org/online-accreditation-importance

Onlineeducation.org. http://www.onlineeducation.org/positive-aspects-online-education

Priebe, T., & Pernul, G. (2000). Towards OLAP security design – Survey and research issues, ACM International Workshop on Data Warehousing and OLAP, 33-40.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problem and current approaches, IEEE, 3-13.

Samarawickrema, G., & Stacey, E. (2007). Adopting web-based learning and teaching: A case study in higher education. *Distance Education*, *28*(3), 313–333. doi:10.1080/01587910701611344.

Scannapieco, M., Missier, P., & Batini, C. (2005). Data quality at a glance, Datenbank-Spektrum, 6-14.

Shaffer, S. (2011). Why do people choose online education? http://www.classesandcareers.com/education/2011/05/26/why-do-people-choose-online-education/

Tavani, H. T. (1999). Information privacy, data mining, and the internet. *Ethics and Information Technology*, ▪▪▪, 137–145. doi:10.1023/A:1010063528863.

Trkman, P., McCormack, K., Valadaresde Oliveira, M. P., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*, 318–327. doi:10.1016/j.dss.2010.03.007.

Wixom, B., & Watson, H. (2010). The BI-Based Organization. *International Journal of Business Intelligence Research*, *1*(1), 13–28. doi:10.4018/jbir.2010071702.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *Management Information Systems Quarterly*, *25*(1), 17–41. doi:10.2307/3250957.