AC 2012-3855: USING CONTENT ANALYSIS TO EVALUATE STUDENT INQUIRY-BASED LEARNING: THE CASE OF HIGH SCHOOL STUDENTS PREPARING FOR A CYBER DEFENSE COMPETITION

Dr. Julie Ann Rursch, Iowa State University

Julie A. Rursch is currently is a Ph.D. candidate in the Department of Electrical and Computer Engineering at Iowa State University. She will graduate with a degree in computer engineering with a focus on secure computing. Her research includes a unique approach to critical infrastructure modeling which provides emergency planners and first responders with resilient and flexible critical infrastructure evaluation in the face of non-recurrent, disruptive events. Her approach creates a new paradigm for modeling critical infrastructure sectors, analyzing real-time physical data, and providing best fit mitigations to impending failures and responses. At Iowa State University, Rursch is very involved in the IT-Adventures high school outreach program, serving as the Assistant Director since the program's inception in 2007.

Dr. Douglas W. Jacobson, Iowa State University

Doug Jacobson is a university professor in the Department of Electrical and Computer Engineering at Iowa State University. He is the Drector the ISU Information Assurance Center, which has been recognized by the National Security Agency as a charter Center of Academic Excellence for Information Assurance Education. Jacobson teaches network security and information warfare and has written a textbook on network security and is completing a second book on computer security literacy. He is director of the IT-Adventures program and oversees the cyber defense competitions hosted at ISU. His research is targeted at developing large scale attack simulation environments and is the Director of the Internet-Scale Event and Attack Generation Environment (ISEAGE) test bed project.

Andy Luse, Iowa State University

Andy Luse is a Ph.D. student in business and technology specializing in management of information technology at Iowa State University. Luse received his Ph.D. in human computer interaction and computer engineering from Iowa State University in 2009. His research interests include computer and network security, visualization for computer and network security, and user interface design and usability. He is a member of the Association for Computing Machinery, the Institute of Electrical and Electronics Engineers, and the Association for Information Systems.

Using Content Analysis to Evaluate Student Inquiry-Based Learning: The Case of High School Students Preparing for a Cyber Defense Competition

Abstract

Inquiry-based learning is a documented, successful method to increase student understanding in science, technology, engineering and mathematics (STEM) areas, as well as a way to facilitate critical thinking. This type of instruction allows students to explore science and resolve problems by reviewing what they know, seeking out additional information about the problem, and coming to a conclusion or resolution based upon evidence they have gathered. This paper evaluates a high school outreach program designed to increase student learning about information technology (IT), in this case specifically information security. The high school students who participated in the outreach program spent the year using the learning materials provided by Iowa State University, asking their own questions about network security and information assurance, exploring additional resources, and determining how to solve the challenges of setting up a secure and viable network. The capstone event for students who participated in the IT club is a two-day cyber defense competition (CDC) on the Iowa State University campus. During the remote setup, the high school students were able to log into a chat room and ask for guidance or clarification from college students supporting the equipment on campus. These chat conversations were logged and this paper utilizes content analysis to quantitatively analyze the chat conversations in terms of the students progressing through Bloom's taxonomy. The results demonstrated that students were in the Applying, Analyzing and Evaluating stages of learning, showing that the students did perform active and complex thinking in designing, configuring, and securing their cyber defense competition networks.

Inquiry-based learning is a documented, successful method to increase student understanding in science, technology, engineering and mathematics (STEM) areas, as well as a way to facilitate critical thinking.[1] This type of instruction allows students to explore science and resolve problems by reviewing what they know, seeking out additional information about the problem, and coming to a conclusion or resolution based upon evidence they have gathered. Inquiry-based learning uses an active learning strategy where students move completely through Bloom's revised taxonomy of cognitive domain.[23] The authors of this paper are involved in a high school outreach program designed to increase student learning about information technology (IT), in this case specifically information security, and wanted to evaluate whether the learning occurring also moved students through Bloom's revised taxonomy. Since this was an after-school, extra-curricular program, traditional measurements used in a classroom such as testing, final reports, or research presentations were not readily available. However, because of the way the program is structured, the authors were able to evaluate the inquiry-based learning taking place using content analysis, a novel, non-intrusive approach.

The high school students who participated in the outreach program were part of an information technology (IT) club formed in the fall of the academic year at their high school to study network and information security using an inquiry-based approach. Students spent the year using the learning materials provided by Iowa State University, asking their own questions about network

security and information assurance, exploring additional resources, and determining how to solve the challenges of setting up a secure and viable network. The capstone event for students who participated in the IT club is a two-day cyber defense competition (CDC) on the Iowa State University campus. In a CDC the students remotely design and configure a set of servers and a network in a secure manner one month prior to the competition and then come to campus for a two-day event where they protect and maintain the equipment from hackers.

During the remote setup, the high school students were able to log into a chat room and ask for guidance or clarification from college students supporting the equipment on campus. These chat conversations were logged and this paper utilizes content analysis to quantitatively analyze the chat conversations. The conversations are analyzed for overarching themes and questions to examine the process the students are going through during the inquiry-based learning of setting up their competition networks. The paper is organized into five sections: Motivation, Project Background, Methodology, Results and Conclusion/Future Directions.

Motivation for Using Content Analysis to Understand Inquiry-Based Learning

Inquiry-based learning is a multifaceted approach that involves reviewing information about what is known about a problem, gathering additional information, proposing solutions or explanations, and communicating or acting on the results. The focus of all activities is on critical and logical thinking, as well as exploration of alternative solutions.[45] It is an effective way to shape students' ability to apply their knowledge, search for new information, and critically synthesize material. When using inquiry-based learning in the classroom, an instructor facilitates their learning, but does not formally structure their experiences. The students are given a problem, given background information, and then asked to find their solution which will include finding new information on their own.

Inquiry-based learning is the backbone of the educational outreach programs offered by the Information Assurance Center (IAC) at Iowa State University. In the IAC's implementation of inquiry-based learning in information assurance and computer/network security education, the students are given access to learning materials, as well as a set of questions to explore. They are then asked to provide solutions to challenges presented to them based upon their own exploration, additional resources they find, and experimentation. The materials provided focus on security-based concepts for designing, implementing, configuring, securing, and protecting a network, its servers, and end users from attacks from the outside world. Although the IAC conducts inquiry-based information assurance and computer/network security programs targeted at college-age students from Iowa State University (see paper by the authors at this conference), community colleges across the state[6] and four year institutions nationwide, this paper focuses on an inquiry-based program developed for and delivered to high school students within the state.

The use of content analysis to understand inquiry-based learning is a novel approach to evaluation that the IAC is undertaking as part of its outreach program. Content analysis is the scientific, systematic, quantitative analysis of messages.[7] It is widely used in analyzing messages in journalism, mass communication, sociology, psychology, and business. Historically, in the mass communication discipline, it has been used to examine word usage, as well as the context of conversations and human interactions in mediated messages transmitted by newspapers, television, and radio. With the advent of electronic mass media, content analysis

has expanded to include the evaluation of web sites and blogs. Additionally, work in content analysis has evolved to include non-mediated messages, such as conversations between individuals. It is these non-mediated messages found in the chat conversations between high school students and the college students supporting their remote setup for a CDC that this paper evaluates.

The authors use content analysis to find themes in the students' inquiry-based learning as they design and configure their competition network. Using these coded themes, the authors examine how the student conversations and questions change over time from the beginning of setup to the end of the configuration period. These coded themes are them examined in the context of Bloom's taxonomy to see if the students are moving through the cognitive learning process with this program.

Bloom proposed a taxonomy of educational learning objectives which was based on a set of conferences in the early 1950's.[8] The taxonomy is composed of a classification of various objectives which can be utilized by educators when developing content for their students. The taxonomy is a tiered model with each level representing a particular degree of cognitive complexity. The levels of the taxonomy build on each other such that functioning by an individual at a particular level within the hierarchy is dependent on mastery of the material at the next lower level.[9]

Anderson and Krathwohl modified Bloom's taxonomy based on input from three stakeholder groups: cognitive psychologists, curriculum theorists, and assessment specialists.[10] This revised model utilizes verbs as opposed to nouns to emphasize the engagement of the individual in the learning process. Furthermore, several name changes were used with regard to categories. Figure 1 shows the revised taxonomy with both the title words as well as supporting words detailed within each level. Also, notice the increasing mastery of the individual with the educational material as you move from bottom to top in the hierarchy.

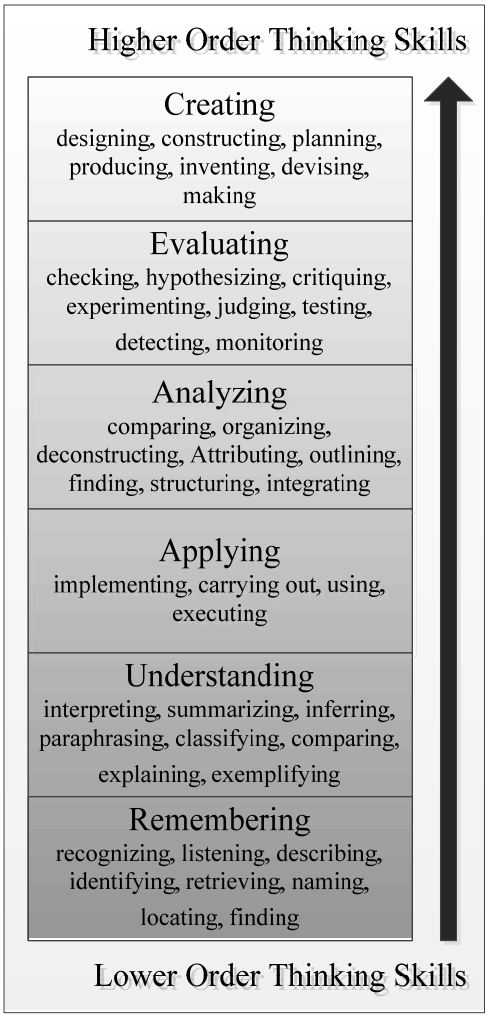

Figure 1. Revised Blooms Taxonomy based on work by Anderson and Krathwohl

This paper utilizes Bloom's taxonomy to help categorize the conversations in the chat logs according to their respective level within the taxonomy. This allows the authors to recognize which type of learning the students are engaging in when they are conversing about a particular portion of the CDC setup process. It is anticipated that students will move through Bloom's revised taxonomy as they move through the setup process. By categorizing the chat logs within the taxonomy the authors can test this anticipated research hypothesis, as well as gain insight into which areas in the structure are receiving sufficient attention and what areas are lacking. This will allow future competitions to better engage the students in the educational nature of the CDC by utilizing all levels of the hierarchy.

Because designing and configuring a network can be thought of as a step-by-step process, the authors anticipate that learning occurs along the steps of the process. Generally, the students first build the servers, then configure the services running on those servers, and get the servers to talk to each other. Once the students have a quasi-functional network, they begin to secure the

servers to protect from vulnerabilities and unwanted exploits launched by attackers during the competition. The authors expected the students' questions and exploration to follow a similar path as they setup their equipment. The authors anticipate the topics in the chat log moving from remembering the words, to understanding the concepts, to applying them, to analyzing when the network/server doesn't work, to evaluating each other's work, and creating their own approaches to their total network design.

Project Background

The IAC's high school information assurance and computer security program centers on the creation of an IT club that is formed at the high school and run as a year-long, after-school, extra-curricular activity. The IT club allows students to explore information assurance and computer/network security in a non-threatening, non-graded environment which allows students the freedom to explore outside of their need to perform in the classroom. The program is not about structured learning or examinations, but exploring networking and security concepts through the provided educational materials and having contact with a local IT professional who serves as a mentor.

Iowa State University provides educational materials such as video-taped lectures and lab experiments for students to watch and try on their own. The students' end goal or "problem" in their inquiry-based learning throughout the year is to compete in a state-wide CDC in the spring. This goal is what guides their progression throughout the months of IT club meetings and experimentation prior to the event. In the inquiry-based learning model, the local IT professional serves as the facilitator, helping students when they hit road blocks or don't understand new information they have found. In a CDC students design, configure, secure, and maintain a set of networked servers in a relatively short, one-month period of time. They then come to campus for the two-day competition where attackers try to exploit vulnerabilities in their systems and gain control of the students' systems and networks. The students' goal during the competition is to prevent, if possible, any security violations or attacks on their network, as well as report and correct any problems that arise. They also must maintain full functionality of their systems for their end users.

Students in the CDC configure their networks as described by a scenario published at the same time as remote setup occurs, one month prior to the competition. The scenario details the services that the students must implement in a short story format. They are told they are the IT support staff for a fictitious company or school and have to implement services such as email, web mail, remote programming, file sharing, and web hosting. They are also told they are responsible for their own Domain Name Service (DNS) and are advised that it would be wise to implement a firewall to help protect their networks. They are also given some service, generally a web server, that is a legacy installation and must be supported in a present state. The legacy system provides inherent security vulnerabilities that they have to protect against because these are out-of-date operating systems or unpatched operating systems. These machines require the students to figure out an alternate way of ensuring protection from the attack team other than updating the operating system.

In addition to defending their network, the students also participate in numerous activities (called anomalies) throughout the competition which are designed to keep them engaged and slightly off

balance just as real IT staffs get engaged in new projects and may overlook intrusions or security risks in new implementations. These requests may run counter to the goal of having secure systems or may open holes in their servers. The students must then decide how, or if, to implement the request on their network, as well as how to implement it securely. It is these activities that help keep the students focused on providing a useable network for their end users, as well as a secure one.

The group who tests each student team's network for vulnerabilities and plays the role of attackers in the competition consists of IT professionals, as well as computer engineering faculty and graduate students specializing in information assurance. Since the competitions occur over a Friday and Saturday, their job is to conduct recognizance work on Friday and early Saturday morning when the teams are still setting up to determine what kinds of networks the teams are running or to carry out social engineering. Then, at the designated time on Saturday morning, the active network scanning and penetration testing begins. Once vulnerabilities are found, they may be exploited to gain control of or break into the servers and/or network. To win a competition, the student teams earn points based upon service uptime, usability, security, and reporting. The teams try to not be the victim of the attackers, but if they are, they need to fix the problem and then report what they did to the judges.

The students' competition networks run in an isolated network environment to avoid the inevitable misconfiguration or unwanted attacks on the real world network. In our CDCs, the Internet-Scale Event and Attack Generation Environment (ISEAGE – pronounced ice-age) testbed creates a virtual Internet for the competition. ISEAGE provides a controlled environment that allows real attacks to be played out against the students' networks and demonstrates to them real world security concepts. The ISEAGE testbed has an air gap proxy server through which students can connect to the Internet to download operating systems and patches or search for additional information about configuration problems, but no other traffic can escape the testbed.

All cyber defense competition student teams remotely configure their competition networks on the ISEAGE testbed. The remote setup begins one month before the actual competition date. A printed document about how to access remote setup is provided to help illustrate the remote access procedures to access their team's rack of equipment in the ISEAGE research facility. The students can remotely load any public domain or Microsoft-based operating system on these computers and can configure the machines in any manner to serve as their competition network. Operating systems are primarily installed through a PXE-boot environment providing operating system images from which to choose. However, there is also some demand by the students for other, lesser known, open source operating systems which require ISO images to be downloaded, CDs or DVDs burned, and then inserted into drives for them.

Since the students have no physical access to the machines during setup, student teams are supported through the use of a chat room, as well as a dedicated cell phone, that is staffed approximately 20-30 hours per week, mainly in the evening, by members of the collegiate Information Assurance Student Group (IASG) and ISEAGE student workers. The type of assistance provided to the student teams through chat and telephone includes such things as answering setup questions, physically assisting them with their racks of computers, installing additional memory or network cards, and rewiring their competition network so that their firewall can be implemented.

Methodology

Log files that record the conversations between the high school students and the college students supporting them are automatically created on the chat room server as part of the software. These text files remain on the server in perpetuity until the server administrator cleans them out. After each CDC, the log files are copied from the chat room server, saved for analysis at a later date, and reset to an empty file for the next CDC. While the CDC for high school students was first run in 2006, the first year that a chat room was implemented to support the remote setup and configuration of student competition networks was for the 2007 CDC. At the writing of this paper, the authors have five years of chat log files archived which can be examined using content analysis. This paper focuses only on the first year in which the authors have data, 2007, and is the first attempt at using content analysis to evaluate the program. Since content analysis is a novel approach for evaluating inquiry-based learning and chat logs, the authors view this paper as a way to frame the use of content analysis in understanding inquiry-based learning programs. As discussed in the Conclusions/Future Directions section, the authors plan to use what is learned in this analysis to expand and enhance the evaluation of the full five years of chat logs, as well as to increase the effectiveness of the high school outreach program.

To conduct the content analysis for this paper, the log file for the 2007 CDC was retrieved from the archive. The text file contains conversations that occurred from April 24, 2007 to midnight on May 17, 2007 and represents the dates the chat room was open to support remote setup for the high school students. April 24 was the date remote access was given for the high school students to begin connecting to the ISEAGE testbed to configure their competition networks. Since the CDC began on May 18, midnight on May 17 represents the time remote support for the students was shut down. There were 27,363 chat lines recorded during this time period. Although technically, the students would have six additional hours of on-site setup time to do last minute configuration when they arrived at the ISEAGE research facilities for the competition, those interactions are not included in this analysis. In person support was provided during the final six hours of setup, therefore, the chat room was not manned and no conversations were recorded.

Because the chat room software logged the student conversations in a line by line fashion recording exactly what the student typed and when he/she pressed the enter key to send, each line of the chat was coded as to its thematic message. The coding scheme was developed by creating an original list of conversation topics that were expected to be found in the log files. The list was developed by one of the authors of this paper who manned the support chat room during the remote setup for the 2007 CDC. The conversation topic codes or themes were developed based upon the author's experience in the types of conversations occurring in the chat room during the shifts she worked.

To test that these thematic codes would be exclusive and able to cover all conversation topics found in the chat records, the first 5000 lines of the log file were utilized. Unfortunately, a random sample of all chat conversation lines would not be used because each line was not an individual thought. Chat rooms are designed to mimic face-to-face conversations. Students type in chat rooms in a conversational style. Therefore, a complete thought many times covered more than one line. Additionally, conversation topics could change quickly and ideas would rotate between adjacent lines of conversation by the same individual, much like a conversation between two friends can jump from one topic to a new topic and then back to the original topic. Also, 5000 lines is a relatively high sample size, representing 18.3% of the entire chat conversation.

However, the authors wanted to make sure that the technical concepts would be covered sufficiently in the codes and expected that the students would spend their first few days in the chat room socializing and trying to figure out how to correctly connect to the ISEAGE testbed to configure their equipment.

After coding the sample chat lines, the conversation topics thematic list was then expanded to include a total of 23 codes. (See Table 1 for a list of the codes, as well as the frequency distribution which is discussed below.) They were designed to be mutually exclusive categories. If two topics were discussed in a line, the line was coded based upon the conversation topic which was first mentioned. All 27,363 lines were coded by one individual. Generally, in content analysis of this size and scale, several coders are trained and an intercoder reliability can be established. In this project, the technical content in the chat conversations is such that the coder has to understand information assurance and computer/network security to be able to accurately code the conversations. Due to the knowledge issue, as well as money and time constraints, only one coder was used for this initial work on content analysis of inquiry-based learning.

Individual students cannot be identified in the chat room conversation logs. All students in the chat room used assumed usernames which are not directly connected to the student. In an effort to keep the logins to the chat room simple for the high school students, as well as the college students supporting them, the students from a team authenticated as their team number (e.g. Team 1) and then were allowed to change their nicknames. While many of them originally just used the Team 1 moniker, as they explored the chat room software and poked at its features, the high school students developed their own naming convention. This naming convention was not enforced by the IAC or the college students doing the support, but which was, interestingly, enforced by the high school students themselves. The high school students would inform each other when they forgot to change the nickname from the default login of team number to their "defacto standard" naming convention. This "de facto" standard was developed using the team number and then the student's first name (T1-Jane). While there was some policing by the high school students of the naming convention, there was also ample playing with the feature which allowed students to change their nicknames. Many of them would play with the nickname changing feature when they first logged into the chat room and were experimenting, as well as when they were bored and waiting on an operating system to install or waiting for their turn to talk to the college students supporting them. The nickname changing phenomenon was discovered during coding scheme testing, but was not determined as important to the setup of the competition network and coded as chat log overhead.

In coding the chat conversation lines as topical themes, a total of 4063 lines were coded as chat log overhead. These lines represented the mechanics of the chat room software recording in the log files when a user logged into the system, logged off the system, was timed out by sitting idle too long, or changed his/her nickname. Since the point of this paper is to understand student learning and examine the conversational themes over time, it was determined that the lines coded as chat log overhead were nothing more than system functions that had no bearing on the students' learning. Therefore, these lines were discarded in the analysis that follows in the Results section. This left 23,300 lines of chat conversation to be analyzed.

Additionally, each line of the chat log was coded as to whether the speaker was a high school student participant or one of the college students who was manning the chat room. Because each

line in the chat logs had a date and time stamp, the number of weeks prior to the CDC event that the conversation occurred could be established. It is this code that allows the authors to look at the progression of conversation themes and to examine whether a progression along Bloom's revised taxonomy is demonstrated in the high school outreach program.

Results

The Results section of this paper serves a very simple purpose. The authors of this paper are trying to demonstrate that the log files from a chat room are a valuable, untapped source for understanding student learning, especially in an inquiry-based environment. Additionally, the authors examine the high school outreach program's ability to move students through the Bloom revised taxonomy in an unobtrusive measurement manner.

Frequencies

The raw frequency distribution of the chat conversation topics is shown in Table 1. While these numbers on their own are not intuitively insightful to the reader, they will be described in more detail as they are classified into the Bloom taxonomy. For the context of this paper, it is sufficient to note that nearly one-quarter of the lines of chat contained messages in which the students talked trash to each other about who was going to win the competition, made fun of each other's mistakes, talked about cars or other topics that came to the top of their mind. Also included in this category were conversations about what homework was due the next day, as well as girlfriends and jobs. An interesting anecdote from the author who conducted the coding is that these seeming random topics would be thrown in among the discussions of installing operating systems or configuring a firewall. As will be discussed in the Conclusions/Future Work section, this could be an additionally rich area to examine at the stream of conscious for the Millennial or Gen-Z students as they learn. However, the social aspect of the chat room, as well as the random jumping from topic to topic, is outside of the scope of this particular paper on inquiry-based learning and will be saved for future work with this data set. The 5540 lines of trash talk and general conversation, as well as the topics which did not represent at least 1% of the conversation lines, were removed from consideration. The remaining 17,210 lines were then used in the remaining analysis.

Table 1. Frequency Distribution of Chat Conversation Topics

Conversation Topic	Frequency	(%)
joking around / trash talk	5540	(23.8%)
connecting to us for setup	2433	(10.4%)
operating system/installation	2345	(10.1%)
networking	1932	(8.3%)
competition /event logistics	1612	(6.9%)
kvms / laggy / complaints about speed	1405	(6.0%)
firewalls	1397	(6.0%)
hello / goodbye	1130	(4.8%)
OS discussion of which is better	851	(3.7%)
hardware	846	(3.6%)
general "I have a question" comments	640	(2.7%)
attack vector discussion	551	(2.4%)
high school/student name	532	(2.3%)
frustration comments	473	(2.0%)
services they have to run	417	(1.8%)
web site installation / running	405	(1.7%)
configuring security	241	(1.0%)
users they have to create	172	(.7%)
college / post-secondary education	160	(.7%)
computer experience	115	(.5%)
programming /programming languages	73	(.3%)
ISEAGE facilities / how it works	30	(.1%)

Of the remaining lines of conversation in the chat log, 77.8% of the lines were made by high school students. The authors know that peer-to-peer learning is a strong resource in student learning in inquiry-based learning and it appears that this is substantiated in the chat conversation analysis. As shown in Table 2, high school students were the ones who were primarily in the chat room. And, anecdotally, it was found that many of the student interactions involved learning from each other and asking questions of each other. Students from different high schools were offering help and ideas to each other and were brainstorming with their "competition." Unfortunately, the authors did not code for that attribute in this first content analysis work. However, as discussed in the Conclusions/Future Work section, that coding is being considered for addition in future content analysis work.

Table 2. Distribution of High School vs. College Student Chat Lines

Student Classification	Frequency	(%)
high school student	13394	(77.8%)
college student	3816	(22.2%)

Table 3 shows the number of chat lines recorded in the four weeks leading up to the CDC event. While there were some students who joined the chat room early to discuss their CDC setup, nearly half of the chat conversation lines (42%) occurred two weeks before the competition, with nearly another third happening the final week of remote setup.

Table 3. Distribution of Chat Lines Based Upon Number of Weeks Prior to the CDC

Number of Weeks Prior to CDC	Frequency	(%)
1 week	4741	(27.5%)
2 weeks	7385	(42.9%)
3 weeks	4473	(26.0%)
4 weeks	611	(3.6%)

Bloom Classification

To evaluate whether the students in the program progressed along Bloom's taxonomy, the conversation topics were classified into one of the six Bloom categories based upon the characteristics of the tasks the students were doing and the types of discussions they had about the topic. The authors concede this is a broad stroke approach to evaluating the program in terms of Bloom's taxonomy. As discussed in the Conclusions/Future Work section, additional coding might be able to be done on each chat conversation line to classify it within each Bloom category. This would further define the students' progressions through the taxonomy. However, as this paper is the first application of content analysis, that approach was not taken and the topics were summarily categorized into one of the six categories. Each of the Bloom categories is listed below and examples of the types of activities that were covered by the students are given. Table 4 provides the frequency distribution of the conversation topics into Bloom's taxonomy.

Remembering*: Learning by retrieving, recognizing, identifying, describing, and naming a particular concept.*

Remembering is the lowest order thinking process and, therefore, several of the very basic conversations that occurred in the chat room were coded into this category. Many of the items included in this category in this analysis have little to do with the technical configuration of computers and networks. Conversations included in this category included both those that were more mechanical in the nature and conversations about the logistics of the CDC. The more interesting types of questions in this category included questions what to expect at the CDC, what time it started, asking for a ruling about whether an operating system or hardware configuration would be allowed, and the location of the event. Since the program was only in its second year in 2007 and the number of high schools and students who participated doubled in the one year period, these types of conversations were expected. The mentors and teachers were new to the high school outreach program and hadn't been to a CDC before so couldn't help their students understand what would happen when they arrived on the Iowa State University campus to compete in the event.

Table 4. Conversation Topics Classified According to Bloom's Taxonomy

Bloom Taxonomy	Frequency	(%)
Remembering	4387	(25.5%)
competition /event logistics		
hello / goodbye		
general "I have a question" comments		
high school/student name		
frustration comments		
Understanding	3279	(19.1%)
connecting to us for setup		
hardware		
Applying	4694	(27.3%)
operating system/installation		
networking		
services they have to run		
Analyzing	1638	(9.5%)
firewalls		
configuring security		
Evaluating	1807	(10.5%)
OS discussion of which is better		
attack vector discussion		
web site installation / running		
Creating	1405	(8.2%)
kvms / laggy / complaints about speed		

Other frequently seen conversation topics which were included in the Remembering category were students from other schools greeting each other as they logged in and logged off the chat room, as well as students wanting to know each other's real names, school names, and school locations. The college students manning the support desk started the process of greeting each team or person logging into the chat by saying "Hello Team X. Do you need any help?" as a way to make the students feel welcome and open the lines of communication. Soon all the high school students were greeting each of the team members as they logged in and also greeted the college students. The college students also let the high school students know that they were going to lunch or leaving the building for the night and the high school students soon started saying good night to each other and the college students as they were logging off for the night.

Additionally, the high school students asked about each other's name and school and would remember this from log in to log in. It became a very social environment for these students who had never met each other until they arrived four weeks later at the CDC. It also may account for some of the joking and trash talk that occurred in the chat room. They felt that they became virtual friends and could joke with each other. While these conversations were not as important

in setting up equipment, it did help to build an online community which it could be argued allowed the students to be comfortable to ask questions about technology within the chat room.

***Understanding**: Learning by interpreting explanations of a concept and then comparing this summarization to other concepts.*

There were two of the conversation themes that were placed into the Understanding category. The first was the most frequency made request in the analysis which was one for help in connecting to the ISEAGE testbed and understanding how to remotely access their equipment to begin configuration. Also included in this theme were requests to reset their remote connection password, as well as requests for the name of the server to connect to. Although there was a well-produced electronic document with pictures and short descriptions of how to make the connections and how to change their password, the students did not read the documentation which was available online.

The other conversation theme that was recorded as Understanding included questions about the hardware specifications (motherboards, memory, hard drive size, clock speed) that the students would be installing their operating systems and then configuring as servers.

Both themes were included as Understanding because the high school students could not do any further work with this information. They simply had to understand the information so they could setup their competition network remotely.

***Applying**: Learning by implementing and using a particular concept.*

Three different themes were included in this category: operating systems and their installation, networking, and which services the students had to run. These three themes actually go together in because the high school students would have questions during the installation about which services to turn on and how to configure their network. Many of these high school students had little to no experience with installing servers and questions about how to install the operating systems, as well as how to answer the questions during operating system installation were included here. Also because of inexperience, many of the high school students needed help with setting their IP numbers, default gateways, subnet masks, and understanding routing tables. Each of these themes were coded as Applying because the students had to learn what they were discussing in the chat to actually install their own servers and make those servers and their services work as a fully-functional network.

***Analyzing**: Learning through comparison and organization to allow for finding, structuring, and integration of information.*

There were two conversation themes that were placed in the Analyzing category. The first topic, the installation and configuration of firewalls is not a topic to be taken lightly. Even the community college and four-year students can have problems with the design, implementation, and successful management of a firewall. Although, the installation of the firewall software is very similar to the installation of an operating system, the configuration depends upon a complete understanding of how all the servers in the network work together and how the traffic needs to pass from the outside world to the inside world and vice versa. That is why the conversations where students were discussing the ingress and egress ports on the firewall, the

TCP/IP port numbers, traffic rules and network rewiring were coded in the Analyzing category. The students had to configure the firewall, troubleshoot what they had done when it didn't work, sometimes explain this information to a college student if they wanted help with a problem, and subsequently implement the necessary changes to get the wanted traffic into their network and keep the unwanted traffic out.

Likewise, configuring security on the servers was placed in the Analyzing category. While the high school students could answer questions during an installation script for an operating system installation, correctly implementing security required additional thought. Similar to the firewall configuration, setting up security on the servers requires students to evaluate the types of requests they wanted their server to allow and which they wanted to restrict or remove. Further, the high school students may have turned more services than they had intended during their operating system installation and securing the servers involved looking at what services were running on their servers and what they might want to remove.

***Evaluating**: Learning by experimentally hypothesizing, then testing and detecting conceptual ideas.*

Three conversation themes were placed in the Evaluating category. Conversations that focused on the strengths and weakness of different operating system were placed in this category. Many of these conversations focused on the differences between Windows, Mac and Unix as operating systems, both for desktop machines and for servers. Some of the conversations were purely antagonistic among various OS camps, however, even these conversations sometimes would develop into critical discussions evaluating the various pros and cons of different OS environments. Another conversation theme regarding the attack vectors that the could be utilized in the competition against them also demonstrated that at least some of the students were able to conceptualize how they might be attacked. Students attempted to hypothesize which attacks would and could be used against their systems and how they could potentially ward off such attacks. Finally, website installation and running was discussed regarding the testing of various configurations to provide service to clients while concurrently securing the website from attack.

***Creating**: Learning through inventive design and planning to construct novel solutions to problems.*

While the authors could argue that the whole CDC and the high school outreach program on a whole is a demonstration of the high school students Creating, in the context of the conversation themes, there was one topic which found students reinventing and circumventing a process we had established for them. The high school students were connecting to the competition hardware through network KVMs, the best of which are notorious for having technical problems and introducing some lag time in response to commands being issued. Since the high school students had never worked with these types of devices, they initially were extremely frustrated with the devices and began complaining about the slow response from the systems. However, the discussions quickly evolved into students designing workarounds to avoid using the networked KVMs and sharing these solutions with other high school students in the chat room. The most frequently mentioned solution was to only use the KVM for installing the operating systems. Once the operating system was installed, the high school students would enable ssh as a service and then connect directly to the server, bypassing the network KVM completely.

Conversation Themes and Bloom's Taxonomy Across Time

Table 5 shows the conversation themes distribution by weeks prior to the CDC. As would be expected, four weeks before the competition, the high school students were discussing how to connect to the ISEAGE research facility. The week prior to the CDC, networking was the most common topic being discussed.

Table 5. Conversation Themes by Weeks Prior to the CDC

	Weeks Before CDC			
	1	2	3	4
Remembering				
	268	632	672	40
competition /event logistics	(5.7%)	(8.6%)	(15.0%)	(6.5%)
	271	451	326	82
hello / goodbye	(5.7%)	(6.1%)	(7.3%)	(13.4%)
	85	125	375	55
general "I have a question"	(1.8%)	(1.7%)	(8.4%)	(9.0%)
	135	250	120	27
high school/student name	(2.8%)	(3.4%)	(2.7%)	(4.5%)
	48	290	135	0
frustration comments	(1.0%)	(3.9%)	(3.0%)	(0%)
Understanding				
	423	1339	448	223
connecting to us for setup	(8.9%)	(18.1%)	(10.0%)	(36.5%)
	260	427	151	9
hardware	(5.5%)	(5.8%)	(3.4%)	(1.3%)
Applying				
	564	1098	675	8
operating system/installation	(5.5%)	(14.9%)	(15.1%)	(1.3%)
	1023	559	350	0
networking	(21.6%)	(7.6%)	(7.8%)	(0%)
	188	63	168	0
services they have to run	(4.0%)	(.9%)	(3.7%)	(0%)
Analyzing				
	215	996	186	0
firewalls	(4.5%)	(13.5%)	(4.2%)	(0%)
	114	101	26	0
configuring security	(2.4%)	(1.4%)	(.6%)	(0%)
Evaluating				
	205	209	313	124
OS discussion of which is better	(4.3%)	(2.8%)	(7.0%)	(20.3%)
	378	123	50	0
attack vector discussion	(8.0%)	(1.7%)	(1.1%)	(0%)
	233	130	40	2
web site installation / running	(4.9%)	(1.8%)	(.9%)	(.3%)
Creating				
	331	592	440	42
kvms / laggy / complaints about speed	(7.0%)	(8.0%)	(9.8%)	(6.9%)

Table 6 shows the conversation themes grouped into Bloom's taxonomy as they were discussed each week prior to the CDC. In the four weeks prior to the competition, more than 70% of the students' discussions were spent in discussing topics in the Remembering or Understanding

category. This is to be expected because the remote setup and chat room had just opened and the high school students had lots of questions on how to get started with their server configurations and setup. Interestingly, 20.6% of the conversation lines were in discussions about operating systems and evaluating which was better for providing what services. This does not necessarily follow Bloom's logical progression up the taxonomy. The high school students were talking to the college students and each other trying to determine what operating systems to select to run on the equipment they were given, many having never installed (or been in the Applying stage) a server.

Two weeks prior to the CDC, the lowest three levels of the taxonomy each accounted for a quarter of the conversation topics. Each had approximately 24%. The final week before setup, we find nearly as many students in the Evaluating stage as we did the first week the setup was open (four weeks before the competition). However, the high school students were not talking as much about which operating system to select, but more about the attack vectors and the insecure web site that they had and how to make it secure. Unfortunately, there were also a large number of students in the Applying stage. This is probably due to the number of students who waited until the last minute to setup their competition network. As discussed in the Conclusions/Future Work section, a way is needed to note the first time a student logs into the chat room which will demonstrate how much time on task he/she has and then look at their conversation's placement in the Bloom taxonomy.

Table 6. Conversation Topics Grouped by Bloom's Taxonomy by Weeks Prior to CDC

	Weeks Before CDC			
	1	2	3	4
Remembering	807	1748	1628	204
	(17.0%)	(23.7%)	(36.4%)	(33.4%)
Understanding	683	1766	599	231
	(14.4%)	(23.9%)	(13.4%)	(37.8%)
Applying	1775	1720	1191	8
	(37.4%)	(23.3%)	(26.6%)	(1.3%)
Analyzing	329	1097	212	0
	(6.9%)	(14.9%)	(4.7%)	(0%)
Evaluating	816	462	403	126
	(17.2%)	(6.3%)	(9.0%)	(20.6%)
Creating	331	592	440	42
	(7.0%)	(8.0%)	(9.8%)	(6.9%)

Conclusions/Future Work

This paper represents a first attempt to use content analysis to evaluate inquiry-based learning couched in Bloom's taxonomy. The authors took this approach to better understand their own high school outreach program, as well as provide a novel method to evaluating inquiry-based learning methods within an active learning environment. While the results demonstrated that students were in the Applying, Analyzing and Evaluating categories, there is enough subjectivity in the Bloom categorization that the authors cannot unilaterally declare their program a perfect example of achieving educational learning objectives in a linear pattern, moving from the lowest order cognition to the highest. Instead, the authors showed that the students did perform active and complex thinking in designing, configuring, and securing their CDC competition networks. The authors also demonstrated that inquiry-based learning is a very good way for students with little to no previous skill with IT to learn about it.

Although the results were positive and the use of content analysis was successful, there are several things that the authors would like to implement in future work and with the full five year data set. The assumption was made when the data analysis began that all students would start their network configuration at the same time, exactly one month prior to the CDC. Therefore, all of the analysis was conducted on time relative to the date the remote setup was opened. However, as can be seen in the Results section, there is a spike two weeks before the competition and again the week before the competition of students asking for connection information and for information about installing operating systems. In other words, there are students two weeks before the event who are just entering the Bloom taxonomy at the Understanding stage or just starting the Applying stage. The authors believe that some students procrastinated on their remote setup and these students couldn't move as far through the Bloom taxonomy as the students who started four weeks prior to the CDC.

To allow the authors to statistically control for the amount of time the student spent on the task of setting up for the CDC, a code needs to be added to mark when the student first logged and was active in the chat room. This will be difficult to do because, as was discussed in the Methodology section, the students log in as a Team and then change their nickname to anything they want. Identifying individual students is impossible. However, as a substitute for the individual student, the first time a team logs in could be recorded and the number of times the team makes an entry in the chat room. This could approximate time on task and could be used in determining if more time on task would show the students progressing farther up the Bloom taxonomy.

Additionally, because a team is tied to a high school, the number of years the high school has participated in the outreach program could be coded to approximate the experience the student has. Granted, individual students may only be first years, but if playing on a team from a high school who has participated for four years, the students have access to other teammates, as well as mentors, who might help them move upward faster in the Bloom taxonomy.

Although this paper took the approach to code the each conversation theme into one Bloom category, in future work the authors may want to code each individual chat conversation line as to its place in the Bloom taxonomy. With each individual line of the log file being coded, it might be possible to more closely track the movement of topics across the categories along time.

While this paper coded whether a college student or a high school student typed the line in the chat room, it would be interesting to add a code that identified if the line of code was a question, a statement or an answer. Further, a code should be added to show if the conversation was between two high school students, two college students or between a college student and a high school student. The original assumption was that the majority of the conversations would be between the high school students as the question askers and the college students who would provide the answers. However, there were more lines of high school students answering each other's questions than was anticipated. Also, the college students used the chat room to communicate to each other about different problems each team was having and providing solutions to each other.

This paper did not attempt to code the rich social networking and conversations that occurred in the chat logs. The authors anticipated some socialization, but found it a heavy component of the chat room. The examination of the social aspect will be examined in future works and maybe found to be synergistic in its effect since a sense of community is established through it.

While coding of content analysis relies heavily upon manual coding which is time consuming, it provides a richness of understanding of student learning and interactions in inquiry-based learning that other methods do not cover. While the authors have five years of chat log data, this paper is an evaluation of the first year of chat logs and constructs the framework and methodology for utilizing content analysis to understand student learning in inquiry-based learning programs. Building this framework allows us, and others, to extend it to the examination of other types of logs where students are communicating to each other both socially and about their coursework and projects. Course management software, as well as social media, has similar log files and, with the appropriate permissions, evaluations of these messages could also provide a valuable look into what students are thinking and discussing while they are learning.

While this is a very simple content analysis, it is provided as a proof of concept that chat logs can be a rich source of information and many educators should be looking at what can be gleaned from such records. Since we have proven that content analysis methods can be used to evaluate student inquiry-based learning using chat logs, we now will look to conduct a much richer content analysis of the remaining five years of data. Information such as peer-to-peer learning, if there is an order in which they learn about information assurance and potentially coding of students talking to the own teams vs. other teams and the college students could be included.

---

[1] Committee on the Development of an Addendum to the National Science Education Standards on Scientific Inquiry, National Research Council. 2000. Inquiry and the National Science Education Standards: A Guide for Teaching and Learning. Washington, D.C.: National Academies Press.

[2] Bloom B. S. 1956. Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

[3] Anderson, L. W., & Krathwohl, D. R. (eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

[4] Committee on the Development of an Addendum to the National Science Education Standards on Scientific Inquiry, National Research Council. 2000. Inquiry and the National Science Education Standards: A Guide for Teaching and Learning. Washington, D.C.: National Academies Press.
[5] Llewellyn, D. 2004. Teaching High School Science through Inquiry. Thousand Oaks, CA: Corwin Press, Inc.
[6] J.A. Rursch and D. Jacobson, "Using cyber defense competitions to build bridges between community colleges and four year institutions: A footbridge for students into an IT program," in Proc. Frontiers in Education, San Antonio, TX, Oct. 18-21, 2009.
[7] Neuendorf, K.A., *The content analysis guidebook*. 2002, Thousand Oaks, Calif.: Thousand Oaks, Calif. : Sage Publications.
[8] Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. (1956). *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain*, New York: Longman.
[9] Anderson , L.W., and Sosniak, L.A. (Eds.). (1994). *Bloom's taxonomy: a forty-year retrospective. Ninety-third yearbook of the National Society for the Study of Education, Pt.2 .*, Chicago: University of Chicago Press.
[10] Anderson, L.W., and Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition*, New York: Longman.