

Methodology

The Internet blocking software study methodology consists of these top-level steps:

- Determine study parameters
- Extract topics from state-mandated curriculums into database
- Generate search strings
- Generate web page lists
- Test web pages against Internet blocking products
- Verify blocking software blocking codes
- Compare and analyze results

Determine Study Parameters

Researchers began by selecting which material to study. Specifically, they chose which state curriculums to study, which search engine to use, and which Internet blocking software packages to test.

- 1) Choose curriculums to study.

Researchers chose curriculums from the following states for the following reasons:

- California has the nation's largest public school system. [15]
- Massachusetts is the state with the first public high school in the United States. [1]
- North Carolina represents a more rural sector of American society than the other two states. [30]

Curriculums from these three states provide a wide cross-section of the American educational system and the values it represents.

In light of the current trend of increasing use of the Internet by young children, researchers chose to study topics from curriculums designed for pre-kindergarten to grade 12.

In order to perform an exhaustive study, researchers included every topic mandated by the state curriculums.

2) Choose search engine.

Researchers chose Google, the largest search engine on the World Wide Web [10]. Google's unique search algorithm produces highly relevant search results.

Because Google accepts no more than ten search terms per search query, researchers had to determine a method for reducing topic texts to ten words or less, as described in "Generate Search Strings" below.

In order to generate a large enough sample size of web addresses to test, the researchers obtained up to 50 search result web addresses for each curriculum topic.

3) Choose Internet blocking software packages.

The researchers chose N2H2's Bess and SurfControl's SurfControl Internet blocking software because the two Internet blocking software packages chosen are reportedly the most widely used Internet blocking software in U.S. schools. The researchers used N2H2 Bess and SurfControl Web Filter 4.0, the server-based product sold to schools.

- A study commissioned by the U.S. Department of Justice determined that N2H2's Bess is the most effective Internet blocking program available. [29]
- SurfControl claims the largest market share of any Internet blocking software company with its SurfControl (formerly CyberPatrol) product. [2]

Extract Topics from State-Mandated Curriculums into Database

The next part of this study examined the chosen curriculums and recorded the observations into a comprehensive database. To transcribe the school curriculums, researchers recorded the topics, each of which included a grade-level designation and a hierarchy of broader categories. The leftmost column of categories contained general subjects such as "science," while further right columns held more specific topics, such as "kinetic energy." Researchers used a maximum of seven hierarchical category levels for each topic. The compiled database both aided in the assessment of the curriculum and serves as an accurate record for any independent audits of this study.

Researchers used this procedure to extract data from each state-mandated curriculum and compile a curriculum database.

1) Find and store a copy of each state-mandated curriculum:

- California curriculum obtained on June 16, 2002, from <http://www.cde.ca.gov/board/> [3]
- Massachusetts curriculum obtained on January 10, 2002, from <http://www.doe.mass.edu/frameworks/current.html> [14]
- North Carolina curriculum obtained on August 26, 2002, from <http://www.ncpublicschools.org/curriculum/> [16]

- 2) Familiarize yourself with the curriculum topic areas such as “Arts.”
- 3) Identify all top-level subject categories in the curriculum.
- 4) Identify subcategories within each curriculum category.
- 5) Identify topics within each curriculum subcategory.
- 6) Record each category, subcategory, and topic combination in a database entry.
 - Use copy and paste functions as much as possible to extract the exact curriculum topic verbatim.
 - Give each category or topic its own cell.
 - Add a row for each category, subcategories, and topic combination as needed.
- 7) Label appropriate grade level corresponding to each curriculum topic.
 - For pre-kindergarten and kindergarten, assign a value of 0 for grade level.
 - For curriculum topics designed for multiple grades, identify the intended range (e.g. assign 5-8 for grades five through eight)
- 8) Moving across the spreadsheet from left to right, the level of detail should become more specific as the reader progresses from category to subcategories to topic (refer to example spreadsheet in Appendix A).

Generate Search Strings

Having compiled the curriculum database, the researchers next created search phrases for every line of the curriculum. These search phrases, or search strings, consisted of the most important features of each topic line. Researchers attempted to limit the search strings to no more than 10 words because Google limits each query to a maximum of 10 words. Placing double quotation marks around a series of words or phrase generates websites that contain only that explicitly ordered combination. For example, searching for “United States” will yield different results than searching for “United” “States”. If the topic were “The History of the United States,” the researchers would include pages discussing “United States History” in the search results by searching for “history” and “United States” rather than placing the entire topic text within the quotation marks.

The researchers introduced a small degree of subjectivity into the study in distinguishing the search strings. However, determining appropriate search strings served the study’s purpose better than entering the full curriculum, which would have in most cases limited the search to search results that used phrasing identical to that of each curriculum. Clearly, students would not necessarily enter searches identical to the state-mandated curriculum topics, but the key words used for the searches should be similar to student searches performed in those topic areas.

The researchers used the following procedure to generate search phrases:

- 1) Choose words from each topic description that convey the essence of each line of the curriculum.
- 2) Limit the search phrase to 10 words by culling the topic text down to that limit.
- 3) Place double quotation marks around any combination of words that would most likely fit within the topic only if they occurred in that specific phrase. Try to limit word combinations to two words (e.g. "United States" or "Molecular Biology," not "social structures of civilized Japan.")
- 4) Collect each search phrase into a single column in the row that corresponds to the category, subcategories, and topic combination from which the researcher generated that search phrase.

Using this system, researchers generated search strings using text culled directly from each state-mandated curriculum.

Generate Web Page Lists

Researchers entered the list of search phrases into a specially developed software application that ran each phrase through the search engine and recorded all of the up to 50 web page search results per topic in a database.

The researchers used the following procedure to generate web page lists:

- 1) Run search strings through the search engine.
- 2) Tabulate web addresses from search results.

If a web page appeared more than once in a set of search results, the researchers entered the page into the database as many times as it appeared in search results.

Test Web Pages Against Internet Blocking Software

Researchers then tested each web address obtained above against the blocking software. They recorded each blocked page alongside the name of the software that blocked that page and the blocking code or codes specified by each Internet blocking software product.

The researchers used the following procedure to refine the web page lists:

- 1) Test web addresses against each blocking software product.
- 2) Record which websites are inaccessible, unblocked, or blocked, and if blocked, note which blocking code or codes the blocking software specified.

In the case of the SurfControl blocking software, the web addresses were tested against a web-based blocking software test tool provided by SurfControl. A SurfControl representative confirmed that the SurfControl test tool provides the same results as the product sold to schools.

For N2H2's Bess product, the web addresses were tested against a Bess installation at a public high school. Schools will install N2H2 Bess in various configurations, so this installation of Bess reflects the settings at that school only, although the researchers speculate that the Bess settings for the school in this study are similar to those of many other schools. Additionally, the blocking codes used by Bess are the same for every installation, just the decision by the school of whether or not to activate particular blocking codes varies according to the installation.

Verify Blocking Software Blocking Codes

In certain conditions, determining which content is objectionable can be subjective. Blocking software manufacturers may assign a web page an appropriate block code but still block the web page incorrectly; conversely, they may assign the wrong block code, even if the blocking software blocks a web page appropriately. For example, material about contraceptives inappropriate for an elementary school student may be entirely necessary for a high school student. [21] Material by the Ku Klux Klan may be hate speech in one context or primary source material in the context of a research report about the history of the American south. Internet blocking companies or schools configuring the Internet blocking software may, and often do, block "controversial" web pages due to political, social, or cultural biases, regardless of whether they fall within the federally mandated guidelines for material that educators must block their students from accessing. Thus, researchers tested two conditions: whether blocking software manufacturers assigned web pages the appropriate blocking code, and whether the blocking software blocked sample web pages appropriately.

To verify that Internet blocking companies assigned the correct block codes to web pages blocked in the study, researchers examined a statistically significant sample of the pages blocked by each blocking product and checked block code assignments using the block code definitions provided by the blocking companies (and included in Appendix B).

The researchers tested for overblocking by checking a sample of pages blocked by each blocking product to determine if a court reviewing a legal challenge would likely agree that CIPA authorized blocking of those web pages. Additionally, the researchers also tested for overblocking by rating the same sample of blocked pages using the Internet Content Rating Association rating system.

To detect potential underblocking, the researchers also ran some of the unblocked web pages through a set of pornographic key word searches in an attempt to find any web pages that the Internet blocking software neglected to block according to its block code definitions related to CIPA requirements.

Compare and Analyze Results

The primary objective of the study was to analyze the differences between the degree of access to curriculum-related web pages on computers running with and without Internet blocking software. The study also determined the percentage of web pages blocked with the blocking software in each major category, each state, and each grade level.

Comparing the samples offered evidence of the ineffectiveness of blocking products at both allowing access to educationally-appropriate pages and at blocking access to pages that are likely legally inappropriate for use in schools, although the courts have not provided much clear direction as to which specific materials are illegal in schools beyond the general categories of verboten content specified in CIPA and other laws. Sometimes, the researchers simply determined whether or not the Internet blocking software blocked web pages as advertised since the block codes were clearly not within the realm of harmful to minors content prohibited by CIPA or other laws addressing information access within schools.

The researchers used the following procedure to analyze the results:

- 1) Determine the statistical significance of differences in results between the samples.
- 2) Determine the percentage of web pages blocked inappropriately with the blocking software in each major category, state, and grade level, as well as by blocking product.
- 3) Sample web pages inaccessible while using blocking software – check for educationally appropriate material (overblocking).
- 4) Sample websites accessible while using blocking software – check for likely legally inappropriate material (underblocking).