

The research results of this study fit into the following categories:

- Overblocking
- Underblocking
- Overall blocking rates
- Blocking by category/topic, including categorization accuracy analysis
- Blocking by state
- Blocking by grade
- Blocking by product
- Bad address and unreachable statistics

Overblocking

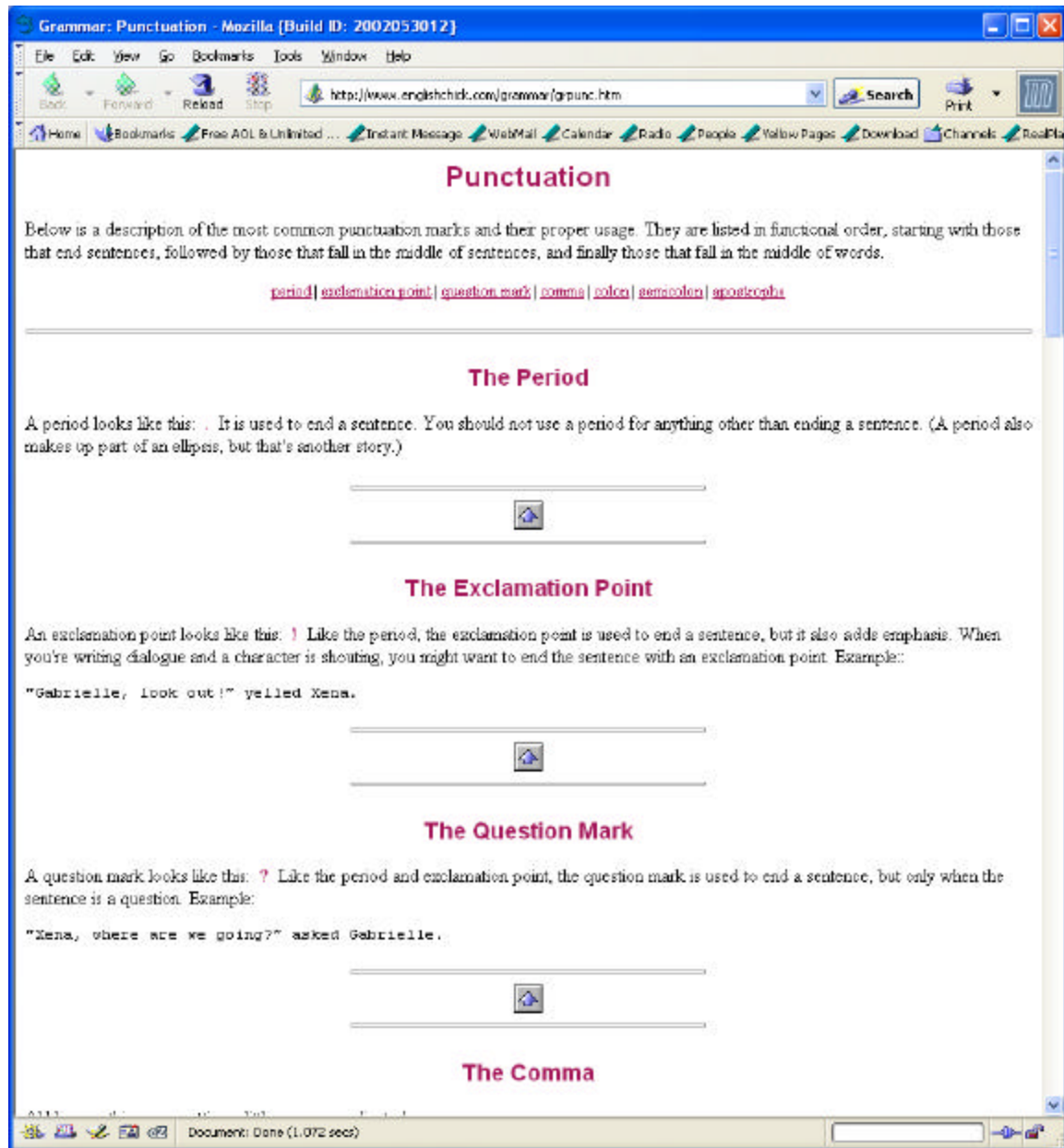
The study found the following results regarding blocking software overblocking:

- For every web page blocked as advertised, blocking software blocks one or more web pages inappropriately, either because the web pages are miscategorized or because the web pages, while correctly categorized, do not merit blocking.
- Schools that implement Internet blocking software even with the least restrictive settings will block at a minimum tens of thousands of web pages inappropriately, either because the web pages are miscategorized or because the web pages, while correctly categorized, do not merit blocking.
- Blocking software products miscategorized many of the web pages they block—assigning the wrong block codes to between a third and a half of the web pages related to state-mandated curriculums blocked depending on the blocking software.
- Of all pages related to state-mandated curriculums blocked by blocking products, the products blocked only 1-3% of those web pages to CIPA's criteria for blocking visual depictions of illegally obscenity, child pornography, or harmful to minors content. That means that of the web pages related to state-mandated curriculums, blocking software products blocked 97-99% of the web pages blocked using non-standard, discretionary, and potentially illegal criteria beyond what is required by CIPA.

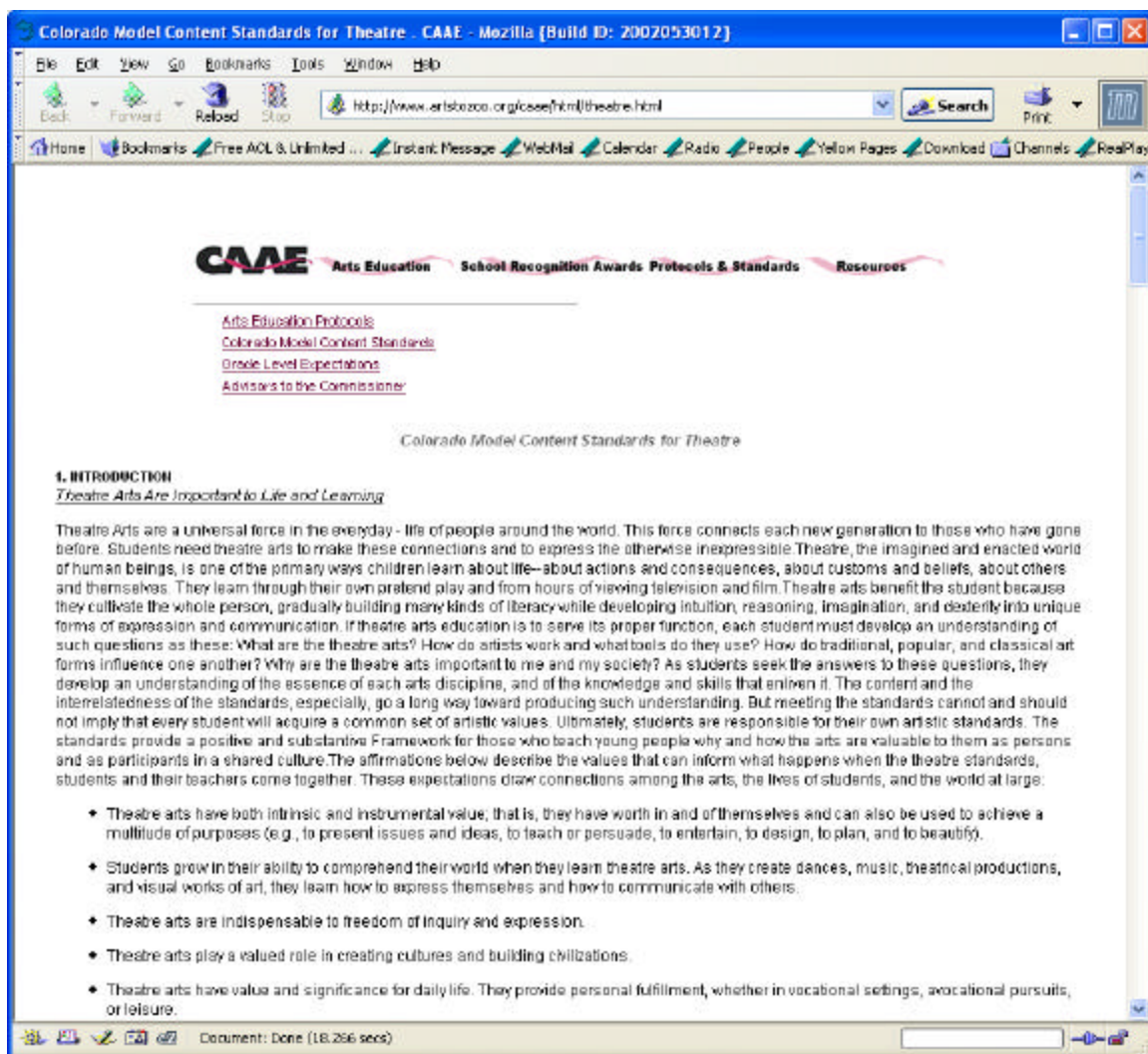
Overblocking Types

Researchers came across two types of overblocking in the research for this study:

- 1) Miscategorization or Inappropriate Coding: Internet blocking company assigned web pages an inappropriate blocking code or codes. For example, the SurfControl Internet blocking software product assigned a Punctuation Primer located at <http://www.englishchick.com/grammar/grpunc.htm> the block code Adult/Sexually Explicit, even though the most controversial topics on that page are the period and the exclamation point. Perhaps “period,” as in menstruation, was the trigger?

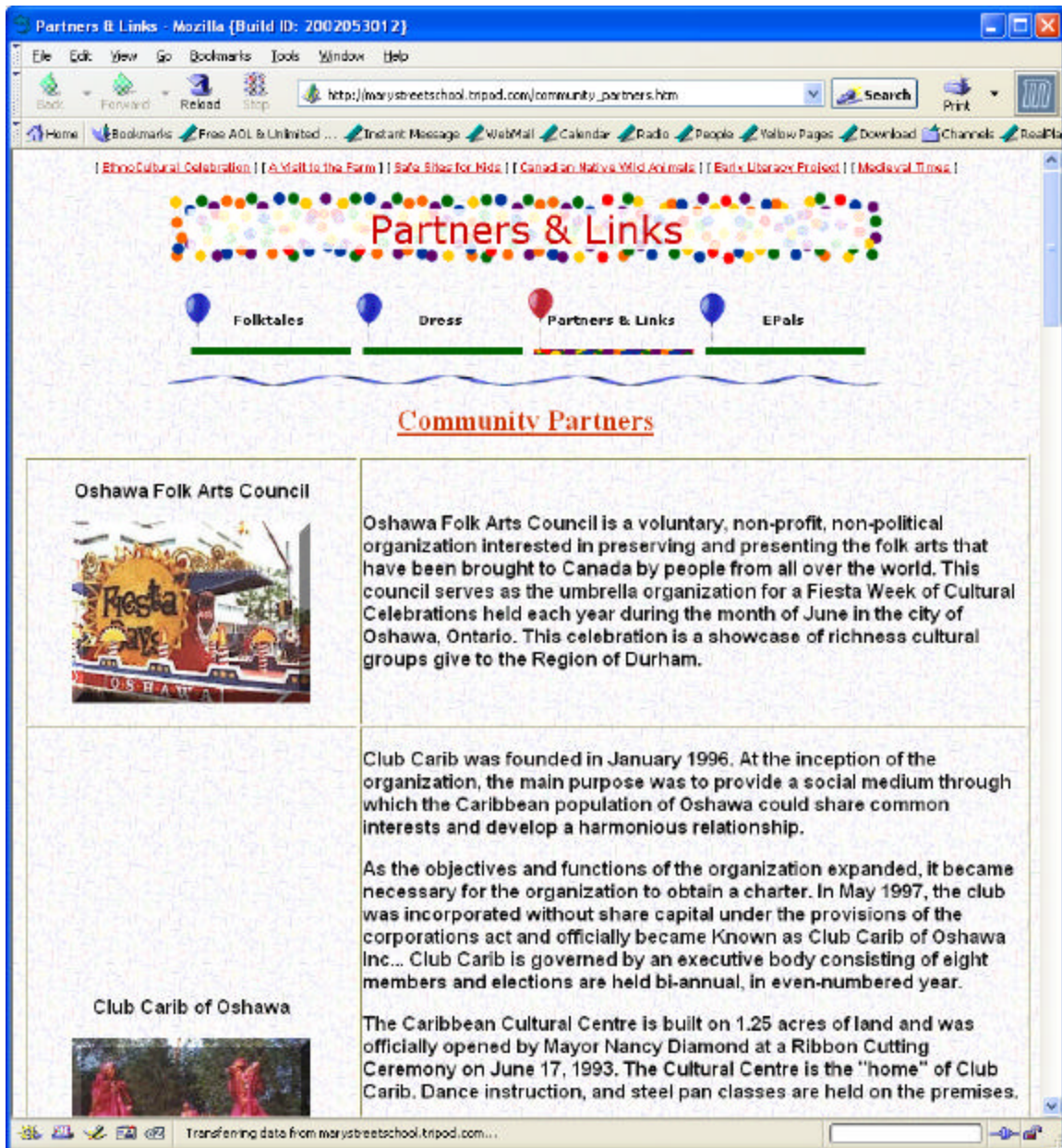


In another example, N2H2 Bess assigned a Colorado Arts Education page on Colorado Model Content Standards for Theatre a Pornography block code:



- 2) Inappropriate Content Evaluation: web pages assigned an apparently correct blocking code or codes, but still not blocked appropriately. N2H2 blocks wide swathes of the Web under the Free Pages, Electronic Commerce, Message/Bulletin Board, and Recreation/Entertainment categories.

For example, N2H2 Bess assigned a page describing community partners of the Mary Street School located at http://marystreetschool.tripod.com/community_partners.htm the block code "Free Pages" because it is hosted for free at Tripod, even though the page includes pedagogically valuable information.



The N2H2 Bess definition for the Electronic Commerce block code seems much more concerned with other content areas than with confronting commercialism within schools:

Electronic Commerce

Sites that allow users to make online purchases. Many e-commerce sites pose a risk to users by offering direct access to items that would normally be filtered under other categories such as Weapons, Profanity, Lingerie, or Pornography.

If the e-commerce sites offer access to Weapons, Profanity, Lingerie, or Pornography, why not block them using those block codes? Or should everything that appears on an electronic commerce site get blocked simply because it might be associated with some other blocking code? For example, N2H2 Bess blocks proceedings of Association for Computing Machinery conferences under the Electronic Commerce code.

Overblocking Standards

To determine overblocked pages, researchers sampled the overall list of blocked pages and evaluated each one to see if the blocking software manufacturer assigned codes correctly, as well as if, even with an apparently correct blocking code, the product should have blocked the web page.

It was difficult to come up with any standard for determining whether or not a blocking product should block a given web page beyond the "I know it when I see it" standard, which was so subjective that the researchers had to reject it for this study.

Instead, the researchers decided to report overblocking rates that reflected what a U.S. court would reasonably find required by CIPA in according to average U.S. community standards as described in this document's Preface. This formulation was sometimes also subjective, although in surprisingly few cases. It provided a much more stable means for measuring overblocking.

The caveat to the CIPA overblocking method is whether it is permissible for schools to use blocking technology to block web pages related to state-mandated curriculums using some standard other than CIPA. The researchers predict that the public policy in this arena will develop further over time.

As an additional independent and less subjective check on overblocking, the researchers rated a sample of the web pages using the Internet Content Rating Association rating system as described below.

Overblocking Rates

When using the criterion of blocking only web pages a court would reasonably find required by CIPA in according to average U.S. community standards, the researchers found that the Internet blocking software has overblocking rates of 97-99% of web pages related to state-mandated curriculums depending on the blocking product. That means that according to the researchers only 13% of the blocked pages related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content. No child pornography was found.

When the definitions of CIPA are extended to include non-visual depictions of illegal obscenity, child pornography, or harmful to minors content, the researchers found that the Internet blocking software has overblocking rates of 95-98% of web pages related to state-mandated curriculums depending on the blocking product. That means that according to the researchers only 2-5% of the blocked pages related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content. No child pornography was found.

N2H2 Bess Overblocking: Non-Visual Depictions

Researchers tested 315 pages spread evenly over the entire sample of web pages blocked by N2H2 Bess and found that of the 294 pages that were accessible, the blocking software overblocked 98.98% $\pm 0.30\%$ (95% confidence interval) (291 web pages) of web pages related to state-mandated curriculums when using the criterion of blocking only web pages a court would reasonably find required by CIPA according to average U.S. community standards. That means that according to the researchers 1.02% $\pm 0.30\%$ of the pages blocked by N2H2 Bess related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content. No child pornography was found.

N2H2 Bess Overblocking: Visual and Non-Visual Depictions

Researchers tested 315 pages spread evenly over the entire sample of web pages blocked by N2H2 Bess and found that of the 294 pages that were accessible, the blocking software overblocked 98.30% $\pm 0.39\%$ (95% confidence interval) (289 web pages) of web pages related to state-mandated curriculums when using the criterion of blocking web pages a court would reasonably find required by CIPA *along with non-visual depictions* of illegal obscenity, child pornography, or harmful to minors content according to average U.S. community standards. That means that according to the researchers 1.70% $\pm 0.39\%$ of the pages blocked by N2H2 Bess related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content along with non-visual depictions of illegal obscenity, child pornography, or harmful to minors content according to average U.S. community standards. No child pornography was found.

SurfControl Overblocking: Non-Visual Depictions

Researchers tested 352 pages spread evenly over the entire sample of web pages blocked by SurfControl and found that of the 324 pages that were accessible, the blocking software overblocked 97.22% $\pm 0.52\%$ (95% confidence interval) (315 web pages) of web pages related to state-mandated curriculums when using the criterion of blocking only web pages a court would reasonably find required by CIPA in according to average U.S. community standards. That means that according to the researchers 2.78% $\pm 0.52\%$ of the pages blocked by SurfControl related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content. No child pornography was found.

SurfControl Overblocking: Visual and Non-Visual Depictions

Researchers tested 352 pages spread evenly over the entire sample of web pages blocked by SurfControl and found that of the 324 pages that were accessible, the blocking software overblocked 95.37% $\pm 0.66\%$ (95% confidence interval) (309 web pages) of web pages related to state-mandated curriculums when using the criterion of blocking web pages a court would reasonably find required by CIPA *along with non-visual depictions* of illegal obscenity, child pornography, or harmful to minors content according to average U.S. community standards. That means that according to the researchers 4.63% $\pm 0.66\%$ of the pages blocked by SurfControl related to state-mandated curriculums fit into the CIPA categories of illegal obscenity, child pornography, or harmful to minors content along with non-visual depictions of illegal obscenity, child pornography, or harmful to minors content according to average U.S. community standards. No child pornography was found.

Verification by Internet Content Rating Association Rating System

To provide more independent verification of the inappropriate blocking, the researchers used the rating system developed by the Internet Content Rating Association (ICRA) to rate each web page according to the ICRA's rating system including blocking codes of Nudity and Sexuality, Violence, Other Topics (such as promotion of tobacco, alcohol, or drug use, discrimination, and gambling), and Chat. [11] Although the researchers in this study retain a healthy skepticism of the effectiveness of the ICRA rating system, many accept it as a standard, including AOL and Yahoo!. [12]

The ICRA data confirmed the conclusion that a relatively small number of web pages returned as search results from state-mandated curriculum topics would have any content for which a court would likely require blocking under CIPA. A maximum of between 5.10% and 9.26% of the web pages rated using the ICRA content rating system could be relevant to blocking as required by CIPA and this represents only an upper bound and not an actual figure, since the ICRA content rating system does not distinguish visual from non-visual depictions and includes items such as "passionate kissing" and "obscured or implied sex," among other specific Nudity and Sexual Material ratings that are not relevant to CIPA.

The researchers rated a distributed sample of web pages for both blocking software products using the Internet Content Rating Association (ICRA) content rating system. Although the researchers collected data using all the ICRA ratings, the study report focuses on the Nudity and Sexual Material rating data since that is the only data relevant to Internet blocking as required by CIPA. The researchers present the ICRA data for other ratings just for reader interest.

Please note that the ICRA content rating system provides for context ratings, such as artistic, educational, or medical "and is suitable for young children" intended to provide exceptions to the Nudity and Sexual Material ratings.

N2H2 Bess ICRA Data

Testing a distributed sample of 294 web pages, the researchers found that 94.56% had no rating related to Nudity and Sexual Material. 4.76% had a rating of "Obscured or implied sex" and 0.68% had a rating of "Passionate kissing." Additionally, 0.34% of the web pages

rated as Nudity and Sexual Material included one or more of the three ICRA context exceptions, meaning that at least 94.90% of the web pages either had no or an irrelevant Nudity and Sexual Material rating. Of the less than 5.10% of the web pages that had a relevant Nudity and Sexual Material rating, many of those web pages had no visual depictions, meaning that a court interpreting CIPA using average U.S. community standards would be likely to require blocking of only a much smaller percentage than 5.10% of the web pages. It was impossible to determine the exact percentage without a much more detailed examination of the data.

This 5.10% figure is significantly larger than the 1.70% figure obtained by the researchers for pages blocked by N2H2 Bess that would fit into the CIPA categories if and only if non-visual depictions not addressed by CIPA are included in both data points. This either means that N2H2 Bess is underblocking a significant portion of the sample or that the ICRA ratings are not specific enough to provide a good basis on which to determine appropriate blocking of combined visual and non-visual depictions of materials covered by CIPA categories.

For the same distributed sample of 294 web pages, the researchers found that 95.24% had no rating related to Violence. Additionally, 0.68% of the web pages rated as Violence included one of the three exceptions, meaning that at least 95.92% of the web pages either had no or an irrelevant Violence rating. Thus, the researchers found that only 4.08% of the web pages rated as Violent, for none of which a court interpreting CIPA using average U.S. community standards would likely require blocking. For the same distributed sample of 294 web pages, the researchers found that 97.28% had no rating related to Language. In fact, only 1.02% of the web pages rated as Language included "sexually explicit" language, but ICRA's "sexually explicit" Language rating does not include any determination of visual depictions, so a court interpreting CIPA using average U.S. community standards would not likely require blocking of any of these pages.

For the same distributed sample of 294 web pages, the researchers found that 96.60% had no rating related to Other Topics, and for those web pages that did rate in Other Topics a court interpreting CIPA using average U.S. community standards would not likely require any blocking.

For the same distributed sample of 294 web pages, the researchers found that 100.00% had no rating related to Chat, and a court interpreting CIPA using average U.S. community standards would not likely require any blocking of those web pages in any case.

SurfControl ICRA Data

Testing a distributed sample of 324 web pages, the researchers found that 81.17% had no rating related to Nudity and Sexual Material. 13.58% had a rating of "Obscured or implied sex" and 0.62% had a rating of "Passionate kissing." Additionally, at least 9.57% of the web pages rated as Nudity and Sexual Material included one or more of the three ICRA context exceptions, meaning that at least 90.74% of the web pages either had no or an irrelevant Nudity and Sexual Material rating. Of the less than 9.26% of the web pages that had a relevant Nudity and Sexual Material rating, many of those web pages had no visual depictions, meaning that a court interpreting CIPA using average U.S. community standards would be likely to require blocking of only a much smaller percentage than

9.26% of the web pages. It was impossible to determine the exact percentage without a much more detailed examination of the data.

This 9.26% figure is significantly larger than the 4.63% figure obtained by the researchers for pages blocked by SurfControl that would fit into the CIPA categories if and only if non-visual depictions not addressed by CIPA are included in both data points. This either means that SurfControl is underblocking a significant portion of the sample or that the ICRA ratings are not specific enough to provide a good basis on which to determine appropriate blocking of combined visual and non-visual depictions of materials covered by CIPA categories.

For the same distributed sample of 324 web pages, the researchers found that 95.99% had no rating related to Violence. Additionally, 2.47% of the web pages rated as Violence included one of the three exceptions, meaning that at least 98.46% of the web pages either had no or an irrelevant Violence rating. Thus, the researchers found that only 1.64% of the web pages rated as Violent, for none of which a court interpreting CIPA using average U.S. community standards would likely require blocking.

For the same distributed sample of 324 web pages, the researchers found that 95.68% had no rating related to Language. In fact, only 3.70% of the web pages rated as Language included "sexually explicit" language, but ICRA's "sexually explicit" Language rating does not include any determination of visual depictions, so a court interpreting CIPA using average U.S. community standards would not likely require blocking of any of these pages.

For the same distributed sample of 324 web pages, the researchers found that 79.94% had no rating related to Other Topics, and for those web pages that did rate in Other Topics a court interpreting CIPA using average U.S. community standards would not likely require any blocking.

For the same distributed sample of 324 web pages, the researchers found that 97.53% had no rating related to Chat, and for those web pages that did rate in Chat a court interpreting CIPA using average U.S. community standards would not likely require any blocking.

Effects of Inappropriate Blocking

If Internet blocking software companies were able to remove all the inappropriate blocks, overall block rates would drop dramatically in this study. This suggests that searches on curriculum-related topics would produce very few attempts to access web pages that would be blocked by Internet blocking software and raises further questions about the effectiveness of its use, especially in light of the extensive collateral damage of overblocking.

For information on web page miscategorization, see the "Blocking by Blocking Product" section.

Underblocking

The study found the following results regarding blocking software underblocking:

- Internet blocking software was not able to detect and protect students from access to many of the relatively small quantity of apparently pornographic sites that appeared in search results related to state-mandated curriculums.

Although the study researchers understood the difficulty of locating web pages that would likely require blocking under the definitions used by CIPA, the researchers made a simple attempt to locate such web pages among the web pages that the blocking software did not block.

The researchers searched the text all unblocked web pages for the following target words: cock, pussy, cunt, fuck, buttfuck, rimjob, muff, blowjob, kink, tits, ass, erotic, cornhole, bestiality, cunnilingus, fellatio, Defecolagnia, dildo, smut, porn, smegma, watersports, bondage, bdsm, masochist, bukkake, cameltoe, twat, pubes, jackoff, xxx, trannie, cocksucker, handjob, lube, dick, spunkloving, cum, hardcore, fisting, fistfucking, and handball.

For any web page that contained three or more of the target words, the researchers examined the web page against the legal standard provided by CIPA and found 11 total occurrences of nine web pages for which CIPA would likely would require blocking, three missed by N2H2 Bess and eight missed by SurfControl. The researchers also found a total of 34 web pages that would require blocking if non-visual depictions were included, 17 for N2H2 Bess and 17 for SurfControl.

The list of web pages likely requiring blocking under CIPA included the following web pages:

- A web page left unblocked by both N2H2 Bess and SurfControl with the description Castle of Torment, Silver Moon identified by the target words cock, cunt, fuck, tits, erotic, cunnilingus, and bondage and including a picture of a naked woman chained by her wrists at <http://www.adultbookshops.com/moon/Chapters/CastleTorment.htm>
- A web page left unblocked by N2H2 Bess that serves as an entry to the Hustler website identified by the target words porn, xxx, and hardcore and including several pornographic pictures at <http://www.pcgreetings.com/>
- A web page left unblocked by N2H2 Bess that contains an ad entitled "Central Texas Gentleman Top ISO Special femsub or Femswitch" on Ambrosio's BDSM website identified by the target words cock, erotic, bondage, bdsm, and xxx and including an image of a man lashing a full-breasted naked woman at <http://www.io.com/~ambrosio/vanity/ad.html>
- A web page left unblocked by SurfControl with the description "The New Supergirl in Town, Chapter 32." identified by the target words cock, fuck, and tits and including an explicit pornographic picture located at <http://www.mfffb.com/julievelor/nstg32.html>

- A web page left unblocked by SurfControl with the description "Online Adult XXX! The 'original' Online Adult Entertainment Portal...with the latest HOT sites. Established 1996." identified by the target words pussy, cunt, ass, porn, and cum, and including explicit pornographic pictures located by auto-referral from <http://www.curableromantic.com/>

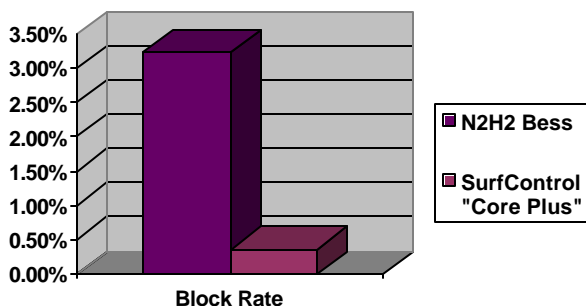
The primitive keyword detection scheme revealed many pornographic web pages that the blocking software products did not block.

However, the researchers did not even attempt to search for "visual depictions" of pornography and, as mentioned by the court declaring CIPA unconstitutional for libraries: "This is of critical importance, because CIPA, by its own terms, covers only 'visual depictions.' 20 U.S.C. §9134(f)(1)(A)(i); 47 U.S.C. §254(h)(5)(B)(i)."

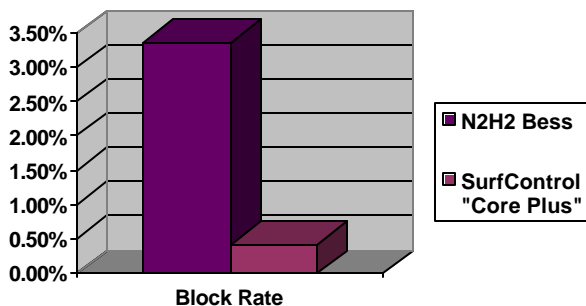
Overall Blocking Rates

At a minimum, Internet blocking software blocks tens of thousands of web pages inappropriately, but because schools administering Internet blocking software can configure the software to restrict access to fewer or greater numbers of web pages by choosing the blocking codes they wish to block, and because Internet blocking companies assign differing numbers of web pages to the block codes in their products, blocking rates can vary widely between various blocking software installations.

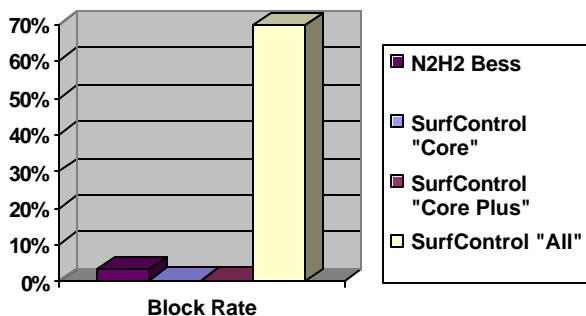
Focusing on the blocking software configurations the researchers speculate are most likely used in schools, the study found overall blocking rates between 0.36% and 3.24%.



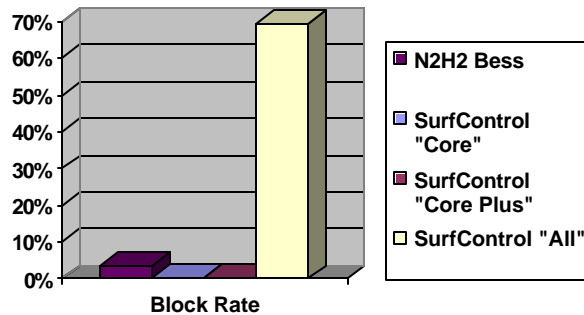
Focusing on the blocking software configurations the researchers speculate are most likely used in schools and removing all web pages that contained “k12” in the web address, the study found overall blocking rates between 0.41% and 3.37%.



The study found overall blocking rates between 0.22% and 71.14% of web pages generated from state-mandated curriculum topic searches, depending primarily on which blocking codes are applied by the blocking product, but also on blocking product and state. However, it is improbable that many schools operate with all blocking codes selected, the level necessary for blocking seventy percent or more of web pages – the researchers speculate that most schools are using blocking codes at least as restrictive as the “core plus” blocking codes for SurfControl.



When removing all web pages that contained “k12” in the web address, the study found overall blocking rates between 0.31% and 69.49% of web pages generated from state-mandated curriculum topics, depending primarily on which blocking codes are applied by the blocking product, but also on blocking product and state.



N2H2 Bess Blocking

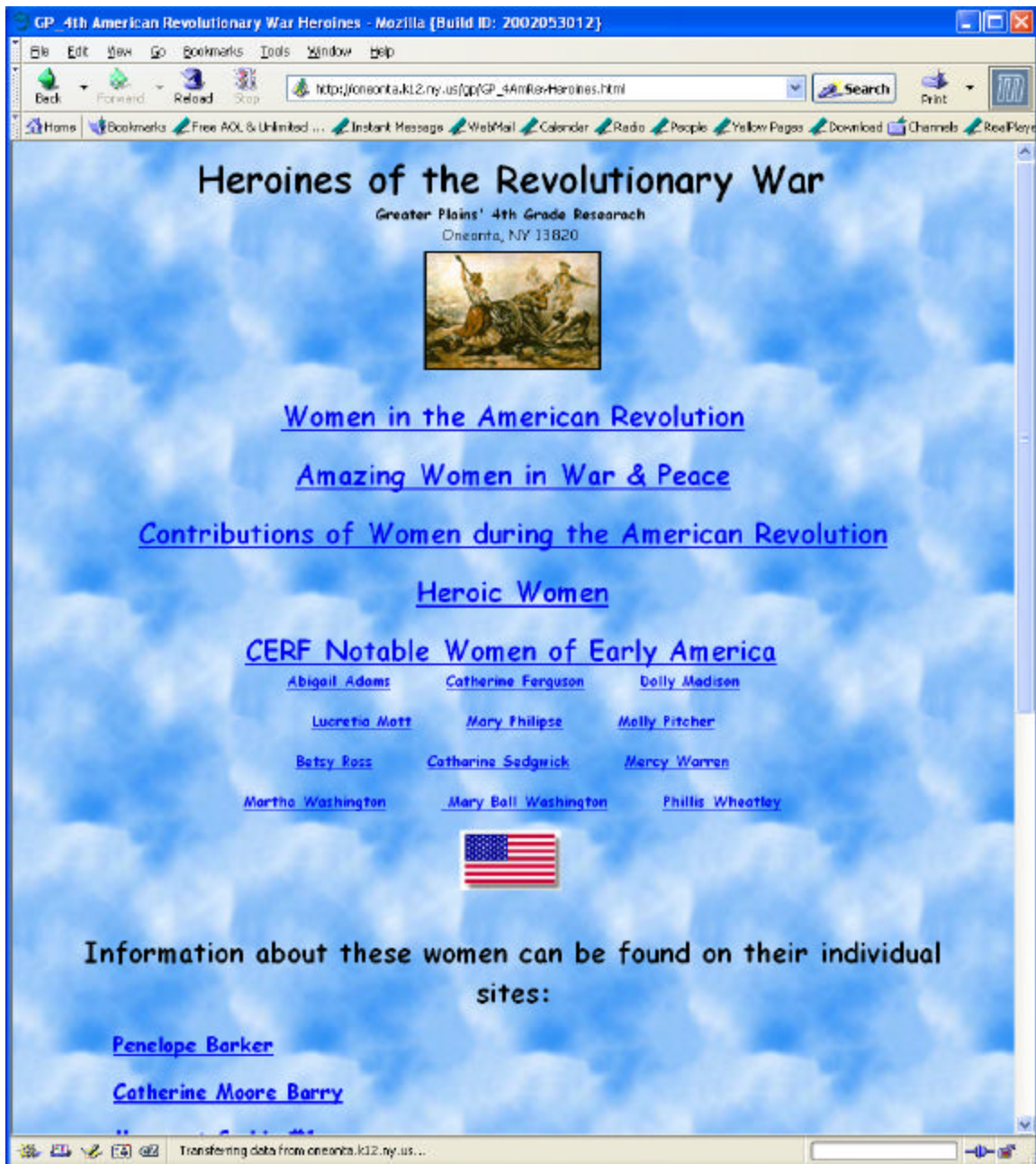
For N2H2's Bess product, this study found an average blocking rate of 3.24% \pm 0.04% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches (a total of 31,549 pages blocked of 973,215 pages checked).

As an example, N2H2 Bess correctly blocked as Pornography a Czech porn web page called Sex Shock at <http://www.sexshock.cz/>



Removing “k12” pages from the sample, this study found an average blocking rate of 3.66% \pm 0.04% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches (a total of 31,480 pages blocked of 860,939 pages checked).

For example, N2H2 Bess blocked as Recreation/Entertainment the “Heroines of the Revolutionary War” web page at http://oneonta.k12.ny.us/gp/GP_4AmRevHeroines.html “created by Mrs. Rees, Librarian, for use by 4th grade students as they work on their American Revolutionary projects in the Library Media Center Computer Lab at Greater Plains...neighborhood school located in the West End of the City of Oneonta,” NY.



N2H2 Bess also blocked as Recreation/Entertainment the same school's "Writing Techniques" web page describing literary terminology at <http://oneonta.k12.ny.us/hs/murphy/terms.htm>



SurfControl Blocking

A SurfControl representative reported that, although the company makes no official recommendation about which blocking codes a school should select, they have noticed that most schools make use of a core set of blocking codes with many of those schools adding a small set of additional blocking codes.

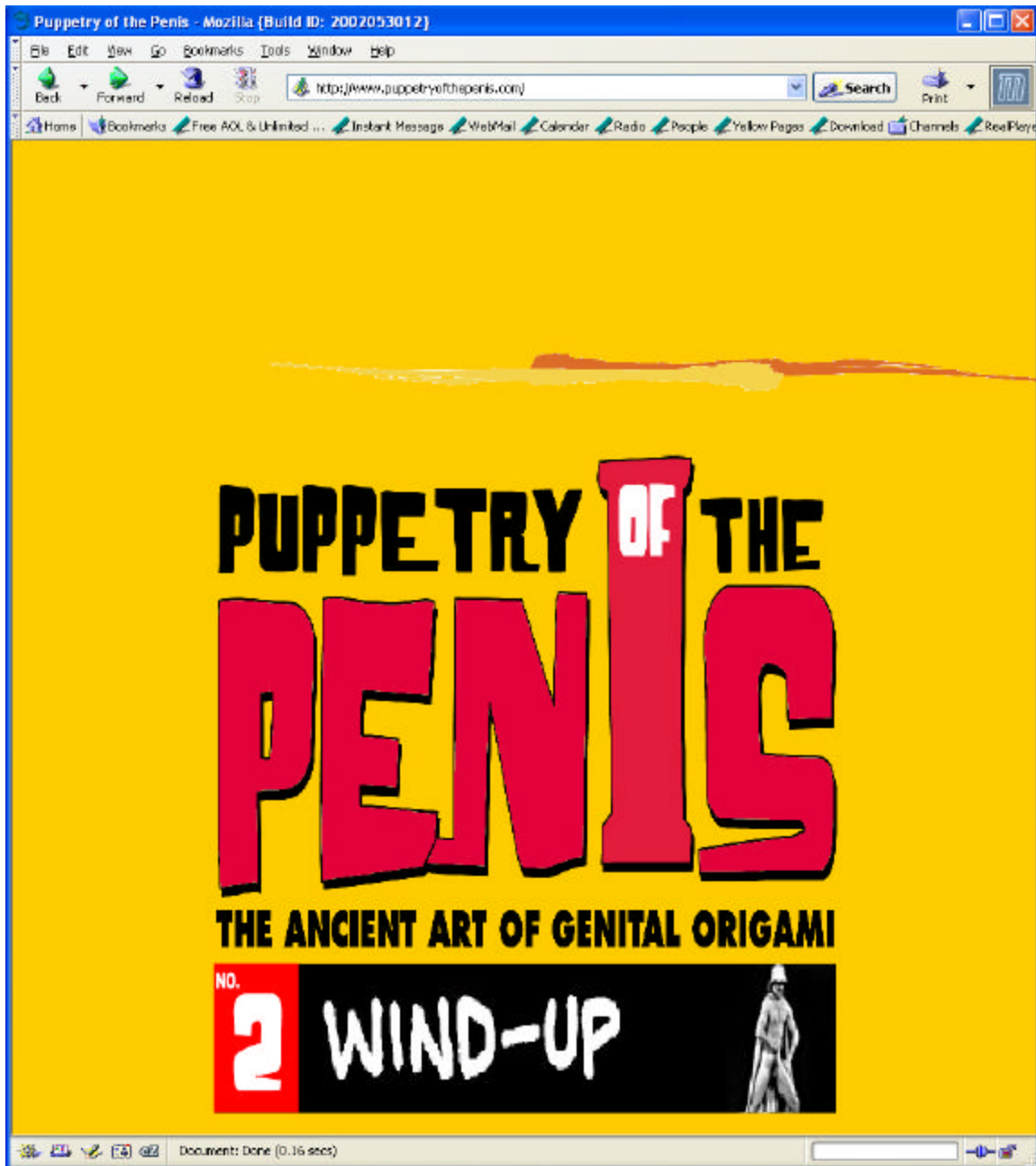
This study found an average blocking rate of 0.28% \pm 0.01% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches for SurfControl “core” blocking codes described by the SurfControl representative as “the Education market’s most important concerns” (a total of 2,682 pages blocked of 973,215 pages checked). SurfControl’s ten “core” blocking codes are: Adult/Sexually Explicit, Chat, Criminal Skills, Drugs, Alcohol & Tobacco, Gambling, Hacking, Hate Speech, Violence, Weapons, and Web-based Email. Many of the codes represent content that is not proscribed by CIPA.

This study found an average blocking rate of 0.36% \pm 0.01% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches for SurfControl core blocking codes plus a small set of commonly used additional blocking codes, i.e. “core plus” (a total of 3,522 pages blocked of 973,215 pages checked). SurfControl “core plus” blocking codes are the “core” blocking codes plus three more blocking codes: Glamour & Intimate Apparel, Personals & Dating, and Sex Education.

It is unlikely that many schools operate the SurfControl software with all blocking codes selected, which is what is required to achieve blocking rates in the 70% range, since some of the blocking codes may be used to include, rather than exclude sites. This study found an average blocking rate of 69.79% \pm 0.09% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches for SurfControl core blocking codes (a total of 679,216 pages blocked of 973,215 checked).

The researchers speculate that most schools are using blocking codes at least as restrictive as the “core plus” blocking codes for SurfControl.

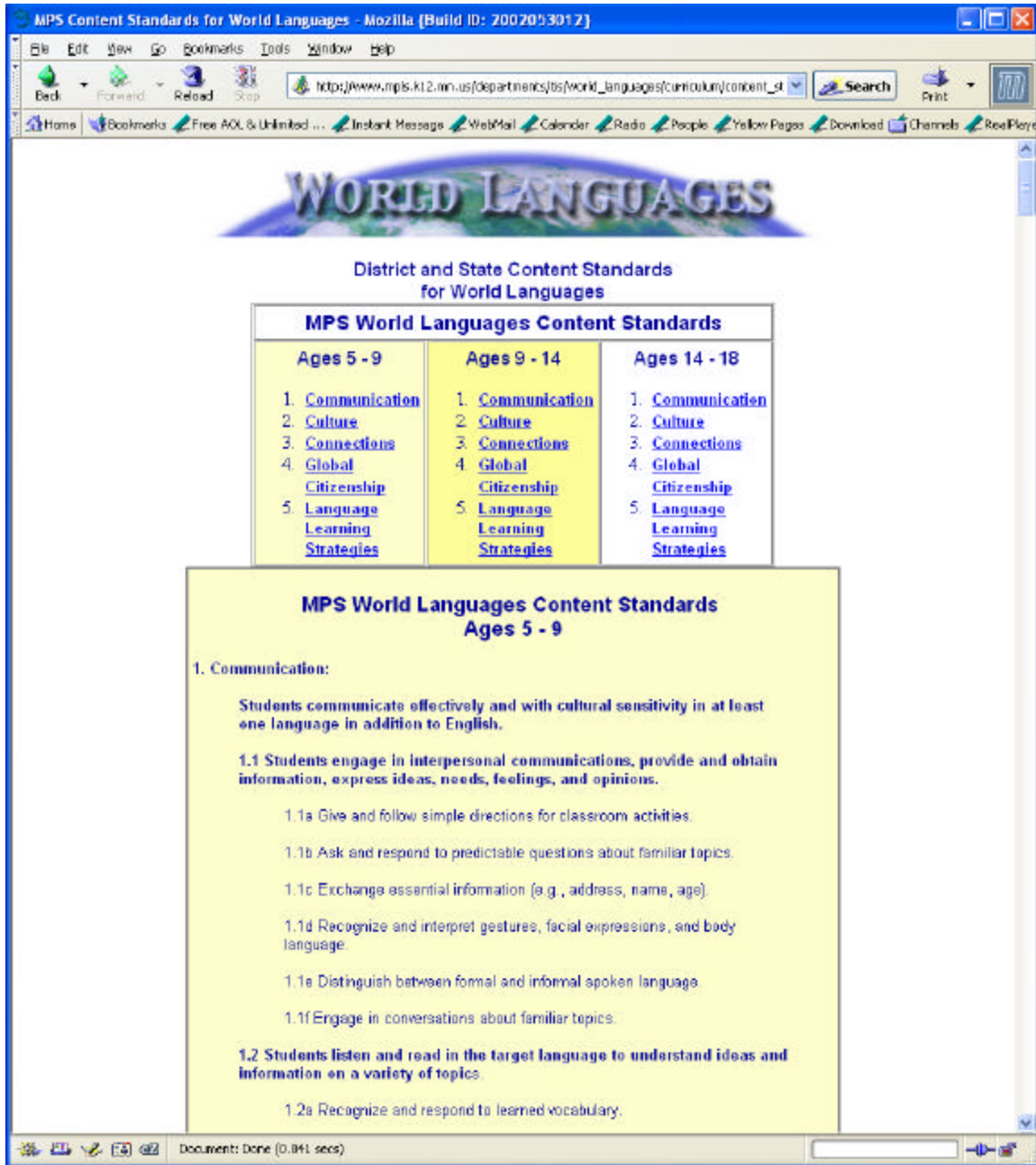
As an example of a web page blocked correctly, SurfControl blocked as Adult/Sexually Explicit the “Puppetry of the Penis” web page which shows a man “winding up” his penis at <http://www.puppetryofthepen.com/>



Removing “k12” pages from the sample, this study found an average blocking rate of $0.31\% \pm 0.01\%$ (95% confidence interval) of web pages generated from state-mandated curriculum topic searches for SurfControl “core” blocking codes (a total of 2,682 pages blocked of 860,939 pages checked), an average blocking rate of $0.41\% \pm 0.01\%$ (95% confidence interval) of web pages generated from state-mandated curriculum topic

searches for SurfControl “core plus” blocking codes (a total of 3,522 pages blocked of 860,939 pages checked), and an average blocking rate of 69.49% \pm 0.10% (95% confidence interval) of web pages generated from state-mandated curriculum topic searches for SurfControl core blocking codes (a total of 598,229 pages blocked of 860,939 checked).

As an example of a “k12” page blocked, SurfControl blocked as News the Minneapolis, Minnesota, “District and State Content Standards for World Languages” web page at http://www.mpls.k12.mn.us/departments/tis/world_languages/curriculum/content_standard_s.html



The absolute number of pages blocked by SurfControl “core” and “core plus” configurations was identical whether or not removing “k12” pages from the sample. The

researchers speculate that the reason for this is that SurfControl assigns all “k12” pages the Education code.

Blocking by Category/Topic

This study demonstrates that Internet blocking software blocks web pages generated from some topic and category searches of the state-mandated curriculums much more than others.

Blocking by Topic

This section details how N2H2 Bess and SurfControl blocked web pages generated from topic searches of the state-mandated curriculum.

N2H2 Bess Blocking by Topic

Topics N2H2 Bess blocked 40% or more of the time included the following:

- 1) Examine the effect of political programs and activities of Populists (100%)
- 2) Odler [sic] adulthood (100%)
- 3) National Labor Relations Board v. Jones & Laughlin Steel Company, Brown... (66.67%)
- 4) Dating (60%)
- 5) Firearms (50%)
- 6) Listen actively and critically by/delving deeper into the topic (48%)
- 7) Pogo-stick (46%)
- 8) Increase sight vocabulary, reading vocabulary, and writing vocabulary through... (44%)
- 9) Comedy (42%)
- 10) Keep hands clean, using appropriate cleaning techniques (42%) [Note: this topic apparently blocked a lot by both products perhaps because of the web pages with the Alanis Morissette song “Hands Clean.”]
- 11) Short problems, emphasizing element force/energy (e.g., swing, melt, explode,... (41.3%)
- 12) Arms (40%)
- 13) Pantoming (40%) [Note: represents a copying error which is amazingly found on many web pages; should have been “Pantomiming”]

- 14) Demonstrate through role-playing appropriate use of formal and informal language... (40%)

SurfControl Blocking by Topic

Topics SurfControl blocked 40% or more of the time in the “core plus” configuration included the following:

- 1) Examine the effect of political programs and activities of Populists (100%)
- 2) Dating (66%)
- 3) History, rules and strategy of sports (43%)
- 4) Firearms (42%)

SurfControl blocked 100% of 102 topics in the “all codes” configuration.

Blocking by Topic Comments

N2H2 Bess blocked 100% of the web pages related to two curriculum topics and SurfControl blocked 100% of the web pages related to one of the two curriculum topics blocked entirely by N2H2 Bess.

The topic most blocked by both products was “Examine the effect of political programs and activities of Populists.” Both products blocked all five web pages associated with that curriculum topic. Examination of the five web pages showed that there were five occurrences of the same web page on National Socialists which researchers agreed was blocked according to guidelines advertised by the blocking companies as “Hate/Discrimination” by N2H2 Bess and as “Hate Speech” by SurfControl.

N2H2 Bess also blocked all five occurrences of the oddly misspelled topic “odler adulthood” 100% of the time. Examination of the five web pages showed that there were five occurrences of the same web page on “Raising a Teenager” which researchers marked inappropriately blocked as “Free Pages” by N2H2 for two reasons: 1) customers have to pay to create a website on AOL, so N2H2 did not assign this web page correctly according to their published block code definitions, and 2) the content on this web page, which is a brief book review, does not merit blocking in schools.

The topics “Dating” and “Firearms” were both extensively blocked by N2H2 and SurfControl.

For example, N2H2 blocked 30 of 50 sites on “dating” with the following block codes: Adults Only, Nudity, Jokes, Personal Information, Personals, Pornography, and Sex. SurfControl blocked 33 of 50 sites on “dating” with the following block codes: Adult/Sexually Explicit, Glamour & Intimate Apparel, and Personals & Dating. Examination of a sample of 63 of the “dating” web pages blocked showed that the Internet blocking companies assigned the wrong block codes 23.81% \pm 10.73 (95% confidence interval) of the time (for 15 out of 63 pages) but only blocked the pages wrongly 6.35% \pm 6.14% (95% confidence interval) of the time (or 4 out of 63 pages) because they could have blocked many of the pages with the wrong block codes using another one of their block codes had they assigned those pages the correct block code.

Blocking by Category

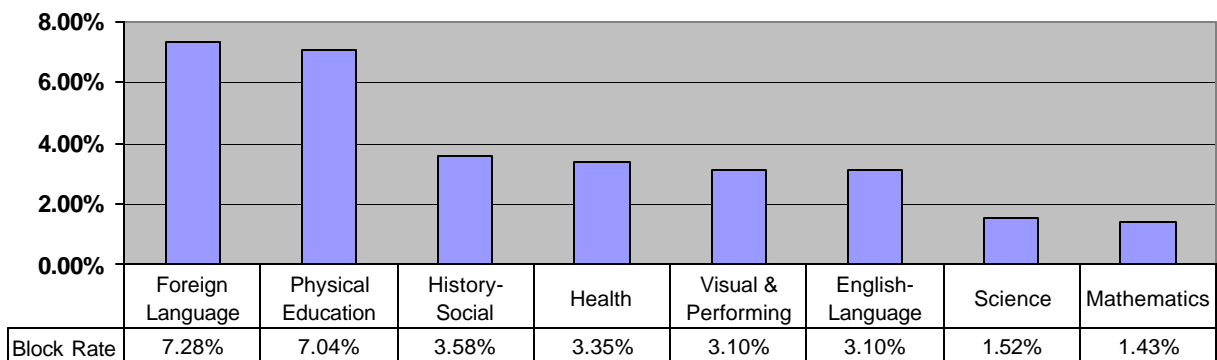
The Internet blocking software products tended to block more web pages associated with the curriculum categories of Physical Education, Health, Foreign Language, Latin, and History – Social Science. They tended to block fewer web pages associated with the Math(ematics) curriculum category.

The blocking software products may have prevented access to some foreign language sites because the blocking company employees could not determine what the material was rather than because of the presence of anything which the law would find objectionable.

N2H2 Bess Blocking by Category

N2H2 Bess tended to block some categories of web pages significantly more than others.

For the California curriculum, N2H2 Bess blocked the following top-level categories at these block rates:



N2H2 blocked the following second-level categories from the California curriculum at rates of 5% or more:

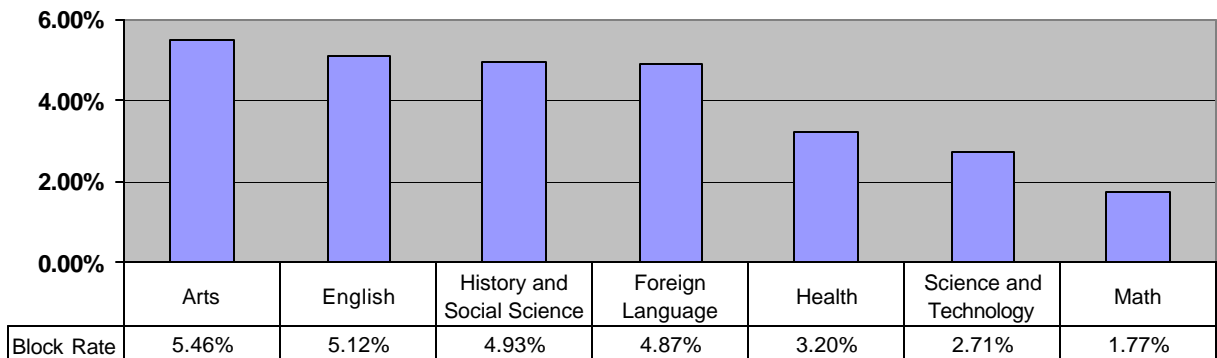
- 1) Physical Education: How I Move in My Environment (10.00%)
- 2) Physical Education: My Partner and I - How We Move in Space (9.38%)
- 3) Physical Education: Moving Through Space and Time (8.14%)
- 4) Physical Education: Working Cooperatively to Achieve a Common Goal (8.10%)
- 5) Foreign Language: Language Learning (7.29%)
- 6) Physical Education: Meeting Challenges and Making Decisions (7.23%)
- 7) Physical Education: Developing a Personalized Fitness Program for a Healthy Lifestyle (7.12%)

- 8) Physical Education: Manipulating Objects with Accuracy and Speed (7.03%)
- 9) Physical Education: Continuity and Change in Movement (6.49%)
- 10) Physical Education: Manipulating Objects in and Through Space (6.24%)
- 11) Physical Education: Working as a Team to Solve Problems (5.63%)
- 12) History -Social Science: World History and Geography Ancient Civilizations (5.18%)

N2H2 blocked the following selected third-level categories from the California curriculum at high rates:

- 1) Speaking Applications (Genres and Their Characteristics)/descriptive presentations/speaker's point of view (14.00%)
- 2) Writing Applications (Genres and Their Characteristics)/biographical or autobiographical narratives or short stories/scenes and incidents (12.00%)
- 3) Federal civil rights and voting rights/women's right movement (12.00%)
- 4) identity $\cos^2(x) + \sin^2(x) = 1$ (12.00%)
- 5) U.S. Constitution and other essential documents/character of American democracy (12.00%)
- 6) factor small whole numbers/numbers 2 3 5 7 11 do not factors except 1 themselves numbers are called prime numbers. (10.00%)
- 7) political, social, economic, technological and cultural developments of the 1920s/18th Amendment and Volstead Act(prohibition) (10.00%)
- 8) tables graphs rules solve problems involving rates proportions/convert one unit measurement another (feet miles centime-ters inches). (10.00%)
- 9) Self Image and Personal Development (9.44%)
- 10) Reconstruction (8.78%)

For the Massachusetts curriculum, N2H2 Bess blocked the following top-level categories at these block rates:



N2H2 blocked the following second-level categories from the Massachusetts curriculum at rates of 5% or more:

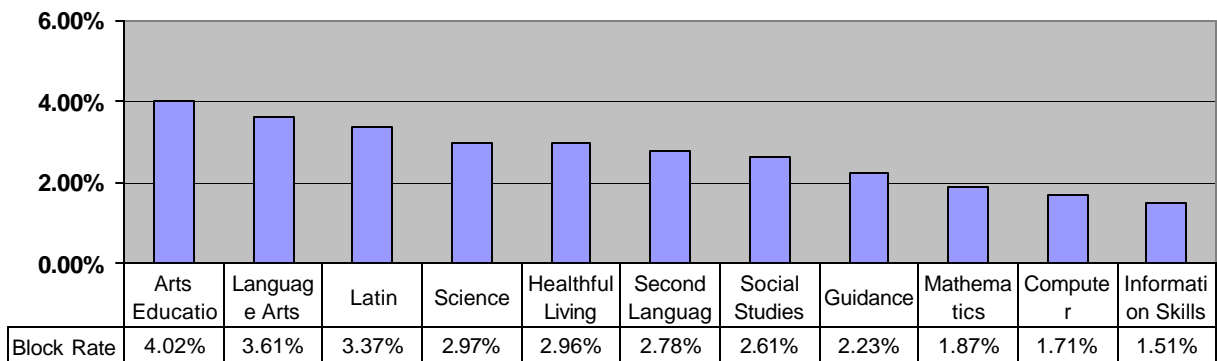
- 1) Theatre (6.89%)
- 2) History and Social Science: Core Knowledge (6.33%)
- 3) Foreign Language: Comparisons (6.32%)
- 4) Music (6.12%)
- 5) Reading and Literature (6.07%)
- 6) History (5.63%)
- 7) Foreign Language: Communication (5.32%)
- 8) Dance (5.06%)

N2H2 blocked the following selected third-level categories from the Massachusetts curriculum at high rates:

- 1) interviewing one person about his or her occupation or interests; (24.00%)
- 2) conversing with speakers of the target language; (14.00%)
- 3) Mapping the Earth (13.00%) [Note: this topic blocked at a high rate by both blocking products because the word “models” appeared in the detailed topic listing.]
- 4) Tell time at quarter-hour intervals on analog and digital clocks using a.m. and p.m. (12.00%)

- 5) Know addition facts (addends to ten) and related subtraction facts, and use them to solve problems (10.00%)
- 6) Identify parts of the day (e.g., morning, afternoon, evening), days of the week, and months of the year. Identify dates using a calendar. (10.00%)
- 7) Classification of Organisms (9.00%)
- 8) Dramatic Literature (8.88%)
- 9) Interdisciplinary Learning Religion, Ethics, Philosophy and Literature in History (8.30%)
- 10) Identify the value of all U.S. coins, and \$1, \$5, \$10, and \$20 bills. Find the value of a collection of coins and dollar bills and different ways to represent an amount of money up to \$5. Use appropriate notation, e.g., 69¢, \$1.35. (8.00%)

For the North Carolina curriculum, N2H2 Bess blocked the following top-level categories at these block rates:



N2H2 blocked the following second-level categories from the North Carolina curriculum at rates of 5% or more:

- 1) Build an understanding of the actions of objects (10.40%)
- 2) Build an understanding of the concepts of sound (6.40%)
- 3) US History (5.46%)
- 4) Build an understanding of plant and animal life cycles (5.33%)
- 5) Build an understanding of solid earth materials (5.20%)
- 6) Theatre (5.18%)

7) Read and Write (5.00%)

8) Build an understanding of electricity and magnetism (5.00%)

N2H2 blocked the following selected third-level categories from the North Carolina curriculum at high rates:

9) Listen actively and critically by/delving deeper into the topic. (48.00%)

10) Keep hands clean, using appropriate cleaning techniques (42.00%)

11) Learn how to make and keep friends. (24.00%)

12) Make informed judgments about/propaganda. (20.00%)

13) Describe meanings of traffic signs and signals (14.00%)

14) Analyze the parts of a light bulb (14.00%)

15) Recognize and seek help for depression (12.00%)

16) Use capital letters to write I and own name (12.00%)

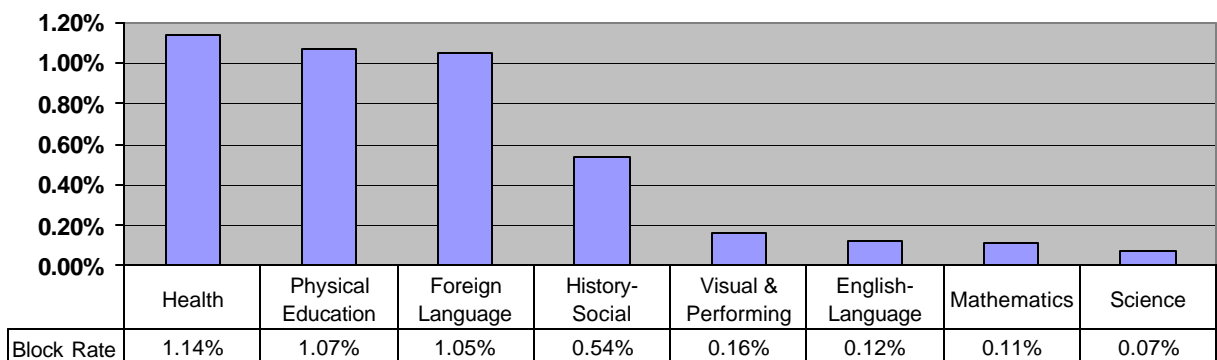
17) Evaluate a variety of public documents by/comparing the argument and counter-argument presented. (12.00%)

18) Recognize two appropriate sites on the body to monitor the heart rate (12.00%)

SurfControl Blocking by Category

SurfControl “core plus” configuration tended to block some categories of web pages significantly more than others.

For the California curriculum, SurfControl “core plus” configuration blocked the following top-level categories at these block rates:



SurfControl “core plus” configuration blocked the following second-level categories from the California curriculum at rates of 1% or more:

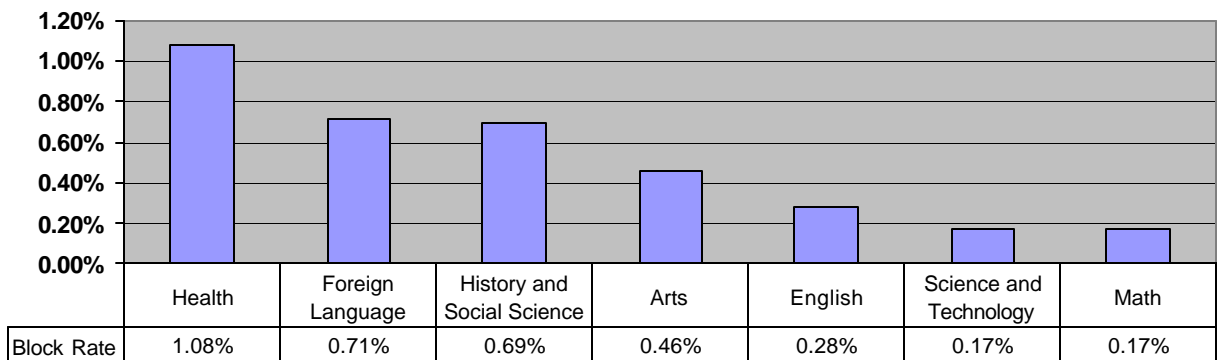
- 1) Health: Informed use of health-related information, products, and services/known (2.00%)
- 2) Health: An understanding of the process of growth and development (1.93%)
- 3) Physical Education: Developing a Personalized Fitness Program for a Healthy Lifestyle (1.82%)
- 4) Physical Education: How I Move in My Environment (1.61%)
- 5) Physical Education: Working as a Team to Solve Problems (1.50%)
- 6) Health: Acceptance of Personal Responsibility for Lifelong Health (1.36%)
- 7) Physical Education: Continuity and Change in Movement (1.30%)
- 8) Health: Respect for and promotion of the health of others (1.26%)
- 9) Physical Education: Moving Through Space and Time (1.24%)
- 10) Foreign Language: Language Learning (1.11%)
- 11) Physical Education: Manipulating Objects with Accuracy and Speed (1.03%)

SurfControl “core plus” configuration blocked the following selected third-level categories from the California curriculum at high rates:

- 1) identity $\cos^2(x) + \sin^2(x) = 1$ (6.00%)
- 2) Writing Applications (Genres and Their Characteristics)/job applications and resumés/conventional style (6.00%)
- 3) Reconstruction (5.74%)
- 4) Students will understand their developing sexuality, will choose to abstain from sexual activity, and will treat the sexuality of others with respect. (5.51%)
- 5) model solve problems representing adding subtracting amounts money/solve problems combinations coins bills. (4.00%)
- 6) political, social, economic, technological and cultural developments of the 1920s/18th Amendment and Volstead Act(prohibition) (4.00%)
- 7) unique roles and responsibilities of three branches of government/identify current representatives in legislative branch (4.00%)

- 8) origins, characteristics, and development of different political systems across time/forms of illegitimate power that 20th century African, Asian, and Latin American dictators used to gain and hold office (4.00%)
- 9) major social problems and domestic policy issues/significant policy speeches of Truman through Clinton (4.00%)
- 10) Federal civil rights and voting rights/diffusion of civil rights movement from rural Southern churches to urban North (4.00%)

For the Massachusetts curriculum, SurfControl “core plus” configuration blocked the following top-level categories at these block rates:



SurfControl “core plus” configuration blocked the following second-level categories from the Massachusetts curriculum at rates of 1% or more:

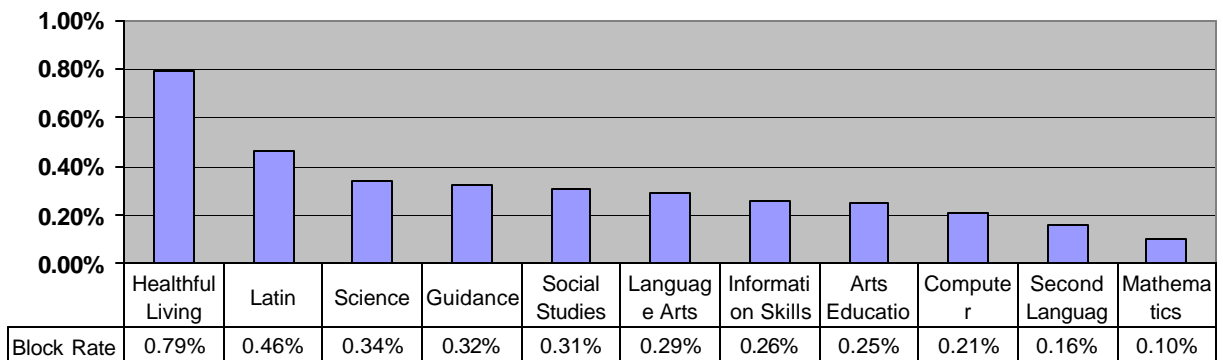
- 1) Health: Social and Emotional Health (1.66%)
- 2) Health: Physical Health (1.19%)
- 3) Foreign Language: Communication (1.15%)

SurfControl “core plus” configuration blocked the following selected third-level categories from the Massachusetts curriculum at high rates:

- 1) Mapping the Earth (8.00%) [Note: this topic blocked at a high rate by both blocking products because the word “models” appeared in the detailed topic listing.]
- 2) Identify odd and even numbers and determine whether a set of objects has an odd or even number of elements. (8.00%)
- 3) Represent the possible outcomes for a simple probability situation, e.g., the probability of drawing a red marble from a bag containing three red marbles and four green marbles. (6.00%)

- 4) Relate geometric ideas to numbers, e.g., seeing rows in an array as a model of repeated addition (4.00%)
- 5) interviewing one person about his or her occupation or interests; (4.00%)
- 6) Select and use appropriate operations (addition, subtraction, multiplication, and division) to solve problems, including those involving money) (4.00%)
- 7) Reproduction (3.90%)
- 8) Interpersonal Relationships (3.61%)
- 9) Tobacco, Alcohol, and Other Substance Use (2.25%)
- 10) Interpersonal Communication (2.00%)

For the North Carolina curriculum, SurfControl “core plus” configuration blocked the following top-level categories at these block rates:



SurfControl “core plus” configuration blocked the following second-level categories from the North Carolina curriculum at rates of 1% or more:

- 1) Healthful Living: Choose not to participate in substance use (1.96%)
- 2) Science: Build an understanding of the actions of objects (1.80%)
- 3) Science: Build an understanding of solid earth materials (1.47%)
- 4) Healthful Living: Interpret health risks for self and others and corresponding protection measures (1.28%)
- 5) Science: Build an understanding of the Solar System (1.09%)
- 6) Science: Build an understanding of technological design (1.00%)

SurfControl “core plus” configuration blocked the following selected third-level categories from the North Carolina curriculum at high rates:

- 1) Demonstrate strategies in a variety of games and sports (38.00%)
- 2) Refine skills and strategies for remaining or becoming abstinent from sexual intercourse (12.00%)
- 3) Explain the effectiveness and failure rates (some studies indicate failure rates range from 2% to 30%) of condoms as a means of preventing sexually transmitted diseases (12.00%)
- 4) Demonstrate skills and strategies for remaining or becoming abstinent from sexual intercourse (12.00%)
- 5) Affirm choice not to use tobacco or look alike products (10.00%)
- 6) Explain reasons not to use tobacco products (8.00%)
- 7) Describe normal weight gain and body changes during puberty (8.00%)
- 8) Explain the risks of premarital sexual intercourse (8.00%)
- 9) Demonstrate how to get help in an emergency (6.00%)
- 10) Keep hands clean, using appropriate cleaning techniques (6.00%)

Blocking by State

This study addressed Internet blocking of web pages related to state-mandated curriculums in California, Massachusetts, and North Carolina.

The study results show that Massachusetts had the highest rate of overblocking, followed by California, then North Carolina, for both blocking products studied as described below.

The research demonstrated that blocking software blocked web pages related to the Massachusetts state-mandated curriculum more often than those of California or North Carolina regardless of blocking product or blocking codes selection. In most cases, the research also demonstrated that blocking software blocked web pages related to the California state-mandated curriculum more often than those of North Carolina.

One possible explanation for the differing blocking rates by state could be the sample size. The states that had the most topics had the lowest block rates. Another possibility is that the blocking software may have blocked some curriculum-related web pages more due to the values expressed in the topic choices made by the authors of the curriculum. Overall, Massachusetts is likely the most “liberal” and North Carolina the most “conservative” of the three state-mandated curriculums addressed by this study. See “Blocking by Category/Topic” for more analysis related to curriculum topics.

N2H2 Bess Blocking by State

This section describes N2H2 Bess blocking by state with a focus on N2H2 Bess overblocking.

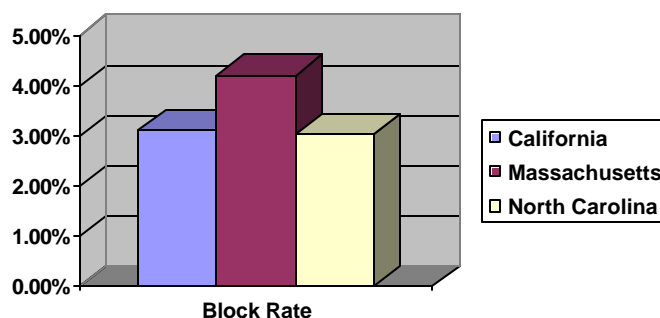
N2H2 Bess Overblocking by State

Researchers using the criterion of blocking only sites a court would reasonably find required by CIPA in according to average U.S. community standards tested 315 pages spread evenly over the entire sample of web pages blocked by N2H2 Bess and found that of the 294 pages that were accessible, N2H2 Bess blocked 98.98% $\pm 0.30\%$ (95% confidence interval) (291 web pages) inappropriately.

A breakdown by state shows that N2H2 Bess blocked inappropriately 98.90% $\pm 2.18\%$ (95% confidence interval) (90 web pages) of web pages related to the California curriculum, 100.00% $\pm 0.00\%$ (95% confidence interval) (61 web pages) of web pages related to the Massachusetts curriculum, and 98.59% $\pm 1.97\%$ (95% confidence interval) (140 web pages) of web pages related to the North Carolina curriculum.

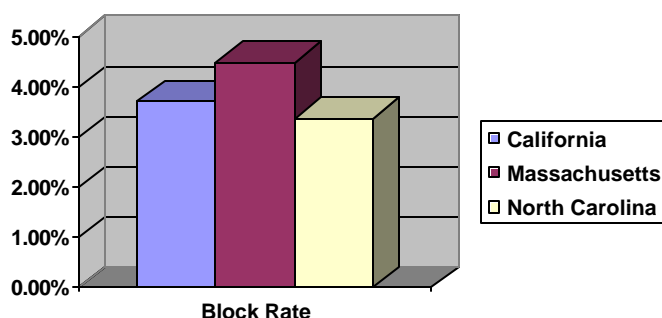
N2H2 Bess Blocking by State

N2H2 Bess blocked 3.10% $\pm 0.06\%$ (95% confidence interval) of web pages related to the California state-mandated curriculum (a total of 10,161 pages blocked of 327,918 checked), 4.21% $\pm 0.10\%$ (95% confidence interval) of web pages related to the Massachusetts state-mandated curriculum (a total of 6,338 pages blocked of 150,648 checked), and 3.04% $\pm 0.05\%$ (95% confidence interval) of web pages related to the North Carolina state-mandated curriculum (a total of 15,050 pages blocked of 494,649 checked). Although the difference between California and North Carolina block rates is not statistically significant, there are statistically significant differences between Massachusetts block rates and the other two states' block rates (5% significance level, χ^2 test with 2 df).



When removing all pages containing "k12" in the web address from the sample, N2H2 Bess blocked 3.70% $\pm 0.07\%$ (95% confidence interval) of web pages related to the California state-mandated curriculum (a total of 10,158 pages blocked of 274,827 checked), 4.49% $\pm 0.11\%$ (95% confidence interval) of web pages related to the Massachusetts state-mandated curriculum (a total of 6,337 pages blocked of 141,278 checked), and 3.37% $\pm 0.05\%$ (95% confidence interval) of web pages related to the North Carolina state-mandated curriculum (a total of 14,985 pages blocked of 444,834 checked).

Differences between N2H2 Bess non-“k12” block rates for all three states’ are statistically significant (5% significance level, χ^2 test with 2 df).



SurfControl Blocking by State

This section describes SurfControl blocking by state with a focus on SurfControl overblocking.

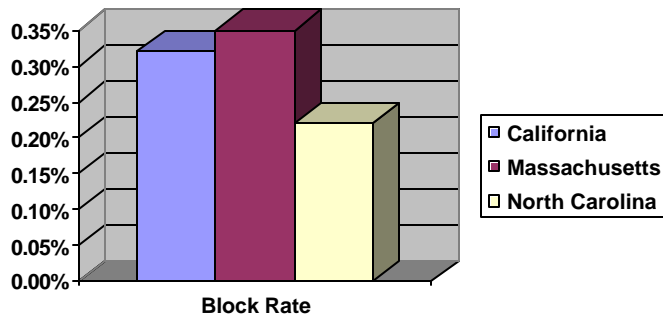
SurfControl Overblocking by State

Researchers using the criterion of blocking only sites a court would reasonably find required by CIPA in according to average U.S. community standards tested 352 pages spread evenly over the entire sample of web pages blocked by SurfControl and found that of the 324 pages that were accessible, SurfControl blocked 97.22% \pm 0.52% (95% confidence interval) (315 web pages) inappropriately.

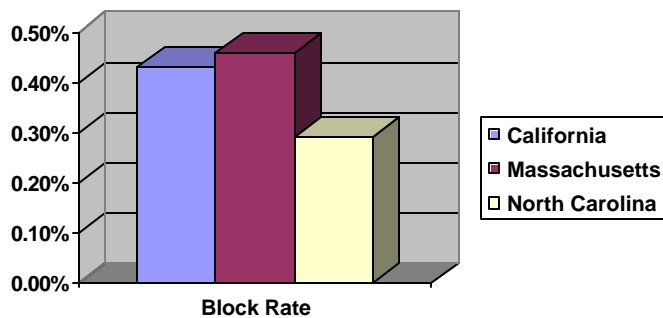
A breakdown by state shows that SurfControl blocked inappropriately 97.76% \pm 0.57% (95% confidence interval) (131 Web pages) of web pages related to the California curriculum, 100.00% \pm 0.00% (95% confidence interval) (65 Web pages) of web pages related to the Massachusetts curriculum, and 95.20% \pm 1.29% (95% confidence interval) (119 Web pages) of web pages related to the North Carolina curriculum.

SurfControl Blocking by State

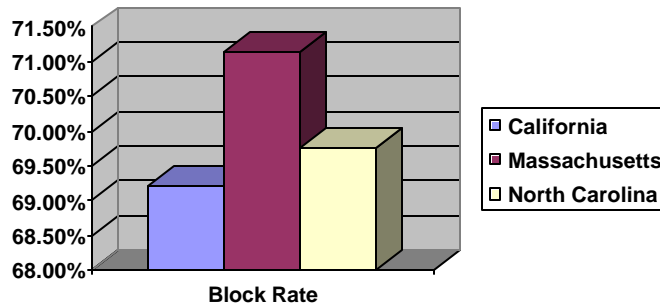
SurfControl “core” settings blocked 0.32% \pm 0.02% (95% confidence interval) of web pages related to the California state-mandated curriculum (a total of 1,053 pages blocked of 327,918 checked), 0.35% \pm 0.03% (95% confidence interval) of web pages related to the Massachusetts state-mandated curriculum (a total of 521 pages blocked of 150,648 checked), and 0.22% \pm 0.01% (95% confidence interval) of web pages related to the North Carolina state-mandated curriculum (a total of 1,108 pages blocked of 494,649 checked). Although the difference between SurfControl “core” California and Massachusetts block rates is not statistically significant, there are statistically significant differences between North Carolina block rates and the other two states’ block rates (5% significance level, χ^2 test with 2 df).



SurfControl “core plus” settings blocked 0.43% \pm 0.02% (95% confidence interval) of web pages related to the California state-mandated curriculum (a total of 1,418 pages blocked of 327,918 checked), 0.46% \pm 0.03% (95% confidence interval) of web pages related to the Massachusetts state-mandated curriculum (a total of 694 pages blocked of 150,648 checked), and 0.29% \pm 0.01% (95% confidence interval) of web pages related to the North Carolina state-mandated curriculum (a total of 1,410 pages blocked of 494,649 checked). Although the difference between SurfControl “core plus” California and Massachusetts block rates is not statistically significant, there are statistically significant differences between North Carolina block rates and the other two states’ block rates (5% significance level, χ^2 test with 2 df).



SurfControl “all” settings blocked 69.21% \pm 0.16% (95% confidence interval) of web pages related to the California state-mandated curriculum (a total of 226,960 pages blocked of 327,918 checked), 71.14% \pm 0.23% (95% confidence interval) of web pages related to the Massachusetts state-mandated curriculum (a total of 107,173 pages blocked of 150,648 checked), and 69.76% \pm 0.13% (95% confidence interval) of web pages related to the North Carolina state-mandated curriculum (a total of 345,083 pages blocked of 494,649 checked). Differences between SurfControl “all” block rates for all three states’ are statistically significant (5% significance level, χ^2 test with 2 df).

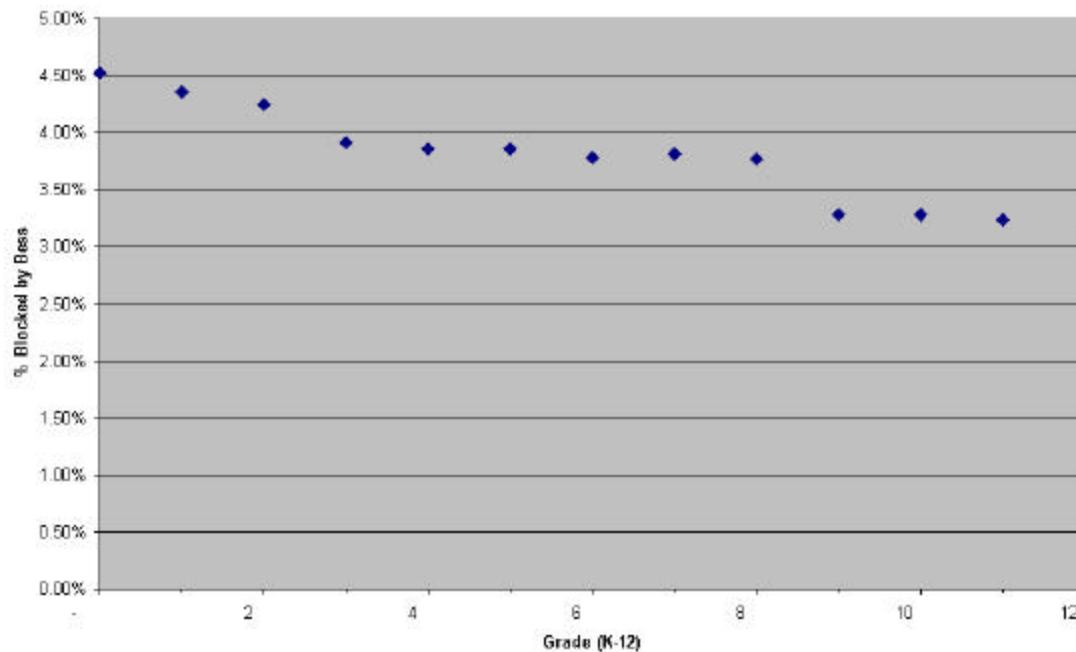


Blocking by Grade

Because the state-mandated curriculums generally had more topics listed in higher grade levels of the curriculum, more blocking of curriculum-related web pages occurs at the higher grade levels, particularly grades 9 to 12. However, the rates of blocking of curriculum-related web pages are fairly constant across all grade levels.

Note: Grade-specific blocking rates correspond to overall blocking rates reported above, but appear on average higher because many topics appeared in multiple grade levels.

N2H2 Bess blocking rates decreased gradually as grade level increased from 4.53% for grades prior to grade 1 to 3.21% for grade 12. With first and lower grades averaging at 4.44% and grades 2-12 averaging 3.52%, the researchers found a statistically significant relationship between grade level and blocking rate for the N2H2 Bess data at the 5% significance level. When regressing the blocked percentage vs. the grade level as a continuous factor, the researchers found that the grade level is a statistically significant explanatory factor for variation of blocking rates between the grades which explains 92% of the variation in blocking rate between grades. With an overall average of blocking rate of 3.60%, the N2H2 Bess blocking rate falls on average 0.107% per grade.



SurfControl “all” blocking rates remained fairly constant from 70% - 71% for all grades.

Blocking by Product

In studying the blocking performance by blocking product, researchers tallied the number of blocks that occurred in each blocking software block code. The study also examined how often blocking companies miscategorized web pages into incorrect block codes.

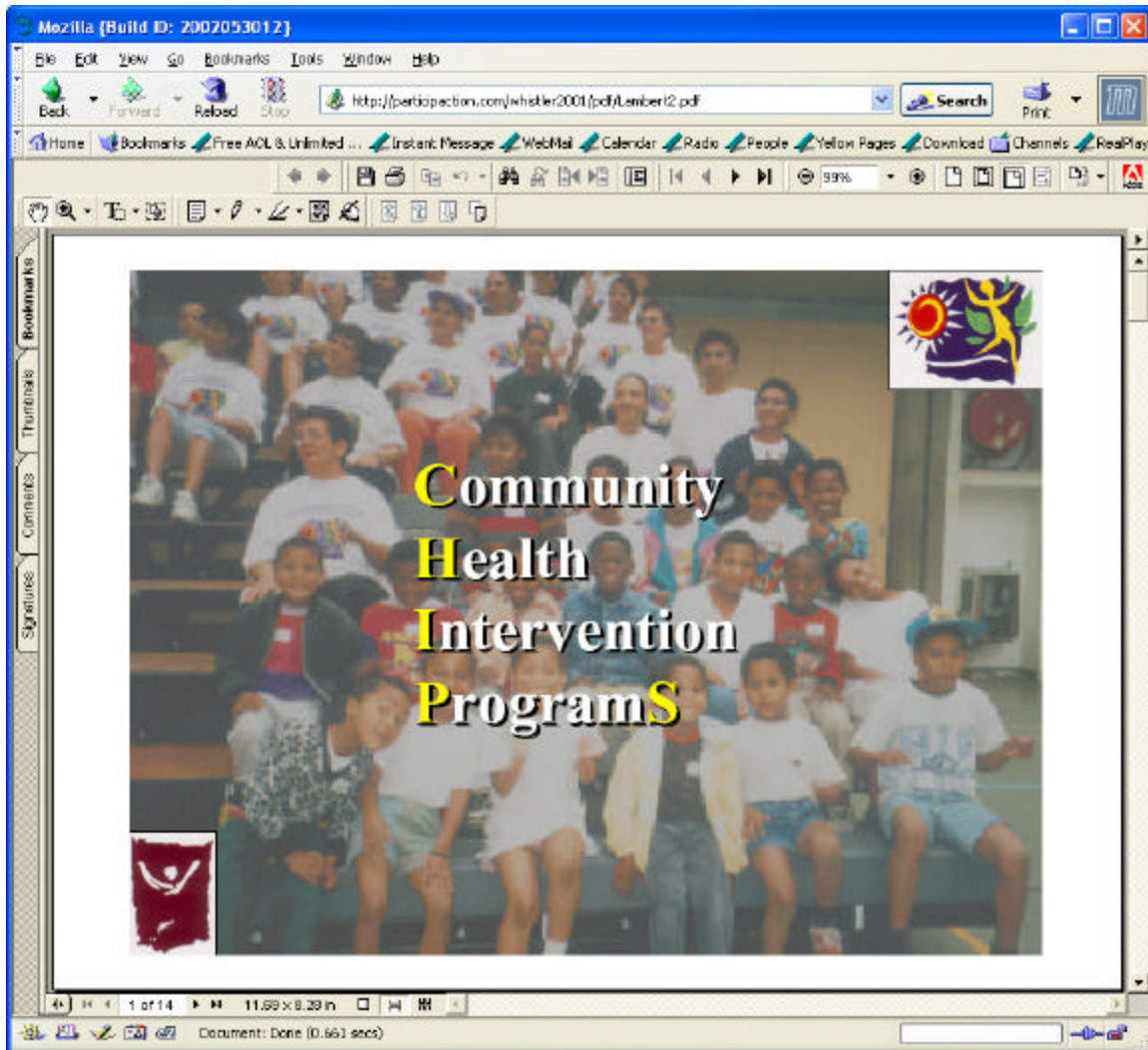
Because the two blocking products assign different blocking codes to the web pages they block, it is difficult to do a product-to-product comparison of the types of codes used. See the “Considerations” section for more information on correspondences between blocking codes of the Internet blocking products. Also, since the study set different numbers of blocking codes on each blocking product in various scenarios, comparisons of overall quantities of blocking are not particularly enlightening.

Note: Both products permit use of some of the block codes, such as “Education,” to be used as “allow” or “exception” codes, explicitly allowing access to pages assigned the “allow” code, rather than restricting access to those pages. N2H2 clearly identifies which codes are block codes and which are “allow” codes, whereas a SurfControl spokesperson refused to indicate which codes are intended for which purpose, so researchers had to make some assumptions about likely uses of the “allow” codes.

Determining whether the Internet blocking company assigned the correct block code for each web page meant reading the block code guidelines published by each Internet

blocking company and making a determination whether or not each web page fit into the block code(s) assigned by the company. In most cases, this was a simple task, but in some cases, there was a bit of subjectivity involved.

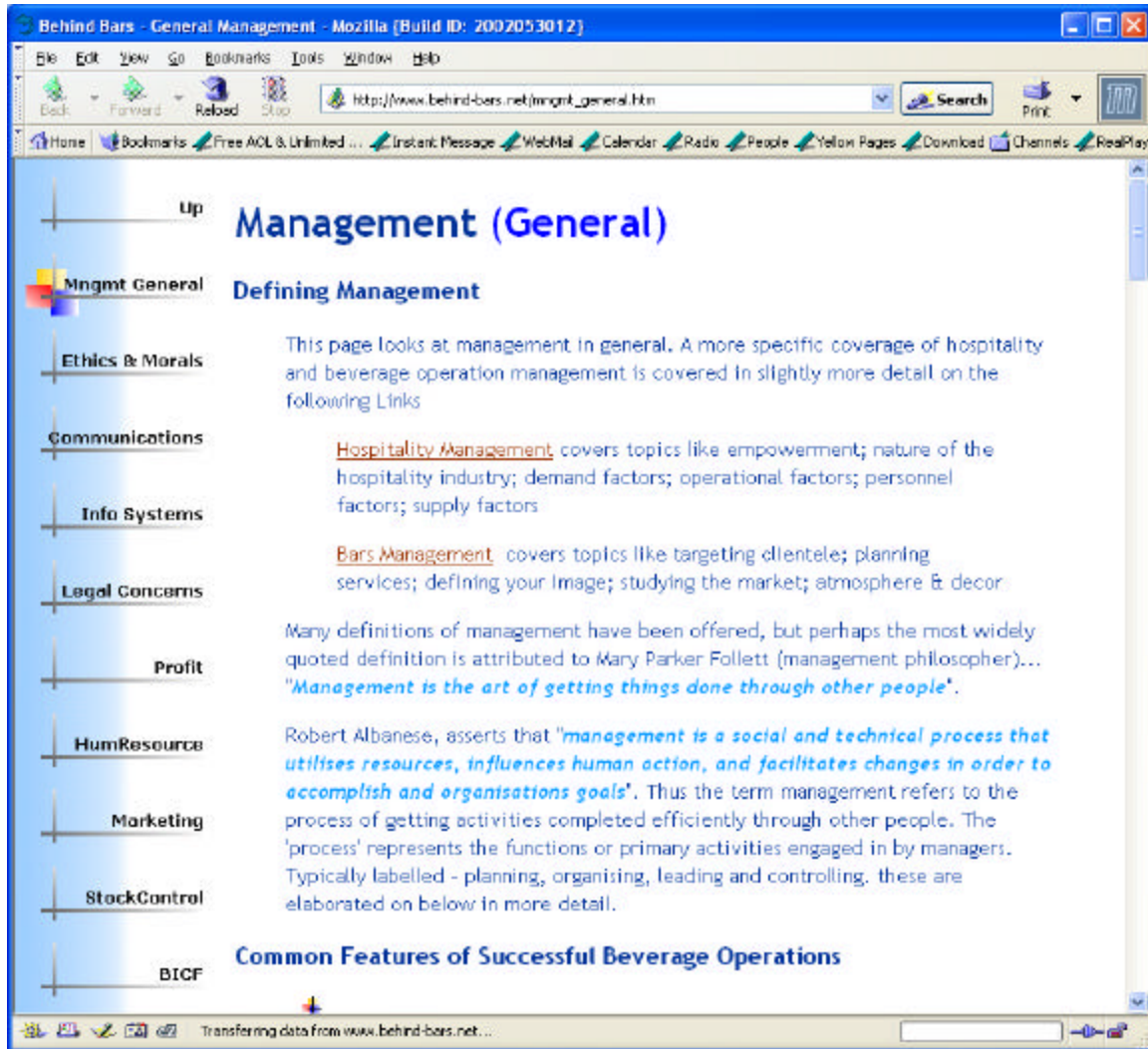
For example, N2H2 Bess assigned “CHIPS – Community Health Intervention Programs” from Dr. Vicki Lambert of Cape Town, South Africa, a Weapons block code, even though there was not a single mention of weapons on the web page.



In another example, N2H2 assigned a Pornography block code to a page about “Large movement skills and why they’re important” at <http://www.parent-education.com/e4.html>.



Sometimes, it's difficult to determine if the blocking company correctly assigned a block code. For example, SurfControl assigned a page on General Management of bars a Drugs, Alcohol, & Tobacco code, even though the page is about hospitality management, not about promoting drinking. In these cases, the researchers had to make a judgment call about whether or not the blocking company assigned the code appropriately.



In fact, the blocking companies used some block code definitions that the researchers found misleading at times. N2H2 defined its Pornography block code as:

“Sites that contain material that are intended to be sexually arousing or erotic. This includes photos, animation, cartoons, and stories. This also includes child pornography.”

The researchers found determining what is “intended to be sexually arousing or erotic” to be extremely subjective in borderline cases. Notice that “stories” do not fit within the visual depictions aspect of the web pages for which CIPA would likely require blocking.

Overall Miscategorization Rates

Overall, researchers found that blocking companies assigned blocked web pages the wrong blocking code between 29.70% and 58.00% of the time depending on the blocking product. Miscategorization rates by blocking product appear below.

Verification by Internet Content Rating Association Rating System

To provide more independent verification of the block code miscategorization, the researchers used the rating system developed by the Internet Content Rating Association (ICRA) to rate each web page according to the ICRA's rating system including blocking codes of Nudity and Sexuality, Violence, Other Topics (such as promotion of tobacco, alcohol, or drug use, discrimination, and gambling), and Chat. [11] Although the researchers in this study retain a healthy skepticism of the effectiveness of the ICRA rating system, many accept it as a standard. [12]

Likelihood of Human Review

A side-effect result was the discovery that researchers verifying block codes assigned by blocking software companies required one day's work for each 100 pages. Cyveillance reported on July 10, 2000, that the World Wide Web had 2.1 billion unique, publicly available pages and that the World Wide Web was growing at a rate of more than 7 million pages each day. [4] Google estimates they index more than three billion web addresses. [8] The federal district court decision striking the library portion of CIPA estimates 1.5 million new pages per day. [28]

SurfControl lays claim on its website to a block list of:

4.5 Million Sites, Covering More Than 800 Million Web Pages - Content is sourced by 40+ team of professional researchers, state-of-the-art automated tools, and customer submissions.

Although the number of pages cited could represent 38% of the web as of July 2000, it is not clear from SurfControl's marketing materials if the company performed a human review of all of the pages or even of just one page on each website.

And also from SurfControl's website:

Daily Updates to Customers

We keep our customers current with the rapid changes on the Internet.

* Daily updates to the SurfControl URL Category List representing an average of 25,000 new sites a week

N2H2 claims on its website:

The N2H2 Human Review Advantage

N2H2 employs a full-time staff to compile its extensive categorized database of Web content. While others rely solely on technology to detect and harvest Web content, N2H2's proprietary process uses a unique combination of technology and human review. This process reduces frustrations associated with "keyword blocking" methods including denied access to sites regarding breast cancer, sex education, religion, and health. Effective human review - like the processes employed by N2H2- is the only way to ensure accurate categorization of Web content.

Assuming Cyveillance's conservative estimates of Web size, assuming blocking software companies really are using human review to assign block codes to web pages, and assuming that blocking software company employees work with the same diligence as the study researchers, it would have required 57,534 people working eight hours with occasional breaks every day with no weekends or holidays off for the entire year of 2000 to assign block codes to the existing web pages on the Internet. It would require another 70,000 people working each day to keep up with the growth rate of the Internet in 2000. Even if Internet growth slowed substantially and even if blocking software company employees have developed techniques to review web pages an order of magnitude faster than study researchers, there is no practical way for companies of the size of current blocking software companies to engage in any meaningful kind of human review to provide block code assignments for all or even a significant portion of the web pages on the Internet.

N2H2 Blocking Codes

N2H2 Bess blocked more with the Free Pages and Electronic Commerce block codes by far than with the other block codes. The product therefore blocked vast swathes of the Internet simply for publication on a web page that the publisher may or may not have paid to publish and those published as commercial web pages, regardless of whether the page actually offered anything for sale.

N2H2 Blocking Code Distribution

Out of 31 total blocking codes encountered, the top ten N2H2 Bess codes that blocked web pages most frequently in the public school installation researched by this study were:

- 1) Free Pages (block, 38.68% of all pages blocked, 13,193 pages blocked)
- 2) Electronic Commerce (block, 21.04%, 7,177)
- 3) Message/Bulletin Boards (block, 7.04%, 2,402)
- 4) Recreation/Entertainment (block, 6.27%, 2,140)
- 5) Pornography (block, 4.09%, 1,396)
- 6) Games (block, 3.07%, 1,048)
- 7) Sex (block, 2.96%, 1,011)
- 8) Profanity (block, 2.92%, 997)
- 9) School Cheating (block, 2.07%, 706)

10) Nudity (block, 1.76%, 599)

N2H2 lists the following block and allow codes as “CIPA-compliant,” in effect pointing out that use of other categories would clearly step beyond the blocking required by the Children’s Internet Protection Act:

1) Pornography (block, 4.09% of all pages blocked, 1,396 pages blocked)

2) Sex (block, 2.96%, 1,011)

3) Education (allow, 0.06%, 22)

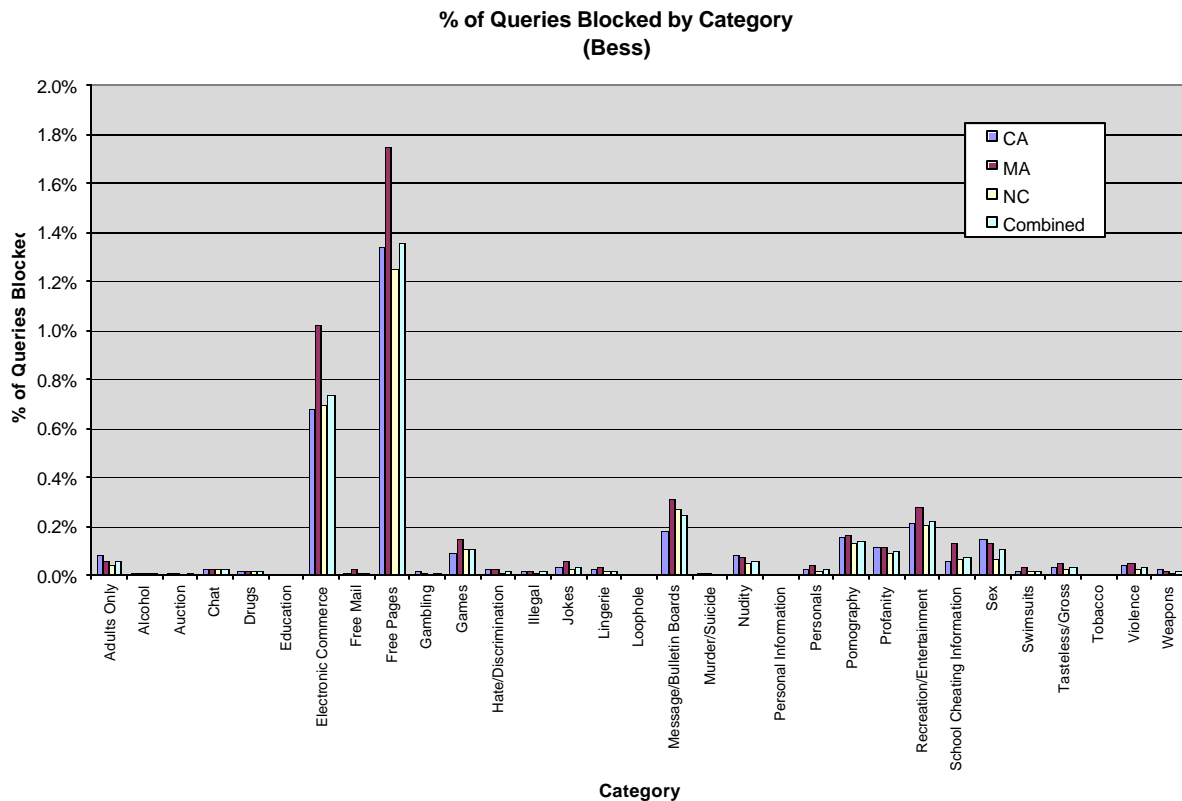
4) History (allow, 0.00%, 0)

5) Medical (allow, 0.00%, 0)

6) Text/Spoken Only (allow, 0.00%, 0)

It is interesting to note that of more than 31,549 curriculum-related web pages blocked by N2H2 Bess in this study, apparently only 22 of the pages, or 0.07% of those curriculum-related pages otherwise blocked by N2H2 Bess, were in fact allowed by the Education “allow” code.

Furthermore, as described in the “Considerations” section, the block codes N2H2 identifies as “CIPA-compliant” also clearly step beyond the blocking required by CIPA.



N2H2 Block Codes Related to CIPA

How does CIPA compliance enter into the picture? If we accept at face value N2H2's own definitions, the researchers found that no more than 2,407 pages, that is no more than 7.05% of all pages blocked, are blocked with block codes that N2H2 has indicated are the codes required for CIPA compliance, that is 4.09% from Pornography and 2.96% from Sex. However, N2H2 Bess blocking codes include not only visual depictions but also non-visual depictions in their definitions, unlike what CIPA requires.

Focusing specifically on web pages blocked in the N2H2 Bess Adults Only, Pornography, and Sex block codes for which CIPA would likely require blocking, the researchers found that out of the sample of 294 accessible pages of 315 pages tested of the total 31,549 curriculum-related web pages blocked by N2H2 Bess in this study, of the 34 pages blocked using the three block codes, N2H2 overblocked 94.12% $\pm 7.63\%$ (95% confidence interval) (32 web pages), and more specifically:

- Of 5 pages blocked as Adults Only, N2H2 Bess blocked 100.00% $\pm 0.00\%$ (95% confidence interval) (5 web pages) inappropriately.
- Of 20 pages blocked as Pornography, N2H2 Bess blocked 90.00% $\pm 13.00\%$ (95% confidence interval) (18 web pages) inappropriately.
- Of 9 pages blocked as Sex, N2H2 Bess blocked 100.00% $\pm 0.00\%$ (95% confidence interval) (9 web pages) inappropriately.

N2H2 Blocking Code Miscategorization

In a sample of the first 266 of the overall 31,549 web pages blocked by N2H2 Bess in this study, N2H2 assigned 79 pages or 29.70% $\pm 5.58\%$ (95% confidence interval) to the wrong block code. In the case of the pages blocked by the Pornography and Sex codes advertised as facilitating CIPA compliance, N2H2 miscategorized every page (100% miscategorization) in the sample.

Focusing specifically on block code miscategorization of those web pages blocked in the N2H2 Bess Adults Only, Pornography, and Sex block codes most pertinent to CIPA, the researchers found that out of the sample of 294 accessible pages of 315 pages tested of the total 31,549 curriculum-related web pages blocked by N2H2 Bess in this study, of the 34 pages blocked using the three block codes, N2H2 Bess miscategorized 85.29% $\pm 11.49\%$ (95% confidence interval) (29 web pages), and more specifically:

- Of 5 pages blocked as Adults Only, N2H2 Bess miscategorized 100.00% $\pm 0.00\%$ (95% confidence interval) (5 web pages).
- Of 20 pages blocked as Pornography, N2H2 Bess miscategorized 90.00% $\pm 13.00\%$ (95% confidence interval) (18 web pages).
- Of 9 pages blocked as Sex, N2H2 Bess miscategorized 66.67% $\pm 31.02\%$ (95% confidence interval) (6 web pages).

SurfControl Blocking Codes

SurfControl blocked more web pages with the Education block code by far than with the other block codes.

SurfControl Blocking Code Distribution

With all 40 blocking codes activated, the top ten SurfControl codes that blocked web pages most frequently in the public school installation researched by this study were:

- 1) Education (50.96% of all pages blocked, 318,049 pages blocked)
- 2) Government & Politics (9.08%, 56,647)
- 3) Arts & Entertainment (7.63%, 47,611)
- 4) Computing & Internet (6.58%, 41,069)
- 5) Health & Medicine (5.23%, 32,655)
- 6) Reference (3.27%, 20,404)
- 7) Hosting Sites (3.20%, 20,001)
- 8) News (2.05%, 12,821)

- 9) Shopping (2.02%, 12,623)
- 10) Lifestyle & Culture (1.92%, 12,010)

The 10 “core” SurfControl block codes blocked curriculum-related web pages with the following frequencies:

- 1) Adult/Sexually Explicit (0.17% of all pages blocked, 1066 pages blocked)
- 2) Chat (0.02%, 144)
- 3) Criminal Skills (0.02%, 144)
- 4) Drugs, Alcohol & Tobacco (0.07%, 406)
- 5) Gambling (0.03%, 215)
- 6) Hacking (0.00%, 27)
- 7) Hate Speech (0.03%, 212)
- 8) Violence (0.02%, 134)
- 9) Weapons (0.04%, 240)
- 10) Web-based Email (0.02%, 94)

And the three additional SurfControl block codes for the “core plus” group blocked curriculum-related web pages with the following frequencies:

- 11) Glamour & Intimate Apparel (0.03% of all pages blocked, 196 pages blocked)
- 12) Personals & Dating (0.03%, 180)
- 13) Sex Education (0.07%, 464)

SurfControl Block Code Related to CIPA

Focusing specifically on web pages blocked in the SurfControl Adult/Sexually Explicit block code for which CIPA would likely require blocking, the researchers found that out of the sample of 324 accessible pages of 351 pages tested of the total 3,522 curriculum-related web pages blocked by the SurfControl “core plus” configuration in this study:

- Of 93 pages blocked as Adult/Sexually Explicit, SurfControl blocked 93.55% \pm 4.37% (95% confidence interval) (87 web pages) inappropriately.

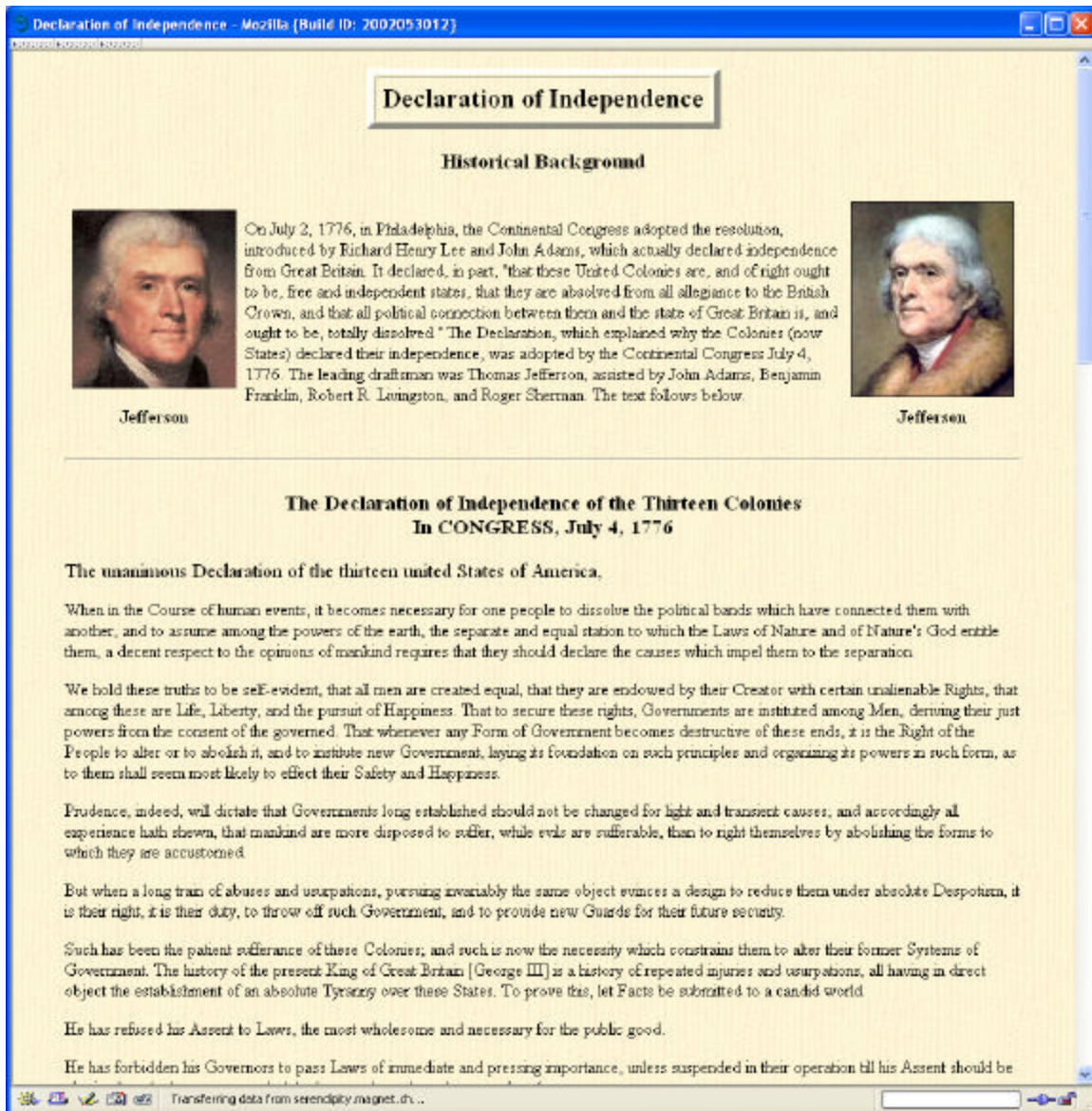
SurfControl Blocking Code Miscategorization

In a sample of the first 265 of the overall 3,522 web pages blocked by SurfControl “core plus” configuration in this study, SurfControl assigned 65.66% \pm 5.61% (95% confidence

interval) (174 web pages) to the wrong block code according to SurfControl's published blocking code definitions.

Researchers also tested 352 pages spread evenly over the entire sample of 3,522 web pages blocked by SurfControl "core plus" configuration and found that SurfControl assigned 54.94% \pm 31.02% (95% confidence interval) (178 web pages) the wrong block code according to SurfControl's published blocking code definitions.

For example, SurfControl assigned the Drugs, Alcohol and Tobacco code to the "Declaration of Independence" page blocked containing the historical background and text of the Declaration of Independence located at <http://serendipity.magnet.ch/jsmill/decl.html>



Focusing specifically on block code miscategorization of those web pages blocked in the SurfControl Adult/Sexually Explicit block code most pertinent to CIPA, the researchers found that out of the sample of 324 accessible pages of 351 pages tested of the total 3,522 curriculum-related web pages blocked by the SurfControl “core plus” configuration in this study:

- Of 93 pages blocked as Adult/Sexually Explicit, SurfControl miscategorized 78.49% $\pm 7.32\%$ (95% confidence interval) (73 web pages).

Bad Address and Unreachable Statistics

A small percentage of sites found by the Google search engine were unavailable, either because of temporary problems with their hosts or Internet traffic or perhaps because their owners had removed them from the Internet entirely. Researchers controlled for these sites in the study without significantly impacting the statistical analysis.

For initial blocking testing, the researchers found very low rates of bad addresses and unreachable web pages. This is most likely because researchers tested the web addresses against the blocking products relatively soon after generating search results from the state-mandated curriculum topics on Google.

Overall, the average rate of web pages with bad addresses or that were unreachable started at was 0.24% $\pm 0.01\%$ (95% confidence interval) (4,688 pages out of 1,946,430 tested) for both blocking products combined.

For N2H2 Bess, the average rate of bad addresses was 0.28% $\pm 0.01\%$ (95% confidence interval) (2,716 pages out of 973,215 tested) and the average rate of unreachable pages was 0.13% $\pm 0.01\%$ (95% confidence interval) (1,227 pages out of 973,215 tested) for an overall average rate of 0.41% $\pm 0.01\%$ (95% confidence interval) for web pages with bad addresses or that were unreachable (3,943 pages out of 973,215 tested). Removing “k12” pages from the sample, the average rate of bad addresses was 0.30% $\pm 0.01\%$ (95% confidence interval) (2,555 pages out of 860,939 tested) and the average rate of unreachable pages was 0.13% $\pm 0.01\%$ (95% confidence interval) (1,084 pages out of 860,939 tested) for an overall average rate of 0.42% $\pm 0.01\%$ (95% confidence interval) for web pages with bad addresses or that were unreachable (3,639 pages out of 860,939 tested).

SurfControl did not provide messages that enabled the researchers to distinguish between pages that had bad addresses or were unreachable, so the average rate of both bad addresses and unreachable pages combined was 0.08% $\pm 0.01\%$ (95% confidence interval) (745 pages out of 973,215 tested). Removing “k12” pages from the sample, the average rate of both bad addresses and unreachable pages combined was 0.08% $\pm 0.01\%$ (95% confidence interval) (670 pages out of 860,939 tested).

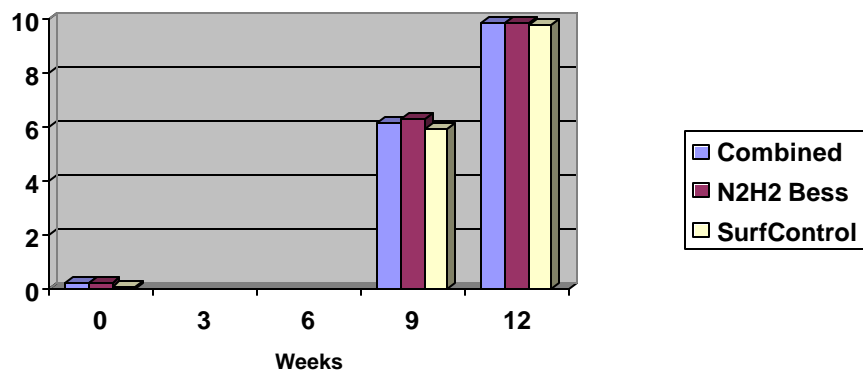
Researchers performed the SurfControl testing the week prior to the N2H2 Bess testing, which may account for the higher rate of bad addresses and unreachable pages for N2H2 Bess.

Overblocking testing took place 9 weeks after the original testing and the average rate of web pages with bad addresses or that were unreachable increased to an overall rate of

6.16% \pm 1.84% (95% confidence interval) or 41 of the sample of 666 web pages tested out of all 35,071 web pages tested for overblocking for both products. For N2H2 Bess, the average rate of web pages with bad addresses or that were unreachable increased to 6.35% \pm 2.73% (95% confidence interval) or 20 of the sample of 315 blocked web pages out of all 31,549 web pages blocked by N2H2 Bess. For SurfControl, the average rate of web pages with bad addresses or that were unreachable increased to 5.98% \pm 2.40% (95% confidence interval) or 21 of the sample of 351 blocked web pages tested out of all 3,522 web pages blocked by SurfControl “core plus” configuration.

Underblocking testing took place 12 weeks after the original testing and the average rate of web pages with bad addresses or that were unreachable increased to an overall rate of 9.86% \pm 0.04% (95% confidence interval) or 84,486 of the sample of 857,055 web pages tested out of all 1,230,648 web pages tested for underblocking for both products. For N2H2 Bess, the average rate of web pages with bad addresses or that were unreachable increased to 9.89% \pm 0.04% (95% confidence interval) or 64,354 of the sample of 650,829 web pages out of all 937,394 web pages tested for underblocking by N2H2 Bess. For SurfControl, the average rate of web pages with bad addresses or that were unreachable increased to 9.76% \pm 0.07% (95% confidence interval) or 20,132 of the sample of 206,226 web pages out of all 293,254 web pages tested for underblocking by SurfControl “core plus” configuration.

If for a moment we assume that rates of bad addresses or unreachable web pages are similar for populations of both blocked and unblocked web pages, then we can chart the increase in bad addresses and unreachable web pages over time.



The chart demonstrates a likely accelerating rate of overall rates of bad addresses or unreachable web pages in this study, increasing approximately 0.67% to 0.83% per week.