Hit Me Again: Design Features that Affect Click-Throughs

Erik C. Hofer, Tom Brinck, and Alfred Speredelozzi
Diamond Bullet Design
315 W. Huron, Suite 140
Ann Arbor, MI 48103 USA
+1 734 665 9307
{ehofer, tom, alfredjr}@diamondbullet.com

ABSTRACT
In this paper we present the results of an empirical study of the relationship between 18 design features, work effort, and two traffic-based metrics of use, visits and pages-per-visit, in a sample of 44 professionally designed web pages. The results highlight important design features to consider in the development of an empirically validated set of design guidelines and underscore the value of design effort. These findings also provide a new perspective for evaluating previous analyses based on expert reviews and user tests of web sites.

Keywords
Guideline Validation, Web Site Design, Log Analysis, Return on Investment, Empirical Studies, World Wide Web

INTRODUCTION
Web site designers would like clear guidance on what effect their design decisions have on the success of a site. Therefore, a few studies have examined the relationship between specific design features and various performance measures of a site, such as expert ratings of the site or task times and completion rates in user testing. In this study, we augment that research with an analysis of the relationships between design features and hit log data gathered across a variety of web sites. This work offers the potential to provide web site designers with design guidelines, helps web site evaluators spot possible problems and predict web site performance (possibly with automated tools), and helps us begin to construct a theory of how web pages are used.

Despite recent efforts to resolve the debate concerning what design guidelines contribute to high-quality web site design, uncertainty regarding which factors influence good design still exists [9]. Two obstacles to the development of web page design guidelines are the lack of consistent guideline scope and validation. While there is no shortage of web design guidelines, many guidelines provide advice at different levels. On one hand are guidelines that provide abstract advice, such as 'establishing levels of importance' [3], while on the other hand are concrete guidelines, such as 'mailto links should explicitly show email addresses' [1]. The differences in implementation ease for these two types of suggestions make guideline comparison extremely difficult. Additionally, the lack of empirical validation of existing guidelines makes evaluating their suggestions difficult.

A few studies contribute to the validation of existing design guidelines or the development of new guidelines from existing collections of pages [6, 10]. Two data sources dominate in these analyses of the design of World Wide Web pages: user testing results and expert reviews. Neither of these data sources considers actual usage, instead focusing on evaluation based on specific task-related or aesthetic criteria and expert ratings. The hit logs from web servers are a nearly unused data source. Tools for analyzing traffic patterns for sites are used primarily for marketing and systems administration purposes, though some studies have used these logs to examine use [2, 4, 5, 8]. These logs of actual use can supplement findings from user testing and expert evaluation by providing insight into the actual browsing behavior of users and how design characteristics of a particular site influence how that site is used.

In addition to relying on data sources that do not reflect actual use of web sites, empirical studies of factors influencing web site usability have considered samples that include sites of varying overall quality. When quality is not controlled for, it is likely that common mistakes in the use of certain design features drive the evaluation of all uses of those features.

In this paper we report the results of an empirical analysis of page-level characteristics and work effort on a set of 44 web pages produced by a single design firm. Eighteen page-level characteristics were measured for each page, ranging from navigation location, to the average number of links used, to the number of different fonts used. Additionally, data from the design firm's timesheet system was used to examine the effects of total time spent on a project and total time spent by usability professionals on a project. The relationships were evaluated between

cite as:
Hofer, E., Brinck, T., and Speredelozzi, A. *Hit Me Again: Design Features that Affect Click-Throughs*, Diamond Bullet Design Report #U1-01-1, Ann Arbor, MI. Sept. 2001.

these factors and two measures of web site performance: visits and average number of pages-per-visit. The results of the analysis allow us to account for nearly 65% of the variance in the average number of pages-per-visit to a site and nearly 32% of the variance in the total number of visits to a site from page-level design features alone. Additionally, we can account for 37% of the variance in the average number of pages-per-visit and nearly 41% of the variance in the total number of visits from the total number of hours spent on a project.

RELATED WORK

Three types of related research are examined: studies based on log analysis, user testing results related to design features, and expert rating results related to design features. In addition, we mention an effort to organize empirically-derived guidelines.

Log analysis

While quantitative analyses of web sites often focus on usage patterns in server log files, no studies have been used to examine the relationship between specific design characteristics and usage patterns. Two types of information are commonly inferred from server log files: traffic based measures, such as pages-per-visit and number of visits, and time-based measures, such as page-view duration. While time-based information provides a useful measure of user interest [4], accurate readings of page-view duration are extremely difficult to capture. Browser-level caching ensures that very little information about the time spent on a given page can be inferred beyond the initial viewing of a page. The use of proxy servers has the potential to introduce error into the interpretation of the logs, but only if two or more users access the same site from behind the same proxy server at the same time, which is unlikely in most settings. User behavior such as hitting the back button cannot be easily recorded or inferred, making it difficult to provide an accurate count of time spent on each page in a site. While pages-per-visit measures may provide slightly less information about user interest, they are much more reliable, as the total set of pages a user views can easily and reliably be recorded, unlike the amount of time spent on each page.

Several studies have focused on examining server logs to focus on patterns of use within web sites. Hochheiser and Shneiderman [5] explain the use of starfield visualization to explore server log data to enable an expert evaluator to examine the relationship between a variety of factors and use patterns. Chi, Pirolli, and Pitkow [2] use a Dome Tree visualization to characterize common use patterns. Once patterns are identified, the site can be examined to provide evidence supporting common traversal paths and uncover patterns of information scent throughout a site. These approaches provide a way to examine the use of a particular site in detail, but are difficult to use. General comparisons of use across a variety of sites are infeasible due to the level of detail and the time required to conduct an analysis of each site.

User studies

Another related approach to empirically developing design guidelines involves conducting user tests on a variety of web sites and examining patterns in user behavior. Spool et al [10] conducted a series of user tests to investigate the impact of design elements on usability and user satisfaction. Nine web sites were tested on a variety of tasks. Analysis of the user tests reveals several patterns in the factors that contribute to usability and user satisfaction. In particular, Spool et al. examine the relationships between several categories of factors and the performance measures, including site navigation, readability, link characteristics and graphic elements. A set of design suggestions results from the observed relationships.

This study is limited in that it examines a relatively small set of web sites whose quality varies (9 sites total). Additionally, the implementation of design elements may vary significantly across sites, so it is difficult to translate problems with a feature in a limited number of contexts to general conclusions about the usability of that feature. For example, one finding suggests that the use of white space on web pages impairs usability. However, there is no discussion of the quality of the use of white space on sites where it was problematic, causing some ambiguity as to whether the use of white space hinders usability or whether bad use of white space is unusable. In addition, while user testing provides valuable empirical data, user behavior may vary from actual usage in important ways. For instance, to the extent the design features affect user motivation, you would expect less of an effect in user testing, where users act so as to complete all tasks.

Expert ratings

Ivory, Sinha and Hearst [6, 7] present an empirical analysis of the influence of page-level characteristics in predicting expert ratings in a large sample of web sites. Specifically, they examine eleven page characteristics, including word count, body text percentage, emphasized body text percentage, text positioning count, text cluster count, link count, page size, graphic percentage, graphics count, color count, and font count [6]. An automated tool was used to compute the metrics, evaluating a rendered web page similar to how Netscape Navigator renders pages. Based on these page characteristics, it is possible to predict with between 65% and 80% accuracy whether a web page will be judged highly by a human judge, depending on whether or not pages are sorted into categories based on content and audience.

In addition to being able to evaluate the quality of pages, the influence of each metric was investigated. When sites are sorted into categories, all characteristics have a significant relationship to the quality of sites, as evaluated by human experts. This finding suggests a set of design characteristics with strong empirical support.

This study has several limitations in that the expert evaluations do not reflect actual use and that the sample reflects sites of varying quality and varying use of design characteristics.

Guideline validation

In addition to the above studies that provide evidence in support of guidelines, Usability.gov [11] hosts an effort by the National Cancer Institute to provide a clearinghouse of research-supported web page design guidelines. The collection of guidelines provides a starting point for determining a valid set of usability guidelines, but does not provide an easy way to subjectively evaluate the studies that contribute and provides a limited indication of the strength of the research supporting certain guidelines.

METHODOLOGY

Our study examines the relationships between page-level characteristics, work effort, and traffic patterns in a sample of 44 web sites designed and hosted by a professional web design company. Site types ranged from e-commerce sites, to marketing sites, to sites providing information about a particular topic, to bank web sites.

Control measures

The sample contained a large number of sites for banks. For this reason, a measure was introduced to control for the effect of sites in this domain.

Page-level Measures

For each of the 44 sites, 18 page-level measures were calculated on the index page for each site. The index page was chosen because it is the primary point of entry for each site, seen by nearly all visitors, and therefore useful for making inferences about page characteristics that influence further browsing behavior. These measures are summarized in Table 1. The measures were calculated by visual inspection on versions of each site that were live for June 2001. In this sample, no major site revisions occurred during that time.

Work-effort measures

The total number of hours spent on 32 of the web sites was obtained from the organization's timesheet system.

Dependent measures

We examined the server logs for 44 web sites from June of 2001. The server logs of a sample of ten sites from previous months were compared with the server logs of June to ensure that the one-month sample accurately reflects long term usage patterns.

The log files were processed to remove requests from automated agents such as search engines and requests for non-page files, such as style sheets and graphics. Only requests for HTML files, ASP/JSP files, PDF files, and scripts that dynamically generate HTML pages (such as search results pages) were included in the traffic measures. The processed logs were analyzed to produce two measures of use: site visits and pages-per-visit.

Page Characteristic	Definition
Navigation Location	Coded as three dummy variables, for navigation at left, right and/or top
Hierarchical Navigation	Dummy variable to indicate whether navigation is presented as on outline or categorized.
Number of Horizontal Rules	Count of horizontal rules (<hr> tag)
Number of Clusters	Count of graphic and text clusters by visual inspection
Number of Link Words	Count of total words in links, both text and image
Number of Links	Count of links
Avg. Link Words	Average number of words per text link, both text and image
Number of Image Links	Count of image links
Number of Text Links	Count of number of links with words, both pure text and image
Number of Embedded Links	Count of links contained within body text
Number of Wrapped Links	Count of links that wrap from one line to another
Number of Graphics	Count of graphics on the page (<img> tag)
Number of Advertisements	Count of advertisements for other sites or services
Number of Colors	Count of different colors used
Number of Fonts	Count of different fonts used
Number of Mouseovers	Count of mouseovers (rollovers) on the page
Number of Animations	Count of elements that contain animation (excludes mouseovers)
Page-size	Dummy variable to indicate if page can be viewed without scrolling on an 800 x 600 screen.

Table 1: Page-level characteristics measured

The site visits metric reflects the number of sessions during which a user at a given IP address requests one or more pages with less than 30 minutes between page requests. Each session can be considered a visit during which a user explores the web site. The visits measure is a count of the total number of visits.

The total number of page requests is also collected. This count is divided by the total number of visits to produce a measure of the average number of pages presented during each visit for a given site. This measure can also be interpreted as a page-conversion rate or a click-through rate. We feel that the pages-per-visit metric reflects user motivation towards a site. This measure is similar to the time-based metrics of user interest [fuller and de graaf], but provides more reliable information due to problems inferring time spent on sites because of browser-level caching.

RESULTS

Linear regression was used to investigate the relationships between the page-level characteristics, amount of work-effort, and the traffic-based measures. This section describes our findings in detail.

Pages-per-visit – Page characteristics

A backwards linear regression was initially conducted to identify factors related to pages-per-visit. In order to test for the effect of the total number of hours spent in the development of each web site, the sample size was reduced to the 32 web sites that we could obtain work-effort data for. The regression was then refined to produce the following model (see Table 2), which allows us to account for 64% of the observed variance (Adj. R^2 = .644, $F(4,31)$ = 15.042, p < .01). In the model, the average number of words per link, the number of embedded links, and the use of hierarchically structured navigation were positively related to the average number of pages-per visit. Additionally, the control measure for banks showed a negative relationship with the average number of pages-per-visit.

Factor	Beta	Std. Beta
Constant	.810	
Bank (control)	-.796**	-.329
Hierarchical Nav	1.906**	.461
Words per link	.743**	.339
Embedded links	.227*	.288
Adjusted R^2	.644	
$F(3, 31)$	15.042**	

Table 2: Regression model for pages-per-visit (* denotes p < .05, ** denotes p < .01)

Visits – Page characteristics

A backwards linear regression was also conducted to identify page characteristics related to the total number of visits to a site. That regression was refined to produce the following model (see Table 3), which accounts for 32% (Adj. R^2 = .319, $F(2,31)$ = 8.263, p < .01) of the total variance in the number of visits to a site. The use of hierarchical navigation showed a positive relationship while the average number of words per link showed a negative relationship.

Factor	Beta	Std. Beta
Constant	7760.85*	
Hierarchical Nav	8010.21**	.491
Words per link	-2938.82*	-.340
Adjusted R^2	.319	
$F(2, 31)$	8.263**	

Table 3: Regression model for visits (* denotes p < .05, ** denotes p < .01)

Pages-per-visit – Work effort

In order to examine the contribution of work-effort to the average number of pages-per-visit, a model was built containing the total number of hours spent on each web site and no other factors (see Table 4). The model was able to account for roughly 37% of the observed variance (Adj. R^2 = .368, $F(1,31)$ = 19.046, p < .01).

Factor	Beta	Std. Beta
Constant	1.74**	
Total hours	.003**	.623
Adjusted R^2	.368	
$F(1, 31)$	19.046**	

Table 4: Regression model for pages-per-visit (* denotes p < .05, ** denotes p < .01)

Visits – Work effort

Total number of In order to examine the contribution of work-effort to the total number of visits, a model was built containing the total number of hours spent on each web site and no other factors (see Table 5). The model was able to account for roughly 41% of the observed variance (Adj. R^2 = .406, $F(1,31)$ = 22.169, p < .01).

Factor	Beta	Std. Beta
Constant	-472.44	
Total hours	13.40**	.652
Adjusted R^2	.406	
F(1, 31)	22.169**	

Table 5: Regression model for visits (* denotes p < .05, ** denotes p < .01)

LIMITATIONS

There are several validity concerns about this study that need to be mentioned, having to do with how the sample of the site was selected, how the dependent variables are interpreted, and possible confounds for these design features. We believe, however, that the core findings are sound, and that the essence of interpreting these is to examine convergence with or divergence from other studies, which also have limitations in their interpretation, and whose differences can be informative in our final conclusions. We also feel that a detailed overview of these limitations is important to understanding the directions future research needs to explore.

Sampling of Sites

Sampling work from a single company has several key advantages, such as guaranteeing that all sites have received a comparable level of professional design and having budget data available, which enables us to consider the impact of different levels of effort on a site's performance. However, this has also limited the size of our data set and limits our ability to generalize conclusions to all types of sites. For instance, none of the sites in this sample were larger than a few hundred pages, and conclusions might be different for larger sites. Because an in-house style has been developed, a variety of design features may co-vary, and a wide range of design alternatives may not have been explored in the sample.

Interpreting the Dependent Measures

While some factors external to the design may play a role in increasing pages per visit, it appears most likely that people are making a decision about clicking through to additional pages based on what they see in terms of design and content on the first page they visit. The alternative is that something attracted high-click-through people to these sites, and that this would co-vary with the factors identified. For example, more complex topics may require more words per link and also attract more motivated users, who spend more time on the site. This potential problem is more obvious in the case of visits, where there is more reason to believe that the count is driven by factors that occur before people even arrive at the site. However, for visits, some of the effect may be attributable to revisits. That is, these design features may be improving the experience sufficiently to encourage users to return to the site.

The other concern with these measures (visits and pages-per-visit) is a question of whether these are the desired goals of web site design. We would like to uncover measures that reflect usability and profitability of a site. The main advantage of these measures is that they reflect actual usage. Visits is an approximation of the overall popularity and the re-use of a site, while pages-per-visit reflects people's desire to continue using a site. The main problem with the pages-per-visit measure is that click-throughs don't necessarily indicate that users were satisfied or completed their tasks, and they may in fact be viewing more pages because of difficulty finding the desired information. However, since typical pages-per-visit were between 2 and 3 per site, it seems likely that this does not reflect people being lost on sites. The other possible problem is that click-throughs are a necessary fact of using certain types of sites. That is, some sites may have target information a level deeper than others, and this may lead designers to create hierarchical navigation and use more descriptive links in these cases.

Co-varying Design Features

The last type of problem is confounds, some of which were mentioned in the preceding paragraphs. For site visits, the main idea is that what design features are requested by the client or what features are selected by the designer may correspond to other behavior that is bringing users to the site, such as marketing plans and site goals. The main factor where this seems like a plausible aspect is the number of graphics used on a page. It is reasonable to imagine that clients who put a higher priority on marketing will both request a more graphical page and spend more on marketing the site.

For pages per visit, the biggest concern is that some other design feature that we did not include in our analysis is responsible for observed differences, and that the other design feature is correlated with the one that appears to have an effect. This concern is made more plausible by the fact that these sites were designed and developed by a small group of people at a single company. An individual designer may have a specific design style that differs from another designer on a number of dimensions, and the style of that designer may be impacting the response to the site in unanticipated ways.

DISCUSSION

The clearest result from this study is that the use of hierarchically-structured navbars leads to greater visits and greater pages per visit. In Spool et al's work [10], hierarchical navbars resulted in confusion for some users in user testing. Spool's sample of sites was relatively small (9 sites), and the difference seems likely due to the different types of sites and design of the hierarchical navigation involved. Most of the sites in Spool's work were quite large, and a clear possibility is that hierarchical navigation is not as successful on very large sites than on smaller ones. Another possibility is how the hierarchical navigation is designed and displayed. Given the greater

consistency in style of navbars in our sample, it is possible that a hierarchical navbar can be designed to be effective, and more work is needed to uncover what makes it effective, such as the selection of categories or the display of relationships between levels of the hierarchy.

The next most consistent finding was the impact of using more words in a link. Greater words per link resulted in more pages-per-visit (but fewer visits). This measure counted words in both graphics and text links. The success in pages-per-visit agrees with Spool et al's work, where they found that longer link names resulted in greater success in finding information in user testing. With these two measures converging on the same conclusion, it seems that this can be a strong design recommendation. This phenomenon is likely because longer link names, when well-designed, provide more information about what they lead to. That is, they have higher quality information scent. In user testing, this should result in greater success at finding the target information in shorter time. In practice, website designers often desire to shorten link names to satisfy page layout constraints, which may contribute to a more desirable site appearance. In fact, the compromise of a less desirable site appearance might be what is responsible for the fewer overall visits (though this would require deeper analysis). Thus, this tradeoff remains pertinent.

More embedded links led to greater pages per visit, but had no significant effect on total visits. The lack of effect on visits seems reasonable, as it wouldn't be expected to be a major factor in attracting people to the site originally or in revisits. This result disagrees with the finding in Spool et al, where they found embedded links to be detrimental to task success in user testing. Given the small sample size in Spool et al's study, we believe the poor performance of embedded links was most likely due to poor usage of embedded links on the few sites that used them. It is also likely that actual use of sites is more likely to involve users actually reading the text of sites and thus encountering embedded links, whereas task-driven user testing would more likely lead to rapid scanning of link options without reading text and noticing embedded links, or not understanding the meaning of embedded links without having read the context.

In summary, the following design guidelines could be drawn from this work:

- except possibly for very large sites, present hierarchical navigation, but design it carefully,
- use longer, more descriptive links to achieve greater task success and usage-per-visit, though this may reduce total visits, and
- embed links into the body text of your site.

In addition to the relationships between design features and the measures of use, we also observe a relationship between the total time spent in the development of a web site and the measures of use. It is important to note that the number of hours spent on a project does not contribute to our understanding of site use when design features are accounted for. This is likely because several factors co-vary with the time spent on a site, including the use of certain design features, overall design quality, and the quality of design decisions. We do feel, however, that this finding is important because it reflects potential return on investment from web design work, as use is desirable to web site owners.

In explaining the relationships between work-effort and visits, two explanations are likely. First, the more time spent on a site contributes to a better site design that is better tuned to the intended audience, which leads to a more popular site and repeat visitors. The second possibility is that the time spent on a site co-varies with the marketing budget of the web site, which leads to more visits. The observed effect is likely a combination of these two explanations, reflecting both better design and a higher marketing budget.

For the observed relationship between the time spent in the development of a web site and the average number of pages-per-visit, it is likely that spending more time on site development enables better design. The more time that designers and usability engineers can spend developing a site, the more likely it is that good design decisions will be made and more alternatives can be evaluated.

While these relationships between the time investment and the use of a site are not fully explained, they provide evidence of return on investment from design work and provide an area in which to direct future work.

FUTURE DIRECTIONS

This work complements a set of studies evaluating the impact of design decisions on web site performance. The overall research agenda needs to explore a variety of performance measures and design parameters and consider other sample sets of web sites. Not all the performance measures should be expected to agree, and in fact, as in our discussion, the disagreements can reveal other possible dimensions that are important factors in how sites are used.

In addition to addressing obvious limitations, a further avenue to explore is to refine our understanding of the variables we see are having an effect. For instance, since we see that more words per link is beneficial, we can begin to explore the cause and the factors involved. What specific guidelines can we give for writing a longer link label? How long is too long? In what contexts would this not apply? When are concerns of visual layout adequate to merit the shortening of labels? For instance, when does reducing the length of labels for the visual layout (e.g. to improve alignment or the density of information) sufficiently improve the clarity of the design to balance the effect of reduced information scent?

Return on Investment

A fundamental issue in design is whether investment in design effort and usability effort is producing a more successful site, that is, a return on the investment. In the current analysis, we had information about how many hours were spent developing a site. Considered by itself, more hours spent on a web project leads to greater visits and greater pages per visit, but unfortunately this result did not remain true in the overall analysis that considered all the design factors. The problem is that several factors co-vary. So for instance, a larger site will have more hours spent developing it and will be more likely to use a hierarchical than a flat navbar. In future analyses, and potentially with larger data sets, we hope to be able to tease these factors apart. In addition, we can potentially evaluate *which* work activities are most productive in creating a better site, and what proportion of different types of work is best. For instance, we'd like to know, for a given budget, what amount of investment in usability evaluation is optimal for producing a successful site.

ACKNOWLEDGEMENTS

Our thanks to our friends and colleagues at Diamond Bullet Design and the University of Michigan, especially Darren Gergle, Seunghee Ha, Mark Handel, Derren Hermann, Stephen Markel, Juliane Morian, Aric Watson and Erik Zempel.

REFERENCES

1. Brinck, T., Gergle, D., and Wood, S.D. (2001). Usability for the Web: Designing Web Sites that Work. Morgan Kaufmann Publishers.
2. Chi, E.H., Pirolli, P., and Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In *Proc CHI 2000*, ACM Press: New York.
3. Detweiler, M.C. and Omanson, R.C. (1996), Ameritech Web Page User Interface Standards and Design Guidelines (www.ameritech.com).
4. Fuller, R. and de Graaff, J.J. (1996). Measuring user motivation from server log files. In *Proc. of HFWEB 2*.
5. Hochheiser, H. and Shneiderman, B. (1999) Understanding patterns of user visits to web sites: Interactive starfield visualizations of WWW log data. Technical Report CS-TR-3989, University of Maryland – College Park.
6. Ivory, M.Y., Sinha, R.R., & Hearst, M.A. (2001). Empirically validated web page design metrics. In *Proceedings of CHI 2001*, ACM Press: New York, pp. 53-60.
7. Ivory, M.Y., Sinha, R.R., & Hearst, M.A. (2000). Preliminary findings on using quantitative measures to compare favorably ranked and unranked information-centric web pages. In *Proc. HFWEB 6*, Austin, Texas.
8. Pitkow, J.E. and Pirolli, P. (1999) Mining longest repeated sequences to predict World Wide Web surfing. *Second USENIX symposium on Internet technologies and systems.*
9. Ratner, J., Grose, E.M. and Forsythe, C. (1996). Characterization and assessment of HTML style guides. In *Proc. CHI '96*, ACM Press: New York.
10. Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C. and DeAngelo, T. (1997), Web Site Usability: A Designer's Guide, North Andover, MA: User Interface Engineering.
11. Usability.gov (2001). Research-based web design and usability guidelines. usability.gov/guidelines.