Design and Formatting in Internet-Based Research

The enormous growth in Internet-based research over the past decade has brought with it numerous questions about the most effective way to administer surveys and experiments via the Web. Some research has already been done examining seemingly minor but methodologically substantial issues, such as the relative effectiveness of different presentation formats. Formats can have an effect on sampling: Buchanan and Reips (2001), for instance, found personality differences between Macintosh and PC users, and educational differences between those who had JavaScript turned on in their Web browsers and those who did not.

One of the few disadvantages of Internet-based research is the difficulty of ensuring understanding of the instructions and materials with remote participants (Reips, 2002c), so every effort should be taken to ensure that Internet-based studies are designed most appropriately to fulfill this prerequisite. The present chapter summarizes what can be done and covers basic information that helps researchers:

- become sensitive to design and formatting issues and identify how they may play a role in their Internet-based research and
- find out how to best create and run their own Web studies in a way that design and formatting do not interfere with the research but rather support its cause.
This chapter will thus be useful to behavioral scientists who are considering Internet-based data collection as part of their research strategy or as a topic of their teaching. It will also be useful to those who are considering setting up traditional laboratory studies with Internet technologies or are creating Web pages to interact with users, clients, or students in any other ways.

Researchers who read this chapter will be able to identify how and why a certain way of designing a Web study may be a useful method for their research and what to look out for when choosing a different format. It will also help in analyzing work by others, as a reviewer of a paper or grant proposal, for example, or as a teacher, and will help in pinpointing likely problems with the design.

Examples of Design and Formatting Issues

Everyday design and formatting issues can be observed in questionnaires on the Internet. Basically, the issues can be grouped into blatant errors and design decisions made consciously but without the designer realizing that the format makes the site less usable or worthless for research purposes. Figure 3.1—a screenshot from a real student project survey submitted to the Web survey list at http://wexlist.net/—shows several blatant errors in survey design and use of Web forms. These are, in order of severity:

- preselected answers in all drop-down menus (see discussion later in the chapter);
- overlapping answer categories (e.g., which option to choose if one regularly has online contact with 20 online friends?),
- size of text to be entered in text fields is not limited (I entered series of 9s in some cases to demonstrate the problem),
- lack of options that indicate reluctance to answer (e.g., “don’t want to answer”),
- all items on one run-on Web page (see discussion below), and
- incorrect writing (e.g., missing comma in the first sentence, other punctuation mistakes, repeated words, and confusing structure of the third item in “Social Networking”).

Furthermore, the URL of the survey (not shown) includes the word student and thus may (correctly) convey the impression that the survey is part of a student project, thereby not requiring the same attention as what the potential participant may deem “serious research.”

When designing a Web study, one is repeatedly confronted with decisions. How many question items may a study have on the Internet? Should question items be all on one page, chunked into groups of items,
Design and Formatting

FIGURE 3.1

Example portion of a questionnaire that recreates a student research project on the Web. It shows several errors in study design and use of form elements that will inevitably result in biased data.
or in a one-item—one-page format? What types of (new) response options are there on the Web, and how do they work? What are the pitfalls of using them? What will happen if you tell participants at the beginning of the study how long it will take, and what if you do not? What if you provide feedback on respondents' progress during the study (e.g., by using a progress bar)? Where should demographic questions be placed, at the beginning or at the end, and how will this affect data quality? What are hidden formatting issues? How much of an issue is order of items or pages for Internet-based research? Once you know what you want to do, how do you create and edit the pages—Which software is available (for free?) and which one best suits your purposes? After reading this chapter you will know how to answer these questions.

Understanding Why Web Design and Format Are Important for Internet-Based Research

At the core of many of the more important methodological problems with design and formatting in Internet-based research are interactions between psychological processes in Internet use and the widely varying technical context (Krantz, 2001; Reips, 2000, 2007; Schmidt, 2007). Data quality can be influenced by degree of anonymity, and this factor as well as information about incentives also influences the frequency of drop out (Frick, Bächtiger, & Reips, 2001; O'Neil & Penrod, 2001). The degree of personalization and the power attributable to the sender of an e-mailed invitation to participate in an Internet-mediated study affect the response rates (Joinson & Reips, 2007). Dillman and colleagues (Dillman & Bowker, 2001; Smyth, Dillman, & Christian, 2007; Smyth, Dillman, Christian, & Stern, 2006) have shown that many Web surveys are plagued by problems of usability, display, sampling, or technology. Design factors such as the decision whether a one item, one screen (OIOS) procedure is applied may trigger context effects that lead to results differing from those acquired with all questions on one Web page (Reips, 2002a, 2007).

DESIGN FACTORS

Design may explain differences between results from online and offline methods (for mixed modes, see Shih & Fan, 2008), and researchers are trying to find ways of bringing established offline methods to the Internet (e.g., to recreate the personal presence of an interviewer by using
response options
are the pitfalls of participant
response efforts at the design of
the high entrance barrier or high-hurdle
technique, a package of procedures
that can be applied to provoke early drop out
to trigger compliance after
someone makes the decision to participate (for a detailed explanation,
see chap. 13, this volume). Several factors that
can lead to a participant keeping motivation to be in the study are
often placed at the beginning of a Web experiment (i.e., on the
general instructions page), for instance,
when researchers

- "tell participants participation is serious, that science needs
good data;
- personalize the research (e.g., by asking for e-mail addresses,
phone number, or both);
- tell participants they are traceable (via their computer's IP address);
- are credible (e.g., by telling participants who the researchers are
and what their institutional affiliation is);
- tell participants how long the Web experiment will take;
- prepare participants for any sensitive aspects of the experiment
(e.g., 'you will be asked about your financial situation');
- tell participants what software they will need (and provide them
with hyperlinks to get it);
- perform Java, JavaScript, and plug-in tests" (Reips, 2000,
pp. 110-111); and
- ensure that the technique can be supported by adjunct procedures
(e.g., a Web design that results in systematic shortening
of loading times).

THE ONE ITEM, ONE SCREEN, DESIGN STRATEGY
AND OTHER ISSUES OF GROUPING AND LENGTH

Even for very long studies with several dozens or several hundred items,
you will be able to find participants on the Internet (see, e.g., chaps. 10
and 11, this volume). Generally, however, about 25 to 30 questions
of medium complexity should be the upper limit (Gräf, 2002; Krasilovsky,
1996; Tuten, Urban, & Bosnjak, 2002). Medium complexity means a level
of complexity between single-item, two-choice questions (i.e., not
complex) and matrix question or lengthy open-ended questions (i.e., highly
complex).

Items can be grouped in thematic chunks of three to four on a
single screen. However, there are several good reasons to always consider
the OIOS strategy:

- Context effects (interference between items) are reduced (Reips,
2002a),
meaningful response times and drop out can be measured and used as dependent variables, and
the frequency of interferences in hidden formatting is vastly reduced (for a clarifying example with radio buttons named “sex,” see Reips, 2002b).

Crawford, Couper, and Lamias (2001), among others, investigated the effects of telling participants at the beginning of a Web study about its duration. If a longer time was given, then fewer persons began the study. However, once they had decided to participate, more finished the study than those in a group that was given a lower duration. Similarly confirming the high-hurdle idea, Ganassali (2008) concluded, “We can say that the decision to quit the survey is influenced by perceived length and by style of wording, on the very first pages of the form.” She found experimental evidence for positive effects of short, direct, and interactive formats in Web surveys. Progress indicators appeared to have a negative effect on completion rate in Crawford et al.’s study. More recent studies have indicated that there may be an interaction with length of study: Progress indicators have a motivating effect in short studies and are demotivating in long studies (see also chap. 15, this volume).

Frick, Bächtiger, and Reips (2001) showed lower drop out and better data quality, that is, significantly fewer missing data, for early versus late placement of demographic questions. They also manipulated whether information about an incentive for participation in the experiment was given at the beginning or the end. Drop out was 5.7% when both incentive information and demographic items were given at the beginning and was 21.9% when both were given at the end, with the other two conditions in between: 13.2% when incentive information was given at the beginning and demographic items at the end, and 14.9% vice versa.

A WORD ON (HYPER)TEXT

Text is more difficult to read on screens than on paper, and handling online documents involves more behaviors than opening books and flipping pages. Thus, keep in mind that Internet participants will more quickly reach their thresholds for perceived burden and attention capacity, and their cognitive processes may more easily get distracted. Studies on knowledge acquisition with hypertexts show that a linear static format better supports understanding for single texts than does an active presentation, in which users have to decide themselves where to move and then scroll or click to do so (Naumann, Waniek, & Krems, 2001). So you fare best to use the linear static format as the default, because most Internet studies are of that type. However, for multiple hypertexts, the construction of a mental model is moderated by task: Active presenta-
tion supports understanding in argumentative tasks and static presentation format does so for narrative tasks (Hemmerich & Wiley, 2002; Wiley, 2001). Accordingly, format your study materials in line with the tasks you are using.

### Formatting Issues in Study Design

In this section, I cover issues in formatting of Web studies, namely, how to avoid pitfalls in creating page titles, page names, and Web forms, and how to use hidden formatting, and response options. Note that many of the recommendations are followed by some, but by far not all software that generates Internet-based studies: Check before you buy.

#### TITLES AND NAMES

Generally, use an informative title on Web pages announcing the study and on the first page of the study to attract participants. For subsequent pages, use uninformative page titles so people will not find these pages via search engines and enter the study somewhere in the middle.

#### HIDDEN FORMATTING

Remember that Web page design is a bit like carving icebergs: Much is happening below the surface. Hidden formatting can become very complex and complicated. Server- and client-side scripting and server-side inclusion are powerful tools that allow Web designers to achieve wonderful things if everything works well. In the interest of space, I do not discuss those advanced possibilities here; for more information on this, the interested reader can read related chapters in this volume and the additional resources at the end of this chapter. Here, I will describe some basics of hidden formatting for illustration.

As an example of the power of hidden formatting, consider open questions. These are created by using text area or text field tags. Because just one of your Internet participants may, for example, copy and paste the entire text of *Harry Potter and the Goblet of Fire* into your carefully formulated text field to crash your server or overwrite your data file, you need to limit the number of characters that can be pasted into text fields.

---

1. For more information on server-side includes see [http://en.wikipedia.org/wiki/Server_Side_Includes](http://en.wikipedia.org/wiki/Server_Side_Includes)
This is done by including a hidden “maxlength” attribute with the text field tag, for example:

```html
<input type="text" name="myopenfield1" size="24" maxlength="30" border="0"/>
```

Here, the text field will be displayed as 24 characters long (i.e., size), but it will take up to 30 characters (i.e., maxlength).

An important technique in Internet-based research is the use of meta tags. Meta tags serve a variety of purposes, for example they keep search engines away from all pages except the first page (so participants do not enter the study on one of the later pages), and they prevent the study materials from being cached. Meta tags are hidden in a Web page’s “head” section and look as follows on WEXTOR-generated (http://wextor.org; Reips & Neuhaus, 2000) pages:

```html
<meta name="ROBOTS" content="NONE">
<meta http-equiv="pragma" content="no-cache">
<meta http-equiv="expires" content="Thursday, 1-Jan-1991 01:01:01 GMT">
<meta http-equiv="content-type" content="text/html; charset=ISO-8859-1">
```

The “ROBOTS” tag is set to “NONE,” because the routines used by search engines to search the Web for new Web pages are named robots (sometimes spiders or crawlers). Thus, the ROBOTS tag informs the routines there is nothing to be catalogued. The two meta tags that follow prevent caches in mediating servers, search engines, and proxy servers from serving old versions of research materials after they have been updated. Caches contain stored files downloaded from the Web, for later reuse. Internet providers and large institutions run computers with large hard disks (mediating servers or proxy servers) to store hypertext markup language (HTML) code, images, and other media. If a page is requested again from within their network, the server quickly checks in the cache if it holds any of the text and media and sends it instead of letting the request go all the way out to the Internet to retrieve the original material. This way, the page can be displayed more quickly, and much unnecessary traffic is avoided. However, the material loaded from the cache may be outdated: If an experimenter finds an error in the material and replaces the Web page, users may continue to be delivered the old version. The meta tags displayed above will prevent this.

Participants may search the Web using keywords and thus find studies that use these terms in a keyword meta tag on the first page. Keywords may be a good way to recruit participants with a particular interest or for long-term Internet-based studies (Birnbaum & Reips, 2005).
CONFIGURATION ERRORS: PITFALLS IN WEB FORMS

Reips (2002b) discussed how to avoid several potential configuration errors with Web study design: Protection of directories, a suitable data transmission procedure, unobtrusive naming of files and conditions, adaptation to the substantial variance of technologies and appearances of Web pages in different browsers, and proper form design (Configuration Errors I–V). Another type of configuration error involves Web forms. Web forms that were available since the World Wide Web consortium (W3C) announced the standard HTML 2.0 in 1995 (Musch & Reips, 2000) constitute much of what can be done interactively on Web pages. Mastering the design of Web forms can be tricky, as can easily be observed on the Web. Figure 3.2 from Reips (2009) shows several problems with a large publishing company’s feedback form:

- Preselected answers (Questions 5 and 7): Skipping the question will show up as having selected the preset option;
- Participant burden (Question 5): Selecting an option other than the default is highly discouraged by signaling further work to the respondent (“Please provide address”); and
- Mandatory responses (Question 6): Even though Question 6 will often need no answer, the survey designers set the field to require a response (for example, a dash), as indicated by “*”.

For more details on configuration errors in Internet research, see chapter 13 of this volume.

RESPONSE OPTIONS

All traditional response option formats have been researched for use on the Web (e.g., open-ended formats versus close-ended formats; Reja, Lozar Manfreda, Hlebec & Vehovar, 2003). Here, I focus on new types of response options. With these, Internet-based research offers types of dependent variables that are not available in paper-and-pencil research and are impractical in offline computer-based research, such as drop-down menus (selection lists) and visual analogue scales (VAS).

Drop-down menus are manually complex response devices. To use them, participants have to click on the one option that is initially displayed. Then, once it expands to show other alternatives, they have to either scroll and click or drag a certain distance and then release at a certain choice, depending on their operating system. Reips (2002b; Configuration Error V) emphasized the importance of not having a legitimate response preselected, so real answers can later be distinguished from failures to respond. Because of the complexities of the device, Birnbaum and
Improper use of Web form elements that results in biased data. The figure is adapted from Reips (2009) and shows a portion of a publisher’s online questionnaire for feedback on book inspection copies.

Reips (2005) thus recommended precaution with drop-down menus and warned of another potential issue with them:

Another problem can occur if the value used for missing data is the same as a code used for real data. For example, the first author found a survey on the Web in which the participants were asked to identify their nationalities. He noted that the same code value (99) was assigned to India as to the preselected “missing” value. Fortunately, the investigator was warned and fixed this problem before much data had been collected. Otherwise, the researcher might have concluded that there had been a surprisingly large number of participants from India. (p. 474)

VASs markedly show how previously known advantages of a measure that was often not used for practical reasons (e.g., burdensome measurement by hand) in offline environments becomes highly valuable when taken online. Reips and Funke (2008) developed VAS Generator
Design and Formatting

(http://vasgenerator.net), a free tool to create VAS for Internet-based research. Their research shows that the online VAS created with the tool produce data on the level of an interval scale, even for extreme scale lengths of 50 and 800 pixels.

The abundance of further issues with these and other types of response options cannot be covered here. Also, much has not yet been researched or will need to be researched again because the Internet environment is constantly changing.

**Formatting Companions:**

**Page Editors**

In Table 3.1, I provide an overview of program and editor options to create and edit Web pages. Even if you use a fully automated Web application for study design like WEXTOR (http://wextor.org), SurveyMonkey (http:// surveymonkey.com), or Unipark (http://unipark.com), you will need one of these to check and understand your study’s format and make minor edits.

The selection displayed and commented on in Table 3.1 is subjective, and many more editors are available for different operating systems. However, this selection was agreed upon by several instructors at the National Science Foundation’s and American Psychological Association’s Advanced Training Institutes “Performing Web-Based Research.”

**Conclusion**

The present chapter will help you in designing and formatting your Internet-based research. Be aware, though, that many factors interact with each other, and both Web technology and user experience are constantly changing. More research is needed in light of the rapid development in Web technology, changes in user experience, and the consequences of formatting and design in Internet-based research.

**Additional Resources**

A portal with a number of useful services in Internet-based research.
### TABLE 3.1

**Recommendations for the Use of HTML Editors**

<table>
<thead>
<tr>
<th>Software, operating system, and availability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notepad Windows Free</td>
<td>Text only. Allows direct editing of HTML. Be careful saving the file or it will add a .txt extension.</td>
</tr>
<tr>
<td>Word Windows and Mac Commercial</td>
<td>Can create Web pages but adds a lot of Microsoft-specific style information that does not do well in some browsers.</td>
</tr>
<tr>
<td>Dreamweaver Windows and Mac Commercial (free trialware; <a href="http://www.adobe.com/go/trydreamweaver">http://www.adobe.com/go/trydreamweaver</a> - $200, educational discount)</td>
<td>Commercial WYSIWYG editor. Allows direct editing of HTML.</td>
</tr>
</tbody>
</table>

**Note.** WYSIWYG = what you see is what you get.


**References**

of HTML. Be careful with syntax like HTML does not do well in other code (not

functions require this.)

 Allows direct

of

Web.


