Internet experiments: Methods, guidelines, metadata

Ulf-Dietrich Reips*[a]

[a]University of Zurich, Department of Psychology, Binzmühlestrasse 14/13, 8050 Zürich, Switzerland

ABSTRACT

The Internet experiment is now a well-established and widely used method. The present paper describes guidelines for the proper conduct of Internet experiments, e.g. handling of dropout, unobtrusive naming of materials, and pre-testing. Several methods are presented that further increase the quality of Internet experiments and help to avoid frequent errors. These methods include the “seriousness check”, “warm-up,” “high hurdle,” and “multiple site entry” techniques, control of multiple submissions, and control of motivational confounding. Finally, metadata from sites like WEXTOR (http://wextor.org) and the web experiment list (http://genpsylab-wexlist.uzh.ch/) are reported that show the current state of Internet-based research in terms of the distribution of fields, topics, and research designs used.

Keywords: Internet-based research, Internet-mediated science, Internet experiment, Web experiment, methodology, guideline, metadata, online, experiment techniques

1. INTRODUCTION

Internet-based research is now widespread and well-established, methods for Internet-based research are currently one of the hot areas in methodology. In 1998 and 1999 Musch and Reips (2000) surveyed all Internet experimenters they could then find – and despite enormous efforts they found fewer than 50 [1]. Within the ten years since, the field has seen a massive increase in the number of studies conducted on the Internet (now in the tens of thousands), marking a grass-root change in how research often is conducted. One of the methods that are being used is the *Internet experiment*. (For examples of Internet experiments see the *web experiment list* at http://genpsylab-wexlist.uzh.ch/). Several terms are used synonymously for Internet experiments: Web experiment, on(-)line experiment, Web-based experiment, World Wide Web (WWW) experiment, and Internet-based experiment.

There are many good reasons to turn to the Internet for experimental data collection, e.g. access, speed, cost, number, and power. One particularly appealing application is large-scale data collection [2][3][4]. As Reips (2007) puts it, there is however an ultimate reason, the “fundamental asymmetry of accessibility (Reips, 2002b, 2006). What is programmed to be accessible from any Internet-connected computer in the world will surely also be accessible in a university laboratory, but what is programmed to work in a local computer lab may not necessarily be accessible anywhere else. A laboratory experiment cannot simply be turned into a web experiment, because it may be programmed in a stand-alone programming language and lack Internet-based research methodology, but any web experiment can also be used by connecting the laboratory computer to the Internet. Consequently, it is a good strategy to design a study web-based, if possible.“ (pp. 375-376)[5].

In the present paper I will describe methods, guidelines, techniques, and tools in Internet-based research, namely the “seriousness check”, control of motivational confounding, the “multiple site entry” technique, the “warm-up” technique, control of multiple submissions via sub – sampling, the “high hurdle” technique, unobtrusive naming, editing procedures, implementation of order, and pretesting. Furthermore, I will report some empirical findings on the methodology of Internet-based experimenting. In the last part of the paper, I will provide metadata from research sites and discuss their implications for research.

* u.reips@psychologie.uzh.ch; phone +41-44-63 57236 or +1 (650) 276-0589; http://www.psychologie.uzh.ch/sowi/reips/reipspers.html

2. METHODS

In the following I will present six methods that are useful or even necessary in Internet-based research.

2.1 Method 1: Seriousness check

Many people surf the Internet for interesting content, and so it happens that in every Internet experiment there will be a sizeable proportion of participants who would just like to look at the study materials, and not really participate in a serious manner. A simple, albeit important method to identify such irrelevant data entries is to ask the participant at the beginning of an Internet experiment for their motivation in accessing the pages (see Figure 1). In recent studies we observed that the answer to the question is the single best predictor for dropout. Of those answering "I would like to look at the pages only" around 75% will drop, while of those answering "I would like to seriously participate now" only ca. 10-15% will drop. Overall, about 30-50% of visitors will fail the seriousness check, i.e. answer "I would like to look at the pages only".

The seriousness check will also allow interested colleagues of a researcher to indicate that they would rather not have their data taken seriously.

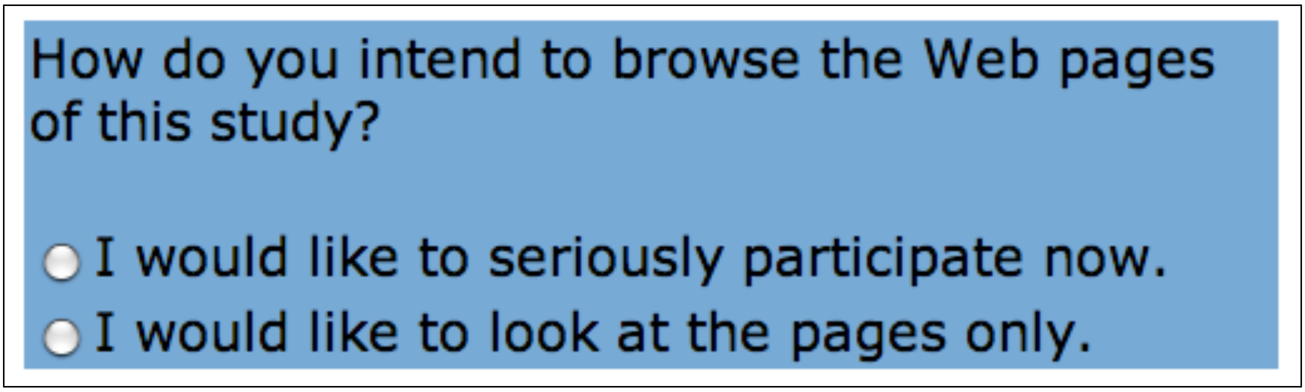

Fig. 1. Implementing the *seriousness check technique*.

2.2 Method 2: Using dropout as a detection device for motivational confounding

Because the experimental setting is highly voluntary there will be dropout in Internet experiments. The dropout is sensitive to variations in motivation, and so it can be used to detect motivational confounding [6][7][8]. For example, if one condition is a less motivating, boring, or very difficult experimental condition then participants might very likely drop out of an Internet experiment (see Figure 2). Because there is usually very minor dropout in offline experiments, the office setting is not as sensitive as the Internet setting in regards to detecting motivational confounding.

2.3 Method 3: Multiple site entry technique

The multiple site entry technique can be used to test if the results depend on sampling in an Internet-based study, and if self-selection rates differ by sample. Several links to the study are placed on different Web sites, in Internet forums, or offline media that are likely to attract different types of participants, or are mailed out to different lists. Placing source identifying information in the published URLs and analyzing different referrer information in the HTTP protocol can be used to identify these sources [9]. Later the collected data sets can be compared for differences in results, and also for differences in relative degree of appeal (measured via dropout), demographic data, central results, and data quality.

A unique string of characters is appended to the URL for each source of recruitment, e.g. "...index.html?source=studentlist" for a list of students who volunteered to participate in studies. The data file will have a column („source") containing an entry of the referring source for each participant ("studentlist"). Figure 3 illustrates the multiple site entry technique.

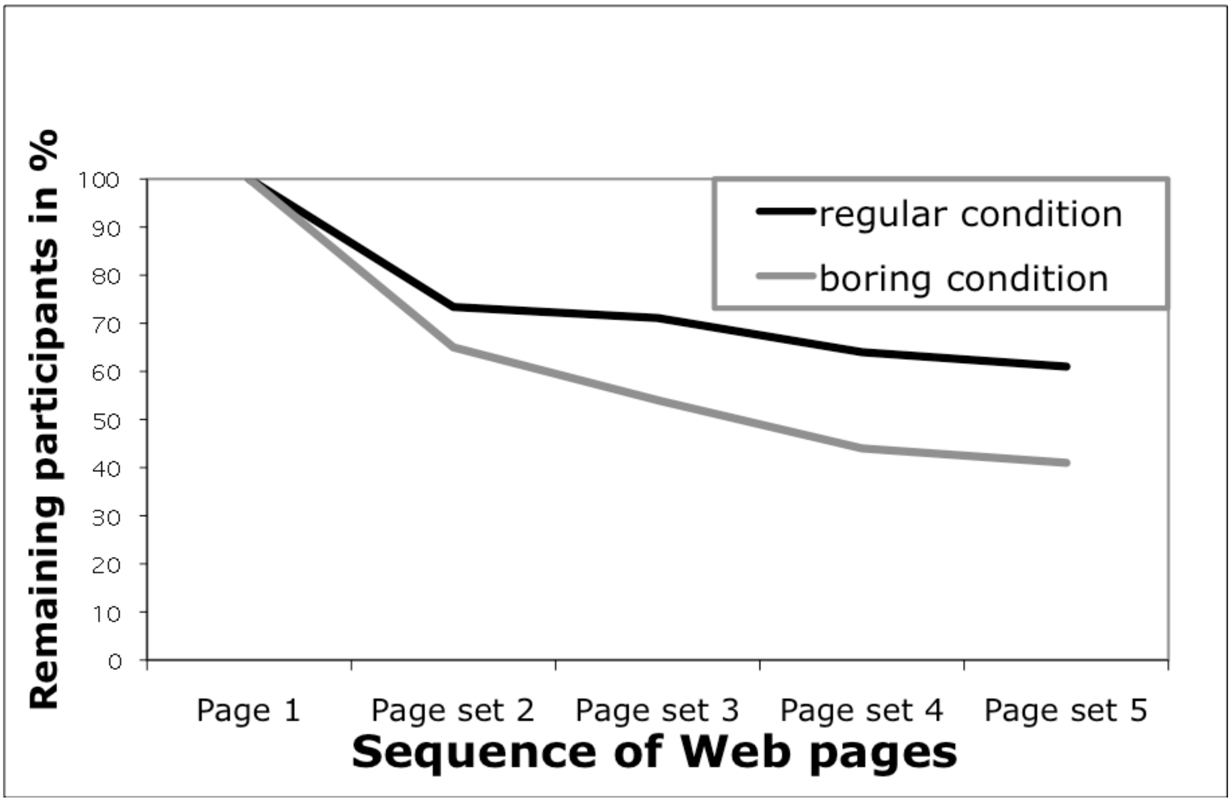

Fig. 2. Dropout in Internet experiments can be used as a detection device for motivational confounding. Figure adapted from [5].

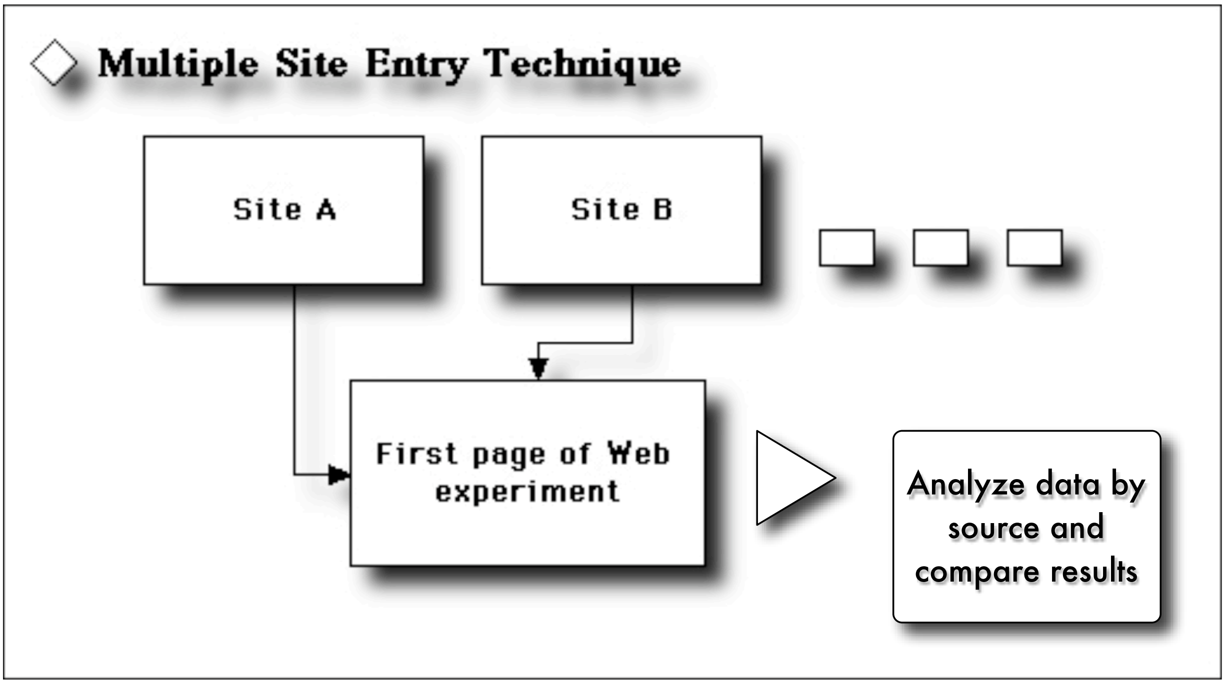

Fig. 3. Illustration of the *multiple site entry technique*. It can be used to estimate the variance of results in different samples. Links to the first page of the study are placed on or posted to different sites, e.g. Web sites, forums, mailing lists, posters, ads etc. (Figure adapted from [6]).

2.4 Method 4: Warm-up technique

The warm-up technique can be used to minimize dropout in the actual experiment. In Internet-based experiments there is almost always a substantial proportion of dropout, because continued participation happens generally on a voluntary basis, with low pre-experimental investment (often no need to sign up, communicate, and physically drive to a laboratory), in subjective anonymity and without presence of an experimenter. Furthermore, there are curious people who are surfing the Web and may just want to take a look at the first few pages of a study (see seriousness check technique above). Most dropout happens at the beginning of an Internet experiment [6][8], the warm-up technique is built upon this observation.

As the technique's name suggests, a period of "warming-up" the participants to the experimental task is inserted before the actual experimental manipulation takes place. Thus, during the actual experiment there is not much dropout. The technique has been shown to work [10]. By applying the warm-up technique Reips, Morger, and Meier were able to reduce dropout to less than 2% during the actual experiment.

2.5 Method 5: Identifying multiple submissions via the sub-sampling technique

A technique that helps control multiple submissions is sub-sampling [6][8]: For a limited random sub-sample from all data sets, every measure is taken to verify the participants' identity, resulting in an estimate for the total percentage of multiple submissions. The method is also useful in estimating the number of wrong answers by checking verifiable responses (e.g., age, sex, occupation). Applications for running Web-based studies may include routines that check for internal consistency and search for answering patterns [11]. Overall, it has repeatedly been shown that multiple submissions are rare in scientific Internet-based research [12][13]. Data quality depends on a number of factors, e.g., whether personal information is requested at the beginning or end of a study [14], on the person who issues the invitation to the study and on how this is done [15]; or whether responding is forced (not allowing participants to leave any items unanswered will cause psychological reactance and thus lead to more random answering) [8][16].

2.6 Method 6: High hurdle technique

Another way of attempting to control dropout is the high hurdle or high entrance barrier technique [6][8][13]. It can be applied to provoke dropout to happen early and ensure continued participation after someone makes the decision to stay. This means bundling of de-motivating factors (i.e. long loading times from large images, long and complicated texts, user-unfriendly design) at the very beginning of an Internet experiment, so visitors with a low motivation for continued participation will likely drop early and those participants who decide to jump the hurdle will feel more committed to finish the study, because according to dissonance theory one is less likely to leave a task the more one has invested [14]. Motivating factors should be implemented increasingly thereafter, enticing participants to continue with the experiment.

While Göritz and Stieger could only find evidence for one of the predictions [17], i.e. higher dropout at the hurdle, our own results indicate that all original predictions for the high hurdle actually work [18]. In long Internet studies, the hurdle will also reduce later dropout and supposedly enhance data quality.

3. GUIDELINES

This section lists a number of recommended guidelines for the design and construction of Internet experiments, the recruitment of participants, the procedures during data collection, and the handling and analysis of data from these experiments. These guidelines are written in a hands-on style, directed to the experimenter.

3.1 Unobtrusive naming

Participants typically make inferences about possible aims and hypotheses of the research they are taking part in. Names of files and folders may deliver subtle clues to that end. Thus, names of files and folders should not be too obvious (configuration error III, [7]). Otherwise, the structure of an Internet experiment may become too visible, causing some participants to switch conditions, skip pages, or correctly infer hypotheses. So, don't name your folders „shock_cond" versus „control_cond", for instance.

3.2 Avoid editing errors

Check for frequent editing errors, among them mixed up or falsely named answer options and missing or superfluous files.

Mixing up answer options: On the Web, checkboxes (square) allow for multiple selections, whereas only one selection can be made in radio buttons (round) that belong to the same group. Belonging to the same group means: In the HTML code the radio buttons have the same "name" (but different values), so on a Web page with several radio button groups a Web browser knows which radio buttons need to be deselected and which don't, if a radio button is clicked.

Wrong naming of answer options: Different groups of radio buttons need different names, each radio button needs a unique value.

Missing index.html etc. files, superfluous files: The experiment folders (directories) contain several important files that need to be present for an experiment to work. Of course, the file structure depends on your particular experiment setup. There is always a default page, e.g. index.html. Frequently, there will also be pages that check a participant's Web browser for compatibility with the scripts used in an Internet experiment, a page that redirects participants whose Web browser fails that test, an informed consent page, etc. Lastly, there shouldn't be any superfluous files, and particularly no files containing participant data or files that reveal background information about the experiment.

Problems on a Web study's last page: The last page is where you thank your participants. Do not forget to provide information on *who* is saying thank you and to whom they can direct questions and commentaries. Provide a name and contact information, preferably an e-mail address and a link to a homepage. However, do not run into configuration error II, whereby participant data may be sent to a different Web site [7].

3.3 Implementing order

The implementation of order depends on the type of design (between-, within- or mixed experimental design). You should consider whether order of items, objects and/or answer options may introduce confounding effects (check out http://www.bolderstats.com/within/index.html for illustrations). Also, contemplate whether order may have a theoretically significant impact on the underlying hypotheses - in this case one would implement order as a factor. If order is not of theoretical significance, it should be implemented by counterbalancing.

3.4 Pretesting

Test by yourself: Check details of all conditions, including all form properties (radio button names and values etc.). Observe two or three people participating, have them speak while they navigate the site and interview them afterwards. Use different Web browsers and operating systems during pretesting. Send the experiment URL to a few colleagues or friends, ask for feedback, check the data, recruit only a part of your sample, wait if any problems surface. Then go for full recruitment.

4. METADATA

The Internet provides researchers with a unique opportunity for easy investigation of behavior by mining data from online interactions, in the traditions of archival research. On a higher level of abstraction, via the same means it is now possible to easily conduct research on the scientific processes in Internet research.

4.1 Metadata from research sites and their implications for research

The recent increase in use of Internet-based research methods is reflected in the strong use of Web services that assist in scientific activities. Web services for Internet-based experimenting such as those listed in the iScience Server at http://iscience.eu naturally collect a large number of data about the Internet experiments that are created or listed on these services. For example, sites for the recruitment of participants like the exponnet site by John Krantz (http://psych.hanover.edu/research/exponnet.html) or the web experiment list (http://genpsylab-wexlist.uzh.ch/) provide the scientific community with data about the relative importance of research areas. Table 1 shows a comparison of data from two studies by Musch and Reips, who surveyed the first Internet experimenters on 35 Internet experiments [1] with data (rounded values) on 250 Internet experiments listed in the web experiment list [19]. For the present paper, I analyzed the current data from this participant recruitment portal (column 4 in Table 1) – it now lists about 600 Internet experiments.

Field	Web survey in 1998/9 (Musch & Reips, 2000)	web experiment list (Reips & Lengler, 2005)	web experiment list February 2009
Cognition	10 (56%)	100 (40%)	230 (39%)
Social Psychology	4 (22%)	90 (36%)	222 (37%)
Perception	1 (5%)	25 (10%)	31 (5%)
Personality	1 (5%)	10 (4%)	19 (3%)
Clinical Psychology	-	10 (4%)	11 (2%)
Internet Science	1 (5%)	10 (4%)	24 (4%)
Applied Psychology	N/A	N/A	16 (3%)
Developmental Psych.	-	3	6 (1%)
Methodology	-	-	10 (1%)
Neurosciences	1 (5%)	2	2
Other	-	N/A	27 (5%)
Sum	18	250	598

Table 1. Fields of study in Internet experimenting. Data from a two-wave Web survey among Internet experimenters in 1998 and 1999 and from the web experiment list. Note. The category „Applied Psychology“ was added after 2004 to the menu of choices in the web experiment list, the 2004 analysis reflects proportions without the „Other“ category.

The results from Table 1 allow us to answer the questions “Which areas of psychological research most frequently use Internet experiments?” (Cognition and Social Psychology) and “Do we see a change in distribution over time?“ (not much). Interestingly, Perception shows the highest relative drop since 2005, and an overall low rate. Not many Internet experiments are published in Perception, with some notable exceptions [3][4][20][21]. The symposium on Internet experiments at the present SPIE conference may hopefully contribute to a turnaround in the use of this method in Perception and the Vision Sciences.

Data from the web experiment list and its sister site, the web survey list, were also used to analyze the frequency of research topics, as reflected in the frequency of terms used in titles of the Internet studies listed. Figure 3 shows a *tag cloud* that was generated from these frequencies. Note that "perception" is among the most frequently used terms, likely due to much colloquial use of the term in the sense "perception of some other psychological phenomenon", similarly to "self-perception" – a term that appears separately. The term "vision" was used only once in a title – as part of "television".

Fig. 4. Tag cloud created from the top 100 of 906 title terms (except common words) with a frequency >2 used in 676 Internet experiments and surveys of English language listed in the combined web experiment and web survey lists. The larger and darker a term, the more frequently it appeared.

So far I have shown how metadata about *fields* and *topics* can be collected. In the following example, I will show how metadata can tell us something about the *typical structure* of Internet experiments. WEXTOR (http://wextor.org) is an experiment generator that is used by about 2'000 researchers and teachers of experimental design. Data from this tool provide us, for example, with a frequency distribution of experimental design types and with information about the use of research techniques (see Figure 5).

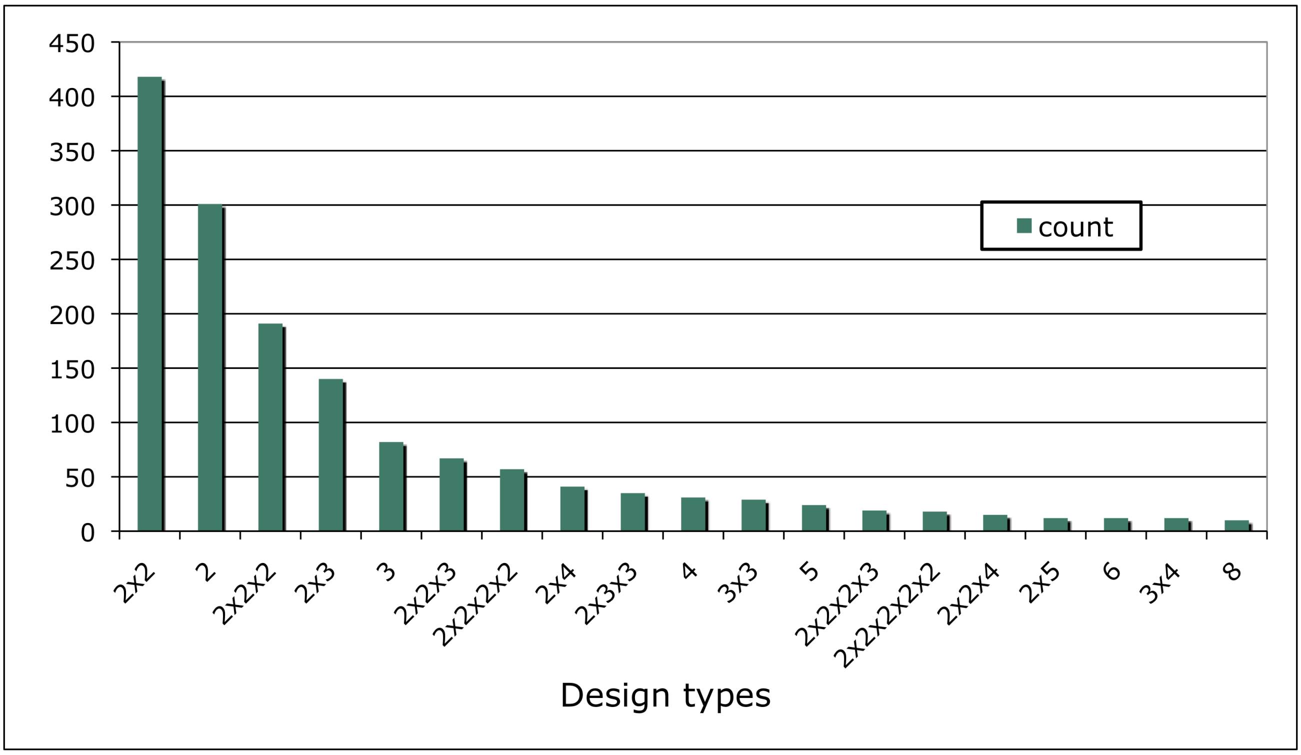

Fig. 5. Frequency of factorial design types in Internet experiments built with WEXTOR (http://wextor.org). November 2007.

5. CONCLUSION

The Internet provides researchers with great new opportunities in conducting experiments and other studies. The present paper described several useful methods and guidelines that are suitable to help in applying the new method, and to avoid common errors. Some typical errors may result from transferring routines that work well in offline experiments to the Internet without making proper adjustment to that situation [5][6][7]. At the root of many of the errors are the characteristics of Internet technology and the methodological consequences that follow from it [8][22]. Modern tools and Web Services for Internet-based research have features built-in that will help experimenters to avoid errors.

Many of the Web services that support Internet experimenters with tasks like the development of an experiment or the recruitment of participants generate metadata. Mining these data was presented as a promising way of identifying structures and trends in experimental research, at least of the portion of research that is conducted via the Internet.

REFERENCES

[1] Musch, J., & Reips, U.-D. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-88). San Diego, CA: Academic Press.

[2] Mangan, M., & Reips, U.-D. (2007). Sleep, sex, and the Web: Surveying the difficult-to-reach clinical population suffering from sexsomnia. *Behavior Research Methods, 39*, 233-236.

[3] Moroney, N., "Thousands of on-line observers is just the beginning," Proc. SPIE 7240, XXX-XXX (2009).

[4] Nakayama, K., "Perceptual experiments on the Web," Proc. SPIE 7240, XXX-XXX (2009).

[5] Reips, U.-D. (2007). The methodology of Internet-based experiments. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 373-390). Oxford University Press.

[6] Reips, U.-D. (2000). The Web Experiment Method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-118). San Diego, CA: Academic Press.
[7] Reips, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20*, 241-249.
[8] Reips, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243-256.
[9] Schmidt, W. C. (2000). The server-side of psychology Web experiments. In M. H. Birnbaum (Eds.), *Psychological Experiments on the Internet* (pp. 285-310). San Diego, CA: Academic Press.
[10] Reips, U.-D., Morger, V., & Meier B. (2001). "Fünfe gerade sein lassen": Listenkontexteffekte beim Kategorisieren ["Letting five be equal": List context effects in categorization]. Unpublished manuscript, available from http://www.psychologie.uzh.ch/sowi/reips/papers/re_mo_me2001.pdf
[11] Gockenbach, S., Bosnjak, M. & Göritz, A. S., "Stereotypic response patterns within matrix questions in Web surveys". *Proc. GOR* (2004).
[12] Voracek, M., Stieger, S., & Gindl, A. (2001). Online replication of evolutionary psychology evidence: Sex differences in sexual jealousy in imagined scenarios of mate's sexual versus emotional infidelity. In U.-D. Reips & M. Bosnjak (Eds.), Dimensions of Internet Science (pp.91-112). Lengerich: Pabst.
[13] Reips, U.-D. (1997). Das psychologische Experimentieren im Internet [Psychological experimenting on the Internet]. In B. Batinic (Ed.), *Internet für Psychologen* (pp. 245-265). Göttingen: Hogrefe.
[14] Frick, A., Bächtiger, M. T., & Reips, U.-D. (2001): Financial incentives, personal information and drop-out in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 209-219). Lengerich: Pabst.
[15] Joinson, A. N., & Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to Web-based surveys. *Computers in Human Behavior, 23(3)*, 1372-1383.
[16] Stieger, S., Reips, U.-D., & Voracek, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology, 58*, 1653-1660.
[17] Göritz, A. S. & Stieger, S. (2008). The high-hurdle technique put to the test: Failure to find evidence that increasing loading times enhances data quality in Web-based studies. *Behavior Research Methods, 40*, 322-327.
[18] Reips, U.-D., & Frauendorfer, D., "The high hurdle technique put to the test: Evidence for less dropout and better data quality in Web-based studies." Manuscript in preparation.
[19] Reips, U.-D., & Lengler, R. (2005). The Web Experiment List: A Web service for the recruitment of participants and archiving of Internet-based experiments. *Behavior Research Methods, 37*, 287-292.
[20] Mulligan, J., "Presentation of calibrated images over the Web," Proc. SPIE 7240, XXX-XXX (2009).
[21] van Ham, F., & Rogowitz, B. E., "Perceptual organization in user-generated graph layouts," *IEEE Transactions on Visualization and Computer Graphics, 14* (6), 1333-1339, 2008.
[22] Schmidt, W. C., "Technical considerations when implementing online research". In A. N. Joinson, K. Y. A. McKenna, T. Postmes & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 461-472). Oxford, UK: Oxford University Press (2007).