

ULF-DIETRICH REIPS
TOM BUCHANAN
JOHN KRANTZ
KEN MCGRAWN

METHODOLOGICAL CHALLENGES IN THE USE OF THE INTERNET FOR SCIENTIFIC RESEARCH: TEN SOLUTIONS AND RECOMMENDATIONS

Abstract

Researchers, editors, and other players in the scientific community have shown and continue to show high interest in and great enthusiasm for Internet-based research methods. However, methodological research has indicated that there are a number of challenges that online researchers must consider. In this article, we provide information about solutions to ten common obstacles in Internet-based research. These are: Repeat or multiple submissions, comparisons between modes, reporting demographics of Internet participants, achieving compatibility for samples when Web and lab samples methods are to be compared, persistency of instrument, international nature of the Internet, possible confounds arising from recruitment techniques, reliance on a single Website or recruitment method to attract participants, dropout and other types of nonresponse, and the preconception that Internet samples invariably lead to different results.

INTRODUCTION

A few years ago, one of the authors of this paper gave an invited talk at another university, describing research that involved data collection via the Internet. In the questions that followed, the first point raised - by an eminent professor of psychology - was "But don't people on the Internet just tell lies?". However, while such suspicions are largely unfounded¹ (and are rarely encountered nowadays, given that the Internet has proved to be an effective research medium), there are methodological challenges to Internet-based research. These need to be understood, and controlled for, both by those using the Internet for research, and those reading the scholarly outputs arising from their efforts. With the proliferation of sophisticated online survey and experiment platforms and Web services that can mine millions of communications and traces of Internet users (e.g. <http://tweetminer.eu>, Reips & Garaizar, 2011), it is easier than ever for individuals to set out to collect data online.

In our experience as reviewers, editors, teachers and readers, this is often done in relative ignorance of the pertinent methodological challenges.

Use of the Internet for the collection of data in Psychology began in 1994 (Musch & Reips, 2000). The new methods were first presented at the Chicago meeting of the Society for Computers in Psychology (SCiP, <http://scip.ws>) in 1996 and at the German Online Research conference in Cologne 1997 (Abstracts available online at <http://www.gor.de/gor97/abstracts.htm>). Much of the early work in the field of online psychological research was methodological in nature, and included efforts by dozens of researchers over a period of many years to empirically test the quality of data collected on the Internet. Examples of this research had appeared in *Behavior Research Methods* (formerly *Behavior Research Methods, Instruments, & Computers*) and *Social Science Computer Review*; special issues on Internet-based research of *Experimental Psychology* (2002, Vol. 49) the *Swiss Journal of Psychology* (2003, Vol. 69); abstracts from the General Online Research conferences available online (<http://www.gor.de>); edited books and reviews like Birnbaum (2004).

Early adopters of Internet techniques raised issues like generalizability, motivation, faking and so on as the few possible problem areas for Internet-based research amongst a plethora of advantages. Early researchers certainly were aware of the methodological challenges presented by Internet-based research (e.g. Reips, 2000). Yet, in most cases, these studies both identified possible issues and also proposed technical and methodological solutions. These researchers (enthusiastic advocates of online techniques) went on to empirically test these possibilities and demonstrate that data quality was not adversely affected if proper techniques are used (for reviews see Krantz & Dalal, 2000; Reips, 2002b). This was driven by the clear need to test the validity of the new additions to psychology's empirical toolbox before adopting them wholesale.

More recent work has added to these investigations, and produced consistent results. Internet-based studies are thus to be trusted. However, in light of previous research we would counsel against overly broad conclusions. Internet researchers have long realized that the Internet is not some monolithic "place" but a collection of locations much like the real world. Different types of people, recruited via different techniques and participating for different reasons may behave quite heterogeneously.

For example, most published work based on Internet-mediated data suggests that people respond seriously (e.g. Voracek, Stieger, & Gindl, 2001). However, there may be cases where people are less likely to provide valid data. For example, Furnham and Buchanan (2005) found very high levels of nonserious responding in a study where people were asked to provide estimates of their intelligence. We therefore urge Internet scientists to consider the problem seriously, and where appropriate implement techniques permitting greater confidence in the veracity of responses (Aust, Diedenhofen, Ullrich & Musch, 2012; Reips, 2002a, 2002b, 2007, 2010; Reips & Birnbaum, 2011).

Similarly, researchers need to be aware that sample characteristics interact with the topic being researched. One illustration is that Krantz and Dalal (2000, Table 3, p. 49) found that the proportion of female participants in Internet samples ranged from 26% to 71% depending on the research project. Web samples are largely

opportunity samples; therefore, it is very important to be intentional about sampling methods. Technology that is used in creating Web-based studies may bias samples, because technology preferences are associated with personality and sophisticated technologies may not be used by people with lower levels of education (Buchanan & Reips, 2001).

Differences between Internet and traditional methods should not lead to the automatic rejection of Internet methodologies. In addition to possible limits on the generalizability of the findings of general (non-Internet) methods, there are several methodological advantages of Internet-based research that should be considered (see Reips, 2000, 2007; Reips & Birnbaum, 2011). While more research is required on the ways that research techniques and topics may interact, much is already known about the effective use of Internet-based research techniques. A number of ‘best practice’ methodological recommendations have emerged from prior work, some of which are outlined below. We urge researchers to make use of Internet-based techniques, in a manner informed by such recommendations.

SELECTED METHODOLOGICAL RECOMMENDATIONS FOR PERFORMING INTERNET-BASED RESEARCH AND COMPARING INTERNET SAMPLES WITH TRA- DITIONAL SAMPLES

Below we outline a number of methodological issues addressed by Internet researchers, and indicate how they might be handled. For a fuller description of these and other issues, see Birnbaum (2001, 2004), Buchanan and Smith (1999), Coomber (1997), Hewson, Laurent, & Vogel (1996), Krantz (2001), Reips (1997, 2000, 2002a, 2002b, 2002c), Reips and Birnbaum (2011), and Schmidt (1997, 2007). Below, problems will be in italics and solutions in normal fonts.

Problem and Solution 1: Repeat or multiple submissions.

Even though this concern is often among the first that come to the minds of those who deliberate moving to the Internet for data collection multiple or repeat submissions are an “old” topic in the literature on Internet-based research methodology. Traditional strategy recommends comparing IP addresses. However, this technique is becoming less satisfactory as it misses one important source of potential error: dynamic IP addressing (e.g. Reips, 2002b). Many Internet providers assign the same IP addresses to different users, or different IP addresses to the same user during the course of one session. Solutions to this problem are the use of other information (e.g., type of Web browser, type of operating system) that can be captured in properly configured server log files and databases or using persistent session keys (Reips & Stieger, 2004; Stieger & Reips, 2010).

In general, possible multiple submissions can be avoided or controlled by collecting personal identification items, by checking internal consistency as well as date and time consistency of answers (Schmidt, 1997), and by using techniques such as sub-sampling, participant pools, or handing out passwords (Reips, 2000, 2002a). There is evidence that multiple submissions are rare in Web experiments (Reips, 1997). (Also see Table 3: Avoidance and control of multiple submissions in Reips, 2002b).

Modern online platforms for surveys (e.g. SurveyMonkey, SurveyGizmo, Qualtrics) or experiments (e.g. WEXTOR.org) often provide proprietary solutions for identifying and eliminating multiple responses. We would advise researchers using such platforms to make themselves aware of how these procedures are implemented, not least so they can be properly reported.

Problem and Solution 2: be wary of inappropriate comparisons, do not compare apples with oranges. For example, it makes no sense to compare race composition of an Internet sample with race composition of the general U.S. population.

In cross-mode comparisons take care to consider mode-specific selection effects. For example, on the Internet technology preferences may determine coverage and sample composition (Buchanan & Reips, 2001). In demographics questions asked to participants, include diagnostic questions that allow you to develop a basis for comparisons, e.g. request the nation of residence of the participant to determine whether it is possible to generalize to a population residing in the same country. But see problem 3 below.

Problem and Solution 3: Reporting demographic data from Internet participants.

When reporting demographic characteristics of a sample remember that these self-report data are not independently verifiable. Therefore, one should write “indicated/reported themselves to be X years old” rather than “were X years old”. It is also important to report distributions, not only means (as for example age distributions may not be normal — bimodal distributions (younger people and the senior “silver surfer” contingent) may often be encountered).

Problem and Solution 4: Achieving compatibility for samples when Web and lab samples methods are to be compared. For example, comparing Internet respondents with University subject pools on measures of neuroticism and introversion. Failure to find a difference tells us nothing unless we are sure that the samples compared really do not differ on the constructs of interest. This is important in light of evidence that a systematic measurement effect may produce elevated scores on Internet-based measures of some psychological characteristics (e.g. indices of negative affect and depression, Buchanan and Joinson, 2004).

Randomly assign participants to Web versus lab condition when performing such comparisons, or use fully matched samples. Also report full data set for all samples.

Problem and Solution 5: Persistency of instrument.

Work with pre-tests. If instruments are changed during a study, the study should be considered to have been completed and any further data acquired with the modified instrument be treated as a separate sample, tested under slightly different conditions. Some of the better modern online survey platforms will automatically warn researchers about this if they make changes to a ‘live’ survey.

Problem and Solution 6: International nature of Internet.

Ensure that measures used are suited to the (probably international) eventual sample. For example, if you are potentially surveying people from outside the USA, do not just list American states as possible locations. Also be aware of cultural and political issues that may arise. For example, one of the authors has been criticized by participants from Taiwan for including it as part of China in a list of participant locations. It is good practice to pretest questions with people from the target audience.

Problem and Solution 7: Be aware of possible confounds arising from recruitment techniques (e.g. in comparison of different tests linked on the same Web page).

The list of linked tests/sites to be compared should be randomly ordered for each access by a new participant. Keep different language versions compatible. Pre-test comparable readability.

Problem and Solution 8: Reliance on a single Website or recruitment method to attract participants. This exacerbates problems of self-selection and potential systematic sampling biases.

Effects of self-selection can be controlled via the *multiple site entry technique* (Reips, 2000, 2002b): Do not recruit from a single Web site or other source. For example, Buchanan and Whitty (in press) recruited participants (victims of a particular crime) from two different sources that were believed to be subject to *different* sampling biases. Replicating findings across the two samples gave added confidence in the findings.

Problem and Solution 9: Drop-out and other types of non-response.

Dropout can be turned into a detection device for motivational confounding. Also, dropout can be reduced by implementing a number of measures, such as the high hurdle and warm-up techniques (Reips, 2000, 2002a), promising immediate feedback, giving financial incentives, and by personalization (Frick, Bächtiger, & Reips, 2001).

Problem and Solution 10: Internet samples invariably lead to different results.

If sample characteristics are thought to be important to a result, collect both traditional and Internet samples. However, it may be that even with large sample differences, the results from traditional and Internet methods may be essentially identical (Krantz, Ballard, & Scher, 1997).

CONCLUSION

There may be some antediluvian defenders of orthodoxy in the psychology laboratory here and there. All of us have encountered rare instances like receiving an NSF Review Panel Summary with the following statement: “The Panel included two journal editors both of whom agreed that they would never publish an article with purely Web-based data.” However, our attempts in establishing a methodology for Internet-based data collection were greeted overwhelmingly with interest and support, if not enthusiasm. Several of our earliest papers have been widely cited, e.g. Buchanan (2000), Buchanan and Smith (1999), Krantz (2000), Krantz, Ballard and Scher (1997), Krantz and Dalal (2000), McGraw et al. (2000), Reips (1996, 1997, 2000, 2002a, 2002b). Newer papers on Internet-based research and applications continue to be met with interest (e.g. Birnbaum, 2004; Buchanan, Johnson & Goldberg, 2005; Buffardi & Campbell, 2008; Freelon, 2010, in press; Joinson, Woodley & Reips, 2007; Wright, 2005).

Since the 1990s, many researchers, students and science managers have been positive about the new methodology and eager to learn best practices. Simultaneously, a new generation of researchers has arisen for whom Internet-based research is an established and accepted tool. However, they too need to be aware of the potential

methodological challenges faced in such work. We will continue to provide them with methods, techniques, guidelines and tools for Internet-based research.

REFERENCES

- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2012). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, Online first. doi 10.3758/s13428-012-0265-2
- Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803-832.
- Buchanan, T. (2000). Potential of the Internet for personality research. In M. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp.121-140). San Diego: Academic Press.
- Buchanan, T., Johnson, J. A., & Goldberg, L. (2005). Implementing a Five-Factor personality inventory for use on the Internet. *European Journal of Psychological Assessment*, *21*, 115-127.
- Buchanan, T., & Joinson, A. N. (2004, March). Are online-offline differences in personality test scores due to increased self-disclosure? Paper presented at German Online Research '04, Duisburg, Germany, 30th March 2004.
- Buchanan, T., & Reips, U.-D. (2001). Platform-dependent biases in Online Research: Do Mac users really think different? In K. J. Jonas, P. Breuer, B. Schauenburg & M. Boos (Eds.), *Perspectives on Internet research: Concepts and methods*. Retrieved February 2, 2013 from <http://www.psych.uni-goettingen.de/congress/gor-2001/contrib/buchanan-tom>
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide Web. *British Journal of Psychology*, *90*, 125-144.
- Buchanan, T., & Whitty, M. T. (in press). The online dating romance scam: Causes and consequences of victimhood. *Psychology, Crime and Law*.
- Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and social networking Web sites. *Personality and Social Psychology Bulletin*, *34*, 1303-1314.
- Coomber, R. (1997, June 30). Using the Internet for survey research. *Sociological Research Online*, *2*, Retrieved June 16th, 2002, from <http://www.socresonline.org.uk/2/2/2.html>
- Frick, A., Bächtiger, M. T., & Reips, U.-D. (2001). Financial incentives, personal information, and dropout in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 209-219). Lengerich, Germany: Pabst.
- Freelon, D. G. (2010). ReCal: intercoder reliability calculation as a Web service. *International Journal of Internet Science*, *5*(1), 20-33.
- Freelon, D. G. (in press). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a Web service. *International Journal of Internet Science*.
- Furnham, A., & Buchanan, T. (2005). Personality, gender and self-perceived intelligence. *Personality and Individual Differences*, *39* (3), 543-555.

- Hewson, M., Laurent, D., & Vogel, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the Internet. *Behavior Research Methods, Instruments, and Computers*, 28, 186-191.
- Joinson, A. N., Woodley, A., & Reips, U.-D. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior*, 23(1), 275-285. doi:10.1016/j.chb.2004.10.012
- Krantz, J. H. (2001). Stimulus delivery on the Web: What can be presented when calibration isn't possible? In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 113-130). Lengerich, Germany: Pabst.
- Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, 29, 264-269.
- Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego, CA: Academic Press.
- McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, 11(6), 502-506. Blackwell Publishing. Retrieved from <http://pss.sagepub.com/lookup/doi/10.1111/1467-9280.00296>
- Reips, U.-D. (1996). Experimenting in the World Wide Web. Paper presented at the 1996 Society for Computers in Psychology Conference. Chicago, IL.
- 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, Germany: Hogrefe.
- Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). New York: Academic Press.
- Reips, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review*, 20 (3), 241-249.
- Reips, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology*, 49 (4), 243-256.
- Reips, U.-D. (2002c). Theory and techniques of conducting Web experiments. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online Social Sciences* (pp. 229-250). Seattle: Hogrefe & Huber.
- 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: Oxford University Press.
- Reips, U.-D. (2010). Design and formatting in Internet-based research. In S. Gosling & J. Johnson, *Advanced methods for conducting online behavioral research* (pp. 29-43). Washington, DC: American Psychological Association.
- Reips, U.-D., & Birnbaum, M. H. (2011). Behavioral research and data collection via the Internet. In K.-P. L. Vu and R. W. Proctor (Eds.), *The handbook of human factors in Web design* (2nd ed., pp. 563-585). Mahwah, New Jersey: Erlbaum.

- Reips, U.-D., & Garaizar, P. (2011). Mining Twitter: Microblogging as a source for psychological wisdom of the crowds. *Behavior Research Methods*, *43*, 635-642. doi:10.3758/s13428-011-0116-6
- Reips, U.-D., & Stieger, S. (2004). Scientific LogAnalyzer: A Web-based tool for analyses of server log files in psychological research. *Behavior Research Methods, Instruments, & Computers*, *36*, 304-311.
- Schmidt, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, and Computers*, *29*, 274-279.
- Schmidt, W. C. (2007). Technical considerations when implementing online research. In Joinson, A., McKenna, K., Postmes, T. and Reips, U. (Eds.), *The Oxford handbook of Internet psychology* (pp. 461-472). Oxford. Oxford University Press.
- Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, *26*, 1488-1495.
- 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, Germany: Pabst.
- Whitty, M. T., Buchanan, T., Joinson, A. N., & Meredith, A. (2012). Not all lies are spontaneous: An examination of deception across different modes of communication. *Journal of the American Society for Information Science and Technology*, *63*(1), 208-216. doi:10.1002/asi.21648
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, *10*(3).

Authors' note

All correspondence concerning this comment should be addressed to Ulf-Dietrich Reips, iScience group, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain. E-mail: u.reips@ikerbasque.org .

Footnotes

- ¹ In fact, it is not just people on the Internet who tell lies. Though many lies are undoubtedly told, recent findings (Whitty, Buchanan, Joinson & Meredith, 2012) in fact suggest that people lie more in other media (most notably by phone).