Conceptual, Methodological, and Ethical Challenges of Internet–Based Data Collection

Jonathan K. Lee
Suffolk University, USA

Peter M. Vernig
Suffolk University, USA

INTRODUCTION

The growth in multimedia technology has revolutionized the way people interact with computer systems. From personal software to business systems and pedagogic applications, multimedia technology is opening up new pathways to increase the efficiency of existing systems. However, the utilization and implementation of new technologies has been occurring at such a rapid pace that theory and research has been unable to keep up. This is particularly evident with data collection methods in the social sciences.

With the growth in use of the Internet passing the 1 billion user mark in 2006 (Internet Usage Statistics, 2006), social scientists are turning to the Internet for data collection purposes in increasing numbers (few, if any, advances have revolutionized data collection more than the use of the Internet). Often referred to as *Internet-based research*, *Web-based research, and cyberresearch,* this mode of data collection refers to the administration of questionnaires and acquisition of response data in an automated manner via the World Wide Web.

Collecting participant responses was a task that once required hours of direct interaction, created problems in scheduling, and limited the diversity of the population being studied. Now data collection can be automated and conducted at any time with increased efficiency (MacWhinney, 2000) and at reduced costs (Cobanoglu, Warde, & Moreo, 2001). Moreover, depending on the nature of the research, the questionnaire can be delivered to Internet users around the world; additionally, research can be targeted to specific populations that have traditionally been underrepresented in the research literature (Im & Chee, 2004; Mathy, Schillace, Coleman, & Berquist, 2002). However, the advantages of using the Internet as a medium for data collection are not without their shortcomings. Theory has not been able to keep up with the proliferation of Internet-based research projects. Further, creative methodologies to facilitate data collection are being proposed in the empirical literature with increased frequency, raising questions about the construct validity of associated studies. The creation of new research methodologies has led to the need for new ethical guidelines for the protection of Internet research subjects, and thus are posing new challenges for research review boards at many institutions (Flanagan, 1999).

Academic disciplines that currently use the Internet as a vehicle for data collection are multifarious, and a discussion of the major issues relevant to each is beyond the scope of this article. Thus, for the sake of brevity, and because it is the discipline of which our knowledge is the most up to date, we will limit our discussion to the specific discipline of psychology, as its wide range of methodologies allow for examples of different challenges relevant to other areas of research. It is important for the reader to note that these issues are similar across the many fields that have implemented Internet-based data collection models for empirical research. This article will provide a summary of the conceptual, methodological, and ethical challenges for the researcher considering the Internet as a tool for data collection and will make suggestions for the ethical and effective implementation of such studies. Due to its brief length, the depth of our discussion is neither expansive nor comprehensive; rather we will focus on what we feel are the key issues for the nascent cyberresearcher. Readers interested in a comprehensive review of this topic are directed to Birnbaum (2000).

CONCEPTUAL CHALLENGES

Of central concern in psychological research is the generalizability of findings—the results of a given study

must hold true in the real world, free from experimental controls. This issue is lessened by Internet-based data collection methods, as it has the potential to reach larger and more diverse groups (Birnbaum, 2004), facilitating multiethnic sampling (i.e., the inclusion of a diverse range of ethnic and cultural populations). However, some theorists (e.g., Brehm, 1993; Civille, 1995; Kester, 1998) argue that questionnaire data collected via the Internet is misrepresentative of the general population. Gonzales (2002) provides a convincing argument against the generalizability of data collected through the Internet, providing statistics on Internet usage among different subsets of the general population. In his review, he highlights that Internet users are (a) predominantly white males, (b) in the middle to upper class in terms of income, (c) under the age of 34, and (d) living in metropolitan areas. According to Gonzales (2002), individuals who complete questionnaires on the Internet are only representative of a small portion of the population.

More recently, Gosling, Vazire, Srivastava, and John (2004) compared data collected through the Internet with traditional methods (i.e., in-person, paper-and-pencil) to test the reliability of Internet data. In this study, data were collected via two questionnaires from two Web sites, and were compared with empirical studies that were published over the course of a year in a peer reviewed journal. While the authors concluded that the representativeness of samples is problematic, it is not an issue that is solely limited to Internet-based data collection (indeed all research studies must address this concern to some degree). Furthermore, the prodigious sample size that a researcher can collect through Internet-based methods is clearly a strength of this methodology. For example, Gonzales (2002) reported that 5% of Asians use the Internet, a substantially lower percentage compared with the 87% of White people also reported; although these percentages represent racial disparities in Internet usage, the precise numbers are hidden behind them. Because we cannot speak to the actual numbers of Asian and White participants reported in Gonzales (2002), we will refer to Gosling et al. (2004). Gosling and colleagues report usage statistics of 7.6% Asian and 76.5% White in their study (a ratio slightly higher than in Gonzales). However, the actual numbers of Asian and White participants in Gosling et al. (2004) are very impressive: 26,048 Asian and 276,672 White. Thus, it is important to note that even

fractions of percentages can reflect very large numbers in Internet-based samples.

Traditional research in psychology has an oft-criticized tradition of utilizing undergraduate college students in research. In fact, 85% of studies reported in Gosling et al. (2004) are comprised of college samples. This poses serious limitations when trying to make assumptions about the general population because college students tend to be younger, better educated, and from a higher socioeconomic group than the population at large. Because the Internet opens up data recruitment to include anyone in the world with access to the Web, data collection through this means has the potential to overcome these barriers, as demonstrated in several studies (e.g. Gosling et al., 2004; Nosek, Banaji, & Greenwald, 2002). For example, Nosek et al. (2002) had over 2.5 million people participate in their study from around the world. Internet-based data collection can also target marginalized groups who are notably absent from the current literature, and ask for sensitive information that a participant may not accurately disclose face-to-face (Gullette & Turner, 2003; Koch & Emrey, 2001; Mustanski, 2001; Whittier, Seeley, & St. Lawrence, 2004). Thus, we believe that the criticism of Internet-based data collection methods on the grounds that the findings are not representative (although technically valid) does not account for the extensive number of participants, and the potential exists to address many of the shortcomings mentioned by Gonzales (2002). That a degree of sampling bias exists in research conducted online is merely reflective of psychological research as a whole, in which true random sampling is never fully achieved.

A common preconception of Internet-based data collection methods is that the findings are not consistent with traditional methods (Cook, Heath, & Thompson, 2000; Couper, 2000). Indeed there are numerous anecdotes about journal reviewers who criticize the validity of Internet-based studies on this assumption. In contrast to this belief, research has demonstrated consistency between Internet-based data collection methods and paper-and-pencil measures (Gosling et al., 2004), traditional phone surveys (Best, Krueger, Hubbard, & Smith, 2001), and mail-in surveys (Ballard & Prine, 2002). In one study that tested demographic effects of individuals participating in Internet surveys and paper-and-pencil methods, Srivastava , John, Gosling, and Potter (2003) matched participants whose questionnaire data were

Related Content

An Improved Arabic Handwritten Recognition System using Deep Support Vector Machines

Mohamed Elleuchand Monji Kherallah (2016). *International Journal of Multimedia Data Engineering and Management (pp. 1-20).*

www.irma-international.org/article/an-improved-arabic-handwritten-recognition-system-using-deep-support-vector-machines/152865

Correlation-Based Ranking for Large-Scale Video Concept Retrieval

Lin Linand Mei-Ling Shyu (2010). *International Journal of Multimedia Data Engineering and Management (pp. 60-74).*

www.irma-international.org/article/correlation-based-ranking-large-scale/49150

Security Issues on Outlier Detection and Countermeasure for Distributed Hierarchical Wireless Sensor Networks

Yiying Zhang, Lin He, Lei Shu, Takahiro Haraand Shojiro Nishio (2012). *Advancements in Distributed Computing and Internet Technologies: Trends and Issues (pp. 182-210).*

www.irma-international.org/chapter/security-issues-outlier-detection-countermeasure/59683

A Novel Approach for Colorization of a Grayscale Image using Soft Computing Techniques

Abul Hasnat, Santanu Halder, Debotosh Bhattacharjeeand Mita Nasipuri (2017). *International Journal of Multimedia Data Engineering and Management (pp. 19-43).*

www.irma-international.org/article/a-novel-approach-for-colorization-of-a-grayscale-image-using-soft-computing-techniques/187138

Constructing and Utilizing Video Ontology for Accurate and Fast Retrieval

Kimiaki Shirahamaand Kuniaki Uehara (2011). *International Journal of Multimedia Data Engineering and Management (pp. 59-75).*

www.irma-international.org/article/constructing-utilizing-video-ontology-accurate/61312