

Exploring technical aspects of developing Internet-based instruments in quantitative research

Dr John Clayton

*Emerging Technologies Centre
Waikato Institute of Technology*

John.Clayton@wintec.ac.nz

Abstract

In the administration of quantitative research instruments it has been customary to use 'pencil and paper' forms, with the researcher, distributing forms, collating the responses and supervising data entry to a structured database. The advancements in connected-computer technologies and statistical software applications offer quantitative researchers a more streamlined method for the collection, storage, manipulation and dissemination of data. These advancements have made it possible to dispose of, in appropriate studies, 'paper-and-pencil' instruments and manual data entry. However, the use of the internet, as with any method of collecting data, does have limitations and many of these are technical. This paper explores the technical issues associated with the development of connected-computer quantitative instruments.

Background

In using 'pencil and paper' quantitative research instruments participants are asked to select from a range of options an appropriate response. The administrator then collates the responses and supervises the entry of data into a structured database (Fisher & Fraser, 1990; Fraser, Giddings, & McRobbie, 1992; Fraser & Walberg, 1995). To aid the process of data entry, instruments are carefully designed with participants responses to specific items aligned to appropriate database rows and tables. The process of 'pencil and paper' instrument administration involves four steps, production, distribution, collection and entry. These four steps are illustrated in Figure 1 below

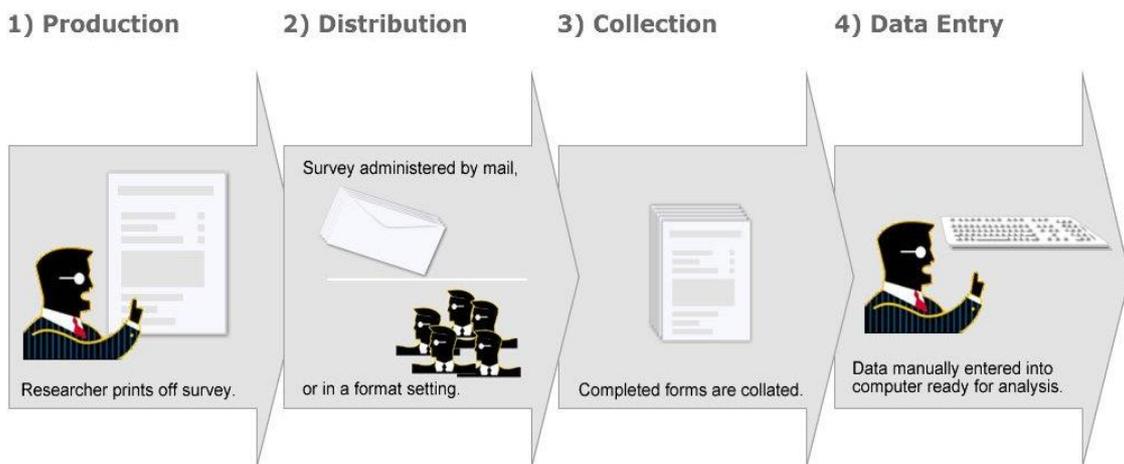


Figure 1. Paper-based survey procedures.

Overview

The advent of the microcomputer in the 1980s, the creation of the Internet, the development of the World Wide Web and Web-page software creation applications have influenced all aspects of modern society including research (Clayton, 2006; Looms, 2002; Reid, 1994; Sangster, 1995). The first attempts at collecting data using networked computers began with relatively clumsy asynchronous email surveys in the mid nineteen-eighties (Andrews, Nonnecke, & Preece, 2003). The establishment of client-server interactions facilitated through the development of web-forms and the development of database technologies led, in the last decade of the twentieth century, to the increasing deployment of dynamic Internet-based surveys (Clayton, 2004; Musch & Reips, 2000; Reips, 2002b). As connected computing technologies have matured and web-page authoring applications for the creation of interactive web forms have been deployed, the use of Internet-based surveying has increased and is now widely used in social science research and educational institutions (Buchanan, Johnson, & Goldberg, 2005; Kraut et al., 2004; Reips & Neuhaus, 2002; Shannon, Johnson, Searcy, & Lott, 2002; Upcraft & Wortman, 2000; Yun & Craig, 2000). The advancements in connected-computer technologies and statistical software applications offer researchers a simpler, more streamlined method for the collection, storage and manipulation of data and have made it possible to dispose of, in appropriate studies, 'paper-and-pencil' instruments and manual data entry (Clayton, 2003, 2005).

It has been argued the Internet offers significant advantages over more traditional survey techniques in the collection, storage and collation of data. For example,

- The use of connected computing technologies offers researchers a simpler, more streamlined method for the collection and protection of data (Buchanan, Ali et al., 2005; Mertler, 2003).
- The often tedious data entry stage is eliminated and there is a greater assurance data acquired is free from common entry errors (Andrews et al., 2003; Schmidt, 1997; Solomon, 2001).
- Since there is no separate data entry phase, tabled results can be available for analysis soon after the data collection phase, speeding up the research process (de Leeuw & Nicholls II, 1996; DePaolo & Sherwood, 2006).
- The costs in terms of both time and money for publishing a survey on the web are low, compared with costs associated with conventional surveying methods. For example, costs of data entry, paper, ink and printing are eliminated, and as a result, research can be much less expensive (Baron & Siepmann, 2000; Benfield & Szlemko, 2006).

Using Web-based Surveys

Although some researchers are cautious of using the internet to gather data (Benfield & Szlemko, 2006), and others have found some instances where the results from internet psychology experiments and tests appear to differ slightly from paper-and-pencil psychological tests (Buchanan, Ali et al., 2005; Buchanan, Johnson et al., 2005), there appears to be general agreement that there is no significant difference in the pattern of responses received from Internet surveys and traditionally-administered pencil and paper forms (Birnbaum, 2000) or postal conducted surveys (Andrews et al., 2003). For example,

- (Carini, Hayek, Kuh, & Ouimet, 2001), in analyzing data collected from first year and senior college students in both web-based and conventional methods, found that when discrepancies existed they tended to be very small.
- (Whelchel & Schechter, 2001), in using both traditional and internet methods in surveying graduating students found, although there were slight differences in gender responses, there was no overall difference in Internet and paper responses.

- (Baron & Siepmann, 2000) found, in a study investigating the determinants of the desire to reduce risks, there was no significant difference in 31 items (from a survey containing 32 items) between 42 subjects completing a paper questionnaire and 49 subjects completing a web questionnaire.
- (Baron & Siepmann, 2000) concluded the very small effects observed for most items in their study should help allay concerns that data gathered via the Web may be very different than that collected from paper.

The process of 'web-form' instrument administration is illustrated in Figure 2 below

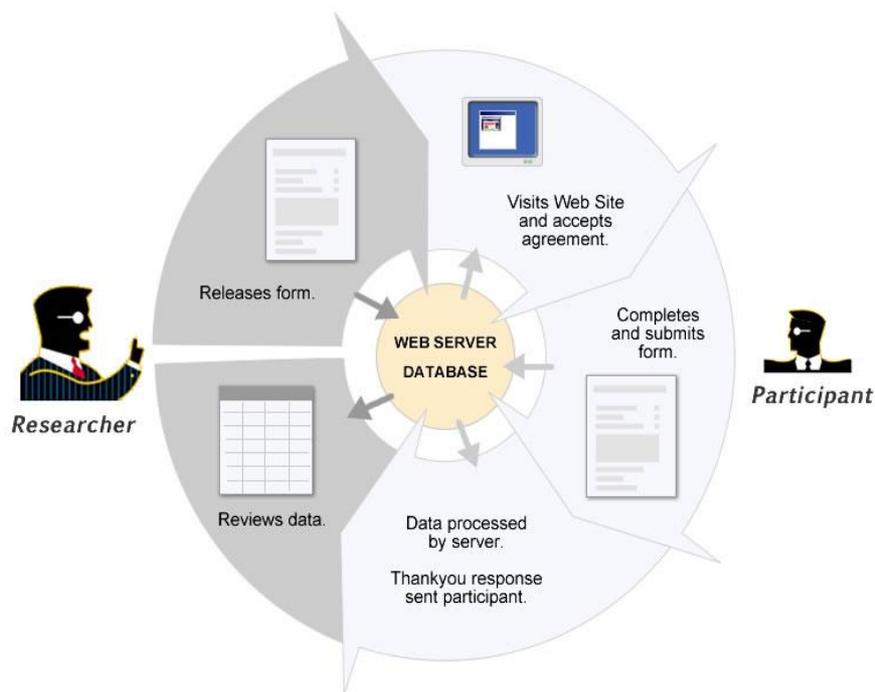


Figure 2. Web-based survey procedures.

Database Procedures

When using dynamic web-forms to generate data to be stored in a digital database, a clear set of processes and procedures needs to be established to ensure data is gathered, stored, retrieved and manipulated in a consistently reliable and robust manner (Clayton, 2003, 2004; Solomon, 2001). Before outlining the procedures to be followed, it would be profitable, at this early stage, to distinguish between the terms 'data' and 'information'.

Data can be regarded as a collection of raw electronic facts stored in isolation. In this raw state data has generally little meaning, for example, the response to a specific item is 5. On the other hand, information can be seen to be a collection of raw pieces of data that have been selected and manipulated to convey meaning; in short, it is useful (Whitten, Bently, & Barlow, 1994). An example could be the mean of 4 was generated from 25 responses to an item. If we accept these definitions are useful it becomes important, when creating an interactive database, to ensure data generated by the participants' completion of a web-based form is stored in a specific place within the database and that this data is secured and able to be easily manipulated to produce information in the form of reports useful to the researcher.

It would appear a relational database, accessible only to the researcher, would meet the requirements outlined above. A relational database stores all its data inside tables. All operations - , (searching, manipulating, analyzing), - on data are done on the tables themselves. The results of these operations are the creation of other tables, which can then be utilized to report on the data collected (Lozano, 1999). If a relationship database is created there is a need to identify the types of tables created and how these tables are internally structured. To best illustrate the procedures to be used, a brief three-step scenario is outlined in Table 1 below

Table 1
Establishing database procedures

Participant Action	Database Procedure
Individual participant accesses a web-form	A numerical identifier, a key, is randomly produced for each participant.
Participant views items.	A separate table row is created for each item and each table-row is uniquely identified.
Participant completes form.	Participant responses for each item are recorded.

Technical Issues

However, the use of the internet, as with any method of collecting data, does have limitations. These limitations have been compounded by the fact the rapid increase in use and deployment of internet-based surveys has been led to some extent by computer programmers, providing technical solutions, rather than by survey methodologists. Many of the solutions are technical and lack, it could be argued, a theoretical foundation (Dillman & Bowker, 2001; Shannon et al., 2002).

As mentioned above the delivery of, and the collection of data from, web-based forms are not unproblematic. Technical problems such as power failure, internet connection loss, internet connection speed can frustrate researchers and participants alike (Benfield & Szlemko, 2006). These technical issues can be created by the researchers themselves in the development of the forms. For example, a number of research experiments and activities conducted on the internet are developed using scripting languages to create interactive web pages. While this enhances the visual presentation of the form, allowing the researcher more control over layout and structure, the use of these scripts becomes problematic when they rely on the type and functionality of the client's web browser (Reips, 2002a). When there is incompatibility there can be computer crashes, network error messages and slow performance, increasing the likelihood of non-completion and drop-out, and in some cases, potentially driving students to disable web-browser functionality to reduce frustration (Buchanan & Reips, 2001). This defeats the purpose of using the scripts in the first instance. These problems can be addressed by using server-side applications like *Common Gateway Interface* (CGI) reducing the dependence upon the web browser being used by the participant and minimizing non-completion and drop-out (Schwarz & Reips, 2001).

The design of web-forms for surveys is a complex task and has led to the development of some survey forms being dictated by computer and programming specialists rather than survey professionals thus potentially reducing the normal academic rigor used in the development of questionnaires (Dillman & Bowker, 2001). To some extent these problems are being addressed with the development of software applications with simple user interfaces which generate forms requiring limited knowledge of technical details (Benfield & Szlemko, 2006; Reips & Neuhaus, 2002; Wright, 2005). These have been complemented by the development of internet-managed survey sites where users can design their survey online and automatically generate database-web form relationships (Macro International, n.d; Questionmark, 2006; SurveyMonkey, 2006). However the use of hosted solutions is not without limitations. Although the protection of data from external unauthorised access has been partially

addressed (Reips, 2002a, 2002b), ethical concerns remain regarding the researchers' lack of assurance regarding security and access to collected data on the host site (Brem, 2002; Wright, 2005). In short, when using external hosts for Internet based research, it should not be assumed participants have the same protections as afforded by traditional non-internet-based research methods.

Often, in collating real-time data generated by Web-based instruments some advanced technical skills are required. For example if the pages are created as active server pages (asp) the technical skills required to ensure the pages function appropriately in real-time is often beyond the basic skills of most HTML developers. Any modifications to the actual number of items presented to participants in the dynamic web-page using asp means the corresponding tables and columns in the database created have to be re-configured manually and this is inefficient and time consuming. Furthermore, the server used to collect and redistribute the data must support asp scripting language and any changes to the measure requires a basic knowledge of this asp scripting language (Clayton, 2007).

Addressing the Technical Issues

In the not to distant past the creation of dynamic web-pages to gather data from participants required developers to acquire sophisticated HTML programming and/or JavaScript language skills (Dillman & Bowker, 2001). However, the ongoing development of user friendly "what you see is what you get (WYSIWYG)" software applications, such as Dreamweaver® and Front Page®, has made the task of creating dynamic web-pages for internet-based research a relatively simple task (Reips & Neuhaus, 2002; Wright, 2005). Web-survey developers using these WYSIWYG software applications can create sophisticated and complex forms with the acquisition of basic word processing skills.

By creating instruments as normal HTML pages the developer can then use the functionality of *web-forms* to collect the data. In essence, the use of web-forms means data can be collected in a comma separated values text file (csv). These csv files can then be directly exported to, or imported by, a number of software and statistical software application packages such as SPSS® or SYSTAT®. Since the data-collected is handled by a form-processor located on the server, the comma separated value nature of the data-collected automatically generates the table and column structure of the database to be used. This eliminates the need for continual manual configuration of the database if the number of items used in an instrument is altered. Individual developers need only to be aware of the URL of the form-processor on the server and do not need in-depth knowledge of specialised scripting languages. In essence, the use of web-forms and associated form-processors can reduce the technological literacy costs of individual developers, ensure modifications to the database structure are always be linked by changes to the HTML web-form and a range of form-processors located in different server-environments can be used in the deployment of quantitative instruments.

Summary

The use of internet technologies in all forms of research is continually expanding. Research into the benefits and limitations of using web-based forms, while still in its infancy, is growing in robustness. It is perceived the costs, in terms of both time and money, for publishing a survey on the web are low compared with costs associated with conventional surveying methods. What is increasingly apparent is the advent of electronic databases and the interconnectivity of dynamic web-forms with these databases, makes the collection, storage and manipulation of data generated by server-based quantitative instruments, an attractive and cost effective option.

However, care must be taken to ensure survey developers have the technological skills and knowledge both to develop the research instruments and also to ensure appropriate procedures are established to prevent the results of the survey from being compromised by poor data management.

References

- Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve internet users. *International Journal of Human-Computer Interaction*, 16(2), 185-210.
- Baron, J., & Siepmann, M. (2000). Using web questionnaires for judgment and decision making research. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 235-267). New York: Academic Press.
- Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities [Electronic Version]. *Journal of Research Practice*, 2(2), Article D1. Retrieved January 31, 2007, from <http://jrp.icaap.org/index.php/jrp/article/view/30/51>.
- Birnbaum, M. H. (2000). SurveyWiz and FactorWiz: JavaScript web pages that make HTML forms for research on the internet. *Behavior Research Methods, Instruments, & Computers*, 32(2), 339-346.
- Brem, S. (2002). Analyzing online discussions: Ethics, data, and interpretation [Electronic Version]. *Practical Assessment, Research & Evaluation*, 8(3). Retrieved December 13, 2006, from <http://PAREonline.net/getvn.asp?v=8&n=3>.
- Buchanan, T., Ali, T., Hefferman, T. M., Ling, J., Parrott, A. C., Rodgers, J., et al. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, 37(1), 148-154.
- Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, 21(2), 115-127.
- 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.), *German online research 2001: Perspectives on Internet Research: Concepts and Methods* (pp. [WWW document] http://gor.de/gor01/index_e.htm). Gottingen: Universitat Gottingen.
- Carini, R., Hayek, J., Kuh, G., & Ouimet, J. (2001, June). *College student responses to web and paper surveys: Does mode matter?* Paper presented at the meeting of the Association for Institutional Research, Long Beach, CA.
- Clayton, J. (2003). Assessing and researching the online learning environment. In M. S. Khine & D. Fisher (Eds.), *Technology-rich learning environments: A future perspective*. (pp. 127-137). Singapore: World Scientific.
- Clayton, J. (2004, September). *Using the internet to collect quantitative data*. Paper presented at the ITPNZ Research Conference: Research that's working, Waikato Institute of Technology: Hamilton.
- Clayton, J. (2005, June 27 - July 2). *The validation and application of an online learning environment instrument (OLLES)*. Paper presented at the ED-MEDIA 2005 World Conference on Educational Multi-media, Hypermedia and Telecommunications, Montreal, Canada.
- Clayton, J. (2006). Education and the Internet. In C. Ghaoui (Ed.), *Encyclopaedia of human computer interaction*. (pp. 175-178). Hershey: Idea Group.
- de Leeuw, E., & Nicholls II, W. (1996). Technological Innovations in data collection: Acceptance, data quality and costs [Electronic Version]. *Sociological Research Online*, 1 (4). Retrieved December 20, 2006, from <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>.
- DePaolo, C. A., & Sherwood, A. L. (2006). Instructional uses of web-based survey software [Electronic Version]. *Journal of Educators Online*, 3(1). Retrieved December 16, 2006, from <http://www.thejeo.com/Archives/Volume3Number1/DePaoloFinal.pdf>.
- Dillman, D. A., & Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. In U. Reips & M. Bosnjak (Eds.), *Dimensions of internet science*. Lengerich: Pabst Science Publishers.
- Fisher, D., & Fraser, B. (1990). *School climate (SET research information for teachers No. 2)*. Melbourne: Australian Council for Educational Research. Retrieved, from
- Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1992). *Assessing the climate of science laboratory classes (What research says, No. 8)*. Perth: Curtin University of Technology. Retrieved, from
- Fraser, B. J., & Walberg, H. (Eds.). (1995). *Improving science education*. Chicago: The University of Chicago Press.

- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Cooper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the conduct of research on the Internet. *American Psychologist*, 59(2), 105-117.
- Looms, P. (2002). Sailing into uncharted waters-The impact of new media use on education. In A. Williamson, C. Gunn, A. Young & T. Clear (Eds.), *Winds of change in the sea of learning: Proceedings of the 19th Annual Conference of Australasian Society for Computers in Learning in Tertiary Education*. (Vol. 1, pp. 5-17). Auckland, New Zealand. 8-11 December 2002: ASCILITE.
- Lozano, F. (1999). *Introduction to relational database design* Retrieved March 15, 2003, from, EDM/2: The electronic developer magazine: <http://www.edm2.com/0612/msql7.html>
- Macro International. (n.d). ORC Macro. Retrieved December 14, 2006, from <http://www.orcmacro.com>
- Mertler, C. (2003). Patterns of response and non-response from teachers to traditional and web surveys. [Electronic Version]. *Practical Assessment, Research & Evaluation*, 8(22). Retrieved December 16, 2006, from <http://PAREonline.net/getvn.asp?v=8&n=22>.
- 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: Academic Press.
- Questionmark. (2006, n.d). Questionmark: Getting results. Retrieved December 14, 2006, from <http://www.questionmark.com>
- Reid, T. A. (1994, December). *Perspectives on computers in education: the promise, the pain, the prospect*. Retrieved March 15, 2003, from Active Learning <http://www.ilt.ac.uk/public/cti/ActiveLearning/index.html>
- 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., & Neuhaus, C. (2002). WEXTOR: A web-based tool for generating and visualizing experimental designs and procedures. *Behavior Research Methods, Instruments, & Computers*, 34(2), 234-240.
- Sangster, A. (1995, July). *World Wide Web - what can it do for education?* Retrieved March 15, 2003, from Active Learning <http://www.ilt.ac.uk/public/cti/ActiveLearning/index.html>
- Schmidt, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, 29(2), 274-279.
- Schwarz, S., & Reips, U.-D. (2001). CGI versus JavaScript: A Web experiment on the reversed hindsight bias. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 75-90). Lengerich: Pabst.
- Shannon, D., Johnson, T., Searcy, S., & Lott, A. (2002). Using electronic surveys: advice from survey professionals [Electronic Version]. *Practical Assessment, Research & Evaluation*, 8(1). Retrieved March 15, 2006, from <http://ericae.net/pare/getvn.asp?v=8&n=1>.
- Solomon, D. (2001). Conducting web-based surveys [Electronic Version]. *Practical Assessment, Research & Evaluation*, 7(19). Retrieved August 8, 2006, from <http://PAREonline.net/getvn.asp?v=7&n=19>.
- SurveyMonkey. (2006). SurveyMonkey.Com. Retrieved December 14, 2006, from <http://www.surveymonkey.com>
- Upcraft, M., & Wortman, T. (2000). Web-based data collection and assessment in student affairs. [Electronic Version]. *Student Affairs On-Line*, 1(3). Retrieved December 16, 2006, from <http://www.studentaffairs.com/ejournal/archive.html>.
- Whelchel, N., & Schechter, E. (2001, May). *Paper or PC? Design, Implementation, and success of a mixed-mode survey*. Paper presented at the Making Connections - Public Opinion Research Professionals and the Public: American Association for Public Opinion Research, Montreal.
- Whitten, J., Bently, L., & Barlow, V. (Eds.). (1994). *Systems analysis and design methods* (3rd ed.). Boston: Irwin.
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services [Electronic Version]. *Journal of Computer-Mediated Communication*, 10(3). Retrieved December 20, 2006, from <http://jcmc.indiana.edu/vol10/issue3/wright.html>.
- Yun, G., & Craig, T. (2000). Comparative response to a survey executed by post, E-mail, & Web form [Electronic Version]. *Journal of Computer Mediated Communication*, 6(1). Retrieved August 12, 2005, from <http://www.ascusc.org/jcmc/vol6/issue1/yun.html>.