Paper 24-25

A Web-Based Survey System for Distributed Learning Impact Evaluation With the SAS® System

Morgan C. Wang, Department of Statistics, University of Central Florida, Orlando, FL
Charles D. Dziuban, Department of Education Foundations, University of Central Florida, Orlando, FL
Joel L. Hartman, Vice Provost, University of Central Florida, Orlando, FL

ABSTRACT

The authors describe a web-based survey system that was developed for evaluating distributed learning at the University of Central Florida. The system structures itself around three components that are connected through SAS® data sets: data collection, data validation, and analysis of survey results. This system differs from other approaches in that it incorporates a participation SAS data set – thereby restricting duplicate and unwanted responses. Before submission, participants are able to validate and modify their responses. A predetermined analysis design generates an immediate set of reports that can be modified according to user specifications. Results of the pilot test (summer 1999) are compared to those of a traditional mailed survey. SAS® products used in this project includes base SAS, SAS/GRAPH, SAS/IntrNet, SAS/STAT, and webEIS software.

INTRODUCTION

The University of Central Florida (UCF) has been developing its web-based distributed learning program since the summer of 1996. An integral part of the program is a comprehensive impact evaluation that indexes cognitive, affective, and behavioral components in the student population. The design calls for a regular administration of self-report mailed questionnaires that are distributed to all students who participate in fully on-line courses. Although the system has been successful (35% to 45% response rates are common) the investigators are attempting to design a web-based survey system with data warehouse storage of results and real-time online display of analyses.

Rationale

Web-based surveys offer several possible advantages over the commonly used mail questionnaire technique. First, web-based surveys can be more cost effective because the investigator is freed from the usual costs associated with duplicating and distributing printed survey instruments. Aside from start-up configuration of the system, the only real cost for web-based surveys is designing and publishing the survey on the web. If the questionnaire remains consistent or undergoes minor modi-fication, the preparation costs are virtually eliminated. Second, because the data are in the electric format web-based surveys eliminate entry errors completely. Third, web-based surveys make the real time analysis and reporting process a by-product of data collection.

On the other hand, web-based survey systems pose several potential problems. First, as with mail questionnaires, web-based surveys usually have no control over who completes the instrument. Thus, an effective online system should have a mechanism that identifies the respondent. Second, the possibility exists for receiving a large number of unexpected responses because the web-based survey questionnaires are posted on the web in a manner in which any "surfer" may obtain access. Accordingly, the system should feature an adequate access control mechanism to eliminate inappropriate responses. Third, duplicate responses, intentional or accidental, from the same respondent can bias the results unless the survey system has adequate mechanisms for dealing with duplicate responses.

The Technology

There are several existing technologies available that partially solve these problems. The three most popular techniques are cookies, server-side passwords, and embedded unique identifiers. Cookies are small files generated by the web server that are installed on the web visitor's computer through the browser. When a respondent completes a survey, a cookie is placed on his/her computer. When the respondent tries to respond a second time, the system sees that the cookie is there and can deny the response. The problem with cookies is that browsers can be set to not allow them to be placed on the computer. Unfortunately, this is the only available mechanism for preventing duplicate responses. If the survey system uses cookies, those who do not want cookies to be placed on their computer will not be able to complete the survey. This will not only limit the number of potential respondents but introduces additional bias (to the extent that people who turn off their cookie capability on their browser have a response pattern that differs from those that do not).

Server-side passwords on the survey questionnaires can protect the respondent from individuals who do not know the password. However, the simple addition of password security does not prevent the same respondent from completing the survey multiple times or prevent someone from sending the password to other people.

Embedding a unique identifier on the URL (Uniform Resource Locator) address of the survey questionnaire and using e-mail to invite people to participate eliminates duplicate responses and denies unwanted responses. Unfortunately, URL links are not supported by all e-mail packages, raising the possibility of additional bias.

Because all three of these technologies pose some drawbacks, the survey system developed by the authors utilized another technology. The technology is based on the development of a survey participant database. This database includes participants' demographic information, e-mail addresses, and participation indicators. Because survey participants are students taking fully online and web-enhanced courses at the university, this database is not difficult to build at the beginning of each semester. The system uses a database approach to solve the three problems of web-based surveys. A detailed description of how to use this database approach is given in the next section.

COMPONENTS OF THE UCF WEB-BASED SYSTEM

The web-based survey system developed by the University of Central Florida research team is designed around three components: data collection, data validation, and data analysis. All three components were developed with SAS/Intranet Application Dispatcher. SAS/Intranet Application Dispatcher is a SAS technology tailored to building dynamic and interactive pages for the World Wide Web. SAS/Intranet Application Dispatcher technology allows users to create dynamic web content through a common gateway interface (CGI) program developed by the SAS Institute. Although it is a thin client technology, the user can add JAVA code and/or JAVA scripts on the client side as well. For example, client side JAVA and/or JAVA Scripts allow the survey participants to examine and modify their answers before clicking on the submit button. After the SAS CGI sends the data collected on the web page to the server, the full SAS computing power becomes available. For example, it is possible to use SAS/SQL to retrieve data from SAS data sets and store data into the SAS data sets. In addition, other SAS

procedures such as SAS/STAT, SAS/BASE, and SAS/GRAPH can be used to perform data validation, data manipulation, statistical data analysis, graphical data presentation, and reporting. The survey analysis component was developed with SAS/WEBEIS technology developed by the SAS Institute. A Dell PowerEdge 6300 workstation with Windows NT Server 4.0 was used as a web server to host both data collection and data validation components. Another Dell personal workstation with Windows 4.0 was used as a SAS/Intranet server to host the survey analysis component.

Data Collection Component

Data collection through web-based survey methodology is one example of self-administrated data collection with survey research. Similar to mailed instruments, respondents are expected to complete the questionnaire by themselves in the absence of assistance. Therefore, web-based survey methods also require questions that have been validated and are unambiguous, straightforward, and logical. One way to achieve this goal is to capitalize questions from other studies that have a history of valid and reliable performance (Bourque and Fielder (1995, Page 39). The survey instrument used in UCF's distributed learning impact evaluation over the past three years (Moskal, Dziuban, and Moskal, 1999) was carefully designed and successfully executed. It was used for the web-based survey. This decision was made to ensure consistency of evaluation. This also allows a comparison between web-based surveys and paper-based surveys.

Converting the questionnaire into dynamic web pages requires three steps. The first is to create the survey questionnaire with HTML (HyperText Markup Language). Five different response icons that include text box, radio button, check boxes, drop down box, and multiple line text box with scroll bars were used in the page. The text boxes are used to collect a single-line text response. For example, we might use a text box to ask survey participants to provide their social security number. This is, however, always optional. Radio buttons are used for single selection questions and check boxes for multiple selection questions. The drop down boxes are used when there are many options and the respondent is required to select on one. For example, we use the drop down boxes to ask the survey participants to provide their academic standing. The multiple line text boxes with scroll bars are used to collect open-ended responses such as difficulties they experienced with web-based learning. The next step is to create the self-examination page with the client-side script. The self-examination page will pop up when the survey participants click the "Validate Data" button on the survey questionnaire. This allows the participants to validate their responses before submission. The last step is to include the server-side script that will send data to the data validation component after the participant clicks the "Submit" button on the questionnaire

Data Validation Component:

The survey participant data set includes students who participated in fully online courses during the summer of 1999. This data set includes students' demographic information, their e-mail address, and a survey response indicator. The survey indicator is set to the number of fully on-line courses taken by the student. For example, the indicator is set to two if the student has taken two fully online courses during the summer of 1999. Through an email invitation, students on this data set are invited to participate in the survey study. To reach the survey questionnaire they can either click through the URL of the survey questionnaire on the e-mail or enter the URL in their web browser. Each student's e-mail address is his or her password to participate the survey. In addition, there are several questions on the survey instrument that ask the respondents to provide their demographic information that may be found on the survey participant data set.

The data validation component performs three types of data checking: identify survey respondents, eliminate unwanted responses, and delete the duplicate responses with SAS data

step programming. The first eight questions on the survey questionnaire are used to collect the demographic information from survey participants. The demographic information can then be compared with the demography on the potential survey participant database, thus ensuring the survey respondents are the individuals selected by the survey administrator. Although the demographic information comparison cannot completely identify the survey respondents, the procedure is preferable to the absence of identification. The system uses server-side passwords and demographic information comparison together, thus guarding against invalid responses. Without knowing the password (e-mail address), the web surfers cannot log on to the questionnaire page. If, by some unforeseeable circumstance, they log on to the survey questionnaire, they are still required to provide demographic information that can be matched with their password before their responses will be accepted. After accepting the responses from each respondent, the participation indicator for each respondent is set to zero to restrict the same respondent from submitting another completed instrument.

After validation, the collected data will be sent to the SAS data library where three data sets are created. The demographic information matrix includes the data from the first series of questions. For the next series of questions, a matrix is created that contains the reactions to their online learning experience. The last series of questions provides data on the self-reported learning style of the respondent. Each of the three sets also includes e-mail addresses and the courses taken by the students as the key to connect three sets.

Data Analysis Component

The data analysis component is designed around a SAS/WEBEIS application. SAS/WEBEIS is a point-and-click GUI (Graphical User Interface) environment that allows the user to perform descriptive data analysis. This application allows the user to look at the raw data in tabular format, to compute numerical measures such as count, frequency, mean, and median, and to view the graphical presentation through pie, bar, organization charts, line graphs, and data plots. Predefined hierarchies allow the administrator to perform drill down analysis to browse the data. For example, the gender-ethnicity hierarchy allows the user to look at the gender difference first and then the ethnicity difference within gender. In addition, the ethnicity-gender hierarchy allows the user to look at the ethnicity difference first and then compare the gender differences. Hierarchical definitions providing a greater the flexibility to explore through the data in several different ways and consequently to discover previous latent dimensions in the data. There are several screens that allow the user to view the raw data, summary statistics, and the graphs simultaneously. One of the analytical screens is shown in the Appendix 1.

CONCLUSION

The SAS/Intranet product brings SAS power into the web. This on-line survey project is the first step to bring online survey capability into distributed learning at the University of Central Florida. The analysis from this pilot study has found that the on-line survey system is at least as good as the traditional mail survey system although the sample size for the data collected in the Summer of 1999 is too small to declare that the online system is better. A full version of the online survey system will be used again in the summer of 2000. Hopefully, a clearer understanding of the advantages and/or disadvantages of using an on-line survey system can be reported.

REFERENCES

1. Bourque, L. B. and Fielder, E. P. (1995) "How to Conduct Self-Administrated and Mail Surveys, Thousand Oaks, CA: Sage Publications.
2. Homer, A. et. al. (1997) "Professional Active Server Pages", Olton, Birmingham: Wrox Press Ltd.
3. Moskal, P. D., Dziuban, C. D., and Moskal, P. J. (to appear)

"Faculty Go On-Line, Impressions of Faculty Teaching Web and Web-Enhanced Courses" submitted.

4. Section on Survey Research Methods (1997a), "How to Collect Survey Data?", Alexandria, VA: American Statistical Association.
5. SAS Institute Inc., Building a Data Warehouse Using SAS/Warehouse Administrator (Cary NC: Author, 1997a).
6. SAS Institute Inc., Introduction to SAS/EIS^{TM} Software and SAS/MDDB® Server, Course notes, (Cary, NC: Author, 1997b).
7. SAS Institute Inc., SAS/EIS Software Reference, Version 6, Second Edition, Course notes, (Cary, NC: Author, 1997c).
8. SAS Institute Inc., SAS Web Tools: Running SAS Applications on the Web, (Cary, NC: Author, 1998).
9. SAS Institute Inc., SAS/STAT, Version 6, Fourth Edition, (Cary, NC: Author, 1994)
10. SAS Institute Inc., SAS/GRAPH, Version 6, First Edition, (Cary, NC: Author, 1991)
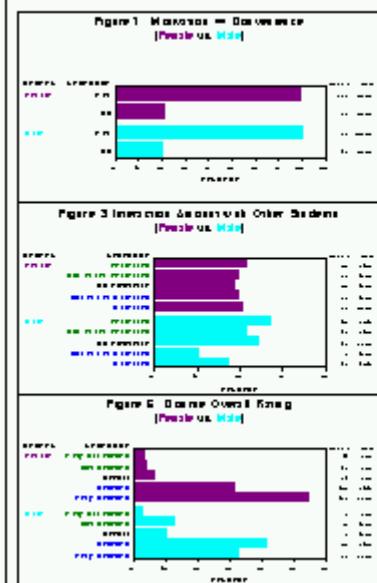
CONTACT INFORMATION
Morgan C. Wang
Department of Statistics
University of Central Florida
Orlando, FL 32816-2370
Work Phone: 407-823-2818
Fax: 407-823-3930
        Email: cwang@mail.ucf.edu
        Web: http://www.pegasus.cc.ucf.edu/~cwang

Appendix 1: Selected Output from the Data Analysis Component

Students' Attitudes By Gender Toward Taking Web—Based Courses

Convenience is the primary motivation for students taking courses on the web. However, after taking web courses, more than 80% of students are either satisfied or very satisfied and they are willing to take another web—based course.

Gender Differences:
(1) More females have changed their learning approach.
(2) More males have reduced their interaction with other students.

Green Text: Negative Direction
Black Text: Neutral Direction
Blue Text: Positive Direction