"Methodological issues in Online Data Collection for Research"

* Upendra Singh Panwar
**Sanjay Maheshwari

Abstract

One of the fundamental questions to be answered when designing a research project is the data collection methodology that will be used. Two methods which are now commonly in use are– Hard copy surveys and Internet surveys; each offers a set of advantages and challenges. Hard copy surveys, in the past or in present, been the method used most by researcher based on positive response rates and broad access to most of the respondent.

Now, with the continuing evolution of online surveys and the increasing level of Internet usage, researchers are using more online research as compared to traditional way of hard copy. But before going on any mode of data collection researcher has to look the pros and cons of each method carefully prior to commence.

This overview discusses a variety of issues and considerations related to the pros and cons of both Hard copy (Paper and pencil) and online survey research.

Online data collection for researcher might be replacing paper-and-pencil surveys or questionnaires in the near future. This research discusses the advantages and limitations of online data collection; Email is generally sent to respondent for contacting and to ask them to visit a designated website in order to complete the questionnaire/response.

There are chances that some problems arose with the use of an online web-based programme for data collection. Among them one can be unreliability of the email address lists and the other can lack of willingness, particularly among students, to participate. The research concludes that while online surveys can access large and geographically distributed populations and achieve quick returns, they may no longer be as universally appealing as was once believed. Reaching to the large population sample remains a problem in online as well as in traditional data collection. Also the response rate is very low. As very few of them turned up with the response.

Keywords: Internet Banking, Online data collection, Customer satisfaction, User studies, Banking.

Asst. Professor, SGSITS, Indore, upendra, zivon@gmail.com
****** Asst. Professor, Vishisht School of Management, Indore, sanjay@vsom.in***

Introduction:

Internet research is the practice of using the Internet, especially the WWW, for research. The internet is widely used and easily accessible to thousands of people in many parts of the globe. It can provide fast information on most of the topics.

Research is a broad term. Here, it is used as "looking something up. It includes any activity where a subject is identified, and an effort is made to actively collect information for the purpose

of furthering understanding. Common applications of *web research* include personal research on a particular subject.

Compared to the Web, prints have physically limited access to information. A book has to be identified, and then actually to be looked for the information. While on the Net, the Web can be searched, and typically hundreds of pages can be found with some relation to the topic, within seconds.

New mode of data collection; Google spreadsheets:

Google spreadsheet is an online spreadsheet application that lets you create and format spreadsheets, charts, and gadgets, and simultaneously work with other people. Here's what you can do with Google spreadsheets:

- Import and convert .xls, .csv, .txt and .ods formatted data.
- Export .xls, .csv, .txt and .ods formatted data and PDF and HTML files.
- Use formatting and formula editing so you can calculate results and make your data look the way you'd like.
- Chat in real time with others who are editing your spreadsheet.
- Create charts and gadgets.
- Embed a spreadsheet, or individual sheets of your spreadsheet, in your blog or website.

Literature Review: The Internet gives researchers access to an unprecedented –and unstable– archive of human activity and its traces (Jones, 1999; Lindlif and Shatzer, 1998).

It differs from other media and communication technologies in a number of ways that affect research design. It reaches across cultural boundaries (Mitra and Cohen, 1999), national borders (Commercenet, 2002) and legal jurisdictions (Geist, 2001).

It integrates multiple modalities of communication and types of content (Di Maggio, Hargittai, Russell-Neuman and Robinson, 2001; While facilitating play between them (Sosnoski, 1999).

It provides technical utilities for efficiently collecting and processing data (Lindlif and Shatzer, 1998) but can obscure correlations by destabilising identity (Allen, 1996; Turkle, 1999).

Researchers from across different disciplines are starting to see the benefits of collecting data using the Internet, and increasingly, journals are publishing data that have been collected online (Schleyer & Forrest, 2000).

In an age of advanced technology, Americans are becoming progressively more computer literate. More and more people have access to the Internet, and the Internet is fast becoming the communication method of choice for many Americans (puffy, 2000).

Rationale:

As the no. of participants enrolling for PhD is increasing, Collection of data is a tuff and major task for them. Online surveys has emerged as a good platform where a researcher can collect data in a less time as well as other advantages of web-based research includes reduced response time, lower cost of data collection, ease of data entry and collection, flexibility of and control over format, advances in technology, recipient acceptance of the format, and the ability to get additional response-set of information. But there is a major problem in collecting data i.e. less response rate. This study is a way to find out the factors which contribute less response rate of the respondents, and by working on those factors response rate can be increased.

Objective:

1. To study the advantages of the online data collection.
2. To study the disadvantages of the online data collection.

Methodology:

a) Sampling Technique: Convenient by mail.
b) Sampling Unit: Respondent who have either responded to web-based questionnaire or who have conducted research by the web-based technology
c) Sampling Size: 31
d) Tools for Data Analysis: Simple Percentage Method.

Method:

The data was collected by the respondents who have either responded to web-based questionnaire or who have conducted research by the web-based technology. 31 responded to the questionnaire.

Advantages of online data collection:

Online surveys have several pros over traditional way of paper- and-pencil surveys that make them particularly beneficial to researchers and reduce response time, lower cost of data collection, ease of data entry and collection, flexibility and control over format, advances in technology, recipient acceptance of the format, and the ability to obtain additional response-set information of respondents.

1. Lowered cost for collection of data: When analyzed the costing for data collection by pen and pencil the cost is much high as it includes the cost of travel/boarding from one place to another for collecting data also costing of zerox etc. When analyzed the entire respondent i.e.

100% accepted that the costing in online data collection is cheaper than the hard copy collection or by paper and pen.

Farmer (1998) argued that Web-based surveys are 50% less expensive to implement than telephone surveys, and 20% less expensive than mail surveys. However there are, some costs associated with Web-based surveys, primarily for programming, using space on a server, and some limited data entry and/or manipulation.

2. Reduced time of response: The response shows that 83.87% responded accepted that the time is reduced, when compared to collection in a hard copy or by pen and pencil.

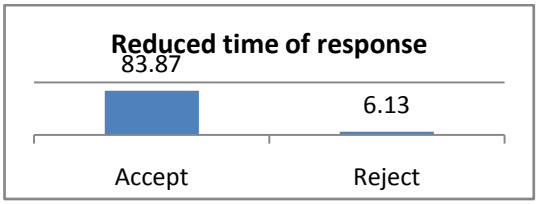


Fig.1

Farmer (1998) reported that typical turnaround time is 4 to 6 weeks with traditional mail surveys, 2 to 3 weeks for telephone surveys, and only 2 to 3 days for Web-based surveys. One of the primary advantages of e-mail and Web-based surveys is that they dramatically decrease response times (Lazar & Preece, 1999), Franceschini (2000) also reported reduced turnaround time. In his research, half of the respondents were sent mail surveys, the other half were surveyed via the Internet. He reported that 21 of the 29 Web-based responses were received before there were any responses to the traditional mail survey.

3. **Easiness of data entry:** Most of the respondent responded that there is ease of data entry by electronic way, out of 31 respondent, 25 respondent i.e. 80.64% feel easiness in entry of data. In traditional way of hard copy surveys, data entries are very expensive and time consuming. Data collection by electronic way can be used to send data to a database or spreadsheet, eliminating the need for manual data entry like by google spread sheet. This also eliminates potential errors in rekeying data or while entry. Automatic data entry is an advantage only by Web-based surveys. For e-mail surveys, when collecting by excel or MS office, data still need to be manually transferred to a database.

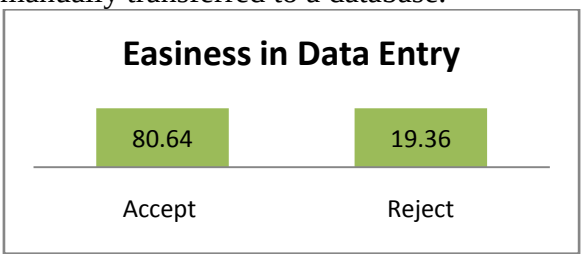


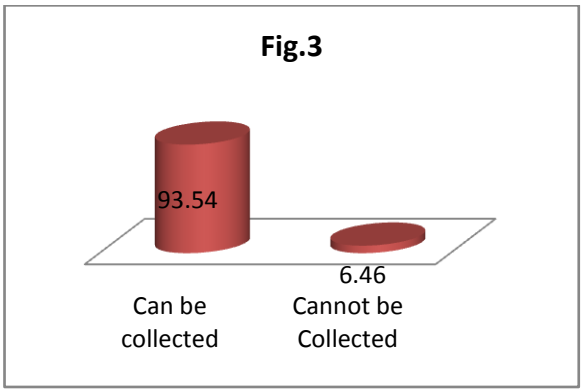
Fig.2

4. Control over format and Flexibility: 50% of respondent responded that electronic way of collection is flexible for collecting data. While half does not feel control over format.

Researchers can control the order in which respondents answer the questions more easily than with paper-and-pencil surveys, which allow respondents to flip back and forth among the pages and change answers [Wyatt, 2000). Using the electronic way it allows researchers to use flexible design formats such as graphics, color, displays innovatively questions, split screens, animation, and sound (Dillman, Tortora, & Bowker, 1999). Moreover, other rules, such as "select one answer only" or "do not leave this question blank," can be enforced with radio buttons (Lazar & Preece, 1999). Although flexibility

of format can be extremely useful, there are no definitive answers as to the psychometric effects of the various Web-based formatting options (Arnau, Thompson, & Cook, 2001).

5. Additional respondent information: 93.54% respondent agrees that while collecting data by online mode one can collect many more information as compared to the traditional way. With traditional paper-and-pencil surveys, researchers can only see the results of the participants' responses. With both e-mail and Web-based surveys, information such as time of day or day of the week of the response can also be tracked.



Participants in online discussion groups often log on in the evening and during the night. One study of participants in an electronic support group found that 31% of postings (i.e., messages or responses placed online) were between 11 p.m. and 7 a.m. (Winzelberg, 1997). Using the Web, researchers can learn about the respondents' answering process (Bosnjak &Tuten, 2001).

6. Acceptance by Recipient: When we talk of sending questionnaire by general post there is no surety of acceptance by the recipient. But in case of electronic way instantly there is a delivery report for delivery. Every respondent i.e. 100% accepted that the acceptance by the recipient is confirmed with in minutes by electronic way.

E-mail surveys tend to be very easily understood and completed by recipients, although there are cautions about reduced self-disclosure for e-mail surveys because the recipient's e-mail address is attached to the response (Harris & Dersch, 1997). There is some evidence that the Internet is becoming more acceptable to respondents as a method of collecting data, particularly for men (Dillman et al., 2001) and for individuals who are college educated (Cartwright, Thompson, Poole, & Rester, 1999; Franceschini, 2000).

Disadvantages of Online Data Collection

With so many advantages of Web-based surveys, concerns about their use have been raised by researchers. These concerns majorly focus on the following limitations: representativeness of the sample, response rates, measurement errors, and technical difficulties.

1. Response rates. The response rate by mail is very low even I have collected data by web-based survey for that I forwarded twenty thousand mail and got only 200 response i.e. only 1% of them responded When asked about the response rate by the respondent the response of 87.7% is that it is very low. And 12.3% of them responded don't know.

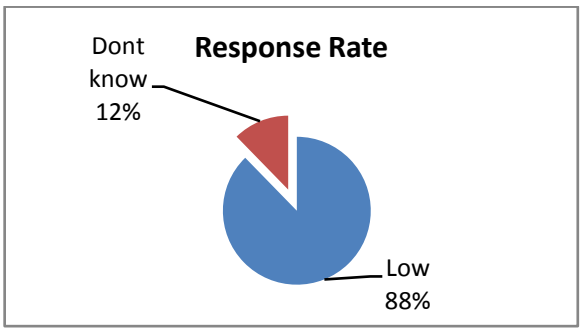


Fig.4

The reasons for low response are: Lack of interest as the respondent don't know who is a researcher and don't have interest in his research, Lack of time as most of people uses only to read mail of the office/ company and revert back to customers only and don't have access of internet at home, other reason can be lack of awareness, lack of trust or misuse of information by the respondent, lack of resources like computer or net connection or for student class who uses net at cyber café by paying money will not respond.

Several studies show that e-mail surveys produce a significantly lower response rate than traditional mail surveys (Bachmann, Elfrink, & Vazzana, 1996; Couper et al, 1997; Crawford, Couper, & Lamias, 2001; Tse et al., 1995). Researchers doing studies using Web-based surveys have also found lower response rates than for traditional mail surveys (Medin, Roy, &Ann, 1999; Nichols & Sedivi, 1998).

2. Representativeness of the sample. 31 respondents all of whom have collected data or responded by web based. All of them don't know who is the respondent or by whom mail is sent for data collection in case of bulk mail. That means while sending bulk mail most of them don't know the respondents age, name, sex etc.

Internet use in the United States is growing at a rate of 2 million new Internet users each month. In 1 month alone [September 2001), 143 million Americans (54%) used the Internet, representing a 26% increase over August 2000 (ClickZ Network, 2002). Despite this growth, access to the Internet remains unequally distributed throughout the U.S. population. Most Web users are White (87.2%), male (66.4%), married (47.6%), and highly educated, with almost 88% having some college education and more than 59% having obtained at least one degree. In addition, 48% of Internet users are 35 years old or younger (Graphics, Visualization, and Usability Center, 1999).

3. Technical difficulties: This is the most problem which is faced in India i.e., most of the people are not well versed with the technology and computer so to reach to this type of respondent is a tough task. Also the technology like software and other programs in computer we are using, some times don't support the questionnaire at the end of respondent. When asked about this to respondent, 93.5% responded for this problem.

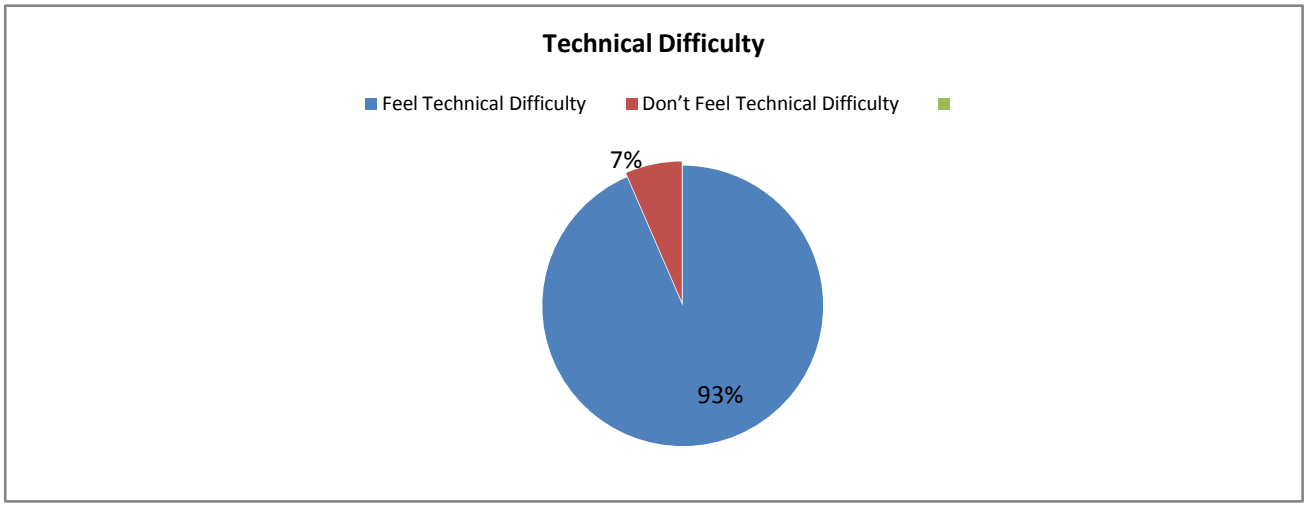


Fig. 5

Also not everyone who completes a Web-based survey will be extremely computer-literate, nor will everyone have access to the most up-to-date technology. Dial-up access is still the most popular method to access the Internet (80%; ClickZ Network, 2002), and 66.5% of Internet users have a connection speed of 56k or slower (Graphics, Visualization, and Usability Center, 1999). As of 1999, most Internet users had as their primary computing platform either Windows 95 (44%) or Windows 98 (18%), and the most commonly used Internet browsers were Navigator, as part of Communicator (Netscape, 45%); Navigator, stand alone (Netscape, 18%); and Internet Explorer (Microsoft, 34%; Graphics, Visualization, and Usability Center, 1999). Researchers using Web-based surveys must ensure that their pages are easily downloaded and maintain their formatting in all types of software and hardware environments. In addition, formatting issues, such as the use of open-ended questions or questions arranged in tables, can lead to higher dropout rates, as can the absence of clear navigational aids on the Web site (Bosnjak & Tuten, 2001).

Findings:

Even if web-based technology has greater advantages over traditional way, then also there is very low response rate for which *major reasons are Lack of interest*, as the respondent don't know who is a researcher and don't have interest in his research 96.77% responded for this reason, *Lack of time* as most of people uses mail only to read mail of the office/ company mail and revert back to customers of company only80.64% responded for this factor. Other factor can be, many people *don't have access of internet at home* 93.54% responded for this factor, other reason can be lack of awareness, *lack of trust* 83.87% responded for this factor and *misuse of information* 32% feel for this factor, lack of resources like computer or net connection or for student class who uses net at cyber café by paying money will not respond response for this factor is 90.32%.

Conclusion:

When we talk about web-based, online data collection method, it creates opportunities to conduct research globally, especially among areas where it is difficult to reach people. But there is limit in this mode that is low response, the major factor responsible for that are *Lack of interest, Lack of time, don't have access of internet at home, lack of trust etc.* So researcher has to look towards this factor before collecting data via web-based survey and find out the substitutes of the above mentioned reasons. However, web-based research requires careful consideration of how the study will be conducted and how data will be collected to ensure high quality data and validity of the findings. Sometimes reminding the respondent to fill data can help in increasing response rate.

References:

Arnau, R. C.,Thompson, R. L., & Cook, C. (2001). Do different response formats change the latent structure of responses? An empirical investigation using taxometric analysis. *Educational and Psychological Measurement*, 61, 23-44

Allen, C, 1996, "What's Wrong with the 'Golden Rule'?" *The Information Society. Issue 12*. Retrieved from Swetsnet Navigator database http://ilm.swetsnetnavigator.com/direct

Bachmann, D., Elfrink, J., & Vazzana, G. (1996). Tracking the progress of e-mail versus snail mail. *Marketing Research*, 8, 31-35.

Bauman, S., Airey, J., & Atak, H. (1998, May). *Effective use of Web-based technology: Using the Internet for data collection and communication applications*. Paper presented at the 53rd Conference of the American Association for Public.

St. Louis, MO. *Opinion Research*, Retrieved from http://www.wirthlin.com/strategic/casefiles/wes.phtml

Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in Web-based surveys. *Journal of Computer-Mediated Communication, 6(3).* Retrieved from http://www.ascusc.org/jcmc/vol6/issue3/boznjak.html

Bowker, D., & Dillman, D. A. (2000). *An experimental evaluation of left and right oriented screens for Web questionnaires*. Retrieved from http://survey.sesrc.wsu.edu/dillman/papers/AAPORpaper00.pdf

Cartwright, D. W., Thompson, R. J., Poole, M. C, & Kester, D. D. (1999). Assessing distance learning using a Website survey. *ERIC Document Reproduction Service No. 433781*.

ClickZ Network. (2002). *U.S. Internet population continues to grow*. Retrieved from http://www.clickz.com/stats/big_picture/geographics/article.php/5911_969541

Conboy, L., Domar, A., & O'Connell, E. (2001). Women at mid-life: Symptoms, attitudes, and choices, an Internet-based survey. *Maturitas, 38*, 129-136.

Couper, M. P., Blair, J., & Triplett, T. (1997, May). *A comparison of mail and e-mail for a survey of employees in federal statistical agencies.* Paper presented at the annual conference of the American Association for Public Opinion Research, Norfolk, VA.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perception of burden. *Social Science Computer Review, 19,* 146-162.

Dillman, D. A., Phelps, G., Tortora, R. D., Swift, K., Kohrell, J., & Berck, J. (2001). Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response, and the Internet. Retrieved from *Washington State University, Social &*

Economic Sciences Research Center, http://survey.sesrc.wsu.edu/dillman/papers/Mixed%20Mode%20ppr%20_with%20Gallup_%20POQ.pdf

Dillman, D. A.,Tortora, R. D., & Bowker, D. (1999). Principles for constructing Web surveys. Retrieved from Washington *State University, Social & Economic Sciences Research Center* Web site: http://survey.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf

Dillman, D. A., Tortora, R. D., Conradt, J., & Bowker, D. (1998). Influence of plain vs. fancy design on response rates for Web surveys. Retrieved from *Washington State University, Social & Economic Sciences Research Center* Web site: http://survey.sesrc.wsu.edu/dillman/papers/asa98ppr.pdf

Di Maggio, P, Hargittai, E, Russell-Neuman, E and Robinson, J, 2001, "Social implications of the Internet." *Annual Review of Sociology. 27*, Retrieved from *Proquest database database* http://global.umi.com.

Duffy, M. (2000). Web-based research: An innovative method for nursing research. *Canadian Oncology Nursing Journal, 10(2),* 45-49.

Farmer, T. (1998, December). *Understanding this thing we call Internet research.* Retrieved from http://infotekonline.com/Info_Online_New/White_paper_on_internet.htm

Franceschini, L. A., III. (2000). Navigating electronic survey methods: Three pilot studies. *ERIC Document Reproduction Service No. 448183*

Graphics, Visualization, and Usability Center. (1999). GVU's tenth WWW user survey. Retrieved from *Georgia Institute of Technology, GVU Center, College of Computing* http://www.gvu.gatech.edu/user_surveys

Harris, S. M., & Dersch, C. A. (1997). Conducting research on the Internet: Potential, concerns, and reflections. *Kappa Omicron Nu Forum: Technology, 11(1).* Retrieved from http://www.kon.org/archives/forum/11-1/harris.html

Joinson, A. N. (1999). Social desirability, anonymity, and Internet-based questionnaires. Behavior Research Methods, *Instruments and Computers, 31*, 433-438.

Joinson, A. N. (2001). Knowing me, knowing you: Reciprocal self-disclosure in Internet-based surveys. *CyberPsychology & Behavior, 4*, 587-591.

Kittleson, M. (1997). Determining effective follow-up of e-mail surveys. *American Journal of Health Behavior, 21*, 193-196.

Lazar, J., & Preece, J. (1999). Designing and implementing Web-based surveys. *Journal of Computer Information Systems, 39(4),* 63-67.

Lindlif, T and Shatzer, M, 1998, "Media ethnography in virtual space: strategies, limits and possibilities." Journal of Broadcasting and Electronic Media. Volume 42, Issue 2. Retrieved from *Infotrac database database*. http://infotrac.galegroup.com.

Lindlif, T and Shatzer, M, 1998, "Media ethnography in virtual space: strategies, limits and possibilities." Journal of Broadcasting and Electronic Media. Volume 42, Issue 2. Retrieved from *Infotrac database database*. http://infotrac.galegroup.com.

Medin, C., Roy, S., & Ann, T. (1999). *World Wide Web versus mail surveys: A comparison and report*. Retrieved from http://www.anzmac99.unsw.edu.au/anzmacfiles/papers.htm

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 27*, 323-339.

Nichols, E., & Sedivi, B. (1998). Economic data collection via the Web: A census bureau case study. Proceedings of Survey Methods Section, 1998 *American Statistical Association Meetings, Dallas, Texas.*

Niles, S. G., Akos, P., & Cutler, H. (2001). Counselor educators' strategies for success. *Counselor Education and Supervision, 40*, 276-291.

Parker, L. (1992, July). Collecting data the e-mail way. *Training and Development, 46*, 52-54.

Panwar Upendra et al. (2012) "Analyzing dimensions of Service quality of government banks in Indore region" *Prabandhan & Taqniki, LNCT Indore*, *Vol.6/Aug.2012,* 367-371.

Panwar Upendra et al. (2012) "Measuring Service Quality in Government Banks with special reference to Indore District" *Business Innovation & Entrepreneurship*, PIMR/ Indore/ 293-302.

Schleyer, T. K. L., & Forrest, J. L. (2000). Methods for the design and administration of Web-based surveys. *Journal of the American Medical Informatics Association, 7*, 416-425

Schmidt, W. (1997). World Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers, 29*, 274-279.

Solomon, D. J. (2001). Conducting Web-based surveys. Practical Assessment, *Research, and Evaluation, 7(19).* Retrieved from http://pareonline.net/getvn.asp?v=7&n=19

Tse, A. C. B., Tse, K. C., Yin, C. H., Ting et al. (1995). Comparing two methods of sending out questionnaires: E-mail versus mail. *Journal of the Market Research Society, 37*, 441-446.

Turkle, S, 1999, "Looking toward cyberspace: Beyond grounded sociology." *Contemporary Sociology. Volume 28, Issue 6.* Retrieved from Proquest database. http://global.umi.com

Web Accessibility Initiative. (2002) *Web Accessibility Initiative (WAI).* Retrieved from http://www.w3.org/WAI/

Wheaton, J. E., & Granello, D. H. (2001). The training needs of employees at the Ohio Rehabilitation Services Commission: Results of a multiphase needs assessment. Unpublished manuscript.

Winzelberg, A. (1997). The analysis of an electronic support group for individuals with eating disorders. *Computers in Human Behavior, 13*, 393-407.

World Wide Web Consortium. (1999). *Web content accessibility guidelines*. Retrieved from http://www.w3.org/TR/1999/WD-WAI-PAGEAUTH-19990217/#transform-gracefully

Wyatt, J. C. (2000). When to use Web-based surveys. *Journal of the American Medical Informatics Association, 7*, 426-430.

Wheaton Joe E, Darcy Haag Granello, Journal of Counseling and Development: *JCD. Alexandria: Fall 2004. Vol. 82, Issue 4*; 387, 1-7

http://en.wikipedia.org/wiki/Knowledge Retrieved 12/03/2013

http://support.google.com/docs/bin/answer.py?hl=en&answer=49008 Retrieved 12/03/2013