

Mixed Mode

Handling method-differences between paper and web questionnaires

Cecilia Bäckström, Christina Nilsson
Department of Information Technology and Media
Mid Sweden University, 831 25 Östersund, Sweden
E-Mail: ceba9100@student.mh.se chni9900@student.mh.se
Web: <http://gathering.itm.mh.se/~cicbac>

Abstract: The purpose of this paper is to point out differences in the answers of paper and web questionnaires and to recommend what to consider when using web questionnaires in mixed mode surveys. These recommendations can be useful when implementing web questionnaires. A problem with mixed mode surveys is that the diverse methods have generated different answers. To investigate this further we have conducted a survey consisting of one paper questionnaire and two web questionnaires, involving 90 respondents. To test our collected data we have used a Test of Homogeneity- Pearson's Chi square test (X^2) and ANOVA tests. The result of this study shows that there are differences between the methods. Based on the differences we present recommendations how to facilitate the use of web questionnaires in mixed mode surveying. The recommendations are based upon the importance of usability and how to design a web questionnaire.

Keywords: Mixed mode, web questionnaires, paper questionnaires, surveying, design

I. INTRODUCTION

Mixed mode surveys constitute a combination of different methods, for example telephone, mail, Web, and/or face-to-face interview procedures, to collect data for a single survey project. Each basic survey method has its own inherent limitations such as problems with representative sampling in web-surveys, and thus researchers are designing and implementing mixed mode surveys in an increasing rate. According to NuStats, this requires special considerations concerning questionnaire design in order to minimize measurement differences [17].

Web surveying over the Internet is now an established part of market research. The need to further study the behaviour of respondents in terms of their interaction with the questionnaires has been shown previously [11]. Taylor et al argues that online surveys tend to produce much more "not sures" and different responses to attitude scales than telephone interviews [19]. It is also known that in surveys over the Internet the respondents avoid extreme values in scaling questions. Today there are no explanations for this behavior, which is why it is important to continue this research. With more research we might have a greater knowledge and understanding how to use the Internet to make both qualitative and quantitative surveys [18]. Prior research in this area has mainly been focused on mixed mode surveys combining telephone and web questionnaires.

Changing from paper-based to electronic media creates new challenges both for survey methodologists and for user-interface designers. The fact is that the use of mixed mode surveys are increasing hence there is a need for further research to facilitate for those who conduct mixed mode surveys in order to collect reliable data.

Grisaffe questions whether web-based surveys will introduce method effects that are not present in other methods [9]. A way to test this method effect is to simultaneously compare web-based surveys in combination with other methods, where equivalence of groups is ensured before administering different survey modalities. Researchers considering online surveys should make a decision whether it is reasonable to assume that there is no pure method effect. [9]

In this paper our aim is to explain and draw attention to differences in the answers between paper- and web questionnaires and to come up with recommendations what to consider when using web questionnaires in mixed mode surveys. These recommendations can be useful when implementing web questionnaires.

In order to fulfill our purpose, the following research questions were formulated: What difference is there in the answers between paper and web questionnaires? How can these differences be explained? What recommendations can we give?

The disposition of this paper is as follows: In section II we describe the method that we used for our survey. Sections III, IV and V consist of the analysis where we present our data with descriptive statistics, analyze the comparisons and present our calculations and conclusions. In section VI we present our conclusions and recommendations. Finally we give some conclusive remarks and a short summary on our research in sections VII and VIII.

II. METHOD

To answer our research questions we constructed three separate questionnaires. One traditional paper questionnaire (PQ), one web questionnaire (WQ1), identical in layout with the paper questionnaire and a second web questionnaire (WQ2) with a different layout and different components. The design of WQ2 is described in detail under part E, "the survey". The questions asked were the same in all three questionnaires, as were the answer alternatives provided.

After completing the questionnaires the respondents were asked to answer some additional questions for evaluation. This evaluation was made in order to help explaining why differences between methods occur. To be able to compare the evaluations towards the methods, the respondents were given an ID-number to be noted on both the questionnaire and evaluation answered. In addition a few respondents were selected to answer some questions in a face-to-face interview and we also held an Ideawriting-seminar in order to explain and generate possible causes to the differences.

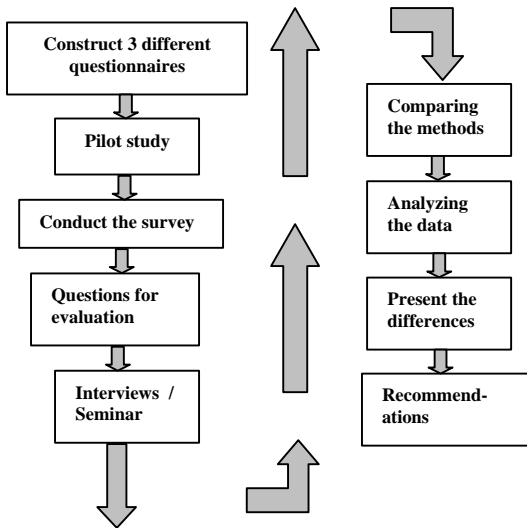


Fig. 1. Method description

A. Relevant variables

In our survey the independent variables (IV) are the questionnaire mode answered, the type of questions and gender. We have selected a sample based on an approximately equal amount of both men and women and made a comparison based on which questionnaire answered. The dependent variables (DV) are the answers of the questionnaires. Other variables apart from the independent and dependent are situation variables, which contain all those variables found in the environment [8]. In an attempt to avoid such differences in our survey, we had the respondents answer the questionnaires in the same environment and at the same time.

B. Population and sample

The population consisted of students at Komvux, an adult education center, in Östersund. This population was chosen to achieve a larger variety in terms of age, computer experience and Internet usage. Out of these students we have randomly chosen 90 persons to participate in our survey, 30 respondents to each questionnaire respectively. These students were selected from three separate classes, each containing an approximately equal amount of men and women.

Since Komvux was supposed to conduct the survey in any case we did not need to inform the respondents in advance what our purpose of the survey was. This was to make sure that the respondents were not to be influenced by the fact that there are differences between the methods. The questions were predetermined by the head master at Komvux.

C. The multigroup design

In our study we assigned the student sample group to answer either a PQ or two different types of WQ's. As mentioned in part B, "population and sample", the respondents were selected from Komvux to make sure that the groups were as equal as possible. We compared our

three conditions using a Chi-square and an ANOVA test. Figure 2 shows a model of our experiment design.

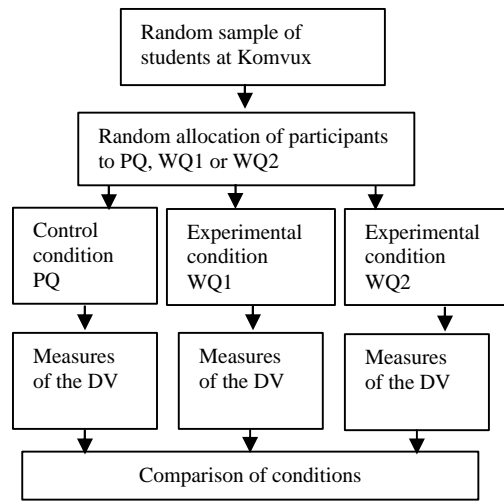


Fig. 2. Multigroup experiment design

D. The comparisons

Out of the collected data we made several comparisons to see if there were differences between the methods and, if so, where they occurred. The first comparisons were made to see whether there were any significant differences between the groups. These comparisons were analysed using a Chi-square test to single out the significant differences between the methods. Then we used a One-way ANOVA test to see if there were any differences between genders and to measure the use of extreme values in the answers due to the different methods. We also calculated the number of used steps of the attitude-questions made by each participant. For this calculation we used a comparison of the means. Based on the answers provided by the respondents the evaluations were calculated in percentage. The methods of the comparisons are illustrated in figure 3.

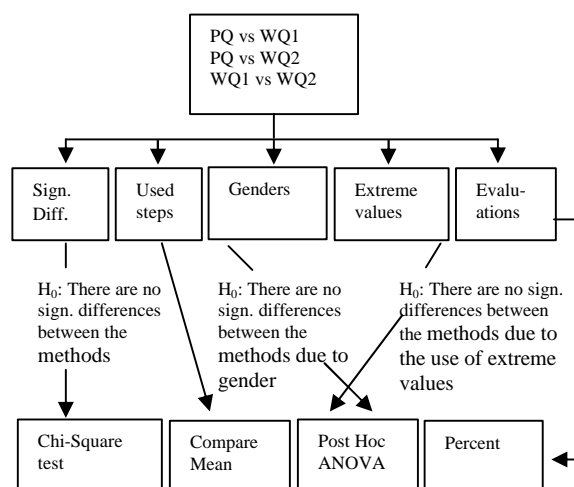


Fig. 3. Comparison model

E. The survey

The survey consisted of a paper questionnaire (PQ) and two web questionnaires (WQ1) and (WQ2). Each contained thirteen fact- and attitude-based questions. The questions were identical and appeared in the same order independently of method. The design of the PQ and WQ1 was as identical as possible. WQ2 was designed with a different layout with the questions displayed one at a time and no possibility to go back in the questionnaire. WQ2 contained a set of different components compared to WQ1, such as drop-down menus and list boxes. The questions categorized as attitude-questions contained seven steps each. We made all of the questions closed-ended to make them quick and easy for the respondents to answer and to make them easy to code and analyze for us. The choice not to use open-ended questions was made in order to have the possibility to measure the use of extreme values.

The survey was conducted in the course of one day at Komvux. The respondents to the questionnaires were divided into three groups containing 30 students in each group. The groups were assigned to answer the questionnaire in different modes.

F. The evaluation questions

In order to suggest why differences in the answers occur, the respondents were asked to answer some additional questions. This questionnaire was solely paper based and consisted of ten questions regarding how the respondents perceived the questionnaires. The answer alternatives were predetermined by us and concerned causable reasons to the occurring perceived differences. The questions were designed with multiple-choice answers as to make them easy to analyze. The lack of open-ended questions was compensated for with face-to-face interviews with some of the respondents.

G. Software

The web questionnaires have been constructed using the Mod_survey software for designing and collecting answers for web surveys. Mod_survey is developed at ITM by Joel Palmius¹.

For the statistical calculations we have used the SPSS² software, developed for data mining technology and analytic solutions.

H. Pilot study

Surveys often have problems with unclear wording, confusing design, questions that are difficult to understand and typographical errors. This can easily be corrected by a pilot test. Therefore we tested our questions on a selected group of people before finalizing the questionnaires. Their feedback helped us clarify and improve our survey and reduce measurement errors. The corrections made concerned unclear wording.

I. Ideawriting-seminar and interviews

In order to suggest possible explanations and solutions we held a seminar and interviewed some of the respondents.

The seminar was based on John Warfield's "Ideawriting" [20]. We invited a group of people with expertise knowledge in areas of interest like statistics, psychology and web design. The triggering question, what fundamental reasons cause the differences between paper- and web questionnaires, was discussed.

The interviews were conducted at Komvux with some of the students who had answered the web questionnaires. The interviews lasted for 20-40 minutes and were recorded. After each interview we listened to the tapes and wrote down the essentials that had come up during the interviews.

J. Methods of Analysis

For the comparisons of the collected data we have used two different tests, Pearson's Chi square test (X^2) [10] and a One-way ANOVA [15]. For each analysis we have chosen a significance level of 0.05. The null hypothesis will be rejected if the p-value is 0.05 or less.

The first null hypothesis was calculated using a Chi-square test. H_0 : there are no significant differences between the survey methods.

The second and third null hypothesis was made using a One-way ANOVA. The hypothesis are: H_0 : there are no significant differences between the groups mean scores due to the use of extreme values and H_0 : there are no significant differences between the methods due to gender. The answers of the questionnaires are our dependent variable. Gender and the questionnaire mode answered are our independent variables.

To be able to make the calculations on the extreme values we recoded the variables of the attitude questions into new different variables which contained only two values instead of seven. The extreme values 1 and 7 were given the value 1 and the values 2-6 were given the value 0. After obtaining a statistically significant test from the ANOVA, we wanted to know which mean values that contributed to the effect, which groups are particularly different from each other. We looked at the results of a Post-Hoc comparison to see exactly which pair of groups that had significantly differences. As expected we found avoidance in the use of extreme values in the WQ's.

The evaluations were analyzed in frequencies and percentage and then compared towards the different methods.

K. Sources of error

When conducting surveys there are always sources of error that might occur. These errors include coverage error, sampling error, measurement error and nonresponse error [7].

In our survey we tried to reduce these errors by selecting the sample at random from the population, in this case the students at Komvux. To reduce the sampling error, data were collected from different classes at Komvux. To reduce the measurement errors we made a pilot study to get

¹ Palmius, J, Mod_survey: <http://gathering.itm.mh.se>

² SPSS, <http://www.spss.com>

feedback on the questions before handing out the final questionnaires.

L. Response rates

Our sample consisted of a total number of 90 people, who were assigned to answer each method respectively. The response rate was high, 85 out of 90 persons asked participated in our survey. There was one person on the paper questionnaire and two persons respectively on the web questionnaires, who did not participate. One of these persons on WQ1 was disregarded as of error due to lack of computer experience.

M. Validity and reliability

While measuring data it is important to consider both reliability and validity. The definitions we use are: Reliability – Refers to the replicability of the measurement procedure to yield consistent results.

Validity – Refers to the extent of which the measurement procedure actually measures the concept that it is intended to measure [2]. Since all three questionnaires contained the same questions we find our measures valid and that the answers would probably be similar if the survey was repeated.

III. RESULTS AND ANALYSIS OF THE QUESTIONNAIRES

A. General comparison of the methods

The first comparison was made to see if there were any significant differences between the methods. For this we used a Chi-square test.

TABLE I
CHI-SQUARE TEST, GENERAL COMPARISON OF THE METHODS

Chi-square test			
Question	PQ vs WQ1	PQ vs WQ2	WQ1 vs WQ2
Q2 Reason for study	0,108	0,000	0,000
Q3 Experiences	0,072	0,818	0,068
Q4 Learning	0,066	0,033	0,539
Q5 Responsibility	0,083	0,116	0,160
Q6 Feedback	0,032	0,052	0,419
Q7 Formal education	0,027	0,684	0,425
Q8 Future studies	0,538	0,000	0,000
Q9 Satisfying studies	0,549	0,547	0,745
Q10 Student guidance	0,325	0,000	0,000
Q11 Study hours	0,676	0,204	0,750
Q12 Buying materials	0,074	0,293	0,089
Q13 Number of breaks	0,577	0,140	0,691

The test showed that there were several significant differences. As shown in table I some of the questions has a p-value under our significance level of 0.05. The null hypothesis is rejected.

PQ vs. WQ1 and PQ vs. WQ2 showed differences in the attitude-based questions. WQ1 vs. WQ2 on the other hand showed differences only in the fact-based questions. The conclusion is that there are significant differences in the attitude-questions between the PQ and the WQ's that are not present between the two WQ's. Wildner and Conklin

claim that Internet users are not a representative sample of the total population, there can be differences in terms of attitude etcetera [21]. This is not valid for our survey since our groups are identical in terms of sociodemographics.

B. The use of extreme values

The second comparison was made using an ANOVA-test to see if there were any significant differences in the use of extreme values between the different methods.

TABLE II
ONE-WAY ANOVA TEST, COMPARISON OF THE EXTREME VALUES

One-way ANOVA		
Question	df	Sig
Q3 Experiences	2	0,336
Q4 Learning	2	0,009
Q5 Responsibility	2	0,035
Q6 Feedback	2	0,001
Q7 Formal education	2	0,099
Q9 Satisfying studies	2	0,224

The test showed that there were several significant differences between the methods. The null hypothesis is rejected. Table II shows in which attitude questions the differences were found (sig. level under 0,05). Table III shows between which methods there were significant differences.

TABLE III
ONE-WAY ANOVA TEST (POST HOC), COMPARISON OF THE EXTREME VALUES

One-way ANOVA Post hoc			
Question	PQ vs WQ1	PQ vs WQ2	WQ1 vs WQ2
Q3 Experiences	0,484	1,000	0,813
Q4 Learning	0,014	0,039	1,000
Q5 Responsibility	0,052	0,115	1,000
Q6 Feedback	0,004	0,004	1,000
Q7 Formal education	0,097	0,707	1,000
Q9 Satisfying studies	0,262	1,000	0,848

As shown in table III, there are differences between the PQ and the web questionnaires. There are no differences between WQ1 and WQ2. The conclusion is that the respondents avoid the use of extreme values when answering a questionnaire on the web, which is not the case on paper.

This agrees with Humphrey Taylors findings that scales may bring forth different response patterns on the Web, meaning that fewer people pick the extremes on scales [18].

C. Differences between genders

In addition we also made an ANOVA test to see if the differences were due to gender. In this test we found out that there were no significant differences and our null hypothesis is retained. With this we can eliminate the possibility that the differences are due to gender in the use

of extreme values between the methods. This contradicts literature in psychology that states while answering attitude questions men tend to avoid the use of extreme values compared to women [3].

D. Used steps in attitude scales

We then measured the number of used steps of the attitude-questions. For this we used a comparison of the means. We made a count of the number of different steps each respondent used during the survey. This is presented in table IV

TABLE IV
USED STEPS IN EXTREME SCALES

Used steps		
PQ	WQ1	WQ2
3,48	3,32	3,5

As shown in the table there are minor differences between the methods, thus we can assume that the methods are equivalent. The number of steps that the respondents used was between 3 and 4.

IV. RESULTS AND ANALYSIS OF THE EVALUATIONS

The comparisons of the evaluations were made by calculations in percentage. On the basis of the calculations we have accounted for those questions that showed the most important differences between the methods. These differences are shown in table V and further explained under parts A-E.

TABLE V
CALCULATIONS IN PERCENTAGE, THE EVALUATION

Calculations in percentage			
Judgement	PQ	WQ1	WQ2
Truthfulness	76%	79%	100%
Complicated	0%	0%	18%
Number of questions	10%	4%	32%
Uncomplicated questions	66%	86%	86%
Comprehensive answers alternative	17%	50%	46%
Time spent on answering the questions	35%	42%	14%
Introduction text	97%	54%	29%

A. Truthfulness

In the first issue, if the respondents felt like they answered the questions truthfully, the results showed that all of the respondents to the WQ2 felt that they had supplied totally honest answers. This might be due to the design of the questionnaire. When answering the PQ and WQ1, the respondents had the possibility to go back in the questionnaire and check the other questions in order to achieve consistency in the answers. This was not possible in WQ2 and this might lead to more honest answers. The

first choice of answer is often the one most close to the truth according to psychologists [3].

B. Complicated and the number of questions

Nearly 20 percent of those who answered WQ2 experienced it as complicated and a third of the respondents felt that it contained too many questions. The reason for this might be caused by the design. According to Murphy et al, usability problems can arise if the design of the user interface imposes cognitive demands that people are unable or unwilling to meet [14].

C. Uncomplicated questions and Comprehensive answer alternatives

Those who answered the WQ's felt that the questions were easier to understand and that the answer alternatives were more comprehensive than those who answered the PQ.

D. Time spent on answering the questions

When it comes to time spent on answering the questions, the respondents to the WQ2 considered themselves to not having answered the questions hurriedly. This might be caused by the fact that the WQ2 was more complicated to fill in and that the respondents only had one chance to answer the each question.

E. Introduction text

The last question, if the respondents had read the introduction to the questionnaire, showed that almost all of the respondents to the PQ had read and understood the purpose of the survey. This was not the case with the WQ's. Only 50% on WQ1 and 30% on WQ2 had read and understood the introduction text.

The introductory information is a critical part of any questionnaire. The purpose of the information is to present the survey in such way that respondents are motivated to complete it. Respondents do not necessarily read the introductory information as intended [12].

V. RESULTS AND ANALYSIS OF THE SEMINARS AND THE INTERVIEWS

During the seminar and the interviews several suggestions to explain some of the differences between PQ's and WQ's came up.

A. Reading patterns

One of the explanations could be the fact that reading on a screen and on a paper differs quite a lot and that might influence the answers. Some of the people we interviewed said that they usually read through a text on paper but scan a text presented on a screen searching for single words. They often experience the need of scrolling as being frustrating and that it is difficult to calculate the length of the text. One of the participants at the seminar pointed out that reading on a screen could be described as looking at a picture. Most people do not read web pages word by word,

they scan the page for individual words and sentences. Many users read only the first sentence of each paragraph.

Nielsen has suggested that a reason for this might be that it is tiring for the eyes and that it may take somewhat longer time to read from computer screens. Another reason could be that the web is a user-driven medium. Users feel that they have to move on and click on things, they want to feel that they are active on the Internet. Modern day is hectic and people don't have the time, web users are under emotional and time constraints [16].

B. Stress and decision making

Through the interviews we found out that a common opinion is that when using the Internet one is under some kind of stress. According to a study made by Morkes and Nielsen everything that you do on the Internet is supposed to go very fast [22]. We assume this could be a reason for making quick decisions in web-based surveys and thereby cause the phenomena of avoiding extreme-values. Another aspect of why respondents answering the WQ2 avoid extreme values might be due to the fact that some of the people saved up on the use of extreme values in order to save these options for subsequent questions. This could be caused by the fact that the respondents did not have an overview over the questionnaire as a whole.

C. The Design and question order effect

The design of the questionnaire is also of importance of the results from a mixed mode survey. As mentioned on the seminar and by some of the people interviewed questionnaires like WQ2 are favorable when conducting surveys. The lack of possibilities to go back and change the answers in order to achieve consistency makes the answers more honest.

It is widely recognized by survey researchers that the context of survey questions can affect the respondents' answers to them. When a question has influence on the response to a subsequent question or questions, the phenomenon is referred to as a question order effect. Two particular types of question order effects have frequently been discussed; assimilation and contrast effects also referred to as consistency and redundancy effects, respectively. Assimilation effects occur when respondents adjust their reply to a later question so that their responses to a set of questions are consistent. Contrast effects occur when, having answered a question on a particular topic, respondent's discount, or subtract, the information used in making that judgment when answering a subsequent, related question [12]. Another possibility could be that the use of a specific media stimulates to give certain answers.

When it comes to the design of web questionnaires, Don A Dillman [5] points out the importance of presenting the questions in such a way that they can be understood and answered accurately by all respondents. He also states that the likelihood of responding should not be affected by people's lack of computer skills or other factors that are related to the study variables of interest.

D. Computer experience and usability

At the seminar all participants agreed that it is important to make sure that the respondents answering a questionnaire in a WQ-mode, have some kind of computer experience or that there are instructions understood by all respondents.

Allwood underlines the importance of the user having adequate understanding, skills and competence to interact with the computer effectively [1]. In a web questionnaire, respondents are often given information and instructions. They must also understand the directions that are meant to guide them through the questionnaire [12]. When a person has been instructed to fill out a questionnaire, that person is stimulated to recall past experiences associated with completing questionnaires, the essential components of the process, questions and instructions, and the response task, read questions and choose an answer. Filling out a web questionnaire requires that the respondent think simultaneously about how to answer a questionnaire and how to operate the computer. Human Computer Interaction (HCI) plays an important role in order to make it as easy as possible for the respondent to complete the task [5].

VI. CONCLUSIONS AND RECOMMENDATIONS

Our conclusion is that there are significant differences in the answers between the methods. When comparing the use of extreme values we found that the significant differences were between the PQ and the two WQ's. There were fewer significant differences between the PQ and WQ2 than between the PQ and WQ1. In the comparison between WQ1 and WQ2 there were no significant differences at all. There were no differences between genders. Table VI shows if the null hypothesis are rejected or retained.

TABLE VI
NULL HYPOTHESES, REJECTED OR RETAINED

COMPARISONS	NULL HYPOTHESES
Methods in general	Rejected
Extreme values	Rejected
Gender	Retained

A. Recommendations

As mentioned earlier another purpose of this study is to recommend what to consider when using web questionnaires in mixed mode surveys. The following recommendations are based on the data collected from the survey, the questionnaires, the evaluations, the seminar, the interviews or literature. The recommendations are put up to reduce, handle or point out the differences between the methods.

1. The use of questionnaires like WQ2 is to prefer in order to get more unbiased answers since the respondent can not go back and change the answers in the questionnaire. The first choice of answer is often the one closest to the truth according to literature in psychology [3]. This recommendation does not address the differences

between the methods, but we consider it to be of importance.

2. When presenting an introduction to a questionnaire, it is important to consider the differences in reading patterns on paper respectively on the web. Being aware of this may help the respondents to read the information needed to fill in the questionnaire. This is based on the findings in our evaluation and according to Nielsen [22].
3. If the questions are displayed one at a time, it is of importance that the questions are numbered in such a way that the respondents are aware of the advancement in the questionnaire. For example Q1 out of Q13. Several of the respondents to WQ2 experienced that the questionnaire was extensive and researchers such as Dillman suggests using a progress bar when implementing WQ's [5].
4. When using options other than checkboxes and radio buttons, it is important to supply complete instructions on how to operate them. This is to reduce the possibility that respondents with less computer experience not will be able to fulfill the questionnaire. This recommendation is based on previous research [5] and the fact that the respondents on WQ2 found it difficult to fill in the questionnaire.
5. In the use of mixed mode surveys it is to prefer that the respondents themselves have the opportunity to chose in which mode they would like to participate. This may increase the respondent's motivation to answer the questionnaire. This is motivated by the varying computer experience among the respondents.
6. The web questionnaire should be adjusted to the use of both keyboard and mouse navigation. This is to make it easy and convenient for the respondents. This is also important according to Nielsen's aspects of usability [16].
7. In the design of a WQ it is important to pay attention to human limitations and expectations. The user-interface should be familiar to the respondent. Several researchers have been discussing the value of HCI and usability concerns in the design [5], [14], [16].
8. By maintaining strict comparability of question wording across modes, you could minimize mode effects. This is to avoid that the respondents interprets the meaning of the questions differently. The evaluation showed that most of the respondents independently of method thought that the questions were easy to understand.
9. It is important to be attentive to the many concerns that must be addressed in planning and implementing mixed-mode surveys that involve populations with big differences in computer experience. This is important in order to avoid that some of the respondents are unable to fill in the questionnaire correctly. In our survey, one respondent failed to complete WQ1 because of the lack of computer experience.

VII. CONCLUSIVE REMARKS AND REFLECTIONS

The objective of our research was to explain and draw attention to differences in the answers between paper- and web questionnaires and to come up with recommendations what to consider when using web questionnaires in mixed mode surveys. Profound psychological aspects could cause the explanations to the differences between the methods but we have chosen to focus on adjacent reasons to the differences.

The result is a number of recommendations to facilitate the use of web questionnaires. Our recommendations have a general point of view in terms of usability aspects. We believe they can be useful in designing web questionnaires in mixed mode surveys in order to make sure that all respondents can participate in the survey and reduce non-response errors.

It is difficult to respond to the question about which answers that is considered to be true, what the respondent believe for now or what most of the respondents believe (in average). We are not able to answer this question but we know that there are differences due to the answers and we have strived to handle and describe these differences.

We believe that it is important to be aware of the possibility that it might be the paper questionnaire that are an unsatisfactory method for conducting surveys. The reason for not recommending solutions for how to handle the paper questionnaires is that it is an established method in surveying and it is hard to make changes in a paper questionnaire.

Based on our results we think that one should consider the use of mixed mode in surveying. This might not be appropriate for all situations, however, if the situation requires the use of a mixed mode including web questionnaires then it might be helpful to consider our recommendations.

A. Method discussion

We believe that our choice of method has been successful and appropriate for this research project due to the time aspects. However, if we were to conduct this research again we would:

- Include more questions in the questionnaire
- Include a greater mix of questions of both negative and positive values.
- Use open-ended questions in the evaluation.
- Conduct more interviews.
- Calculate the correlation between the questionnaires and the evaluations.
- Validate the results by conducting a second survey in order to see if the results would be the same.

It is important to note some of the limitations of this survey. Our sample was selected at random and do not necessarily represent any given population other than the kind found at Komvux.

B. Further research

Some areas of interest concerning further research are conducting a survey where you combine a traditional WQ with the use of a touch screen. Instead of checking the answers with a mouse or a keyboard you just use the hands. It would also be interesting to log the changes that the respondents make in the questionnaire, both on PQ's and WQ's. This is to see if the respondents go back and change their answer in comparison with subsequent questions. An alternative is to construct the survey with an automatically submit when you have chosen an answer as to see if the respondents read instructions and takes time to think about the answers. Another aspect is to conduct a survey based on personalities. First you make a test to determine personalities, impulsive and inhibited persons, then separate them into different groups and compare them against different modes to see if the use of extreme values have any connection to the personality.

Understanding the basis of the differences between the methods should be a high priority for future research.

VIII. SUMMARY

We have in our study come to the conclusion that there are significant differences between paper and web questionnaires. These differences can be handled with the knowledge of how the respondents read a text presented on a computer screen, and the knowledge about how to design a web questionnaire to make sure that all respondents can answer the survey. It is essential that the respondents have the opportunity to choose in which mode they would like to participate. It is vital to recognize the importance of psychological aspects on how the respondents act when sitting in front of a computer and how they tend to think when filling in a questionnaire. In order to fulfill our purpose, the following research questions were formulated: What difference is there in the answers between paper and web questionnaires? How can these differences be explained? What recommendations can we give? These questions were answered by using statistical tests, the interviews, the seminar and literature.

REFERENCES

- [1] Allwood, Carl Martin, *Människa- datorinteraktion – ett psykologiskt perspektiv*, 1998, Studentlitteratur, ISBN 91-44-00587-3
- [2] Andersen, R & Heath, A, *Research Design & Survey Methods*, 2000, Department of Sociology, University of Oxford
- [3] Atkinson, R L, Atkinson, R C, Smith, E E, Bem, D J, Nolen-Hoeksema, S, *Hilgard's Introduction to Psychology*, Thirteenth edition, 2000, Harcourt College Publisher, ISBN 0-15-508044-x
- [4] Couper, Mick P, Tourangeau, R, Givens, J, Ezzati-Rice, Trena M, (2002) *Opportunities and Challenges in Web Surveys*, Draft paper for ICIS, Copenhagen <http://www.icis.dk/Trena%20M%Ezzati-Rice.pdf>
- [5] Dillman, D, Tortora, R D, Bowker, D, (March 1999) *Principles for Constructing Web Surveys*, Social and Economic Research Centre, Washington State University <http://survey.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf>
- [6] Dillman, D, Bowker, D, (May 2000), *An Experimental Evaluation of Left and Right Oriented Screens for Web Questionnaires*, Washington State University <http://survey.sesrc.wsu.edu/dillman/papers.htm>
- [7] Dillman, D & Bowker, D K, (2001) *The Web Questionnaire Challenge to Survey Methodologists*, Social and Economic Research Center, Washington State University http://survey.sesrc.wsu.edu/dillman/zuma_paper_dillman_bowker.pdf
- [8] Dyer, Colin, *Beginning Research in Psychology*, 1995, Blackwell Publishers Ltd, Oxford, ISBN 0-631-18929-7
- [9] Grisaffe, D, (2001), *Weighing your Options on Web Surveying?* Walker Information <http://www.websm.org/> http://www.walkerinfo.com/resources/white_papers/default.cfm?&start_pass=24
- [10] Howell, David C, *Fundamental Statistics for the Behavioral Sciences*, Fourth edition, 1998, ISBN 0-534-35821-7
- [11] Jeavons, Andrew, (2000), *Ethology and the Web: Observing respondent behavior in Web surveys* <http://w3.one.net/~andrewje/ethology.html>
- [12] Jenkins, Cleo R., and Dillman, Don A. (1995, revised June 2000): *"The Language of Self-Administered Questionnaires as Seen Through the Eyes of Respondents"*. In Volume 3, Statistical Policy Working Paper 23: News Directions in Statistical Methodology U.S. Office of Management and Budget, pp. 470-516.
- [13] Jenkins, C R, Dillman, Don A, (February 1998) *Combining Cognitive and motivational research perspectives for the design of respondent-friendly self-administered questionnaires*, <http://survey.sesrc.wsu.edu/dillman/papers.htm>
- [14] Murphy, E, Marquis, K, Hoffman III, R, Saner, L, Tedesco, H, Harris, C, (2000), *IMPROVING ELECTRONIC DATA COLLECTION AND DISSEMINATION THROUGH USABILITY TESTING*, U.S Census Bureau <http://www.census.gov/srd/papers/pdf/sm00-01.pdf>
- [15] Newbold, Paul, *Statistics for Business and Economics*, Fourth Edition, 1995, Prentice Hall International Editions, ISBN 0-13-185554-9
- [16] Nielsen, Jacob <http://www.useit.com/alertbox/9710a.html>
- [17] NuStats <http://www.nustats.com/mixedmode.htm>
- [18] Taylor, H , (2000), *Does Internet research work?*, International journal of market research, (42) 1:51-63
- [19] Taylor, H & Terhanian, G, (2000), *Heady Days Are Here Again – Online polling is rapidly coming of age*, Louis Harris and Associates, Harris Black International http://www.ropercenter.uconn.edu/pubber/pdf/pp104_b.pdf
- [20] Warfield, John. N. "A Science of Generic Design, Managing Complexity Through Systems Design", Volume 1, (1990), Intersystems Publications, California, USA, ISBN 0-914105-49-1
- [21] Wildner, Raimund and Conklin, Michael (2001): "A statisticians view- The importance of sampling in online research". Paper presented at WORLDWIDE INTERNET CONFERENCE AND EXHIBITION Net Effects 4, Barcelona, Spain, 11 - 13 February 2001
- [22] J, Nielsen, J, *Concise, SCANNABLE, and Objective: How to Write for the Web*, 1997 <http://www.useit.com/papers/webwriting/writing.htm>