



Two Experiments with Likert Scale

Author: Wang Lei
Mohammed Azizur Rahaman

Supervisor: Mikael Möller

C-level in Statistics, Fall 2009

School of Technology & Business Studies

Högskolan Dalarna, Sweden

Table of Contents

Abstract	2
1. Introduction	3
2. Background	3
2.1 Likert Scale and its shortcomings	3
2.2 Narrative ways of items	4
2.3 Experiment objective	5
3. Experiment design and implementation	6
3.1 Experiment design	6
3.2 Sampling design and implementation	7
4. Survey results and Data analysis	8
4.1 Experiment A	8
4.2 Experiment B	10
5. Discussion	15
Reference	16
Appendix	17

Abstract

For the traditional Likert Scale's shortcomings of low level differentiation, this paper is an exploration to improve the measurement level. Design two experiments, apply the method of questionnaire survey, collect data based on *the Library Satisfaction Survey*, and analysis the data by comparison test.

From the analysis of experiment results, Experiment A shows that referring to the ways of statements of items, different levels of impact on Respondents' attitudes vary in different situations; Experiment B shows that, compared to the traditional Likert Scale, the improved GE Scale has an advantage in reflecting respondents' satisfactory levels.

Key words: contrast experimentation, Likert Scale, attitude measurement, questionnaire survey

1. Introduction

Social and psychological sciences are not so easily quantified as natural sciences. Hence, the developing speed of those has not been as fast as for the natural sciences.

To investigate mental activities, such as emotions, opinions and attitudes of human beings, the first problem is to find a technique to measure such phenomenon, in order to acquire the basic data for our research. However, to quantify mental phenomenon is not so easy as objective measurements. That is why researching on social and psychological issues stagnated in the stage of qualitative analysis for a long time. The conclusions were difficult to popularize.

Due to the great efforts of psychologists over hundreds of years, there are many methods, which can make comparatively accurate, to measure complex and barely observable mental phenomenon. These methods play a very important role for the data collecting in research of mental phenomenon. In social surveys, especially the satisfaction survey, attitude measurement is usually involved.

The data of satisfaction surveys mostly comes from questionnaires. The quality of the questionnaire design has a direct impact on the evaluation of the results.

However, improvement needed to be made to current research techniques. In this essay, we will discuss the Likert Scale method, which is widely used in attitude evaluating, about its shortcomings and improvement.

2. Background

2.1 Likert Scale and its shortcomings

Likert Scale, which is also called Summated Rating Scale, is named after an American social psychologist, R.A. Likert, who invented this particular scale in 1932. The Likert Scale is the most common form of attitude scale in the field of social surveys and psychological experiments. The body of Likert Scale consists of two parts. One is called items, which are a set of questions or statements about the subject. Each item is assumed to indicate the respondents' attitudes, opinions, evaluations or intentions. The other part is respondent levels. This part has five or more options, where respondents are required to choose one option to stand for their attitudes. Usually, respondent levels are symmetric distributing between a neutral option. For example, a 7-point Likert Scales with grades have following elements: "strongly agree" = 7, "basically agree" = 6, "agree"=5, "Neither agree nor disagree" = 4, "disagree" = 3, "basically disagree" = 2, "strongly disagree" = 1. And the way of grading depends on you.

In practical applications, the 5-point form is commonly used, which gives the respondents five choices of agreement level relative to each item and use the numbers 1-5 to grade the answer. There are other kinds of Likert Scale, such as 3-point, 7-point, even 9-point form in applications. Besides, even-number-point form is also

used in many cases, where the neutral option “Neither agree nor disagree” is omitted. There are arguments all along about how many respondent levels are suitable. But the answer can not be universal.

Add up each grade of items in the scale, then the total score reflects the general attitude of a respondent to some object or subject. The higher is the score, the more positive of a respondent towards the topic. In social surveys, the average attitude of respondents is more concerned. Therefore, the mean value of the total score is used to show the average intentions toward some topic.

However, the Likert Scale’s coverage of attitudes is limited, especially in the positive side. When taking consumer satisfaction surveys, traditional Likert Scale which is always symmetric, does not work so efficiently in comparing two services of which are both well qualified. Therefore, an improved form was created by General Electronic(GE), which aims to increase the differential level of consumers. The GE scale’s respondent levels are non-symmetric. Followings are the translations of each level:

Happy means that “the services I received are better than expected”

Completely satisfied means that “everything is just as good as I expected”

Almost satisfied means that “almost everything is as good as I expected”

Satisfied means that “most of the things comply with my expectation”

Unsatisfied means that “my expectation is not satisfied”

To compare the non-symmetric form with the traditional one, an experiment is to be taken to find out which form more efficient to evaluate the level of consumers’ attitude.

2.2 Narrative ways of items

Another point is how the items should be translated, in positive or negative way. Are there any differences of the impact on respondent answers between those two ways? In the development of questionnaire, several experiments are so famous that you can’t miss them. One of them is a poll after the German invasion of the Soviet Union, by Gallup in 1941. The respondents were randomly divided into two parts, and asked the following questions: right half of the survey respondents to ask question A: "It is said that besides the UK, Germany is also fighting with the Soviet Union, so we do not have to enter the war to help the United Kingdom, do you agree or not agree with."; on the other half of the respondents to ask question B: "It is said that Germany could defeat the Soviet Union in a few weeks, and then back to fully deal with the British, therefore, our need are more urgent than ever before to help the British, do you agree or disagree with."

The result showed that, on question A, 73% said they agreed with the point of view, while, on question B, 71% agreed. Obviously, in the test on this issue, the answers to this question are basically the same.

But there are also examples to the contrary. Another experiment by Roger (Rugg) in 1941, conducted in this regard is the oft-quoted example. Question A was: "Do you think the U. S. should prohibit in public statements against democracy", the

survey findings, 54% of people answered in the positive (Yes); question B was: "Do you think the U. S. should permit public statements against democracy", 75% of the people replied in the negative (No). This experiment reached the opposite conclusion: although both groups didn't agree with the United States openly opposed to democracy speech, but its rates were 54% and 75%, and the difference was very significant.

2.3 Experiment objective

In this paper, for the limitations of the traditional Likert Scale and the uncertainty impact of positive and negative ways on respondents, two experiments are designed to evaluate those differences. Take *the School Library Satisfaction Survey* as a research vehicle, design a sample survey of two questionnaires, use the comparative method to analysis the final findings on the two questionnaires.

3. Experiment design and implementation

3.1 Experiment Design

Experiment A:

The items are stated either in an positive or negative way, is there any significant impact on the outcomes? Select facilities of the library to construct items in Experiment A. Taking into account of the reader's attitudes to these more objective aspects, not subjective factors, the differences between individuals can be ignored. We can try to reduce the other factors' impact and take a good control of variables. Each of the questions 1 to 5 is a statement of some service of the library, and they are set contrastively in Questionnaire I and Questionnaire II, where the items in Questionnaire I are stated in positive way, while the items in Questionnaire II in negative way. Be noted that the analysis of question 6 is not included in Experiment A.

When making the analysis, for Questionnaire I, each of the results give the score as follows:

"Strongly agree" = 5

"Agree" = 4

"Neither agree nor disagree" = 3

"Disagree" = 2

"Strongly disagree" = 1

In Questionnaire II, give the score in a contrary way, for the statements are completely in the opposite way:

"Strongly agree" = 1

"Agree" = 2

"Neither agree nor disagree " = 3

"Disagree" = 4

"Strongly disagree" = 5

Experiment B:

Comparison of scoring with the traditional Likert Scale and the improved GE Scale, Is there a significant difference between the results?

Compared with Experiment A, Experiment B measures the level of respondents' satisfaction about the facilities and services of the library. In both questionnaires, Question 7 to 14 are the main body of the test. The most important point of Experiment B is that the options of scales in two questionnaires are different:

In Questionnaire I, the alternative options are of traditional symmetric Likert Scale:

"Very satisfied"=5

"Happy"=4

"Hard to say"=3

"Not satisfied"=2

"Very dissatisfied"=1

In Questionnaire II, the alternative options are of GE Sclae:

“Very satisfied”=5
“Happy”=4
“Basically satisfied”=3
“Hard to say”=2
“Not satisfied”=1

3.2 Sampling design and implementation

The target population of the Library Satisfaction Survey is all visitors to the library. We divide them into two groups, regarding to our research aim: Group A refers to visitors who read books or do self-study in the library, and Group B refers to visitors who take books out of the library to read.

Since an investigation of group B is more costly we restrict the investigation of group A. On the other hand, the two groups are not strictly separated. Therefore, we choose Group A to be the target population of the survey. These visitors are more familiar with the library's hardware and software facilities, than those of Group B. Our study does not involve the respondents' academic backgrounds, reading preferences and other factors, the library's visitors can be seen as undifferentiated.

After defining the target population we choose the respondents. Since our survey did not involve the analysis of individual differences in reading, different reading rooms are not specified in the survey, The questionnaires were distributed evenly.

Hence our survey program becomes the number of each questionnaires are approximately 50. For each site, the first, second, third reading rooms, magazines and periodicals room, and self-study room the corridor we sample 10 for questionnaire II.

Finally, the total amount of valid returned questionnaire were 116, including 58 copies of Questionnaire I, and 58 copies of Questionnaire II.

4. Survey results and data analysis

4.1 Experiment A:

The experiment was designed to compare the difference between positive and negative way on respondents' answers. According to our assumptions, if the respondents' attitudes were affected by how the possible answers are stated, then, the scores would have significant difference between the same items in Questionnaire I and Questionnaire II. Based on the scores items-, we found:

The mean and the sum value of each item's score of two questionnaires, as well as the difference between the mean of the corresponding items of two questionnaire (ie, the column "mean difference" in the Appendix 1). Roughly speaking, of these six items, the 6th question has the smallest mean difference (0.04), this result is more in line with our expectations. For in two questionnaires, the 6th question's narrative ways are completely consistent, if our assumption - the individual respondents are non-discrimination ones, then the statistical results of two questionnaires obtained should be close.

On the contrary, if we find asinificant difference between a pair of questions from questionnaire I and II then we may say that the way you express your statements are important. Mark the first question's mean of Questionnaire I as μ_{11} , the first question's

mean of Questionnaire II as μ_{21} , and so on. We make 5 tests of the following type:

$$H_0: \mu_{1i} - \mu_{2i} = 0$$

$$H_1: \mu_{1i} - \mu_{2i} \neq 0$$

Do the t-test of two independent samples, and the results as blow:

Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
A1	Equal variance assumed	7.734	.006	4.220	114	.000	.74138	.17567
	Equal variance not assumed			4.220	109.002	.000	.74138	.17567

Table 1 The results of Hypothesis-test to Question 1¹

Firstly, in the F-test of two samples, where F-test is to examine whether the variances of two samples are equal or not, the observation value of F statistics is 7.734, and the p-value is 0.006, that is to say, there are significant differences between

¹ Set A1 as the results of Question 1 in Experiment A. The same as A2, A3 and so on in the following tables.

the variances of the two samples.

Secondly, in the test of two sample means, since the F-test tells that the variances of two samples are not equal, the second row-Equal variance not assumed-shows t-test results: the observation value of statistics t is 4.220, corresponding two-tale p-value is 0.000. If the significant level of the test is 0.05, for p is less than 0.05, we can conclude that two sample means have significant differences. That is to say, there are significant differences between the results of Question 1 of two questionnaires.

The narrative ways of items have impact on the respondents' answers. From Table 1, on Question 1, the score obtained in positive way is higher than the score in negative way.

Below are the results of Question 2 to 5:

Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
A2	Equal variances assumed	.002	.964	-2.135	114	.035	-.37931	.17769
	Equal variances not assumed			-2.135	113.674	.035	-.37931	.17769

Table 2 The results of Hypothesis-test to Question 2

Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
A3	Equal variances assumed	8.220	.005	-2.900	114	.004	-.44828	.15458
	Equal variances not assumed			-2.900	108.456	.005	-.44828	.15458

Table 3 The results of Hypothesis-test to Question 3

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
A4	Equal variances assumed	.113	.737	1.252	114	.213	.24138	.19278
	Equal variances not assumed			1.252	113.649	.213	.24138	.19278

Table 4 The results of Hypothesis-test to Question 4

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
A5	Equal variance assumed	.407	.525	.435	114	.665	.08621	.19836
	Equal variance not assumed			.435	113.740	.665	.08621	.19836

Table 5 The results of Hypothesis-test to Question 5

From the statistical results, the p-value of Question 2 and 3 are both less than 0.05, which means that the null hypothesis can be rejected, the means of two samples have significant differences. But in the test results of Question 4 and 5, the p-value are 0.213 and 0.665 respectively, which are larger than 0.005, means that the null hypothesis can not be rejected, and there is not enough evidence to prove that two sample means haven't significant difference.

To sum up, in Experiment A, of three questions to do two-sample mean test, the result is to reject the null hypothesis, that is there are significant differences in the means; the same time, two questions results can not reject the null hypothesis, can not support the assumptions of our experiment.

The test results above show that the attitudes of the respondents whether are effected by the narrative ways can't be concluded in general, for on some issues such impact is more obvious, while on some other problems the impact is not.

4.2 Experiment B

The aim of test is to examine whether there are differences in attitude measurement between the improved GE Scale and the traditional Likert Scale, and whether detail breakdown of the improved one has advantage to reflect customers' attitude. Analysis method of Experiment B is similar to Experiment A to carry out two

independent sample t-test. However, in Experiment B, the scores to be compared are the sum of 8 items of every respondent instead of score of each item. That is to say, examine respondents' overall attitudes to Question 7-14. Compared to Experiment A, Experiment B is a vertical comparison.

The table blow shows the data obtained of the two questionnaires:

Descriptive Statistics					
	Mean	Mean	StDev	Minimum	Maximum
Questionnaire I	28.552	0.521	3.966	20.000	39.000
Questionnaire II	24.155	0.566	4.312	12.000	34.000

Table 6 Descriptive Statistics: Questionnaire I, Questionnaire II

From Table 6, the mean value of Questionnaire I is greater than 24 ($3 \times 8 = 24$), illustrate this part of the survey, the library's overall service is quite satisfied; if the mean value of Questionnaire II was greater than 24, illustrated the library readers were satisfied. Questionnaire I has greater mean value than Questionnaire II, that means Questionnaire II have a more detail distinction of the satisfaction level.

As can be seen from Table 6, for Question 7-14, the respondent's subjective experience is quite satisfactory. Among them, the minimum score of each question's mean value is 2.5, the maximum is 4.88, the average is 3.58, indicating the majority of respondents to the evaluation of the library's satisfaction level lie between "says no" and "very satisfied"; by the standard deviation of 0.4887, only a few respondents show "not satisfied" tendency. To confirm how much the level of satisfaction much is, the traditional Likert Scale does not give good answers. Hence, based on Questionnaire I, introduce Questionnaire II to widen the scope of satisfaction level, allowing us to clearly reflect the respondents satisfactory to the library. The following table shows the data distribution of scores obtained by the GE improved Scale.

From Table 6, in the improved questionnaire, the minimum score of respondents on the same issues is 1.5, maximum 4.25, average at 3.06, with standard deviation of 0.518. That means the majority of respondents scoring is at 3, which shows that most readers held a "basically satisfied" attitude to the library, very few held "very dissatisfied" attitude, and not too many held "very satisfied" or "very satisfied".

The following are the histograms of summary score of Questionnaire I and II:

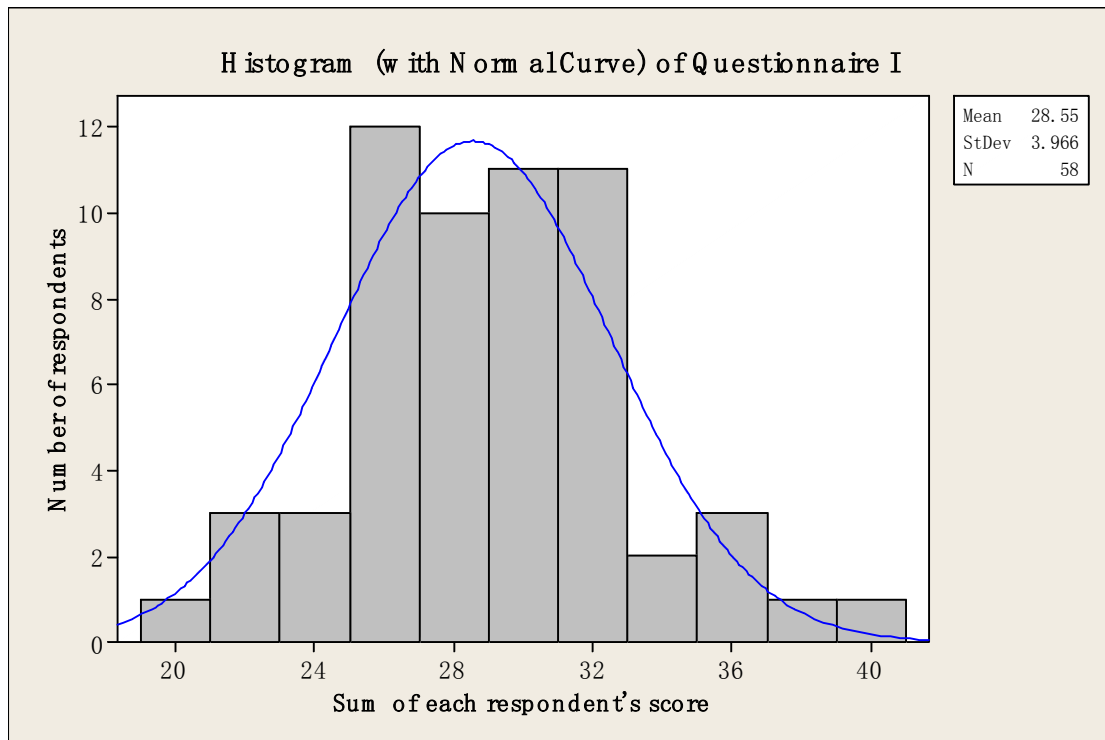


Figure 1 Sum Score Distribution of Questionnaire I

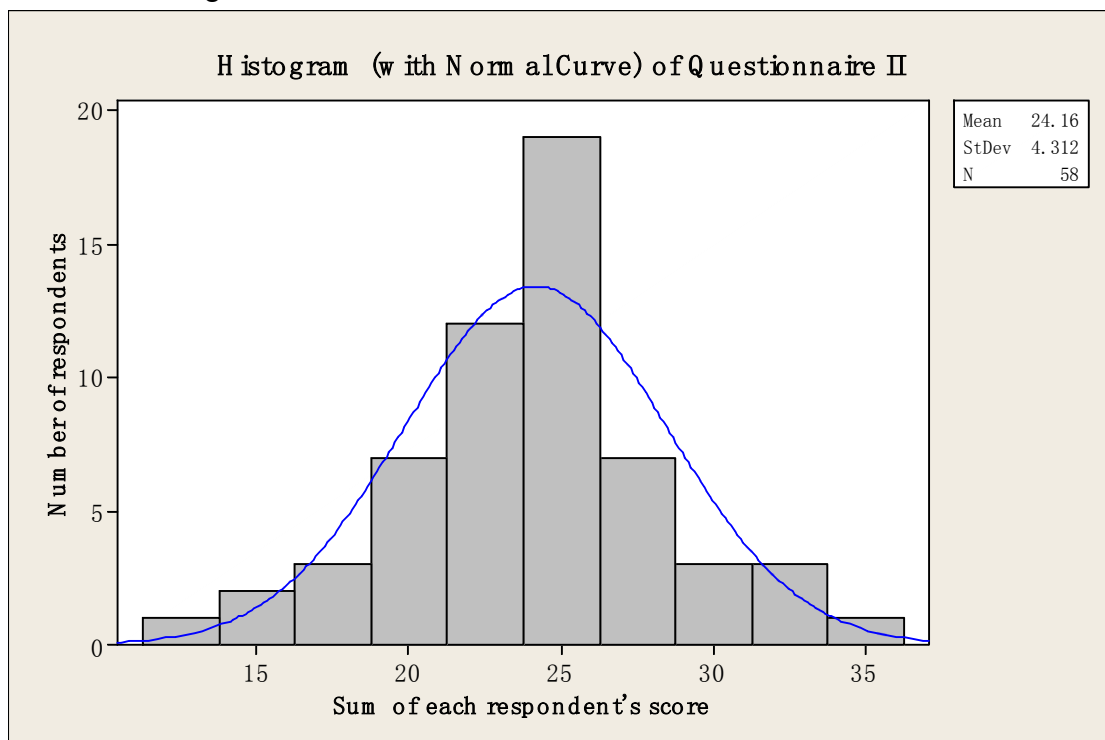


Figure 2 Sum Score Distribution of Questionnaire II

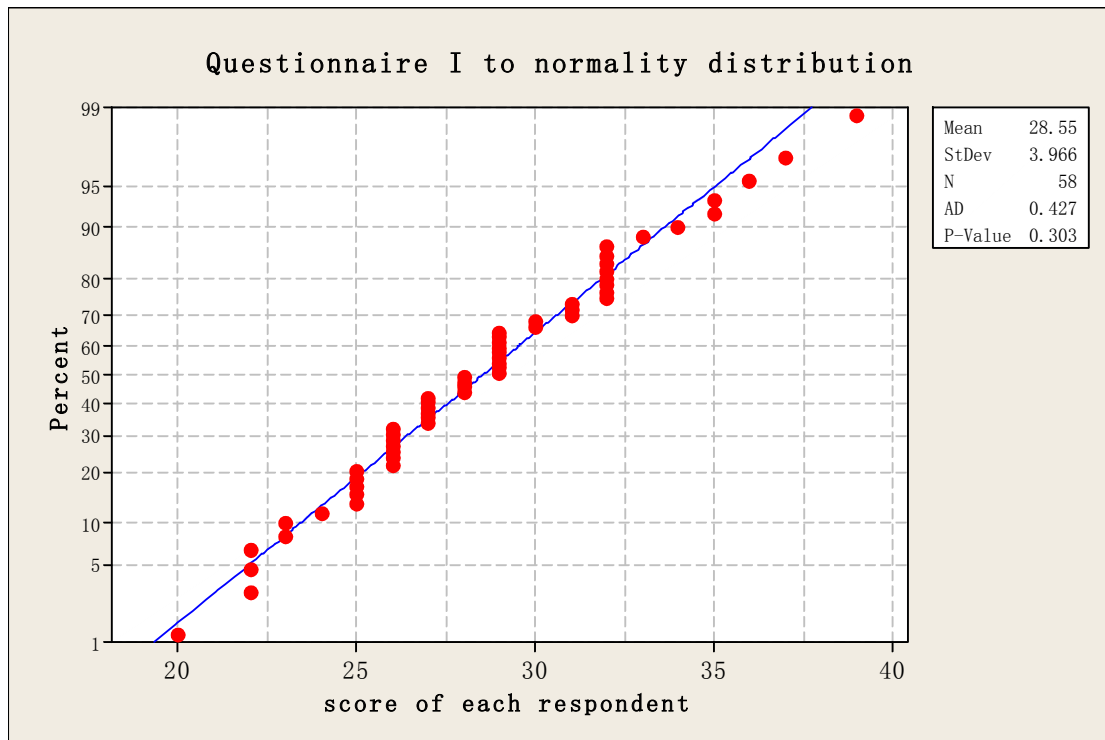


Figure 3 Q-Q Plot of Scores of Questionnaire I

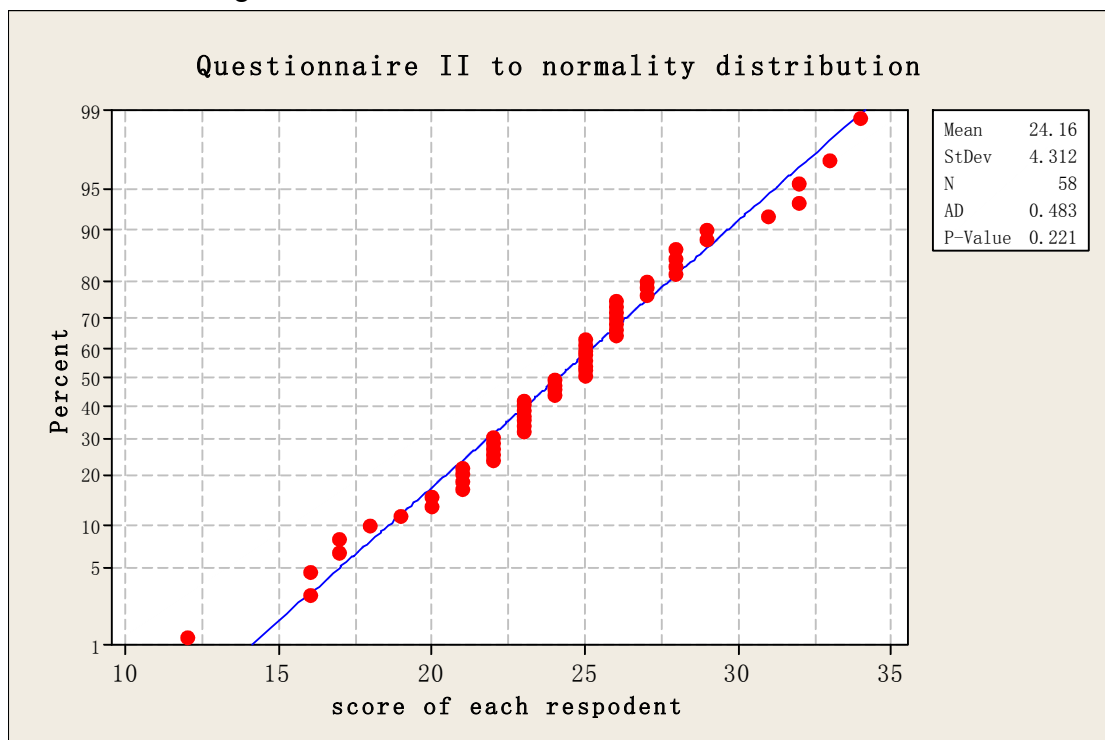


Figure 4 Q-Q Plot of Scores of Questionnaire II

From the data distribution, the Questionnaire II data is closer to normal, generally distributed in both sides of 24; while Questionnaire I is more scattered, shows that the division of attitudes is not enough detail, hence, it restricted the respondent's judgments, and led to blindness and randomness of choices, which can not form a concentrated point of view.

Comparison of statistical results of the two questionnaires shows that, although respondents were satisfied with the library in general, actually, they are not so highly satisfied reflected in the results of Questionnaire II compared with of Questionnaire I. The level of satisfactory only reached "satisfied", which is lower than that of Questionnaire I. As a result, after a more detail breakdown in the level of satisfaction in Questionnaire II, the reflection of respondent's attitude is closer to the truth.

5. Discussion

For Experiment A, the results indicate that on certain questions, the way to ask whether positive or negative, indeed affects the attitude of the respondents, While with some other questions, this effect is not so apparent. It is related to the tendency of a question itself already formed in people's eyes. For those issues which have already formed a relatively consistent positive or negative attitude, respondents will be less affected by the question is stated. While for those which are disagreed in different people's opinions, they would be more vulnerable to the impact of narrative way. It can be easily understood from a psychological point of view. Therefore, the research has to be specialized.

For Experiment B, the GE improved scale actually has obtained more refined grades. Compared with the traditional Likert Scale, the difference of the results is quite significant. However there are certain risks, for choosing the library as an example of the experiment, For it is not generally known whether the library already or not. The aim of the research is to find a better measurement of satisfaction between different agencies which have already provided qualified services, and take a further differentiation of satisfaction level of their clients. There may be some misfit between the target body and the library we chose and more research need to be set.

Reference

- Xiaotian. Feng., 2002. *Questionnaire Design in Social Survey*[M]. Tianjin People Press.
- Robert F. DeVellis., 2003. *Scale Development: Theory and Application, Second Edition*. SAGE Publications, Inc.
- Yu. Liu., 2003. *Measurement of Consumers Satisfaction*. Society Science Document Publishing House.
- Yongjin. Jin., 1995. *Non-sampling Error Analysis*. China Statistics Press.
- Laibin. Qi., 2006. Likert Scale's Statistical Analysis and Fuzzy Synthetic Decision. *Shangdong Science*, April, 2006.
- Yuzhi. Han., 2008. *Modern University Management: Take the American University Student Satisfaction Survey for example*. Zhejiang University Press.

Appendix

Table 1 Summary of Experiment A results

Questionnaire I		μ_{1i}	$n\mu_{1i}$	$\mu_{1i} - \mu_{2i}$	$n\mu_{2i}$	μ_{2i}	Questionnaire II	
1	The variety of books in the library can meet my requirement of reading.	3.78	219	0.75	176	3.03	The variety of books in the library can't meet my requirement of reading.	1
2	The books are so varied that I can find what I want in shelves.	2.98	173	-0.38	195	3.36	The books are so limited that I can't find what I want in shelves often.	2
3	The speed of refreshing new books on the shelf is fast.	3.08	179	-0.45	205	3.53	The speed of refreshing new books on the shelf is slow.	3
4	The results of e-library are accurate, and the books in the results can be found.	3.29	191	0.24	177	3.05	The results of e-library are not accurate, and the books in the results can't be found.	4
5	The category and numbering of the books is scientific, and helpful for looking for books.	3.12	181	0.09	176	3.03	The category and numbering of the books is not scientific, and against for looking for books.	5
6	Some books have only several copies, as a result, I always wait a long time to get it.	3.78	219	0.04	217	3.74	Some books have only several copies, as a result, I always wait a long time to get it.	6

Table 2 Contrast of two questionnaires in Experiment B

Questionnaire I	Traditional symmetric Likert Scale				
	Very satisfied	Happy	Hard to say	Not satisfied	Very dissatisfied
7 Temperature of the library					
8 The lightness of study room					
9 The seat number in each reading room					
10 The attitude of officers					
11 Prolong opening time to Sunday					
12 Restrictions on juniors in Old Collection Room					
13 Allowance of taking with school bags					
14 Same place of borrowing and returning books					
Questionnaire II	Very satisfied	Happy	Basically satisfied	Hard to say	Not satisfied
	Non-symmetric GE scale				