Stated preferences by mail to evaluate public transport passengers' preferences

Sjöstrand, Helena

1999

Paper 1

Stated Preferences by mail to evaluate public transport passengers' preferences

Helena Sjöstrand (1999)

Stated Preferences by mail to evaluate public transport passengers' preferences

Helena Sjöstrand
Lund University, Department of Technology and Society, Division of Traffic Planning

1 Problem

The Stated Preferences method is now an accepted and widely used tool for assessing peoples' valuations. Most surveys have been made as home interviews or hall tests because of the complex nature of the questions. Especially home interviews are expensive and time consuming to make.

To get more detailed information about different groups of bus passengers' preferences many interviews have to be done. This knowledge can be valuable to public transport planners to optimise the bus system or make forecasts.

Some years ago we thought we could let people consider a lot of SP-alternatives and were happy to get a lot of information from every respondent. Research has now showed that the quality of the answers decreases when the respondent gets tired as he does if he have to make many choices, ratings or rankings (see e.g. Widlert, 1994).

Besides that, research on the "repeated measurement problem" tells us that the standard errors of the estimated parameters are more underestimated the more alternatives every respondent has judged (see e.g. Bates and Terzis, 1997). It's therefore better to spread the total number of choices; ratings or rankings over a greater number of respondents than letting fewer respondents make more choices, ratings or rankings.

These three issues speak for preferring a mail survey to home interviews or hall tests if the survey should not be too expensive.

In Stated Preference games some of the attributes need to be customised if the interviewee shall be able to find the alternatives realistic (see e.g. Widlert, 1994). The customisation is very easy to make when computers are used in the interviews. It's also possible to give the respondents different cards or questionnaires as long as personnel are present during the interview. But if the survey is made by mail the customisation must be solved in another way.

2 Aim

This study aimed to see how a Stated Preference-study by mail, adapted to the respondents' conditions, can be carried out. Six different Stated Preference questionnaire types were created to see if the game type had any influence on the response rate and the answers' quality.

At the same time the study gave bus passengers' valuations of a number of bus attributes such as in-vehicle time, walking time, waiting time, bus interchange, information, seating place and low floor buses.

3 Methodology

The study was carried out in the city of Jönköping in Sweden. This is only one of a number of studies our department has conducted in Jönköping to evaluate the city's new bus system, which started 1996.

3.1 Collection of data

As the study concerned bus passengers' preferences the first task was to find the bus passengers to get their addresses and to make it possible to adapt the SP games. About 420 bus passengers were contacted at bus stops while waiting for the bus. They were asked for their name, address and in-vehicle time this time. Travel time on bus was the only attribute chosen to be adapted while other factors weren't assumed to be that individual.

In addition to noting name, address and travelling time on bus the person's age was also estimated. The estimation was rough. People were divided into three groups, young people, middle-aged people and over 65 years old. According to other studies in Jönköping (Johansson and Svensson, 1998), about 10% of the bus passengers are over 65. The selection of bus passengers to this study was made so that 10% should be over 65, that is 7 persons per questionnaire type.

The letters, containing the SP experiments and other questions, were sent out as soon as possible after the contact interview. The reason for this was that the SP questions dealt with the special bus trip, when the person was contacted.

3.2 Creating six different forms

The types of SP that should be tested were choices, rating and ranking. Of each of these three main SP types were two different questionnaires created. All of them should assess valuations of the same attributes. The attributes were

- ticket price: all passengers judged cash paid one way tickets even if they held month tickets or discount tickets.
- in-vehicle time: the passengers were divided into four travel time groups. The shorter the real bus trip, the shorter the bus trip in the SP game, according to the table:

Table 1 Levels of in-vehicle time depending on stated in-vehicle time.

In contact interview stated time	In-vehicle time levels in SP game		
	fast trip	medium trip	slow trip
-12 minutes	8	10	13
13-17	12	15	19
18-27	18	22	28
28-	27	34	43

- walking time to bus stop
- bus frequency
- bus interchange and waiting time while changing buses
- real time information on bus stop
- availability of seating place on bus
- low floor bus

Because of the large number of attributes, the attributes were combined into two SP games, each of them including ticket price. Each alternative then included four attributes, each of them having 2-3 levels. Half of the contacted passengers got seating place as the fourth factor in the second game the other half got the factor low floor bus. This concerned all the six questionnaire types.

Table 2 List of tested SP questionnaires.

type no	main SP type	description
type 1	binary choices	Six binary choices presented on both sides of one paper for each game.
type 2	binary choices	Six binary choices per game, each choice on a separate card.
type 3	rating	Rating of six alternatives per game on a scale between the worst and the best possible alternatives. The alternatives printed on both sides of a paper.
type 4	rating	Rating of six alternatives per game on a scale between 0 and 100, where 0 means “very bad” and 100 mean “very good”. All alternatives printed on one side of a paper.
type 5	ranking	Each of the two games consisted of six alternatives printed on separate labels that should be pasted in preferred order in six empty squares on another paper.
type 6	ranking	Each alternative was printed on a separate card. The six cards in each game should be sorted with the one most preferred on the top and the least preferred on the bottom with paper-clips on and then sent back in an envelope.

All 420 letters sent out included in addition to the two SP games two pages of background questions, pictures illustrating “real time information” and “entrance in low floor bus” and careful instructions to the SP games.

Two reminders were sent out to people that had not yet sent in their answers after two weeks and after four weeks, respectively.

Observations from all six types of questionnaires are analysed in computer package Alogit.

4 Results

4.1 Quantitative tests

To compare the six different SP questionnaires several aspects of their performance were tested. The tests with results are described here.

Response rate

The response rate is the first test to see if the inquiry was interesting at all and possible to send in. If the response rate is high the answers represent the population in a better way than if only a few people have answered.

About 70 questionnaires were sent out of each of the 6 types. The number of 70 was chosen because it was assumed that at least 50 responses per category were needed in the analysis. Since the response rate had been high, around 80%, in previous studies in Jönköping (Johansson and Svensson, 1998), 70 sent-out forms were hoped to be enough.

Only type 3 gave less than 50 responses. Type 1 had the highest response rate (90%) and type 3 the lowest (71%). Still 71% are a rather high response rate. The high rates probably depend on the personal contact interview. People not willing to participate have already refused in the earlier stage on the bus stop.

Table 3 Response rate.

Main type	Type	Response rate
Binary choices	1 (all choices on 1 paper)	90%
	2 (choices on separate cards)	85%
Rating	3 (between the best and the worst)	71%
	4 (between 0 and 100)	74%
Ranking	5 (labels to paste)	75%
	6 (sort cards)	88%

Useful observations

Even if the questionnaire was sent in the SP task was not always solved in the right way. All choices are not made, there are not markings on every scale or all ranking cards are not sent in. This can be due to misunderstanding or fatigue.

A complete binary choice game gives 6 observations per game. Both a rating and a ranking game give 5 observations to be treated by for instance Alogit. Therefore the number of observations are weighted here, to admit comparison.

Table 4 Useful observations rate depending on questionnaire type.

		useful observations	not useful observations	non-response
type1	choices on one paper	86%	4%	10%
type2	choices on cards	82%	3%	15%
type3	rating best-worst	59%	12%	29%
type4	rating 0-100	65%	9%	26%
type5	ranking of labels	74%	1%	26%
type6	ranking of cards	86%	3%	12%

Table 4 shows that type 3 and type 4 loose still more answers when rates of useful observations are compared.

Share of lexicographical answers

If the SP task is too difficult, the respondent tries to simplify it. One way to do this is to sort the alternatives according to only one factor. This is called lexicography. Lexicographical answers are not always wrong (see e.g. Widlert, 1992). One parameter, for example the cost, could be extremely important to the respondent. The design of the game could be unbalanced so that one factor dominates the others. The alternatives could be lexicographically sorted by random.

Share of lexicography is still to be tested in a quality test because the factor that respondents have sorted according to will be overestimated in the analysis, while the other ones will be underestimated.

The rating games had the lowest share of lexicography, while ranking had the highest rate in the first game and rating had the highest in the second game.

Table 5 Share of lexicographical answers by factor in the first game.

Game 1	cost	in vehicle time	walking time	bus frequency	share of lex	no of games
type1 choices	7%	2%	15%	12%	35%	60
type2 choices on cards	10%	3%	18%	8%	39%	61
type3 rating best-worst	0%	2%	5%	0%	7%	41
type4 rating 0-100	2%	2%	2%	7%	14%	44
type5 ranking of labels	6%	4%	14%	29%	53%	51
type6 ranking of cards	10%	3%	11%	18%	43%	61

The factors most sorted after are walking time to bus stop and bus frequency. Since the sorting is spread over all factors the lexicography is no big problem in the analysis (Widlert, 1992).

In the second game this is different, because almost all lexicography lies on bus interchange. This factor will be overestimated in the analysis.

Table 6 Share of lexicographical answers by factor in the second game.

Game 2	cost	bus interchange	info	seat	floor	share of lex	no of games
type1 choices	3%	37%	2%	7%	0%	48%	60
type2 choices on cards	5%	43%	7%	5%	0%	59%	61
type3 rating best-worst	0%	5%	3%	8%	3%	18%	39
type4 rating 0-100	0%	9%	16%	16%	2%	44%	43
type5 ranking of labels	2%	27%	2%	8%	0%	39%	51
type6 ranking of cards	0%	30%	10%	7%	2%	48%	61

Scale factors

The relative size of a model's estimated parameter values can give information about the model's scale factor (Brundell-Freij, 1995). Comparing all models' scale factors is one way to measure which of the models that best fulfil the demands of being explained by the predictors included in the model. If the choices are not explained by the included pre-

dictors the choices are made more affected by some other not included factor. The higher the scale factor, the more are the decisions made influenced by the predictors in the model. When RP-data is used to make models, the scale factors are normally smaller, because peoples' choices are based upon factors that aren't presented in the model.

The sizes of the estimated parameters in each model from the Alogit-analyses have been compared. To make it possible to compare each questionnaire's three models' all parameters at the same time an average scale factor for each type was calculated.

Table 7 Scale factors, average over all parameters in each type.

Type	scale factor
1 (all choices on 1 paper)	0.69
2 (choices on cards)	0.53
3 (rating best - worst)	0.22
4 (rating 0 – 100)	0.60
5 (ranking of labels)	0.53
6 (ranking of cards)	0.36

Types 1 and 4 have the highest scale factors as a result of having the largest parameter values. This means that these two types give models in which peoples' choices are made due to the factors included in the models. Models from types 3 and 6 give models where choices aren't based on the presented attributes in the same range as in the other models.

Are the valuations reasonable?

Valuations of each of the eight assessed standard factors are calculated as the factors' estimated parameter divided by the cost's estimated parameter. In table 8 values of time of different parts of the bus trip are shown estimated from each SP type.

Table 8 Average value of time, SEK per hour.

	in-vehicle time	walking time to bus stop	bus frequency time	waiting time by bus interchange (s)	waiting time by bus interchange (f)
type 1	17	31	22	41	35
type 2	16	28	20	42	52
type 3	26	21	23	35	61
type 4	23	20	23	29	40
type 5	43	46	77	35	30
type 6	34	0	27	43	53

Value of waiting time by bus interchange is estimated in two ways for each questionnaire type. This is because this factor is estimated together with seating place (s) on half of the questionnaires and together with low floor bus (f) on half of the questionnaires.

As expected all valuations are positive meaning that longer time is less comfortable. The only exception is walking time to bus stop in type 6, where the value is 0.

If you want to compare the bus frequency time with the more often used waiting time, the bus frequency time should be multiplied by 2, if you assume the average waiting time to be half the bus interval.

The binary choice games give the most reasonable values compared to other studies. But even the rating and ranking games, with some exceptions, give values not far from that.

In the second game the comfort factors are judged. Bus interchange and real time information are valued twice, as their values are estimated both in the seating place game and in the low floor bus game. Therefore these factors have two values each in table 9.

Table 9 Estimated values, SEK per trip.

	bus interchange (s)	bus interchange (f)	information (s)	information (f)	seating place	low floor bus
type 1	4.97	3.08	2.82	0.80	5.79	1.72
type 2	2.97	4.05	1.38	-0.02	3.75	1.52
type 3	0.30	0.84	0.86	0.85	2.76	1.55
type 4	2.03	1.16	1.81	1.16	5.15	1.57
type 5	1.95	3.08	0.27	-0.01	3.17	1.41
type 6	1.51	3.30	1.67	-0.19	2.65	1.15

The valuation of bus interchange of type 3 is unlikely low. The number of observations here is however small and the value is very imprecise. A negative valuation of real time information is not reasonable. More information should not make the bus trip worse.

It seems reasonable that getting a seating place is more worth than that the bus has a low entrance, which is the case in all the 6 types. The reason for this is that all passengers are interested in getting a seating place, but only a few are ready to pay for getting a low entrance in the bus.

The estimated values' precision

Another way of measuring the quality of the answers is comparing the size of the standard errors. Of course the size of the standard errors not only depends on the type of SP questionnaire, but also on number of observations and homogeneity among respondents.

For values of time in the first game (in-vehicle time, walking time, bus frequency time) type 5 clearly gives the largest standard errors, while types 1 to 4 give the smallest. For waiting time by bus interchange, questionnaire types 4 and 5 have the smallest standard errors.

Table 10 Standard errors of estimated values of time. SEK per hour.

std.error	in-vehicle time	walking time	bus frequency time	waiting time by bus interchange (s)	waiting time by bus interchange (f)
type1 choices	3	5	2	19	8
type2 choices on cards	4	7	3	13	23
type3 rating best-worst	5	5	2	16	49
type4 rating 0-100	5	4	2	7	12
type5 ranking of labels	49	49	91	9	5
type6 ranking of cards	11	11	6	17	16

The factor waiting time by bus interchange has two different standard errors because it has been estimated twice, together either with seating place (s) or with low floor bus (f). The same goes for bus interchange and real time information system below.

When it comes to the standard factors bus interchange, real time information, seating place and low floor bus type 3 gives the largest standard errors and types 4 and 5 give the smallest. To make this comparison possible the standard errors of estimated values are added. This isn't the case in reality but as they've about the same size, the addition allows for a comparison.

Table 11 Standard errors of estimated values of standard factors. SEK per trip.

std.error	bus inter-change (s)	bus inter-change (f)	info system (s)	info system (f)	seat	floor	total
type1 choices	1.97	1.01	1.40	0.46	1.82	0.54	7.20
type2 choices on cards	0.92	1.86	0.64	0.95	0.74	1.05	6.16
type3 rating best-worst	1.38	2.03	1.24	1.97	1.21	2.55	10.38
type4 rating 0-100	0.56	0.76	0.58	0.76	0.80	1.02	4.48
type5 ranking of labels	0.63	0.53	0.76	0.55	0.66	0.59	3.72
type6 ranking of cards	1.00	0.85	0.97	0.92	1.06	1.03	5.83

Thus, questionnaire type 4 turned out best in this test. Types 3 and 5 showed to have the largest standard errors and are therefore not recommendable according to this test.

Regression analysis

The rating games, types 3 and 4, were also analysed by regression. Both direct regression on marked utility and regression on differences in utilities and levels were made (see e.g. Prather Persson, 1998). The regression on differences adjusts for different people having different personal scales. The regression analysis gave two more valuations per standard factor to compare with the ones from the logit analysis.

In the same way as before, some factors have two values because they are judged both together with seating place (s) and low floor bus (f). The values of time are pretty much the same either they are analysed by logit or regression. The only exception concerns type 4, when analysed by ordinary regression. This is natural because type 4 admits individual zeros, which type 3 doesn't.

Table 12 Valuations estimated by logit analysis and regression analysis, SEK per hour.

	in-vehicle time	walking time to bus stop	bus frequency time	waiting time by bus interchange (s)	waiting time by bus interchange (f)
type3 logit	26	21	23	35	61
type3 regr	18	20	24	28	65
type3 regdiff	25	13	21	28	56
type4 logit	23	20	23	29	40
type4 regr	-4	52	37	31	45
type4 regdiff	21	25	27	34	44

For the rest of the factors it can be said that neither ordinary regression analysis nor regression on differences give valuations similar to the ones from the logit analysis for type 3. The valuations of type 4 are about the same either they are analysed by logit analysis, ordinary regression or regression on differences.

Table 13 Valuations estimated by logit analysis and regression analysis, SEK per trip.

	change-s	change-f	info-s	info-f	seat	floor
type3 logit	0.30	0.84	0.86	0.85	2.76	1.55
type3 regr	0.98	-0.43	1.28	0.36	3.56	-0.87
type3 regdiff	0.81	0.23	1.40	0	3.65	-0.28
type4 logit	2.03	1.16	1.81	1.16	5.15	1.57
type4 regr	2.35	1.89	1.89	1.31	5.78	2.56
type4 regdiff	2.51	1.99	1.79	1.46	6.03	2.60

When comparing the confidence intervals nearly always estimations from the logit analysis have the smallest confidence intervals. Among the regression analyses the one made on differences gives the most confident estimations.

4.2 Qualitative tests

The qualitative tests were made to hear the respondents' opinion of the forms. None of the respondents has seen more than one of the types though, so they are haven't been able to compare them.

Telephone interviews

One of the questions among the background questions was "Can I call you if I wish further information? If so, please give your phone number...". Out of the respondents who had done this, up to 8 persons per type were interviewed over telephone. It has not been able to phone people who haven't given their phone numbers, and unfortunately not people who haven't sent in any answers at all.

The interviews started with "what did you think about the choices/ ratings/ rankings?" If the respondent said it was difficult, he/she should tell in what way it was difficult. The interviews were made as soon as possible after the answer had returned so that the respondent should remember the inquiry well.

The telephone interviews show that rating between the best and worst alternatives was most difficult, this type didn't get any positive comment. Binary choices work well for most people, while ranking has got mixed opinions. More than one person considered the pasting of labels as funny.

Spontaneous comments

Some of the respondents have written spontaneous comments on the forms sent in. SP type 3 has got four such comments, concerning the used method. All of them were negative stating for example "I'm sorry, but I didn't understand the questions at all" and "I found the X-ing contradictory wherever I put the X".

Type 5 got one comment, "this was not as easy as it seemed". None of the other questionnaire types got any such comments, which points to the fact that type 3 was the hardest type.

5 Conclusions

This study has showed that it is possible to carry out a SP survey by mail and get a fairly high response rate. The response rate is probably affected by the personal contact interview and by the design of the inquiry.

The test of six different questionnaire types with the same SP design showed that binary choices are to be used, not only in computer interviews. The binary choice games appeared to be easiest for the respondent to understand, gave reasonable valuations and acceptable confidence intervals.

Rating on a scale between 0 and 100 has favours like low lexicography, high scale factors and small confidence intervals but lacks in response rate. Rating on a scale between the best and the worst alternative had the lowest share of lexicographical answers but no other advantages.

Ranking of cards had a high response rate and ranking of labels to be pasted was experienced fun.

If rating data is to be analysed by regression analysis, regression on differences of utilities is recommended.

In future studies the binary choice method is recommended. The work will go on by making SP interviews by mail to be able to find assessments of different groups of bus passengers. For example different age groups, persons with different travel purposes, persons travelling during different time periods or with different service standards will be studied.

References

Bates, John och Terzis, G. Stated Preferences and the ”Ecological fallacy”. Proceedings of Transportation Planning Methods, PTRC, 1997.

Brundell-Freij, Karin. The logit model applied to the modal split of regional commuting. Department of Traffic Planning and Engineering, University of Lund, Bulletin 123, 1995.

Johansson, Stina och Svensson, Helena. Vad tycker resenärerna i Jönköping om trafikomläggningen? Department of Traffic Planning and Engineering, University of Lund, Bulletin 160, 1998.

Prather Persson, Cristina. The railway station and the interregional traveller. Department of Traffic Planning and Engineering, University of Lund, Bulletin 157, 1998.

Widlert, Staffan. Trafikantvärderingar vid regional kollektivtrafik. TFB-rapport 1992:2. Transportforskningsberedningen, Stockholm, 1992.

Widlert, Staffan. Stated Preference Studies – The design affects the results. Paper submitted to the 7th International Conference on Travel Behaviour, Santiago, Chile, 1994.