Journal of Choice Modelling, 2(1), pp. 8-28
www.jocm.org.uk

A pilot study into the perception of unreliability of travel times using in-depth interviews

Yin-Yen Tseng[1,*] Erik Verhoef[1,2,†] Gerard de Jong[3,4,∓]
Marco Kouwenhoven[3,§] Toon van der Hoorn[5,¶]

1Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands
2 Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam, The Netherlands
3 Significance, Koninginnegracht 23, 2514 AB The Hague, The Netherlands
4 Affiliated to ITS Leeds, Leeds LS2 9JT, The United Kingdom, and to NEA, PO Box 276, 2700 AG Zoetermeer, The Netherlands
5 Dutch Ministry of Transport, Rijkswaterstaat and University of Amsterdam

Received 16 November 2007, received version revised 21 June 2008, accepted 4 September 2008

Abstract

Transport investments normally reduce travel times, but may also reduce unreliability. Conventional time gains can be evaluated in cost benefit analysis using standard values of time. For valuing reliability gains, however, no standard measures are readily available. The Dutch Ministry of Transport has commissioned a project to design a Stated Preference methodology. Reliability is a complex "academic" concept and it may be difficult to present and explain it to respondents. Therefore, a pilot study using in-depth face-to-face interviews has been carried out, in which various SP choice designs have been presented to the participants to test their understanding of the questions. In the end, we found a "best" design. But still, much care is needed for an acceptable response rate and to prevent illogical and inconsistent responses.

Keywords: travel time unreliability, perception, stated preference design, value of reliability

* Corresponding author, T: +31-20-598-6098, F: +31 20-598-6004, ytseng@feweb.vu.nl
† T: +3120-598-60-94, F: +31 20-598-6004, everhoef@feweb.vu.nl
∓ T: +31-70-31-21-530 F: +31-70-31-21-531, dejong@significance.nl
§ T: +31-70-31 21-544, F: +31-70-31-21-531, kouwenhoven@significance.nl
¶ T: ++31-88-798-2372, F: +31 88-798-2999, toon.vander.hoorn@rws.nl

ISSN 1755-5345

1 Introduction

In The Netherlands, travel time benefits resulting from transport investments are assessed in a cost-benefit analysis (CBA), using standard values of time for the conversion of time into money units. But transport investments can also reduce the unreliability of travel times. For valuing such travel time reliability benefits, new empirical research in The Netherlands is required.

Significance (since 1 January 2007; in 2006: RAND Europe), the VU University Amsterdam and John Bates have carried out a preliminary design project for the Netherlands Ministry of Transport Public Works and Water Management, under contract to the AVV Transport Research Centre and the KiM Netherlands Institute for Transport Policy Analysis. Subsequent work, including the main fieldwork and estimation of discrete choice models, will be organised later, in the form of a separate project.

In the first phase of the preliminary design project, in consultation with a group of national and international experts, decisions were taken on which modes and transport market segments to include. In the second phase, face-to-face interviews were carried out with 30 travellers to test their understanding of and preferences for different ways of presenting unreliability in stated preference (SP) surveys (e.g. clock-face, bar chart, several ways of phrasing it). In the third phase pilot interviews were held to test the questionnaires and to check whether respondents understood the experiments. This was a real pilot, to test not only the correct flow and understanding of the questionnaires but also to obtain sufficient information for initial model estimation trials to check whether the attribute levels presented were the right ones.

This paper will mainly deal with the second and third phase. The structure is as follows. In Section 2, the interview questionnaires and the eight presentation formats for unreliability are described, together with the recruitment process. In Section 3 the outcomes of the tests of the different formats for presenting unreliability to respondents are presented. We give our general impressions and formulate a final preference for one of the eight presentation formats. The set-up and the results of the full pilot interviews are discussed in Section 4. Finally, Section 5 gives the conclusions from this project. The Appendix includes – as an example – the complete eight formats used in the face-to-face interviews for car non-scheduled trips (translated from Dutch).

2 Design of the face-to-face interviews

2.1 Motivation and objectives

It is widely accepted that travel time reliability is an important factor in travellers' decision-making, and that the benefit of reliability improvement has to be taken into account in current cost and benefit analysis (CBA) frameworks. Nevertheless, travel time reliability represents virtually the most difficult attribute in stated preference (SP) experiments. Unlike travel time and cost attributes, which usually relate to a specific trip in a stated choice experiment, reliability often relates to the traveller's experience over multiple trips. The concept of travel time reliability involves two dimensions: (i) a frequency or probability dimension – how often the delay occurs – and (ii) a magnitude dimension – how big the delay is when it occurs. The amount of information contained in the reliability attribute is much larger than the information in the other attributes. As a result, the way of presenting travel time reliability in the literature varies considerably across studies because of the different thoughts and considerations of the researchers.

To reduce the amount of information presented in the reliability attribute and to facilitate the respondents' task of digesting information on probability, many researchers have used "5 or 10 possible travel/delay times with equal chance" to describe the travel time reliability (see, e.g. Small et al., 1995, 1999; Bates et al., 2001; Hollander, 2005). Some researchers used verbal description to illustrate the reliability attribute, while some thought that graphical presentation can help the respondents to understand the attribute. Some researchers even believed that the respondents are capable of understanding a complicated histogram of travel time distribution if good instruction is provided in the computer-aided personal interviewing (CAPI) program (e.g. Copley et al. 2002). We summarize the reliability presentation formats that have been used in the empirical SP studies in Table 1.

The different presentation formats have their own strengths and weaknesses, and may therefore be preferred on different criteria. The research team was not able to find any systematic evidence (by literature study and by consulting the experts in this area) that respondents would prefer, or better understand certain presentation formats over other ones. Therefore, 30 face-to-face interviews were organized to fill this gap in the present knowledge. The objectives of the face-to-face interviews were as follows:

Table 1: Presentation of travel time reliability in SP studies for passenger transport

No.	Country of study	Author	Presentation of reliability	Graphical or verbal
1	US	Small et al. (1995)	5 individual travel times, each has an equal chance	Verbal
2	US	Small et al. (1999)	5 individual arrival times, each has an equal chance	Verbal
3	US	Small et al. (2005)	Frequency of unexpected delays of 10 minutes or more	Verbal
4	UK	Bates et al. (2001)	10 possible delay times are presented graphically in a circular, clock-face format	Graphical
5	NZ	Hensher (2001)	Uncertainty time is presented in a vertical bar	Graphical
6	UK	Copley et al. (2002)	Graphical presentation of journey time information as a distribution histogram	Graphical
7	UK	Hollander (2005)	5 possible arrival times (equal chance) are shown graphically by a series of vertical bars	Graphical
8	US	Bhat and Sardesai (2006)	Maximum and usual door-to-door travel time	Verbal
9	US	Tilahun and Levinson (2005)	Travel time distribution graphically	Graphical
10	NL	Tseng et al. (2005)	Maximum and minimum time in congestion	Verbal
11	NL	Tseng et al. (2006)	2 mass points of arrival (delay) times and their associated probabilities	Verbal

- Test the respondents' understanding of different reliability presentation formats;

- Investigate the respondents' impressions of these presentation formats in several aspects, e.g., clearness, ease of understanding, and visual appearance attractiveness;

- Collect the respondents' preferences on the presentation formats from several aspects. In the analyses of these interviews, we paid special attention to the effect of education, to see if it affects the respondent's ability to understand the questions and especially the presentation of unreliability.

In summary, the main emphasis here is on the understanding and appreciation by respondents of rather difficult interview questions. Checking how people deal with (un)reliability in actual trip making, or investigating the relationship between objective and subjective experienced travel time variability, was not the aim of this study. There is a fair amount of literature on the comparison of objective and subjective travel times, e.g. Rietveld et al. (1999), but not so much yet on a similar comparison for travel time reliability. This may be a good subject for further research, e.g. by using GPS logging devices.

2.2 Description of the interview questionnaire

It is likely that the valuation of unreliability depends on the question of whether the timing of a traveller's arrival and departure matter, in addition to the duration of the travel time. If the timing (departure and/or arrival times) of the trip is important for the traveller, this trip is called a 'scheduled trip'. On the other hand, the trip is called 'non-scheduled' if the trip timing is less relevant to the traveller. The interview starts with some brief screening questions, where we can distinguish four types of trip for the interviewees:

1. Car non-scheduled trips (CN): travelling by car as a driver without scheduling consideration (i.e. trip timing is not important);

2. Car scheduled trips (CS): travelling by car as a driver with scheduling consideration (i.e., trip timing is important);

3. Public transport non-scheduled trips (PN): travelling by public transport without scheduling consideration;

4. Public transport scheduled trips (PS): travelling by public transport with scheduling consideration.

After the screening questions, the interviewer would ask the interviewee to imagine that a certain type of trip was to be considered during the whole interview. The rest of the interview questionnaire is set up as described in Table 2.

2.3 Presentation formats tested

In the selection process of the presentation format, the starting point was the list of reliability SP studies in the Start Memorandum of the study (available from the authors on request). According to Hamer et al. (2005), a series of possible travel times (or arrival times)

Table 2: Face-to-face interview set-up

Section	Content
I	General information of the reliability perception
II	Test questions for the reliability presentation format
III	General impression of the reliability presentation format
IV	Preference of the reliability presentation format
V	Individual socio-demographics

is recommended to describe the travel time unreliability in the current Dutch value of reliability (VoR) SP survey. Thus, we only tested the formats that are presented by a series of possible travel times, and discarded the formats using different types of definition of travel time reliability (e.g. Hensher, 2001; Tseng et al. 2005; Bhat and Sardesai, 2006).

In Black and Towriss, 1993 (cited from Bates et al., 2001 and Small et al., 1999), it was suggested that people could interpret a 5-point (compared with 10-point) distribution of travel times reasonably well. The research team found this plausible, and also found that 5-point distribution of travel times were relatively more widely used than other ones (Small et al., 1995; Small et al., 1999; Hollander, 2005). Consequently, we adjusted all the tested presentation formats in such a way that in every format reliability is presented as a distribution of 5 possible travel times.

We selected eight presentation formats, namely Format A to H, to be tested. A brief description of these formats is as follows.

- *Format A*: the verbal description (without any graph) of 5 possible travel times in 5 different lines. This format is adopted from Small et al. (1999).

- *Format B*: the clock-face presentation of 5 possible travel times. This presentation format is a variant of the one in Bates et al. (2001), where there are 10 possible travel times in the clock-face circle.

- *Format C*: 5 'bars' present 5 possible travel times. The lower end of the bar gives the departure time and the top end of the bar gives the arrival time. Travel time duration is therefore implied by the length of the bar. This presentation format was used by Hollander (2005).

- *Format D*: Format D is similar to Format C. The only difference is that the sequence of the 5 possible travel times is ordered by the lengths in Format C, while the sequence is randomized in Format D.

- *Format E*: histogram of the travel times distributions. The horizontal axis represents the travel time/arrival time, and the vertical axis represents the percentage of trips. This format is used by Copley et al. (2002).

- *Format F*: Format F is similar to Format E. The only difference between these two is that Format E uses percentage (XX%) to represent the likelihood of certain travel time, while Format F uses frequency (× out of 5 trips) to represent the same information.

- *Format G*: Format G is a new format. In this format, we show some possible travel time/arrival time intervals and the associated frequencies to the respondents. The consideration here is that these possible arrival time intervals are closer to what the reality is, and may be more representative for the concept of unreliability.

- *Format H*: Format H is a combined version of Format A and G. The 5 possible travel time/arrival time intervals are shown in 5 different lines.

Note: although each individual travel time is given an equal chance, in the design some individual values may be included several times. That results in some values having a higher chance than others. Example: "the following travel times occur with equal chance: 30 min, 40 min, 40 min, 40 min, 45 min". The eight presentation formats are given in the Appendix for the car non-scheduled (CN) trips.

2.4 Recruitment and interview process

The 30 interviewees were recruited from the personal network of VU University Amsterdam, such as relatives and friends. To enhance the credibility of the interview outcomes, we paid special attention to selecting the representative interviewees in different education levels as well as in the types of trips (CS, CN, PS, and PN) they made. All the interviews were done during October - November 2006. An interview usually took 40-60 minutes, while a few took more than one hour. The respondent was compensated by 15 Euros as a reward for participating.

3 Results from the face-to-face interviews

The number of interviews was 30 in total, with 8 respondents in CN, 10 respondents in CS, 3 respondents in PN, and 9 respondents in PS versions. In the following subsections, we will present the results, with particular focus on the group of lower educated people.

3.1 General information of the reliability perception:

Before giving any description or explanation of unreliability (uncertainty) of travel time to the interviewees, we tried to ask the respondents what is in their mind when they are confronted with unreliability/uncertainty of travel time in practice.

Table 3: QI(1) Can you indicate which factors you consider in practice when deciding about trips and travel times are uncertain? (more than one option can be chosen)

	Lower education (freq.)	**Higher education (freq.)**	**Full sample (freq.)**
Average travel time	11	10	21
Maximum travel time	4	7	11
Minimum travel time	1	8	9
Probability	2	9	11
Don't know	0	0	0
Other	0	1	1
Total	18	35	53

Table 4: QI(2) How complicated do you find it to think about uncertainty of travel times?

	Lower education (freq.)	Higher education (freq.)	Full sample (freq.)
Not at all complicated	5	5	10
A bit complicated	3	10	13
Complicated	4	3	7
Total	12	18	30

Note: Complicated" in this paper comprises reasonably/considerably/very complicated in the questionnaire. Same for other concepts like "helpful"

As expected, 'average travel time' got more votes than the other elements (e.g. "my travel time varies, but on average it is 30 minutes"). The elements of maximum /minimum travel time (e.g. "my travel time varies between 20 and 40 minutes") and probability (e.g. "my travel time is 20 minutes with probability 10%, 30 minutes with probability 50%, and over 30 minutes with probability 40%") accounted for around equal share of the choices. In principle, average travel time, maximum and minimum travel times, and probability are the four most important factors that the travellers would think of to describe the travel time uncertainty in practice. It is interesting to note that higher educated people tend to think more in terms of 'probability' than lower educated people.

Next, the interviewers would give some description and explanation for the uncertainty of travel time. In particular, a dice-throwing example was given to the respondents to help them to think of the unpredictability of travel times and the probability associated with the realization of a certain travel time event. Opinions were then probed about the complexity and helpfulness of this example. From these, it can be concluded that it is useful to keep a text like this in the main questionnaire.

3.2 Test questions for reliability formats

The main objective of this face-to-face interview was to test which reliability presentation format is best understood by the respondents, and to see how the respondents interpret these presentations. This section will discuss these two issues.

Table 5: QI(3) How complicated do you find the description of uncertainty of travel times in the text (throwing dice example)?

	Lower education (freq.)	Higher education (freq.)	Full sample (freq.)
Not at all complicated	2	4	6
A bit complicated	6	9	15
Complicated	4	5	9
Total	12	18	30

Table 6: QI(4) How helpful do you find the example of throwing a dice in the text?

	Lower education (freq.)	Higher education (freq.)	Full sample (freq.)
Not at all helpful	2	5	7
A bit helpful	2	4	6
Helpful	8	9	17
Total	12	18	30

The test questions were designed to check to what extent the respondents have the "correct" perception of reliability, i.e., the same as expected by researchers, for each format. To prevent people getting familiar with the numerical example, and basing answers on previous formats, we varied the attributes (time, cost, and reliability) levels across these eight presentation formats. Furthermore, we randomised the order of showing the formats, to reduce the possible impact of learning or fatigue effects. In other words, formats A to H were shown in different sequences to different respondents. The results of the test questions are summarized for scheduled and non-scheduled trips versions separately.

3.2.1 Non-scheduled trips

There are two test questions for the non-scheduled trip interviewees (versions of CN + PN = 11 respondents). To compare respondents' ability of understanding the eight formats, we computed the score rate – defined as the percentage of correct answers given by the respondents – for each format. Tables 6 and 7 summarize the results of these two test questions.

Though the sample size is rather small, we deemed the score rates of lower educated people as satisfactory, and in many cases they are even higher than the ones for higher educated people.

Table 7: QII(3) The travel time for trip 1 is more uncertain than trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	No	Yes	Yes	No	Yes	No	Yes	No
Lower education								
Score rate (%)	100	100	100	80	80	100	100	100
Higher education								
Score rate (%)	83	66	66	83	66	83	50	83
All respondents								
Score rate (%)	90	81	81	81	72	90	72	90

Note: Score rate is defined as the percentage of correct answers given by the respondents.

Note: Here, "more uncertain" can be interpreted either by a larger standard deviation of travel time or a wider range between maximum and minimum travel times (amount of uncertain time). In this interview experiment, we deliberately assigned the reliability level (a series of five possible travel times) in such a way that a more unreliable alternative has larger standard deviation as well as larger uncertain time (difference between maximum and minimum travel times).

Table 8: QII(4) With trip 1, I arrive, on average, earlier than with trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	Yes	No	Yes/No	Yes/No	Yes	No	No	Yes
Lower education								
Score rate (%)	80	100	-	-	80	60	100	100
Higher education								
Score rate (%)	83	83	-	-	83	66	83	83
All respondents								
Score rate (%)	81	90	-	-	81	63	90	90

Note: Since the two alternatives in Format C and D have the same average travel time, there is no correct answer for these two formats.

3.2.2 Scheduled trips

There are four test questions for the interviews with scheduled trips (versions of CS + PS = 19 respondents). The results are summarized as follows.

Table 9: QII(1) With trip 1, I have a greater probability of arriving earlier than I want than with trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	No	Yes	Yes	No	Yes	No	Yes/No	Yes/No
				Lower education				
Score rate (%)	85	71	100	71	71	71	-	-
				Higher education				
Score rate (%)	100	50	91	83	83	83	-	-
				All respondents				
Score rate (%)	94	57	94	78	78	78	-	-

Table 10: QII(2) With trip 1, I have a greater probability of arriving later than I want than with trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	Yes/No	No	No	Yes	No	Yes	Yes/No	Yes/No
				Lower education				
Score rate (%)	-	85	100	71	71	57	-	-
				Higher education				
Score rate (%)	-	58	91	91	75	58	-	-
				All respondents				
Score rate (%)	-	68	94	84	73	57	-	-

Table 11: QII(3) The travel time for trip 1 is more uncertain than trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	No	Yes	Yes	No	Yes	No	Yes	No
				Lower education				
Score rate (%)	57	42	28	42	28	42	42	71
				Higher education				
Score rate (%)	100	66	50	83	66	83	50	83
				All respondents				
Score rate (%)	84	57	42	68	52	68	47	78

Table 12: QII(4) With trip 1, I arrive, on average, earlier than with trip 2

Format	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Correct answer	No	Yes	Yes	No	Yes	No	Yes/No	Yes/No
				Lower education				
Score rate (%)	85	85	100	71	85	85	-	-
				Higher education				
Score rate (%)	83	50	83	91	83	91	-	-
				All respondents				
Score rate (%)	84	63	89	84	84	89	-	-

The score rates of test question QII(3) (Table 11) are generally lower than the other question, especially for the lower educated people. One reason could be that the respondents did not know how to interpret 'uncertainty' exactly and these respondents just gave up in answering this question (the non-response rate is higher in this question). Another possible reason is that the 5 possible travel times are shown together with 5 resulting possible arrival times in the format, so people may be inclined to link the uncertainty to the level of lateness (though the arriving probability question was asked in QII(2)).

The score rates of Format B are relatively lower than the other formats for these test questions. As for the level of education, we do not find any strong evidence that lower educated people have lower score rates.

3.3 General impression of reliability presentation formats

In interview Section III, we asked the interviewees to indicate how they perceived the levels of clearness or difficulty in each presentation format. This is described by five different indicators:

- Clearness of the presentation of reliability;
- Ease of making choice between two alternatives/trips;
- Ease of considering all information/attributes;
- Attractiveness of the visual appearance;
- Ease of answering the test questions in Section 3.2.

To make the answers more comparable between formats, we assign scores to different levels of understanding (or difficulty). Thus, we can rank these eight presentation formats according to the average scores they get in the questions. Here we use 1 for 'very unclear', 2 for 'unclear', 3 for 'not very clear', 4 for 'clear', and 5 for 'very clear'. A format with higher score will therefore be more preferred.

Figure 1 below summarizes the average scores (the average of 30 respondents) in these eight formats based on these five indicators.

In general, Format A is ranked first in most cases. The next is Format C, which also performs well in the scores, and the score differences between Format A and C are small in many cases. It is obvious that Format G and H are among the worst for most of the cases. Since the reliability information given in these two formats is more complex (showing the travel time intervals rather than some certain travel times), the respondents seem to have much more difficulty in reading them.

3.4 Preference of reliability presentation formats

In interview Section IV, we showed all eight presentation formats once more to the interviewees. The interviewers placed these eight formats on the desk (randomly), and then asked the interviewees to indicate once again their preferences on the format, according to the five indicators discussed in Section 3.3 (clearness of the presentation; ease of making choice between two alternatives; ease of considering all information in the alternatives; visual appearance attractiveness; ease of answering the test questions).

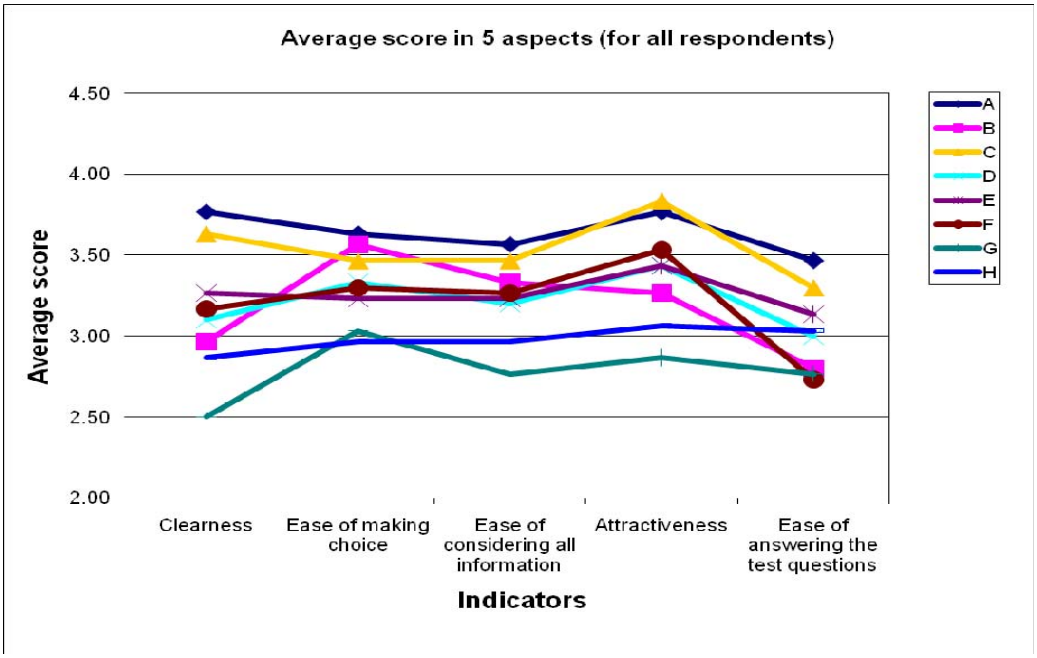


Figure 1: The average scores of 5 indicators for Format A to H: all respondents

The respondents indicated both the most preferred and the least preferred format according to these five aspects. The results of the choice frequency are shown graphically in Figure 2 to Figure 6. Here we combine the most and the least preferred format into one graph. Thus, the positive part of the vertical axis represents the frequency that the format is chosen as favourite format; while the negative part represents the frequency that it is chosen as the least preferred format.

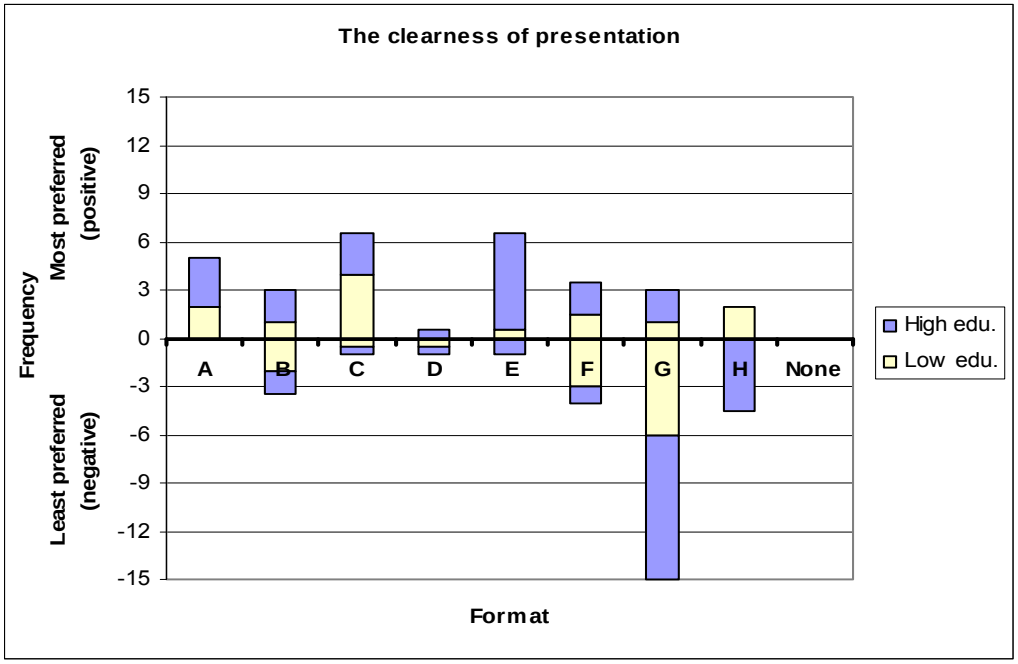


Figure 2: The frequency of the clearness of the format

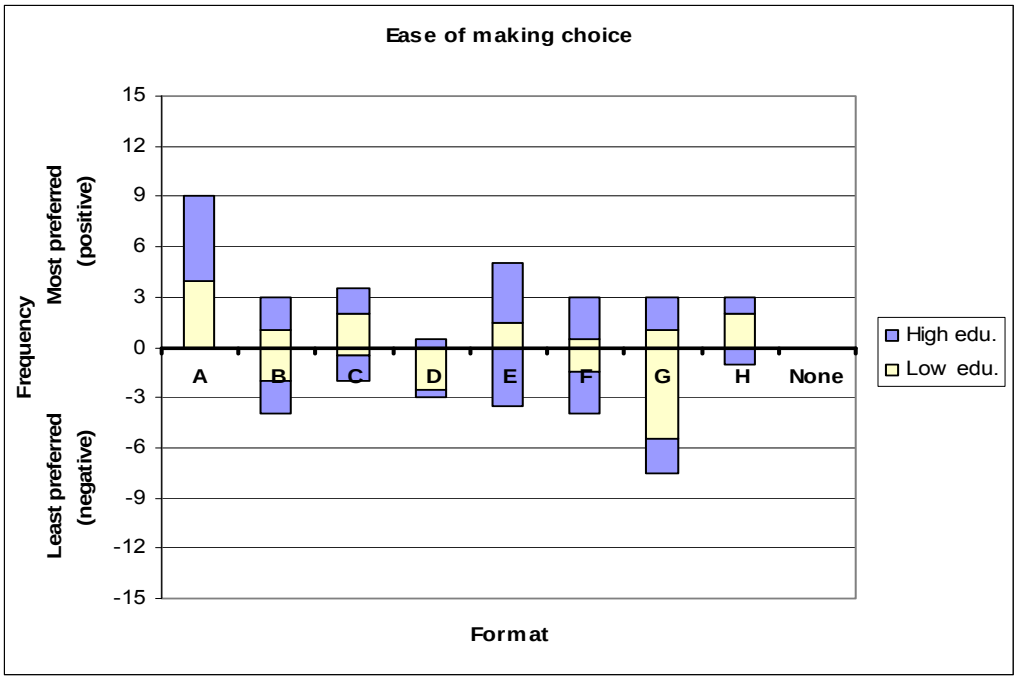


Figure 3: The frequency of the ease of the format in making choice

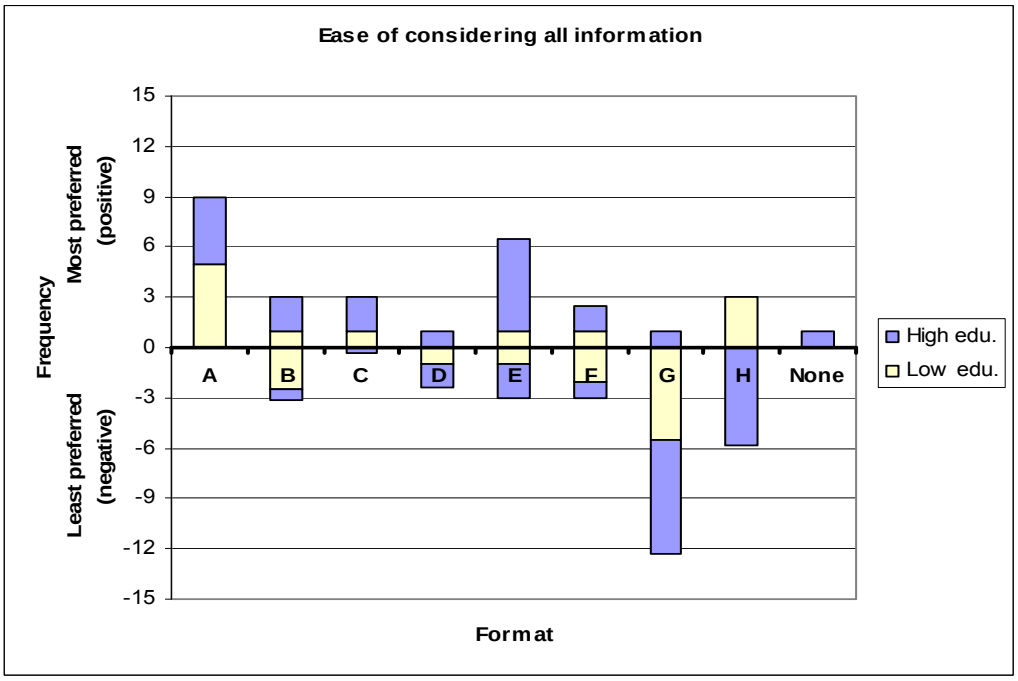


Figure 4: The frequency of the ease of the format in considering all information

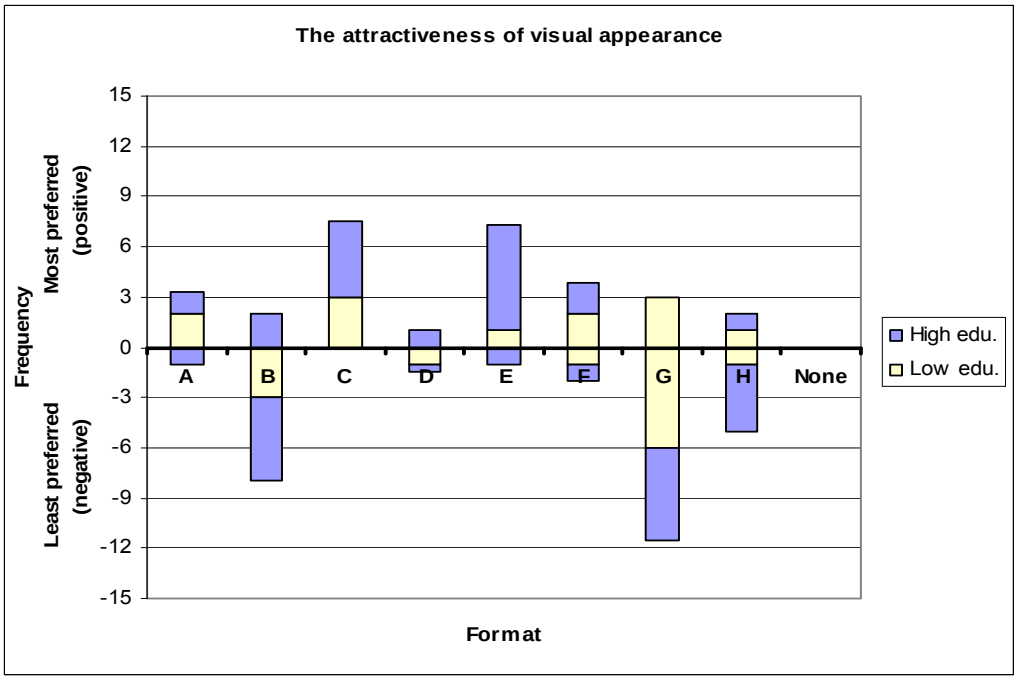


Figure 5: The frequency of the format in the visual appearance attractiveness

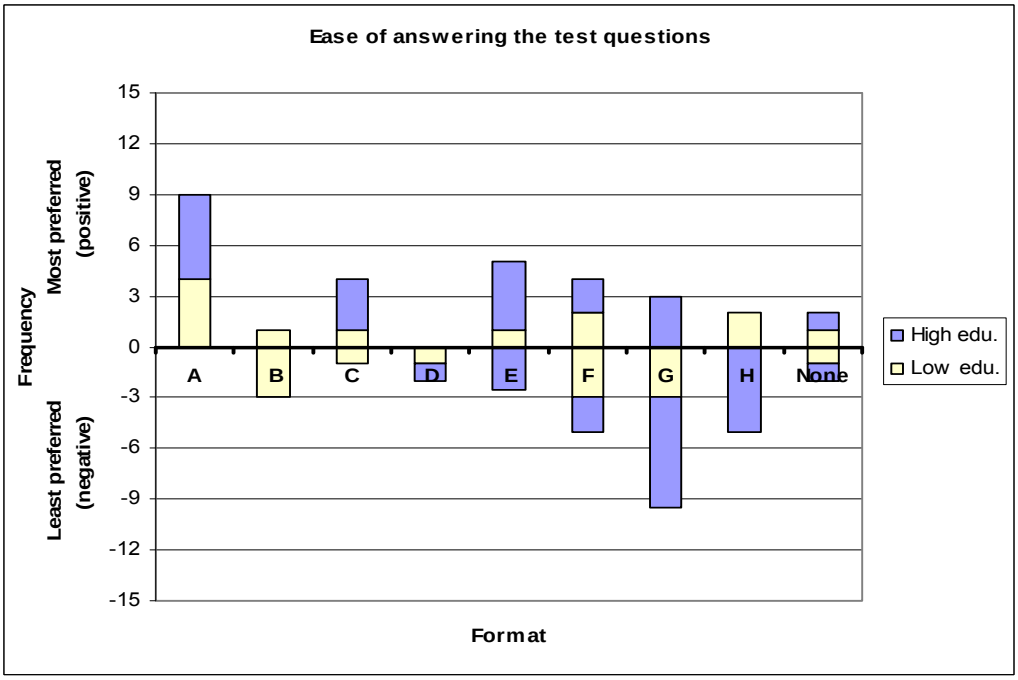


Figure 6 : The frequency of the format in the ease of answering the test questions

From these results we can conclude that the preferences of these respondents are quite diverse. None of the formats is favoured (most) by more than half of the respondents in every situation (clear/easy/attractive, …, etc.). Nevertheless, Format A performs well consistently, compared to the other formats, and scores relatively high on criteria that are

more likely to directly affect the quality of responses for repetitive SP questions (ease of making choice, ease of considering all information, ease of answering test questions).

4 The pilot to test the full surveys

The objectives of the pilot interviews were to test the full survey design and to finalize the survey design based on the results from the tests. Four different segment questionnaires, car, public transport, freight transport, and recreational navigation, were tested. For each of these 4 segments, 20 self-completion paper questionnaires, and 4 face-to-face interviews were planned. These face-to-face interviews are the ones we reported above, and were mainly carried out to check whether the phrasing and presentation of attribute were properly understood by the respondents. The "self-completion" questionnaires were based on Format A as discussed above. After having analyzed the face-to-face results, we can get some further insight into the respondents' understanding of the questionnaire. We will discuss the results briefly in this section, and we will not deal with the pilots for freight and recreational navigation here.

Since the pilot was done by using the paper-based questionnaire, it was essential to have a separate recruitment questionnaire first, to collect some information on a respondent's recent trip that is in scope. This information was then used to customize the main questionnaire (e.g. to set the base level for the attributes, and the changes relative to this reference level), and the main questionnaire was mailed to the recruited respondent. As in the face-to-face interviews, to encourage respondents to participate in the pilot, we provided a reward of 10 or 15 Euros.

For car drivers, the respondents were mainly recruited during the morning peak from parking garages in the Amsterdam metropolitan area. For public transport users, the recruitment took place in the shopping centre and train stations around the Amsterdam Bijlmer/Arena area and Almere, and it was done between 10 am and 3 pm. In the end, we successfully recruited 24 car drivers and 25 public transport travellers who were willing to participate in the main survey in the next stage. For the face-to-face interviews in car and public transport, we then contacted the respondents from the previous face-to-face interviews.

All in all, we found it harder than anticipated to recruit respondents for this study, despite the fact that a compensation of 10 and 15 Euros was offered. We attribute this, at least in part, to the fact that people may feel overloaded with questionnaires, and possibly requests by other people (e.g. charity collectors) while on the street. It is clear that a low response rate may cause all sorts of biases, in particular if it is correlated with the value of time. We therefore advise seriously considering the use of an existing household/consumer panel, or at least a professional market research firm for the recruitment, when doing the formal questionnaire. It may prove cheaper in the end, and also may give more possibilities of stratifying the sample.

In this pilot we did a few face-to-face interviews as discussed above, but larger numbers of pilot interviews were carried out with printed mail-back questionnaires that were personalized on the basis of answers from a recruitment interview. Over and above the SP questions, each questionnaire also contained three sets of 'check' questions that were intended to check whether the respondent understood the concept of reliability in the format that was presented to them. This included questions as to which alternative (in the view of the respondent) had the highest reliability, or had the highest probability of arriving on or before a certain time. In addition, one of the choice pairs contained a "dominant" alternative

(i.e. an alternative that is superior in all aspects and, therefore, logically should be chosen), as in the earlier face-to-face interviews.

The average correct response to the check questions (about 55%) appears to be much lower than during the face-to-face interviews that were held to test the presentation of the reliability attribute. This might be attributed to the fact that most of the interviews in the pilot were mail-back questionnaires, so the amount of time taken to understand the concept of reliability might have been less. However, also the respondents in the pilot that did a face-to-face interview performed poorly.

Space constraints prevent us from presenting further results of this pilot, and more details can be found in de Jong et al. (2007). In brief, the outcomes of the pilot were used to estimate discrete choice models. These models do not have any practical value yet, their only goal is to test whether the survey design works well. It turned out that for passenger transport (car, public transport, recreational navigation) the estimations gave quite acceptable results (VoTs and VoRs in plausible ranges), given the small sample sizes. Ideally, we would have liked to compare models for the total group with models only for respondents who successfully answered the check questions, but the sample size is too small for such an exercise.

On the basis of the pilot we conclude that the proposed passenger surveys, in combination with the proposed sample sizes and interview methods, will most likely provide data that will allow estimation of a variety of models for the determination of VoTs and VoRs.

5 Conclusion

In the *face-to-face interviews*, we first collected the information of respondents' perception about travel time unreliability. We then examined several reliability presentation formats that have been used in previous empirical studies. To have a thorough assessment of these presentation formats, we not only asked questions about the respondents' subjective preferences, but also questions that can be tested objectively to see if these respondents perceived the unreliability in a consistent and logical way. The interview results were analyzed separately for lower and higher educated people.

In conclusion, we recommend using the verbal description, Format A, to represent the travel time reliability in the VoR SP experiment. In many indicators of the interviews, Format A is favoured by most (a relatively high proportion) of the respondents. Furthermore, respondents' preference for Format A is rather consistent between lower and higher educated people. Histograms with unequal probabilities (Formats E and F) are less encouraged by the research team, since we find that some people have difficulty in reading the probability from the graph, especially lower educated people. It is also interesting to note that the preferences between Format C & D vary a lot in some cases, although the two formats present reliability information in a very similar way. For this reason, although Format C is ranked second in many cases, we do not recommend this format here.

One proviso about these interviews is that the test formats are conditional on the 5 possible travel times with equal probability. Thus, the above conclusion is contingent on this particular situation and may not apply to all cases. In some cases, when very small differences in probabilities (far below say 20%) in one of the tails of the travel time distribution (often the right tail) are likely to be decisive for choice behaviour, the presentation format suggested in this study may no longer be adequate. A trip to the airport is such an example. Because the consequence of missing the flight is usually very severe, even a small probability may play an important role in a traveler's decision making. In such

a case, the presentation format E could be an alternative; nevertheless, some proper training in reading the histogram may be required and would be helpful for the respondents.

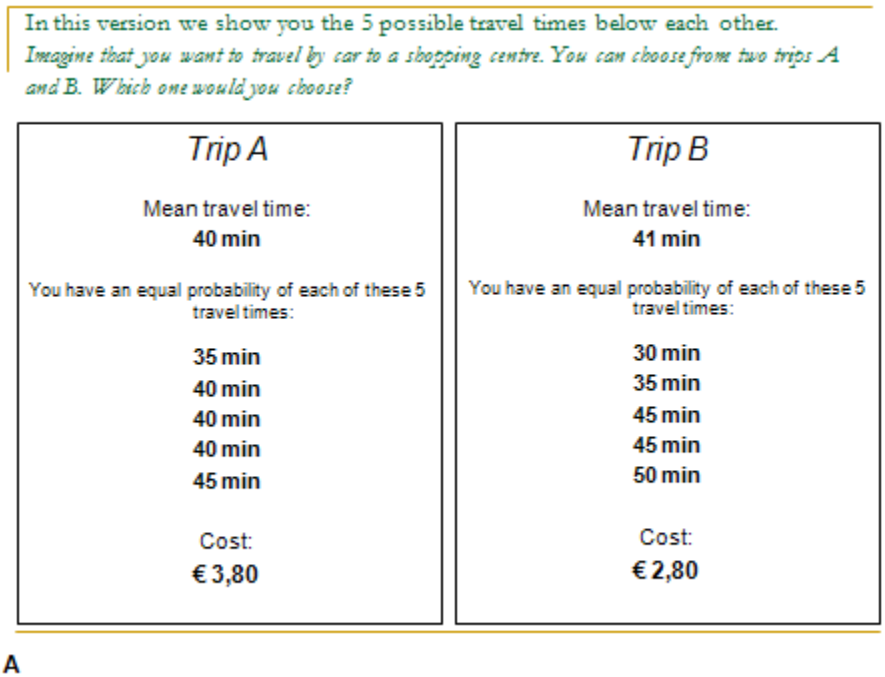
The objective of the pilot interviews was to test the full survey design based on the results from the face-to-face tests. We found it harder than anticipated to recruit respondents for this study, despite the fact that a compensation of 10 and 15 Euros was offered. People may feel overloaded with questionnaires, and possibly requests by other people [e.g. charity collectors] while on the street. We therefore advise seriously considering the use of an existing panel, or at least a professional market research firm for the recruitment, when doing the formal questionnaire.

The average correct response to the check questions was much lower than during the face-to-face interviews that were held to test the presentation of the reliability attribute. This might be due to the fact that most of the interviews in the pilot were mail-back questionnaires, so the amount of time taken to understand the concept of reliability might have been less. However, respondents in the pilot that did a face-to-face interview did not answer these check questions more correctly than the mail-back respondents. So, this remains a point of concern.

The outcomes of the pilot were used to estimate discrete choice models, not for their own merits, but just to test whether the design works well and whether the project is on the right way. This gave quite acceptable results (VoTs and VoRs in plausible ranges), given the small sample sizes.

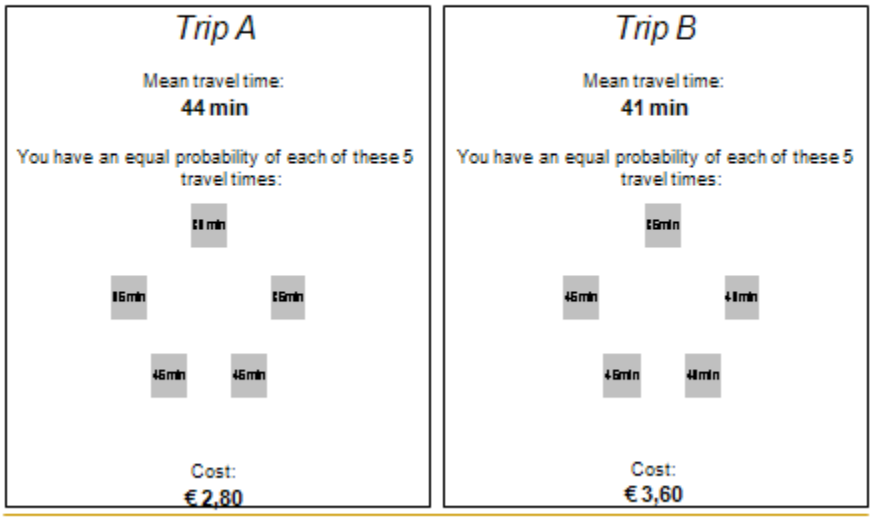
Provided that the recruitment is done by a professional market agency and/or by using a panel, we expect that the passenger surveys in the forthcoming main study will supply data that will allow estimation of models with significant VoTs and VoRs.

Appendix: Reliability presentation formats for car non-scheduled trips



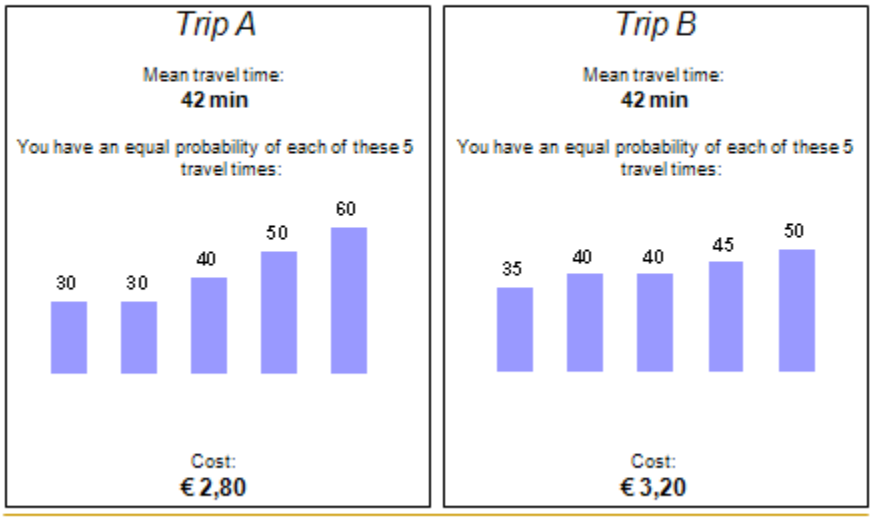
A

In this version we show you the 5 possible travel times as points on a circle.
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



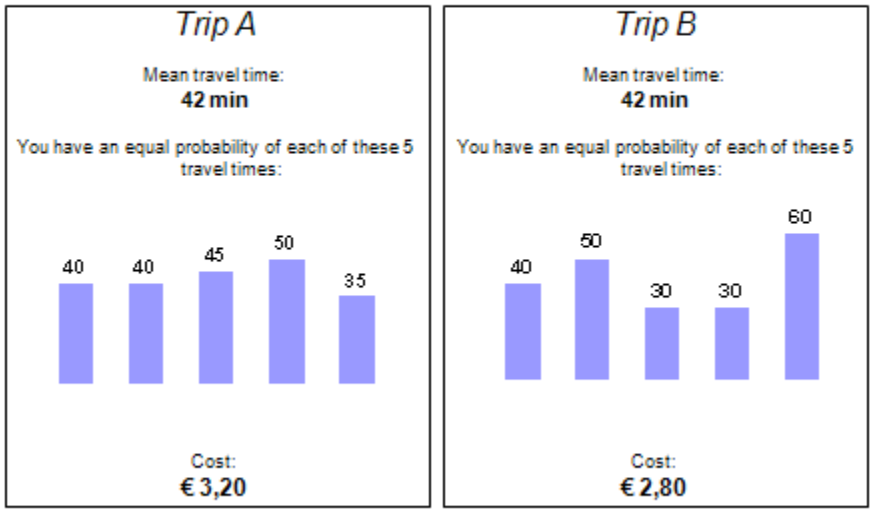
B

In this version the 5 possible travel times are illustrated by the height of the bars.
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



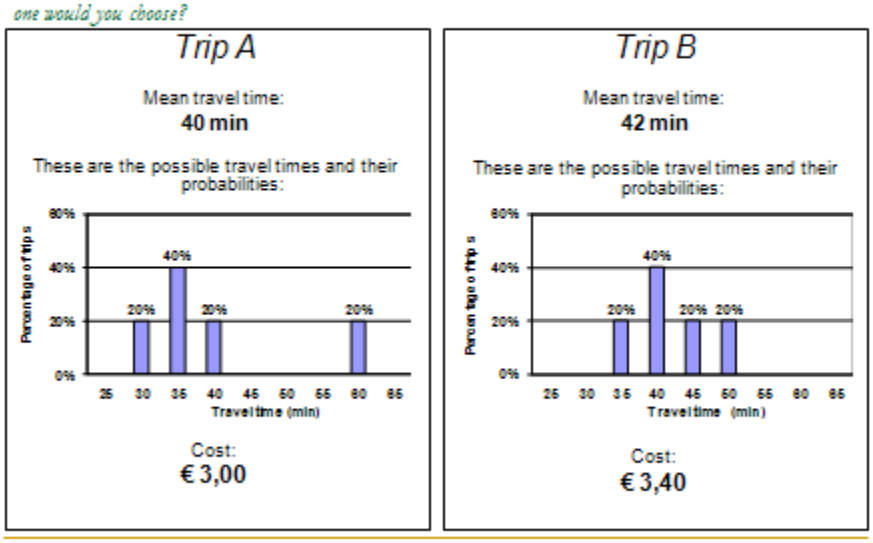
C

In this version the 5 possible travel times are illustrated by the height of the bars.
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



D

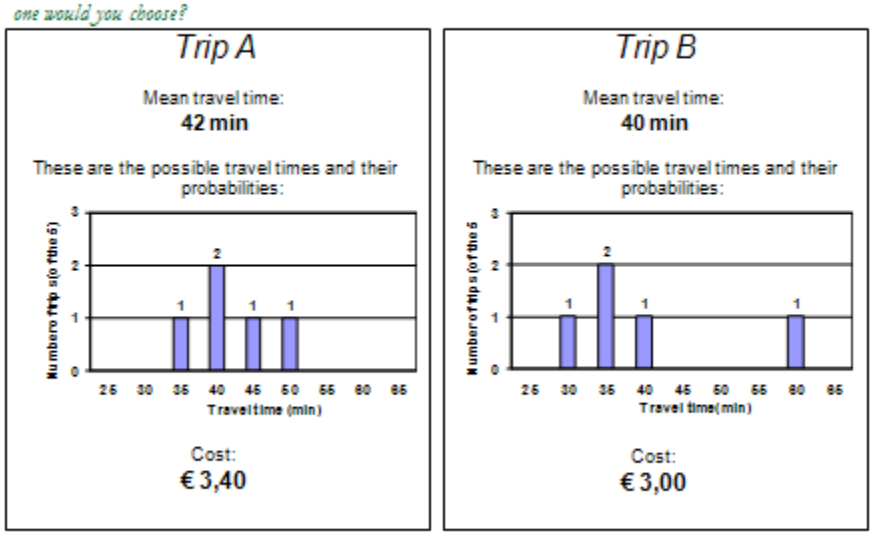
In this version the 5 possible travel times are illustrated by the height of the bars. (probabilities as percentage).
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



E

In this version the 5 possible travel times are illustrated by the height of the bars. (probabilities as number of times out of 5)).

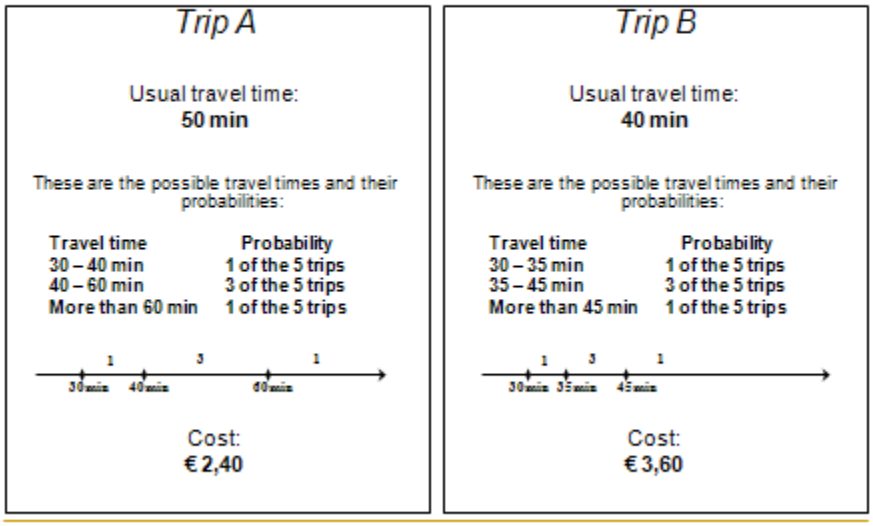
Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



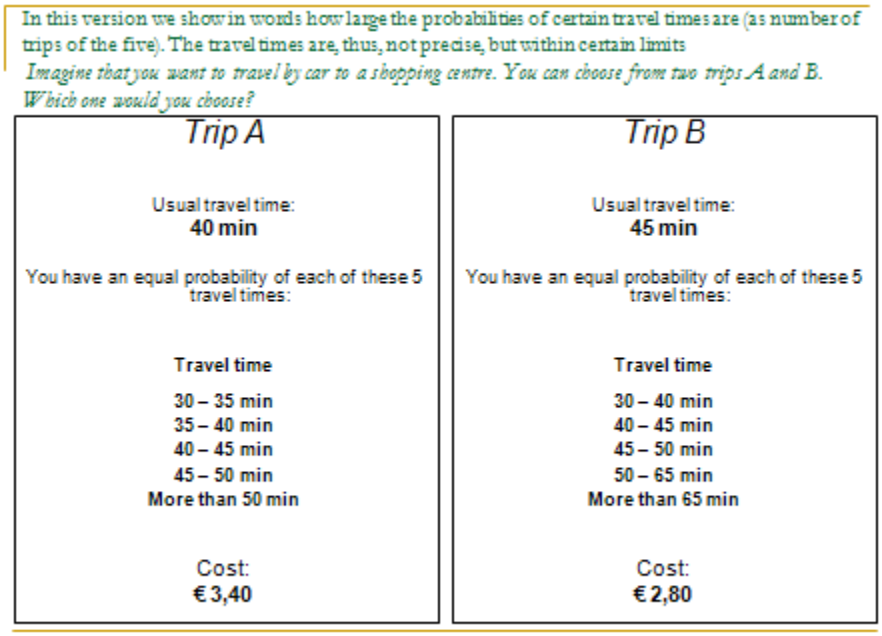
F

In this version we show in words and along a time axis how large the probabilities of certain travel times are (as number of trips of the five). The travel times are, thus, not precise, but within certain limits

Imagine that you want to travel by car to a shopping centre. You can choose from two trips A and B. Which one would you choose?



G



H

References

Bates J.J., Polak J., Jones P., Cook A.J., 2001. The valuation of reliability for personal travel. Transportation Research Part E, 37(2-3), 191-230.

Bhat, C.R., and Sardesai, R., 2006. The impact of stop-making and travel time reliability on commute mode choice. Transportation Research B, 40(9), 709-730.

Black, I.G., and Towriss, J.G., 1993. Demand effects of travel time reliability. Center for Logistics and Transportation, Cranfield Institute of Technology.

Copley, G. Murphy, P., Pearce, D., 2002. Understanding and valuing journey time variability. Proceedings of the European Transport Conference, Cambridge. PTRC, London.

John Bates Services, VU University Amsterdam. Leiden (the Netherlands)

Hamer, R.N., de Jong, G.C., Kroes, E.P., Warffemius, P., 2005. The value of reliability in transport. Outcomes of an expert workshop. RAND Europe Report TR-240-AVV, Leiden (the Netherlands).

Hensher, D.A., 2001. The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications. Transportation, 28(2), 101-118.

Hollander, Y., 2005. The Attitudes of Bus Users to Travel Time Variability. Proceedings of the European Transport Conference, Strasbourg. PTRC, London.

Jong, G. de, Tseng, Y., Kouwenhoven, M., Verhoef, E., Bates, J., 2007. The Value of Travel Time and Travel Time Reliability, Survey Design. Final report prepared for The Netherlands Ministry of Transport, Public Works and Water Management. Significance, John Bates Services, VU University Amsterdam. Leiden (the Netherlands).

Rietveld, P., Zwart, B., Van Wee, B., Van der Hoorn, A., 1999. On the Relationship between Travel Time and Travel Distance in the Netherlands. Annals of Regional Science, 33: 269-287.

Small, K. A., Noland, R. B., Koskenoja, P., 1995. Socio-economic attributes and impacts of travel reliability: a stated preference approach. University of California, Irvine, California PATH research report.

Small, K.A., Noland, R.B., Chu, X., Lewis, D., 1999. Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation. NCHRP Report 431, Transportation Research Board, National Research Council.

Small, K.A., Winston, C. and Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. Econometrica 73: 1367-1382.

Tilahun, N., and Levinson, D., 2005. Travel Time Reliability in Route Choice using Adaptive Stated Preference. Working paper, Department of Civil Engineering, University of Minnesota.

Tseng, Y. Y., Ubbels, B., Verhoef, E., 2005. Value of time, schedule delay, and reliability – Estimation results of a stated choice experiment among Dutch commuters facing congestion. Paper presented at the 45th Congress of European Regional Science Association.

Tseng, Y.-Y., Verhoef, E. T. and Rietveld, P., 2006. Valuation of travel time reliability for railway passengers - A stated choice experiment for Dutch railway travellers. Paper presented at the seminar Infrastructure Reliability TU Delft in June.