Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

# Comparison of Response Bias in an Intercultural Context – Evaluation of Psychological Items in Travel Behavior Research

Miriam Magdolen[a]*, Sascha von Behren[a], Jessica Hobusch[a], Bastian Chlond[a], Peter Vortisch[a]

[a]*Karlsruhe Institute of Technology, Kaiserstrasse 12, 76137 Karlsruhe, Germany*

**Abstract**

The inclusion of attitudinal questions in surveys to determine the effects of the psychological dimension is becoming increasingly important in travel behavior research. However, the occurrence of response bias must be taken into account when determining the psychological characteristics. Especially unmotivated participants have the option to run different answer strategies in item sets that query attitudes. With this study, we analyze variations of response bias and carry out an international comparison to identify cultural differences. For this purpose, we use a standardized survey design conducted in Shanghai (China), Berlin (Germany) and San Francisco (USA) and investigate response behavior to psychological questions. Different measurement methods are used to identify certain response patterns. By modifying an existing algorithm, we combine several aspects of response bias into one indicator. Based on an ordered logit regression, we analyze factors influencing response behavior. As a central outcome, people from Shanghai and San Francisco tend to show more suspicious response behavior than people from Berlin. Since the transferability of surveys has to be seen against the background of different linguistic and cultural interpretations, we investigated these peculiarities through a qualitative follow-up study in China. We conclude that international studies must also take into account social and cultural conditions, since a mere translation of attitudinal questions is not enough. This paper underlines the importance of such studies of response behavior, especially in an international comparison, in order to allow interpretations of survey results. Hints and recommendations for further research are given.


Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

*Keywords:* Response bias; Attitudes; Intercultural comparison; Straightlining; Likert scale

* Corresponding author. Tel.: +49-721-608-47738; fax: +49-721-608-48031.
*E-mail address:* miriam.magdolen@kit.edu

2352-1465 
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

## 1. Introduction

In travel behavior research, the inclusion of attitudes in surveys to determine the impact of the psychological dimension is becoming increasingly common. Furthermore, by means of advanced analytical methods such as Structural Equation Models (SEM) and Hybrid Choice Models (HCM), the relevance of attitudes and norms gains importance. However, the occurrence of response bias must be taken into account when determining and analyzing the psychological characteristics.

Often participants have to answer attitudinal items using Likert Scales, i.e., given options. Since such items are commonly queried in batteries, a grid structure applies. This structure enables the participants to answer the questions quickly and easily. Nevertheless, this also allows quick answering without reading the questions, which leads to falsified data and low quality. Various methods are used in research to identify suspicious response behavior. For example, a frequently used hint to identify bad data and response bias is “straightlining”. Each statement is given the same response category on the Likert Scale, which results in a vertical line. Although it is relatively simple to identify, the occurrence and effects of straightlining are ”understudied” (Kim et al. 2018). In general, there are several methods to measure straightlining, which differ with regard to the definition of non-differentiation and the complexity of the calculation. As far as to our knowledge, only a few surveys identify response bias of people in an international context and therefore investigate this issue in an intercultural comparison.

This paper meets this research gap and presents an analysis of a survey that was undertaken in cities in three different countries: San Francisco (USA), Berlin (Germany), and Shanghai (China). With this research, we aim to investigate whether possible response biases result from different response strategies or if any differences between the cultures can be observed. It is of further research interest to identify the factors influencing response bias. With the standardized psychological item set, which we use for our further analyses, the respondents’ attitudes towards different transport modes as well as their motives for car use were surveyed. By analyzing the available data, we calculate and use indicators for identifying response bias in the attitudinal item set and offer a comparison of response behavior in an intercultural context. Due to the cultural conditions, including languages as well as values in the respective societies, we expect differences in the response behavior. Both content-related peculiarities, such as the tendency to agree or reject all statements, as well as visual bias, e.g., straightlining, are of interest. For this purpose, we present basic measurement methods such as standard deviation as well as methods that combine several types of response bias in one value. Furthermore, the extent of the identified bias is analyzed and discussed with regard to the three different countries. A regression analysis helps to identify characteristics that show effects on the response behavior. Since we expect the greatest difficulties in the survey in China with regard to the intelligibility of the questioned attitudinal items, we investigated this issue in a detailed qualitative analysis. This allows a link between the understanding of the questions and the emerging response bias.

Our paper is structured as follows: First, we review the literature to find out how to measure bias and which studies exist that compare response behavior in different cultural contexts. Second, we calculate various measurements with the data to identify suspicious response behavior. We also compare response behavior between the surveyed cities. Third, we describe the application of a regression analysis to identify the factors influencing response bias. Finally, for a deeper insight into the response behavior of people from China, we present a qualitative follow-up study.

## 2. Literature review

For almost one century, researchers have used Likert Scale formats to quantify the respondents’ attitudes and behaviors in a scientific and validated way (Joshi et al., 2015). In this format, the various items are arranged one below the other, with the different response categories - typically “does not apply” to “applies” - displayed in columns next to them (Loosveldt and Beullens, 2017). The grid format of questions formulated in a Likert Scale might affect respondents to answer the questions of an item block in a particular visual pattern, independent of the item’s content. Consequently, different types of response bias might arise (van Vaerenbergh and Thomas, 2013). In line with this, Weijters et al. (2010) found an impact of scale formats on response behavior and “misresponse to reversed items” when investigating the effects of numbering and labeling of response categories. Further, Schonlau and Toepoel (2015) examined the impact of the participant’s experience itself on straightlining. Respondents who are familiar with answering surveys might either learn how to interpret questions over time and therefore make fewer mistakes or even show more response errors because of a more strategical answer style. They found evidence for the latter hypothesis: straightlining increases with the respondent's experience in answering surveys (Schonlau and Toepoel, 2015). In addition, Loosveldt and Beullens (2017) investigated the impact of interviewers on straightlining and non-differentiation and found evidence of a significant influence of the interviewers in most of their observed European countries. Consequently, besides the respondent's enthusiasm in answering a survey, response style is also biased by situational factors such as face-to-face interviews.

However, in 1988 Flaskerud questioned if the Likert Scale format could also be culturally biased (Flaskerud, 1988). Since then there have been some studies with cross-cultural comparisons between Eastern and Western countries, especially within East Asian and North American cultures. For instance, Chen et al. (1995) surveyed Japanese, Taiwanese, US American and Canadian students to test the relation between the cultural dimension and the response style. Lee et al. (2002) argued that the "construct validity of a Likert-measured variable might be restricted to specific cultures" for which they examined response behavior among Californians who self-identified as Chinese, Japanese and Americans as well as the impact of the number of scale choices. Both found some evidence that East Asian respondents have a greater tendency to choose the midpoint and a weaker tendency to choose extreme responses whereas American respondents were more likely to use extreme values (Chen et al., 1995; Lee et al., 2002). Both studies reinforced a relationship between culture and response style, although Chen et al. (1995) found no evidence that the use of the midpoint or extreme value provides a significant representation for cross-cultural differences. After Hofstede conducted a country-level factor analysis with IBM-employees to examine the level of aggregation in different cultures in the late 1960s, he derived six dimensions that are known as Hofstede's Model. Since then, researchers draw upon the third dimension of Hofstede's Model (Individualism/ Collectivism) to explain the above-mentioned cross-cultural heterogeneity. To characterize the level of people's inclusion into groups, this dimension distinguishes between individualistic and collectivistic cultures (Hofstede, 2011): According to the model, individualists are expected to rather look after themselves. Hence, their self-conception is independent of the social context which often leads to answers in extreme values (Hofstede, 2011; Watkins and Cheung, 1995). By contrast, collectivistic societies are characterized by strong, cohesive in-groups who prefer to rate harmony and modesty (Hofstede, 2011; Watkins and Cheung, 1995). Individual opinions are predetermined by peers which leads to a higher tendency to choose the midpoint in a Likert Scale (Chen et al., 1995; Hofstede, 2011; Watkins and Cheung, 1995). Smith (2004) compares some cross-cultural studies on response tendencies and considers them in the context of communication styles.

While several studies on either the response behavior and its measurement or cross-cultural differences have been conducted, to our status of knowledge there have been little empirical investigations into the combination of both. As psychological item sets are increasingly used in various studies, and international comparisons are carried out, it becomes relevant to compare response data in different cultural contexts on the one hand and to conduct studies on response behavior on the other. This study attempts to close this knowledge gap and thereby provides helpful insights for prospective surveys.

## 3. Data and survey design

The analysis presented in this paper is based on a unique data collection approach, especially in terms of capturing comprehensive information related to individual attitudes and travel behavior. The study was undertaken in Berlin, San Francisco and Shanghai with a principally identical questionnaire and deals with the comparison of urban mobility. Our survey approach used comprises two survey components: travel behavior and psychological factors. The survey collects information on typical activities and mode choices using a travel skeleton approach. To create a cost-effective survey alternative to trip diaries and to keep respondent burden on a reasonable level, we developed a travel skeleton, which focuses on the collection of typical elements of everyday travel. The skeleton provides a reasonable compromise between the level of detail needed and the required effort to survey travel behavior. A quota sampling regarding age, gender, household size, and net income was conducted to develop a representative survey group for each captured city. An access-panel with telephone screening was used to recruit persons. An on-street recruitment was additionally applied in Berlin and San Francisco. The survey aimed to capture behavior and psychological factors of individuals above the age of 17 and, as far as possible, for the whole household. For detailed information as well as applications of the survey data, please refer to von Behren et al. (2018) and Magdolen et al. (2019). In addition, we used standardized psychological item sets to investigate respondents' attitudes towards different transport modes as well as motives for car use. The survey was implemented as a Computer Assisted Personal Interview (CAPI), while the respondents themselves answered the psychological item sets on attitudes and motives. In this section, we describe these psychological items, which are the basis for our subsequent analyses.

Essential elements of mobility cannot be explained solely by observing travel behavior. Our assumption, based on existing research, is that travel behavior can be better understood when we know more about people's attitudes towards different transport modes, and their motivations for car usage. In our study, we used 27 items on attitudes, norms and control beliefs developed and applied by Hunecke et al. (2007). In addition, we used a standardized set of items with eleven questions on motives for car use by Steg (2005). Altogether 38 psychological questions were asked (see Table 1), based on a Likert Scale format. The questions were translated and pre-tested by a professional market research firm (Spiegel Institut). The later field work was carried out by this firm, as well.

Table 1. Psychological Items

| **Used psychological items** | | |
|---|---|---|
| *Psychological constructs* | *Items* | *Questions* |
| Public transportation privacy (PTP) | PTP1 | In public transportation people sometimes come too close to me in an unpleasant manner. |
| | PTP2 | In public transportation my privacy is restricted in an unpleasant manner. |
| Public transportation autonomy (PTA) | PTA1 | I can structure my everyday life very well without a car. |
| | PTA2 | I can take care of what I want to with public transportation. |
| | PTA3 | It is difficult for me to travel the ways I need to go in everyday life with public transportation instead of by car. |
| | PTA4 | If I want, it is easy for me to use public transportation instead of a car to do my things in everyday life. |
| Public transportation excitement (PTE) | PTE1 | I appreciate public transportation, because there is usually something interesting to see there. |
| | PTE2 | I can easily use the traveling time on the bus or train for other things. |
| | PTE3 | I like to ride buses and trains, because I don't have to concentrate on traffic while doing so. |
| | PTE4 | I can relax well in public transportation. |
| Public transportation intension (PTI) | PTI1 | It is my intention to use public transportation instead of a car for the things I do in everyday life. |
| | PTI2 | I have resolved to travel the ways I need to go in everyday life using buses and trains. |
| Subjective norm (SN) | SN1 | People who are important to me think it is good if I would use public transportation instead of a car for things I do in everyday life. |
| | SN2 | People who are important to me think that I should use public transportation instead of a car. |
| Personal norm (PN) | PN1 | Due to my principles, I feel personally obligated to use eco-friendly means of transportation for the things I do in everyday life. |
| | PN2 | I feel obligated to make a contribution to climate protection via my choice of transportation. |
| Car excitement (CE) | CE1 | Driving a car means fun and passion for me. |
| | CE2 | Driving a car means freedom to me. |
| | CE3 | When I sit in the car I feel safe and protected. |
| | CE4 | Being able to use my driving skill when driving a car is fun for me. |
| Perceived mobility necessities (PMN) | PMN1 | My everyday organization requires a high degree of mobility. |
| | PMN2 | I constantly have to be mobile in order to comply with my everyday obligations. |
| Bicycle excitement (BE) | BE1 | I like to be out and about by bike. |
| | BE2 | I can relax well when riding a bike. |
| | BE3 | I ride a bicycle because I enjoy the exercise. |
| Weather resistance (WR) | WR1 | I don't like to ride my bike when the weather is cool. |
| | WR2 | I also ride my bike when the weather is bad. |

| | | |
|---|---|---|
| Car use motive (CM) | CM1 | I feel free and independent when I drive a car. |
| | CM2 | A car can communicate status and prestige. |
| | CM3 | The characteristics of a car can show who and what I am. |
| | CM4 | It doesn't matter to me what vehicle type I drive. |
| | CM5 | The functioning of a car is more important to me than the make of a car. |
| | CM6 | A car is primarily a means to an end for me. |
| | CM7 | I like to drive a car. |
| | CM8 | There are dream cars that I would like to drive once. |
| | CM9 | You can draw conclusions about a person from the car. |
| | CM10 | The make of a car is important to me. |
| | CM11 | I only use a car to get from A to B. |

Likert Scale: 1 = does not apply; 2 = rather does not apply; 3 = applies in part / does not apply in part; 4 = rather applies; 5 = applies

As we want to compare the data from the three cities with each other in the following, the different characteristics of the samples have to be taken into account. In Table 2 sociodemographic characteristics of the participants from San Francisco, Berlin and Shanghai, are analyzed. In terms of gender, we see a smaller number of male participants in Shanghai compared to the other two cities. Nevertheless, the level is roughly the same. Regarding age, the sample is younger in Shanghai. In San Francisco, in contrast, 25% of the participants are 60 years or older. In Berlin, a particularly large number of participants with a Bachelor’s degree or higher is noticeable. When comparing the occupation, we see high rates of full-time and part-time employment in all three cities. In the sample from Shanghai, we find the most retired people, whereas in Berlin the proportion of students is the highest. Differences in the characteristics of the participants from the three cities become clear. The particularities of the samples should be considered in the interpretation of the following analyses.

Table 2. Sociodemographic characteristics of the sample

| | | **Shanghai** | **San Francisco** | **Berlin** |
|---|---|---|---|---|
| *Sample Size* | | *600* | *600* | *600* |
| **Characteristics in %** | | | | |
| Gender | Male | 44.33 | 50.50 | 48.50 |
| Age | Younger than 35 | 34.67 | 28.50 | 33.50 |
| | Older than 59 | 14.00 | 25.00 | 23.00 |
| Education | Bachelor degree or higher | 67.17 | 51.83 | 75.50 |
| Occupation | Full-time + part-time employed | 62.83 | 56.33 | 54.17 |
| | (Early) Retired | 25.50 | 19.50 | 18.33 |
| | Student | 7.17 | 7.67 | 12.50 |

## 4. Identification of response bias

With the available set of 38 items, we are able to calculate indicators that quantify straightlining in response behavior. It is important not only to identify pure straightlining but also generally suspicious response patterns. For example, a person who answers 37 out of the 38 items on the same response category shows no pure straightlining. However, this response behavior is unlikely and should be classified as poor data quality. Furthermore, the consideration of reversed questions in our data set is essential. These reversed questions, which address the same aspects as the corresponding questions, are helpful in identifying contradictions as they are formulated in an opposite manner compared to the original questions. As the survey was conducted in the three cities Berlin, Shanghai and San Francisco, the data offer the prerequisites for a direct intercultural comparison. In the following, we present a descriptive analysis of the response behavior in the three cities. First, we examine the mean values of each item in the cities to identify indications for differences between them. The used indicators for measuring response bias and their calculation are described in the following section. Based on this, we compare the response behavior of the people from the different cultural contexts and discuss main results.

### 4.1. Differences in the responses of the items

An indication for the usefulness of our study, especially the comparison of responses in the three different cities, is represented in the means of the answers per item. Divergent values of the participants of one city in comparison to the others serve as a signal of a different understanding and interpretation of the questions. However, different answers can also be related to certain response behavior and response patterns. In this case, the content may not play the primary role in answering the attitudes. By examining the mean values per item, we determine whether differences between the cities are observable and whether an international comparison is a suitable approach in the following. The mean values of each item are shown in Table 3. The items were surveyed by using a 5-point Likert Scale, in which 1 indicates “does not apply” and 5 “applies”. Therefore, the higher the mean value, the higher the agreement to the statement in the respective city. “No Answer” was not taken into account in the calculation of the mean values. As a result, the sample sizes vary between 337 and 599 within the cities, depending on the item and city.

The list reflects the original order in the survey, which is adapted from the studies by Hunecke (2007). In the item set, several questions refer to the same psychological construct. For example, the items WR1 and W2 both represent the construct *Weather resistance* (see Table 1). In the standardized questionnaire, the items of one psychological construct are not queried directly one after another. This serves on the one hand to maintain the attention of the participants and on the other hand to prevent the reversed questions, which can be used for control purposes, from being too obvious. The mixed order does not apply to the eleven motives for car use. These were asked directly one after another in a separate item battery. Since the attitudes and motives were listed in sequence in the questionnaire, we assume a joint grid structure for further analyses.

Table 3. Mean values of the items in each city

| **Items** | | **Shanghai** | **San Francisco** | **Berlin** |
|---|---|---|---|---|
| Perceived mobility necessities 2 | PMN2 | 4.05 | 2.64 | 3.42 |
| Public transportation autonomy 1 | PTA1 | 3.63 | 3.25 | 3.81 |
| Subjective norm 1 | SN1 | 3.42 | 2.35 | 2.96 |
| Public transportation excitement 2 | PTE2 | 3.26 | 2.73 | 3.43 |
| Public transportation privacy 1 | PTP1 | 3.12 | 2.25 | 3.06 |
| Bicycle excitement 1 | BE1 | 3.43 | 2.14 | 3.70 |
| Public transportation excitement 3 | PTE3 | 3.22 | 2.69 | 2.76 |
| Car excitement 1 | CE1 | 2.99 | 2.93 | 2.96 |
| Public transportation intension 1 | PTI1 | 3.12 | 2.57 | 3.29 |
| Bicycle excitement 2 | BE2 | 3.50 | 2.20 | 3.40 |
| Public transportation autonomy 2 | PTA2 | 3.63 | 2.92 | 3.73 |
| Perceived mobility necessities 1 | PMN1 | 3.88 | 2.69 | 3.44 |
| Public transportation excitement 1 | PTE1 | 3.51 | 2.73 | 2.91 |
| Weather resistance 1 | WR1 | 3.82 | 1.98 | 3.83 |
| Car excitement 2 | CE2 | 3.75 | 3.84 | 3.50 |
| Public transportation intension 2 | PTI2 | 3.57 | 2.56 | 3.02 |
| Public transportation excitement 4 | PTE4 | 3.53 | 2.63 | 3.48 |
| Car excitement 3 | CE3 | 3.25 | 3.11 | 3.27 |
| Public transportation autonomy 3 | PTA3 | 2.89 | 2.48 | 2.23 |
| Personal norm 1 | PN1 | 3.90 | 2.79 | 3.25 |
| Car excitement 4 | CE4 | 3.07 | 2.92 | 3.06 |
| Public transportation privacy 2 | PTP2 | 2.89 | 2.15 | 2.73 |
| Public transportation autonomy 4 | PTA4 | 3.56 | 2.98 | 3.92 |
| Bicycle excitement 3 | BE3 | 3.63 | 2.13 | 3.63 |

| | | | | |
|---|---|---|---|---|
| Personal norm 2 | PN2 | 3.92 | 2.82 | 3.26 |
| Weather resistance 2 | WR2 | 2.49 | 1.65 | 2.58 |
| Subjective norm 2 | SN2 | 3.29 | 2.29 | 2.64 |
| Car use motive 1 | CM1 | 3.31 | 3.28 | 3.30 |
| Car use motive 2 | CM2 | 3.48 | 2.82 | 3.28 |
| Car use motive 3 | CM3 | 3.34 | 2.58 | 2.33 |
| Car use motive 4 | CM4 | 2.89 | 2.60 | 3.33 |
| Car use motive 5 | CM5 | 3.81 | 3.61 | 4.20 |
| Car use motive 6 | CM6 | 3.78 | 2.86 | 4.15 |
| Car use motive 7 | CM7 | 3.01 | 3.40 | 3.64 |
| Car use motive 8 | CM8 | 3.43 | 3.37 | 2.78 |
| Car use motive 9 | CM9 | 3.37 | 2.79 | 3.14 |
| Car use motive 10 | CM10 | 3.40 | 3.02 | 2.36 |
| Car use motive 11 | CM11 | 3.14 | 3.06 | 3.87 |

Items in original order of the survey

When looking at the mean values in Table 3, there is a variation of the values across the 38 examined items and the three cities. The very first item PMN2 (*Perceived mobility necessities*) shows large differences. In Shanghai, there is a high agreement with the statement while people from San Francisco tend to reject this item. We see similar results at PMN1, PN1, and PN2 (*Personal norm*). Overall, it appears that lower categories were chosen in San Francisco. CE2 (*Car excitement*) is the only item, where people from San Francisco have the highest mean value compared to Shanghai and Berlin. We also observe very low values for WR1 and WR2 (*Weather resistance*) in San Francisco. The comparatively low values for BE1, BE2, and BE3 (*Bicycle excitement*) are therefore not surprising. In Berlin, we mainly find averages above 3.0, which indicates a tendency towards more agreement with the statements. The highest values are CM5 and CM6 (*Car use motives*), the lowest values are CM3 and PTA3 (*Public transportation autonomy*).

The diverging mean values illustrate differences between the three surveyed cities. This may be due to cultural influences and socialization, e.g., the status of the car in the USA. In addition, differences in the sociodemographic characteristics should be considered (see Table 2). Different answers can also be based on the wording of the questions and the translation from German into English and Chinese, which may involve a change in meaning. A further reason could also be different response behavior in the three countries. People from one country may tend to specific response patterns independent of the content of the statements. In order to investigate these possible explanations, we present in the next sections an approach for the detection of response bias in an international context.

### *4.2. Measurement of response bias*

The grid format of a Likert Scale makes it easy to reveal special response patterns. There are numerous response patterns, including straightlining, which is defined as the tendency to answer all items within a block of questions identically – independent of the item's content (Loosveldt and Beullens, 2017). In this respect, straightlining can arise in different patterns and strengths. Pure straightlining "occurs when the selected answers are in a perfect vertical [straight] line", representing no dissimilarity (Kim et al., 2018; Loosveldt and Beullens, 2017; Schonlau and Toepoel, 2015). For the present research, we decided to focus primarily on three variations of straightlining: *Acquiescence*, which means that the respondent chose response category 5 (applies) all the time; *Disacquiescence*, which is the opposite of *Acquiescence*; and *Central Tendency*, which is defined as a vertical line in the absolute middle of the response categories. As it is highly unlikely that respondents answer 38 items identically, we broaden the conditions mentioned above to identify even weaker signs of response bias. Response categories were clustered together and named *Near Acquiescence* (4+5), *Near Central Tendency* (2+3+4) and *Near Disacquiescence* (1+2). Table 4 gives an overview of the observed response styles.

Table 4. Response styles

| **Response Style** | **Definition** | **Response Category** | **Pattern** |
|---|---|---|---|
| Disacquiescence | Tendency of total disagreement | 1 | ●○○○○ |
| Near Disacquiescence | Tendency of mild disagreement | 1+2 | ●●○○○ |
| Central Tendency | Tendency to choose the middle response category | 3 | ○○●○○ |
| Near Central Tendency | Tendency to never choose extreme response categories | 2 + 3 + 4 | ○●●●○ |
| Acquiescence | Tendency of total agreement | 5 | ○○○○● |
| Near Acquiescence | Tendency of mild agreement | 4+5 | ○○○●● |

We assume that a straightlining pattern does not automatically imply a response bias but might also reveal the true respondent's attitude. Schonlau and Toepoel (2015) distinguish between plausible and implausible straightlining. The former occurs when the given answers within a question block may indeed be appropriate for the respondents. However, some item blocks contain reversed questions, which are formulated in the opposite way as the previous ones. In that case, a straightlining pattern is implausible, especially for the response styles *Acquiescence* and *Disacquiescence*.

According to Loosveldt and Beullens (2017), there is "no evident simple and univocal measurement" for straightlining. Therefore, we decided to apply various simple and complex methods and indicators to quantify response bias. Table 5 gives an overview of the indicators applied. In the following, it has to be differentiated between two types of indicators: indicators that are calculated with the value of response, e.g., *Standard Deviation*, and indicators that are based on a visual pattern in the grid, e.g., *Maximum Sequence*. The latter describes differences between the previous, the present and the following answer. For all measurements and subsequent analyses, we used tools and procedures given in SAS 9.

To identify response bias, we calculated different indicators (see Table 5). The first indicator measures the *Standard Deviation* for each respondent. A low *Standard Deviation* indicates straightlining: the selection of the same response category across all items results in a *Standard Deviation* of 0. The second indicator *Deviation Previous* describes the average distance between two subsequent answers. A result of 0 demonstrates that the respondent answered all items the same (Loosveldt and Beullens, 2017). To examine the tendencies to agree or disagree, we quantified as a third indicator for each person how often he or she chose each response category. It was calculated for each response style described in Table 4. For example, the response style *Central Tendency* investigates the proportion of items, where response category 3 was chosen. Therefore, the order of the responses was not considered. A high value in a response style indicates a higher tendency towards straightlining and therefore response bias.

Leiner (2013) developed an algorithmic measurement in the manner of a utility analysis based approach to detect visual patterns in Likert Scale formats by a scoring system. For the measure, he assigns "(penalty) points" to each person. The individuals get one point if two subsequent items receive the same answer, which detects straightlining patterns. Further, individuals get one more point if the change between subsequent items is the same as the recent change to detect diagonal lines. Finally, half a point is distributed if the change is the same as the next-to-recent change to detect left-right clicking. The points are accumulated for each person. In our study, the maximum score for the indicator *Algorithmic Measure* is 38 points. Since some of the items in our questionnaire represent reversed questions, we also took them into account when identifying response bias. Therefore we developed a modification of the *Algorithmic Measure*. In order to punish respondents who answered the integrated reversed questions in an opposite way than the corresponding question, additional penalty points were awarded. This kind of biased answering indicates a poor reading or understanding of the statements. In the case of reversed questions straightlining is largely implausible and indicates poor data quality (Schonlau and Toepoel, 2015). The items within the constructs *Weather resistance* (WR1 and WR2) and *Public transportation autonomy* (PTA2 and PTA3) are formulated reversely. The agreement on one item indicates a rejection of the other and vice versa. Therefore, respondents that agreed or disagreed both items within the constructs received 2 penalty points. 5 penalty points were distributed if the items within the psychological constructs *Subjective norm* (SN1 and SN2), *Personal norm* (PN1 and PN2) and *Public transportation privacy* (PTP1 and PTP2) were not both agreed or disagreed. In these constructs, the questions have the same orientation. Therefore, both items should be agreed or rejected for a consistent response. With the extra punishment, we extended the *Algorithmic Measure* with 19 additional points. The maximum value of the *Modified Algorithmic Measure* is therefore 47. The last indicator, *Maximum Sequence*, counts the maximum number of subsequent items with the same response for each individual. Thus, the absolute value is used, and the order of responses is considered (Loosveldt and Beullens, 2017). A result of 38 indicates the same response of all items, no matter which response category was selected.

Table 5. Indicators used to identify response bias

| **Measurement** | **Description** |
|---|---|
| Standard Deviation | Degree to which respondents differentiate between questions |
| Deviation Previous | Average distance between two subsequent answers. The order of responses is considered |
| Tendency to Agree/ Disagree | Frequencies of the categories chosen per respondent |
| Algorithmic Measure | Algorithmic measure based on a scoring system developed by Leiner:<br>1 point if two subsequent items receive the same answer<br>1 point if the change between subsequent items is the same as the recent change<br>0.5 points if the change is the same as the next-to-recent change |
| Modified Algorithmic Measure | An expanded scoring system by penalty points for control questions, in addition to Leiner's algorithmic measure described above:<br>2 penalty points if the items WR1 & WR2 and PTA2 & PTA3 were not answered in a consistent manner<br>5 penalty points if the items SN1 & SN2, PN1 & PN2 and PTP1 & PTP2 were not answered in a consistent manner |
| Maximum Sequence | Maximum number of subsequent items with the same response. The order of responses is considered |

Sources: (Kim et al., 2018; Leiner, 2013; Loosveldt and Beullens, 2017)

One main issue in examining response bias in Likert Scale item sets is the response category "No Answer", which is given in our survey as the sixth category next to the 5-Point Likert Scale. When analyzing the visual patterns of the responses, the contextual significance of "No Answer" plays a minor role. Primarily, it is another response category in which a certain pattern may occur. Therefore, the indicators that examine visual patterns (*Deviation Previous*, *Algorithmic Measure*, *Modified Algorithmic Measure* and *Maximum Sequence*) include "No Answer" as a sixth category. For indicators that consider the values of the response category, the option "No Answer" was ignored in our analysis. The total exclusion of respondents from our analysis who chose "No Answer" was not useful, as a high number of respondents would have been eliminated. In addition, respondents with high proportions of "No Answer" could be important for our cross-cultural comparison later in this study.

### *4.3. Comparison of measured response bias between the cities*

We applied the indicators and measurements described above to our data set. As a result, each individual obtained a value for each indicator. In order to determine whether there are differences between the response behavior of the participants from Shanghai, San Francisco and Berlin the mean value of all participants within each city was calculated. Table 6 shows the results. In addition to the mean value, the 25% and 75% quantiles are given for each city to reflect the dispersion of the values. Furthermore, a cut off is specified for each indicator. The cut off defines the threshold at which the answers are likely to be of low quality (Leiner, 2013). In our case, we chose the median value of the total sample as a cut off value. This does not apply to the *Standard Deviation* (cut off < 1.0) and the *Mean Response of all Items* (cut off < 3.0). For each city, the percentage of the 600 respondents with a worse value than the cut off is given.

Both the *Standard Deviation* and *Deviation Previous* are lowest among people from Shanghai. In comparison, people from Berlin show a higher variance in their answers. The examination of the *Maximum Sequence* identifies participants from San Francisco with the highest average value. 59% have more than four identical successive responses. In Shanghai it is 51%, and in Berlin it is only 27% of the participants.

Table 6. Response bias indicators

| | Shanghai | | San Francisco | | Berlin | |
|---|---|---|---|---|---|---|
| *Response Bias Measurements* | Mean / Proportion | 25% Q1 - 75% Q3 | Mean / Proportion | 25% Q1 - 75% Q3 | Mean / Proportion | 25% Q1 - 75% Q3 |
| Standard Deviation | 1.04 | (0.81; 1.27) | 1.4 | (1.20; 1.68) | 1.39 | (1.17; 1.61) |
| Cut Off < 1.0 | 42.50% | | 14.33% | | 8.83% | |
| Deviation Previous* | 1.24 | (0.95; 1.51) | 1.34 | (1.03; 1.73) | 1.54 | (1.27; 1.81) |
| Cut Off < 1.41 | 64.33% | | 49.00% | | 34.33% | |
| Maximum Sequence* | 5.85 | (3; 8) | 7.27 | (4; 9) | 3.78 | (3; 4) |
| Cut Off > 4.0 | 51.00% | | 59.00% | | 22.17% | |
| Mean Response on all Items | 3.45 | (3.14; 3.76) | 2.75 | (2.37; 3.16) | 3.25 | (2.97; 3.53) |
| Cut Off < 3.0 | 15.17% | | 65.00% | | 26.67% | |
| Algorithmic Measure* | 15.68 | (12.5; 19.0) | 17.28 | (12.5; 21.0) | 12.71 | (10.5; 14.5) |
| Cut Off > 14 | 56.33% | | 60.50% | | 30.67% | |
| Modified Algorithmic Measure* | 18.54 | (14.5; 22) | 20.85 | (15.5; 25.25) | 14.87 | (12.0; 17.5) |
| Cut Off > 17 | 56.00% | | 62.83% | | 26.00% | |
| *Tendencies to Agree or Disagree (Average % of all Items)* | | | | | | |
| Acquiescence | 14.96% | (3%; 24%) | 21.85% | (8%; 32%) | 26.62% | (13%; 37%) |
| Cut Off > 30% | 15.83% | | 28.17% | | 41.17% | |
| Near Acquiescence | 49.34% | (37%; 61%) | 32.85% | (21%; 45%) | 46.30% | (37%; 55%) |
| Cut Off > 30% | 87.00% | | 56.17% | | 89.83% | |
| Central Tendency | 17.25% | (5%; 26%) | 16.19% | (3%; 24%) | 18.80% | (11%; 26%) |
| Cut Off > 30% | 18.83% | | 17.33% | | 16.00% | |
| Near Central Tendency | 62.05% | (47%; 79%) | 40.89% | (18%; 61%) | 51.98% | (34%; 68%) |
| Cut Off > 30% | 88.17% | | 62.00% | | 78.67% | |
| Disacquiescence | 8.77% | (0%; 11%) | 33.27% | (13%; 50%) | 17.54% | (8%; 26%) |
| Cut Off > 30% | 8.67% | | 50.17% | | 17.00% | |
| Near Disacquiescence | 19.17% | (8%; 29%) | 46.90% | (32%; 61%) | 31.07% | (21%; 39%) |
| Cut Off > 30% | 21.17% | | 77.83% | | 50.00% | |
| No Answer | 14.20% | (0%; 29%) | 4.07% | (0%; 0%) | 3.91% | (0%; 3%) |
| Cut Off > 30% | 21.17% | | 5.17% | | 2.50% | |

* No Answer (NA) is included as a sixth response category

We also examined the average answer to all 38 items within each city. For San Francisco, a value of 2.75 indicates the tendency to reject the questioned statements. The values above 3.0 in Shanghai and Berlin indicate a tendency to agree with the attitudinal items. The *Algorithmic Measure*, as well as the *Modified Algorithmic Measure*, also identify people from San Francisco with the highest values. Both indicators combine different aspects of response bias into one value (see Table 5). In the *Algorithmic Measure*, 14 penalty points out of 38 possible points are seen as cut off. In the *Modified Algorithmic Measure*, it is 17 out of 47. The results are, as expected, very similar. More than 60% of the people from San Francisco hit these cut off values. While it is also about 56% of the respondents from Shanghai, the most participants from Berlin have values below the cut off and show on average the fewest penalty points. The *Modified Algorithmic Measure* is of special interest. In addition to the more or less visual response bias (e.g., identical successive responses), it takes contradictive answers on similar statements into account. The additional penalty highlights the differences between the cities. We see an increase in values, which is greatest in San Francisco. There, the participants answered the most contradictorily compared to Shanghai and Berlin. The smallest increase in the *Modified Algorithmic Measure* is observed for people from Berlin, who continue to show the lowest level of penalty points. The differences in the mean values of the measurement indicators may also be due to different market research methods and sampling strategies in the three cities. However, these effects cannot be separated from those of the cross-cultural impacts.

The second part of Table 6 shows the *Tendencies to Agree or Disagree* with the statements. The terms are defined in Table 4. 30% was chosen as cut off for all response tendencies, as we regard this as a critical value. People with a value above 30% have given an identical answer to at least 12 of the 38 questions. The results indicate a *Central Tendency* of the people from Shanghai. On average, 17.25% of the responses are given in category 3. Taking the responses in the categories 2, 3 and 4 (*Near Central Tendency*) into account, the mean proportion is over 60%. People from Shanghai, therefore, tend to select the extreme response categories rarely. The participants from the USA show the largest values for *Disacquiescence* and *Near Disacquiescence* in the international comparison. The participants tend to reject the questioned statements. In Berlin, we find the largest proportion of "Yes-Sayers". On average, 26.62% of the given responses are in category 5. Examining the indicator *Near Acquiescence*, we see a strong increase in Shanghai. With an average of 49.34% of all answers, the value is on the same level as in Berlin (46.30%). This reflects a high selection of category 4 as a response in Shanghai. In addition to the five categories (does not apply - applies), there was also the option "No Answer". We also examined how often this answer was chosen on average. The results show the highest share in Shanghai. On average, 14.20% of the responses are "No Answer". In Berlin and San Francisco, it is only about 4%.

The examination of *Acquiescence, Central Tendency* and *Disacquiesence* identifies a certain tendency for each city. The tendency of the participants from Shanghai to select central categories, as well as the tendency of the people from San Francisco to select extreme response categories (*Disacquiescence*) are in line with the results from previous literature (Chen et al., 1995; Lee et al., 2002). The calculated indicators to identify response bias in the item set show people from Shanghai and San Francisco to be more likely to follow certain patterns. Above all, we see higher values than in Berlin for the *Modified Algorithmic Measure,* which takes different types of response bias into account.

## 5. Regression analysis

After calculating and interpreting different measures, we consider the *Modified Algorithmic Measure* for more in-depth analysis, as it accounts for different types of response bias. For the analysis, we split up the *Modified Algorithmic Measure* value range into "low", "medium" and "high". We named the resulting variable *Leiner Value*. For a high *Leiner Value,* respondents need a *Modified Algorithmic Measure* score greater than 26.0 and for a low *Leiner Value* a score less than 19.0. We applied an ordered logit model to examine influencing factors on response behavior. In our case, the dependent variable is the calculated *Leiner Value*, which describes suspicious response behavior comprehensively. As the *Leiner Value* is an ordinal scale, the application of an ordered logit model ensures interpretability. We analyzed the effects of person and household related variables (i.e., gender, age, occupation status, educational level as well as household income, car availability, household type) coded as dummy variables (see Table 7). The variable *LowEducated_dummy* describes people with a vocational qualification or less. *HouseholdType1_dummy* represents a household with 1 or 2 persons, with at least one professional and no children. *HouseholdType2_dummy* has the same characteristics but without employed people in the household. In addition, we also consider a city-specific variable and examine whether it has an influence that people come from San Francisco or Shanghai in comparison to Berlin. We have also considered an interview specific variable that describes how many interviews an interviewer has conducted in the survey. Interviewers with more than 80 interviews are more practiced and get a value of 1 at *InterviewerExperience_dummy*.

Table 7 illustrates results including estimates, significance levels, and goodness-of-fit values. Positive estimates imply that high variable values increase the utility and tend to a lower *Leiner Value*. Negative estimates imply that high variable values reduce the utility and thus a higher *Leiner Value* is likely. As in the descriptive analysis, the significant role of the city-specific context becomes evident in the regression model. Both variables: *Shanghai_dummy* and *SanFrancisco_dummy* have a negative significant value and increases the probability of suspicious response behavior with a high *Leiner Value* in comparison with Berlin.

Table 7. Ordered logit results

| **Variables** | **Estimates** | | **Standard Error** | **Wald Chi-Square** |
|---|---|---|---|---|
| ***Intercepts*** | | | | |
| Intercept 1 | 0.4372 | * | 0.2103 | 4.3232 |
| Intercept 2 | 2.2717 | *** | 0.2171 | 109.4728 |
| ***Countries*** | | | | |
| Shanghai_dummy | -2.0189 | *** | 0.1978 | 104.2245 |
| SanFrancisco_dummy | -2.554 | *** | 0.202 | 159.849 |
| ***Person-specific variables*** | | | | |
| Retired_dummy | -0.9298 | *** | 0.1582 | 34.5354 |
| Male_dummy | 0.1731 | | 0.1016 | 2.903 |
| Ageunder35_dummy | 0.2473 | * | 0.1226 | 4.0675 |
| LowEducated_dummy | -0.39 | | 0.2426 | 2.5837 |
| Smartphone_dummy | 0.3916 | * | 0.1537 | 6.4921 |
| ***Household-specific variables*** | | | | |
| MiddleIncome_dummy | 0.292 | * | 0.1184 | 6.0819 |
| HouseholdType1_dummy | 0.425 | ** | 0.1294 | 10.7866 |
| HouseholdType2_dummy | 0.9006 | *** | 0.1672 | 29.0259 |
| CarOwnership_dummy | 0.5197 | *** | 0.116 | 20.0575 |
| ***Interview-specific variables*** | | | | |
| InterviewerExperience_dummy | 1.0192 | *** | 0.1691 | 36.3322 |
| **Goodness-of-fit values** | | | | |
| McFadden Pseudo $R^2$ | 0.2 | | | |
| McFadden Adjusted Pseudo $R^2$ | 0.24 | | | |
| AIC | 3,268.94 | | | |
| N | 1,800 | | | |

***,**,* = significance at 0,1%, 1%, 5%

When considering the person-specific variables, results show a higher probability of a high *Leiner Value* for retired persons. Older people may not completely understand some attitudinal questions about modes (e.g., bicycles) because they cannot use them anymore or have been socialized without bicycle use. However, due to social pressure in the interview, they may still answer the questions. In our study, gender and educational level have no significant influence on response behavior. People from households without employed people have a positive effect. Suspicious response behavior is less likely in their case. An interpretation of the result could be a high motivation by the incentives given to participate in the survey. Finally, we examine the interview-specific variable. Interviews with an experienced interviewer have a positive influence on the response behavior. One reason could be that they make the interview less monotonous and the respondents remain more motivated. Another reason could be the support, as the interviewers are more helpful with difficult attitudinal questions. Further effects in combination with the interviewer and completion time as well as the survey methods used cannot be considered in our analysis. The respective data are missing.

## 6. Qualitative evaluation of response bias in China

Based on the descriptive analyses and the regression model, a particular response behavior is visible in Shanghai. In contrast to the USA, surveys conducted by foreign institutions are less tested in China. We see the possibility that the Chinese may not have understood the questions correctly by reasons of a not unambiguous or misleading translation. For the answers to be internationally comparable, it must be ensured that the questions are also understood in China. A unique requirement for surveys in China is the symbolism in the Chinese language and the associated meaning for people. These linguistic peculiarities make it challenging to translate and adopt the meaning of the question. For this reason, we have decided to conduct a follow-up study in China.

From the results, we conclude additional understanding issues with the attitudinal questions in China. To investigate this, a survey was carried out in which the respondents were asked if they understand the questions or not (Wang, 2018). This survey contains 55 participants. The 38 psychological questions were given to the participants. The respondents completed paper or digital questionnaires by themselves. These surveyed Chinese people were not involved in the international survey from Shanghai and were randomly chosen. A certain variance in age and gender is present in the sample. However, no representative conclusions can be made on the sample size.

Based on an adapted 5-Point Likert Scale from “do not understand at all” to “understand clearly”, the people had to provide information if they understood the respective question. In our analyses, we investigated an average intelligibility degree. Questions with a low degree are difficult to understand for Chinese people. Questions with a value near 5 are easy to understand. Results show a low intelligibility degree for *Subjective Norm* (SN) and *Perceived mobility necessities* (PMN). These items are not easy to understand and have different meanings for Chinese people. SN1 has a mean value of 3.75 and SN2 has a value of 3.91. These values are between “half/half” (3.0) and “can understand” (4.0). PMN1 and PMN2 also have both comparatively low intelligibility degrees: 3.96 and 3.93.

The low values can have several causes. Cross-cultural misunderstandings of the questionnaire could be summarized in three categories: semantic inequivalence (e.g., non-equivalent words), idiomatic inequivalence (e.g., long and winding sentences), and experimental inequivalence (e.g., situations not appropriate to the Chinese cultural context). To get more detailed information, 24 people from the survey sample were asked in a web-based interview, why they do not understand the questions in detail. This qualitative analysis provides further insights. Altogether 17 of 24 respondents chose at least one of the misunderstanding categories for SN1: “People who are important to me think it is good if I would use public transportation instead of a car for things I do in everyday life”. In the interview, the participants pointed out why they did not understand this question. Five people could not capture the meaning of "people who are important to me". Seven people were completely confused at the sight of this long and complex sentence. In addition, seven respondents chose the answer “can understand”, but considered this point to be different from Chinese social and cultural practice. The respondents assign parents, close relatives and good friends as "to me important people". However, there is no causal relationship between the influence of these people and the respondents’ travel mode choices. In China, parents in general have high importance for their children, but not especially for their mode choice decisions. Based on the qualitative analysis it can be concluded that some questions were not clearly understood. This can cause bias in the response behavior. In particular, questions on norms are understood differently by the Chinese, since the social norm has a fundamentally different effect on people than in Germany and the US. This qualitative analysis is intended to provide a small insight into the degree of intelligibility. For a comprehensive analysis, this paper is not sufficient, and we propose this issue for future research. However, it becomes clear that certain survey formats and questions are not “transferable” from one cultural context to another and therefore have to be compared with a certain critical distance.

## 7. Conclusions

The study of attitudes in travel behavior research has become increasingly important in recent years. As already mentioned, this development is supported by various model applications such as HCM or SEM. However, the use of items sets with Likert Scales for measuring psychological factors must also be considered critically, as their structure may cause bias in response. Especially unmotivated participants can drive different answer strategies in these item sets. However, the bias in responses may also be due to different cultural conditions and contexts. This should be taken into account especially in international comparative studies. With our study, we aimed to analyze these issues in more detail. For this purpose, we used a standardized survey conducted in Shanghai, Berlin and San Francisco and investigated response behavior to psychological questions in an international comparison.

Different measurement methods were used to identify certain response patterns. This allowed us to measure various degrees of response bias. Visual response patterns, as well as discrepancies regarding the given responses within psychological constructs and on reversed questions, were found. By modifying Leiner's algorithm (2013), we were able to combine multiple aspects of response bias into one indicator. This, as well as other examined indicators, show different values in the three cities. Participants from San Francisco and Shanghai show more suspicious response patterns than people from Berlin. In addition, the study shows the highest level of *Central Tendency* among people from Shanghai. Participants from San Francisco tend to disagree with the surveyed items, while the participants from Berlin tend to agree with them.

To understand which characteristics influence the response behavior, we used an ordered logit regression. Again, people from Shanghai and San Francisco tend to show more suspicious response behavior than people from Berlin. A consideration of sociodemographic factors reveals a higher probability for response bias among retired person. An interesting result is also made visible by the influence of the interviewer. Interviewers with more than 80 interviews in the survey have a positive influence on data quality. Data from these interviewers are less likely to be suspicious of response bias. However, we cannot verify why the interviewers conducted different numbers of interviews. Some interviewers may have been removed from the survey due to lower response quality.

As a result, our research underlines the importance of such studies about response bias, especially the international comparison of response behavior, which is necessary to allow for comparisons of the survey results between cities from different countries. With our outcomes, we provide a further contribution to the understanding of data quality in international surveys. We were able to identify differences in the response behavior of people from different cultural backgrounds. Although the results cannot be generalized at country-level, as the selected cities are not representative for these countries, they show strong international differences. Cultural influences, socialization, and social desirability are reasons for this. Furthermore, pure translations of questionnaires may not reflect the content of the questions in other languages. Surveys should therefore also be adapted to the cultural characteristics in the surveyed countries. This includes thorough pretests of the survey instruments. In addition, it is necessary to consider different response behavior in each country and culture, since it can lead to bias in the answers. Sampling approaches and survey procedures, e.g., incentives, vary between countries: Although the same market research company was in charge of the surveys, an effect of different methods in the respective cities is likely. We also assume different levels of familiarity with surveys in general in different cultural contexts. However, these various effects are difficult to identify and to separate from each other.

Altogether, it must be noted that further research is necessary for international comparative studies. It should also be investigated how incentives work in different cultural contexts. Do they have a similar influence on the response behavior? In addition, the response time of the participants should also be compared, as already observed in other previous studies. This can provide an indication of response bias, as reading and answering the questions take some time.

## Acknowledgements

This paper presents analyses based on the Urban Travel Monitor (UTM) funded by the BMW Group, Munich. Zi Wang supported the implementation, execution, and analysis of the qualitative follow-up study as part of her master thesis.

## References

Chen, C., Lee, S.-y., Stevenson, H.W., 1995. Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students. Psychol Sci 6 (3), 170–175. 10.1111/j.1467-9280.1995.tb00327.x.

Flaskerud, J.H., 1988. Is the Likert Scale Format Culturally Biased? Nursing Research 37 (3), 185-186. 10.1097/00006199-198805000-00013.

Hofstede, G., 2011. Dimensionalizing Cultures - The Hofstede Model in Context. Online Readings in Psychology and Culture 2 (1). 10.9707/2307-0919.1014.

Hunecke, M., Haustein, S., Grischkat, S., Böhler, S., 2007. Psychological, sociodemographic, and infrastructural factors as determinants of

ecological impact caused by mobility behavior. Journal of Environmental Psychology 27, 277–292. 10.1016/j.jenvp.2007.08.001.
Joshi, A., Kale, S., Chandel, S., Pal, D., 2015. Likert Scale: Explored and Explained. BJAST 7 (4), 396–403. 10.9734/BJAST/2015/14975.
Kim, Y., Dykema, J., Stevenson, J., Black, P., Moberg, D.P., 2018. Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. Social Science Computer Review 74, 089443931775240. 10.1177/0894439317752406.
Lee, J.W., Jones, P.S., Mineyama, Y., Zhang, X.E., 2002. Cultural differences in responses to a Likert scale. Research in nursing & health 25 (4), 295–306. 10.1002/nur.10041.
Leiner, D.J., 2013. Too Fast, Too Straight, Too Weird: Post Hoc Identification of Meaningless Data in Internet Surveys. SSRN Journal. 10.2139/ssrn.2361661.
Loosveldt, G., Beullens, K., 2017. Interviewer Effects on Non-Differentiation and Straightlining in the European Social Survey. Journal of Official Statistics 33 (2), 409–426. 10.1515/jos-2017-0020.
Magdolen, M., von Behren, S., Chlond, B., Hunecke, M., Vortisch, P., 2019. Combining attitudes and travel behavior - A comparison of urban mobility types identified in Shanghai, Berlin and San Francisco, in: TRB 98th Annual Meeting Compendium of Papers, Washington, D.C.
Schonlau, M., Toepoel, V., 2015. Straightlining in Web survey panels over time. Survey Research Methods 9 (2), 125–137. 10.18148/srm/2015.v9i2.6128.
Smith, P.B., 2004. Acquiescent Response Bias as an Aspect of Cultural Communication Style. Journal of Cross-Cultural Psychology 35 (1), 50–61. 10.1177/0022022103260380.
Steg, L., 2005. Car use: Lust and must. Instrumental, symbolic and affective motives for car use. Transportation Research Part A: Policy and Practice 39 (2-3), 147–162. 10.1016/j.tra.2004.07.001.
van Vaerenbergh, Y., Thomas, T.D., 2013. Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. International Journal of Public Opinion Research 25 (2), 195–217. 10.1093/ijpor/eds021.
von Behren, S., Minster, C., Magdolen, M., Chlond, B., Hunecke, M., Vortisch, P., 2018. Bringing travel behavior and attitudes together: An integrated survey approach for clustering urban mobility types, in: TRB 97th Annual Meeting Compendium of Papers, Washington, D.C.
Wang, Z., 2018. Analyses of psychological items about mode choice in a travel survey in Shanghai, China – An intercultural interpretation and analysis of intelligibility. Master Thesis, Karlsruhe.
Watkins, D., Cheung, S., 1995. Culture, Gender, and Response Bias - An analysis of responses to the self-description questionnaire. Journal of Cross-Cultural Psychology 26 (5), 490–504. 10.1177/0022022195265003.
Weijters, B., Cabooter, E., Schillewaert, N., 2010. The effect of rating scale format on response styles: The number of response categories and response category labels. International Journal of Research in Marketing 27 (3), 236–247. 10.1016/j.ijresmar.2010.02.004.