# Quality assessment of airline baggage handling systems using SERVQUAL and BWM

Rezaei, Jafar; Kothadiya, Oshan; Tavasszy, Lori; Kroesen, Maarten

**DOI**
10.1016/j.tourman.2017.11.009

**Publication date**
2018

**Document Version**
Final published version

**Published in**
Tourism Management

Tourism Management 66 (2018) 85–93

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

# Quality assessment of airline baggage handling systems using SERVQUAL and BWM

Jafar Rezaei*, Oshan Kothadiya, Lori Tavasszy, Maarten Kroesen

*Faculty of Technology, Policy and Management, Delft University of Technology, 2628 BX Delft, The Netherlands*

## HIGHLIGHTS

- SERVQUAL model is proposed to assess service quality of baggage handling system.
- A list of criteria per dimension of the SERVQUAL model is made based on literature.
- The best worst method (BWM) is used to calculate the weights of the criteria.
- The data for BWM are collected via a sample of passengers from different nationalities.
- Reliability is perceived as the most important dimension followed by responsiveness.

## ARTICLE INFO

*Article history:*
Received 5 July 2017
Received in revised form
19 November 2017
Accepted 24 November 2017
Available online 1 December 2017

*Keywords:*
Quality
SERVQUAL
Baggage handling
Best worst method
BWM

## ABSTRACT

The baggage handling system is a crucial part of ground handling operations, which significantly contributes to the overall satisfaction of passengers. Although several studies have investigated airline service quality, little attention has been paid to this crucial part of the system. This study proposes the SERVQUAL model to assess the perceived quality of service for the baggage handling system. A literature review provides a list of criteria per dimension of the SERVQUAL model. The best worst method (BWM) is used to calculate the weights of the criteria. The data for the BWM are collected via a sample of passengers from different nationalities. It is found that 'reliability' is perceived as the most important dimension followed by 'responsiveness'. The 'assurance' criterion is third closely followed by 'tangibles' and lastly 'empathy'. A cluster analysis further sheds light on how passengers might have different service quality priorities.



## 1. Introduction

Since its deregulation in the 70's, the airline industry has grown exponentially. But, from the perception of the passengers services have not improved much (see e.g. Barnett, Curtis, Goranson, & Patrick, 1992). As the competition has grown over the years, costs and prices have been reduced significantly and profit margins have gone down to 3% or less. In response, the airline industry has started to focus again on service quality to maintain their passenger numbers and competitive edge (IATA, 2015). In the context of the competitive environment, pricing and frequent flyer programs had become new policy spearheads of the industry (Zhang & Round, 2011). In order to strive for better service and reduced costs, airlines created strategic alliances to give global connectivity and to share networks. To be more customer friendly, airlines are providing complete travel solutions, from travel to airport, hotel booking, sightseeing, and more. Through new websites on the internet, the customers can look for cheaper services. Presently, to provide better service, the carriers decouple services that in the past were joined in one sale. This allows customers to plan more flexibly, thereby improving overall quality (Hadjetian, 2015). The ground services of any airport or airline are known to be an important factor determining overall service quality by customers (Tsaur, Chang, & Yen, 2002). As a result, there is a constant need for the aviation industries to improve these services.

In the past the airlines used to rely on point-to-point flights making baggage handling fairly easy as the data was to be shared with the personnel within the organization itself or with the security personnel. However, with time the number of alliances

* Corresponding author.
*E-mail addresses:* j.rezaei@tudelft.nl (J. Rezaei), oshan.kotahdiya@gmail.com (O. Kothadiya), l.a.tavasszy@tudelft.nl (L. Tavasszy), g.vandekaa@tudelft.nl (M. Kroesen).

https://doi.org/10.1016/j.tourman.2017.11.009
0261-5177/

between airlines has grown drastically leading to an increase in the transfer passengers flying to some other destinations. This increase in transfer passenger means that there is an increase in the transfer baggage, which increases the physical and informational load on baggage handling system. In addition, most airlines experience congestion during peak hours, creating a possibility of mishandled bags or even lost bags in worst case scenario (International Airport Review, 2014).

The need for innovation in this area can be seen from the SITA baggage report (SITA, 2013). Between 2007 and 2014 the amount of mishandled baggage has reduced from 19 bags per 1000 to 7.3 bags per 1000. This 61% reduction is equivalent to $18 billion saving and a 43% reduction of mishandled baggage costs. The observed reduction in mishandled baggage can be credited to the automation of baggage handling processes at airports, which has brought down the turnaround time for an aircraft to 30 min (UKEssays, 2015). Another reason is the wide usage of barcodes and RFID (Radio Frequency Identification) tags in the airline industry over the past decade. Despite this reduction, mishandled baggage costs are still substantial, as is their impact on customer satisfaction. As per the SITA baggage handling report it can be seen that 80% of mishandled bags were delayed bags (which accounts for 19 million bags), 14% accounted for damaged ones or pilfered bags (3.4 million bags) and 5.5% were stolen or lost bags (1.3 million bags). In general, a mishandled bag costs around $0.73 per passenger – a small amount, relative to the average operating cost of $216 per passenger, but more important compared to the industry average profit per customer of $11.6. Hence improvements in baggage handling services should have a significant positive impact on the profit margin.

Besides a reduction of the number of mishandled bags, various service elements around baggage handling such as information provision, staff attitude and mishandling response time are important to provide a satisfactory overall service to the customer. For an airport, to improve their baggage handling system, it is important to understand the importance of these factors contributing to the overall quality experience of the customer. This study aims to provide such evidence. Firstly, by reviewing the literature and by doing some focus group interviews, we identify the most relevant factors (criteria) considering the main dimensions of SERVQUAL. Secondly, we acquire data from a sample of travelers and use a multi-criteria weighting method called best worst method (BWM) to find the importance of different criteria. Although some previous studies have used SERVQUAL to measure airline service quality, this study is unique with respect to its focus and methodology. Our focus on measuring the service quality of baggage handling systems has not been considered like this in previous studies. The weighting method BWM has not been used for this problem before and has unique characteristics which makes it a proper method for this study.

The paper is organized as follows. In the next section, we present the results of our literature review. In Section 3, the methodology is proposed. The results are provided in Section 4, which is followed by some discussions and finally the paper is concluded in Section 5.

## 2. Literature review

In this section, first we define service quality and its elements. Then we discuss issues regarding service quality for baggage handling.

Quality has been seen as a mode of defense by organizations in the past, but now it is perceived to be strength for an organization in order to gain market share (Davis, Yoo, & Baker, 2003). Over the past years a lot has been studied on the term quality but, intangibility of services is one of the problems with measurement (Joseph, Sekhon, Stone, & Tinson, 2005). Also, it has been established that in the service sector there is a possibility that production, delivery and consumption happen together or at the same time.

Service quality is defined in different ways, but one of the most commonly used definitions is the amount to which the service addresses the customer needs (Lewis, 1989). In fact, service quality is the difference between the customer expectations of service and perceived service. Hence, service quality is based on the evaluations made by customers (Brown, Gumesson, Edvardsson, & Gustavsson, 1991; Grönroos, 1984; Zeithaml, Parasuraman, & Berry, 1990). Service companies cannot rely on their service standards as that may not be to the expectations of the customer. When the expectations of the customer are more than the quality of service delivered, it is then that dissatisfaction occurs (Lewis, 1989; Parasuraman, Berry, & Zeithaml, 1991). "Perceived service quality is a form of attitude, related but not equivalent to satisfaction, and results from a comparison of expectations with perceptions of service performance" (Zeithaml et al., 1990). Alternatively, an empirical study carried out by Lin (1996) regarding domestic airline sector managers, customers and service executers. This study had a holistic view of all stakeholders of the aviation industry. The study concluded that the reason for a contrast in the service quality is due to the difference in the promised service quality by the service providers and actual service delivered to the customers. This gives rise to customers' expectations and the real service they get. Service improvement is considered a profitable strategy as it helps in retaining customers, reducing lost customers and creating inflow of new customers. The costs of marketing also reduce as less convincing is required for the service to flourish (Anderson, Fornell, & Lehmann, 1994; Boulding, Kalra, Staelin, & Zeithaml, 1993; Buzzell & Gale, 1987; Khatib, 1998; Rust and Oliver, 1994). Service can be said to be an orchestrated act, which not only is evaluated on its outcome but, also on the way in which the process is provided. These two aspects are called technical quality and functional quality (Bowen & Schneider, 1988; Grönroos, 1982). For an organization, in order to improve service quality, the first and the most important step is to measure the quality. It is only after the measurement that the company might improve the quality by focusing on those aspects of the quality in which they perform less well and/or on the aspects which are more important than the others.

There are numerous different conceptual models to measure service quality (see, for instance, Becker & Wellins, 1990; Cronin & Taylor, 1992; Grönroos, 1988; Haywood-Farmer, 1988; Rust and Oliver, 1994, and Zeithaml et al., 1990, 1993). One of the most commonly used models is the SERVQUAL model (Seth, Deshmukh, & Vrat, 2005), which is used in our study and described below. As per Berry, Zeithaml, and Parasuraman (1985) and Parasuraman, Zeithaml, and Berry (1985), in the past researchers relied on the PZB service quality model and the basis of SERVQUAL service quality evaluation scale (Parasuraman, Zeithaml, & Berry, 1988), or the PZB service quality expansion model to prove the difference between service quality from the perspective of customers, managers and service providers.

SERVQUAL is a more generic model which provides a measurement system for perceived service quality. This model has been well debated, but it cannot be denied that a large number of studies have been carried out in the past (Augustyn & Seakhoa-King, 2005; Philip & Hazlett, 1997). Previously, the SERVQUAL model was known as the Gap Model, it presented the 5 service gaps that the company should always avoid. SERVQUAL is based on its past writings of expectancy-disconfirmation theory. The difference between the expectations and evaluation of performance was the measure of quality. The level of satisfaction can be found out once the outcome is categorized with confirmation or else disconfirmation (Parasuraman et al., 1985). The inclusion of disconfirmation

theory shows a close relation between service quality and satisfaction. Here, we define service quality as a characteristic of the service offered while satisfaction is the reaction that is given by the customer on receiving the service (Kasper, Van Helsdingen, & de Vries, 1999).

Originally there were ten dimensions, for evaluating service quality but they were condensed down to five dimensions for the SERVQUAL model (Zeithaml et al., 1990). These dimensions are named and described below.

- Tangibility – Appearance of physical facilities, appearance and communication of the personnel in the service process and type of equipment provided in the service process.
- Reliability – The ability of an organization to do a task or service as promised is called as reliability.
- Responsiveness – The willingness of service provider to help the customers. Making an effort sincerely to provide prompt service to customers.
- Assurance – Ability of service provider to give a sense of trust and security to the customers.
- Empathy – Ability of service providers to communicate with customers and provide individualized attention to them.

Despite its popularity, the SERVQUAL model has been criticized for its lack of focus on technical dimension (Augustyn & Seakhoa-King, 2005). As a result, this model is said to neglect a rounded approach to the management of services. In this study we use the SERVQUAL model as it is a multi-dimensional model covering more aspects of service quality in the airline industry.

The SERVQUAL model has been applied to several airline services problems. For instance, Pakdil and Aydın (2007) used SERVQUAL and factor analysis loadings as the weights in order to measure the quality of services of a Turkish airline and found several interesting results including 'passengers past experience' as the most important factor in selecting an airline. Chou, Liu, Huang, Yih, and Han (2011) proposed a fuzzy weighted SERVQUAL model and applied to a Taiwanese airline. They found 'reliability and assurance' and 'responsiveness' as the first two important dimensions followed by 'empathy' and 'tangibles' and finally 'flight pattern'. Basfirinci and Mitra (2015) investigated airline service quality attributes and their effect on customer satisfaction using the SERVQUAL and the Kano model. They collected data from Turkish and American travelers and found that except for the tangibles dimension, the average gap score ratings of Turkish respondents were significantly higher than that of the Americans.

As in this study we use a multi-criteria decision-making (MCDM) method to measure airline service quality, we also review the applications of MCDM methods in airline service quality. Kuo (2011) used VIKOR, gray relation analysis, and interval-valued fuzzy sets to evaluate service quality of Chinese airlines. Liou, Tsai, Lin, and Tzeng (2011) used a modified VIKOR method and SERVQUAL to improve service quality of Taiwanese airlines. Liou, Hsu, Yeh, and Lin (2011) used a modified grey relation method to evaluate service quality of four Taiwanese airlines. They found 'cabin service' as the most important factor, and 'baggage claim' as the least important factor for the passengers who are served by domestic flights. Chen (2016) used DEMATEL and ANP to select airline service quality improvement criteria for Taiwanese airlines. The study suggest top five service quality which should be used to improve the service quality of airline: 'enhancement of customer relationship management', 'low accident rate control', 'differentiated service', 'full support from top managers for front-line service', and 'access to information about service negligence and compensation'. Tsaur et al. (2002) used AHP to evaluate the dimensions of SERVQUAL and TOPSIS for ranking the airlines with respect to their overall service quality. They found 'tangibility' as the most important dimension (weight = 0.245) followed by 'reliability' (weight = 0.231), 'responsiveness' (weight = 0.189), 'assurance' (weight = 0.170), and 'empathy' (weight = 0.165).

## 3. Methodology

As mentioned before, in this study we use the SERVQUAL model for quality measurement, which has five dimensions. These dimensions are considered as the main criteria for an MCDM problem.[1] Looking at the literature we have also identified some sub-criteria per main criterion. A qualitative study is first conducted in order to check if the sub-criteria extracted from the literature are relevant for our study. Then we collect data from a sample of travelers and analyze the data using BWM to find the weights of the criteria. All the steps are explained below.

### 3.1. Service quality dimensions (SERVQUAL)

Here the main dimensions of the SERVQUAL model which are considered as the main criteria of the MCDM problem are discussed.

#### 3.1.1. Tangibles

Physical facilities, equipment and personnel are what come under the tangibles dimension. For ground services, the variables that are important from the tangibles dimension perspective are the cleanliness of the facilities, quality of equipment provided, modern equipment as well as appearance of the airline crew and staff.

#### 3.1.2. Reliability

Consistent service, first time right and inspections and announcements come under the reliability dimension. Consistent service delivery means that the same service is provided for similar cases repetitively. First time right means that the service promised by the airline is provided at the first opportunity. Inspections and announcements are an important attribute of this dimension because issues like terror threats and hijacking can be averted.

#### 3.1.3. Responsiveness

The responsiveness dimension concerns the efficiency with which the service is provided. Here, attributes like willingness to help, efficient guidance, prompt service delivery and prompt handling of complaints/requests are included.

#### 3.1.4. Assurance

The assurance dimension has attributes like trustworthiness of crew, responsiveness and courteousness with which the crew responds to passengers. Trustworthiness is an important need of passengers, who hand over control of baggage to airline crew and the baggage handling system. There have been instances of theft from checked-in baggage by baggage handlers making passengers question the crew and staff.

#### 3.1.5. Empathy

There are three attributes that come under the empathy dimension, namely personalized attention, best interest at heart and understanding needs. Attributes below this dimension can be

[1] It should be mentioned that in this study we use BWM to find the importance of the dimensions of SERVQUAL to evaluate the quality of a baggage handling system and we do not conduct a full gap analysis contrasting the perception of the passengers to their expectations.

hard for airlines to deliver, as service is provided to customers of different cultures. According to Kasper et al. (1999), for customized people management, courteousness is also an important attribute of this dimension.

### 3.2. Interviews to determine service attributes

Interviews were held to validate the criteria for the above dimensions. Since the nature of the research is exploratory, semi-structured interviews are appropriate. In order to have credible data from interviews, ideally 4–10 participants are required with sound knowledge regarding the topic of discussion so that a significant contribution can be made in the preliminary research. We chose a heterogeneous group with equal gender share, holding independent interviews (Blumberg, Cooper, & Schindler, 2005; Cohen & Crabtree, 2006). Six passengers were selected for interviews. These participants are of Indian, Chinese, European and American nationality. The ages within the group ranged between 20 and 38 years. This age group represents a large market share and also for the future the age group is more representative (Pearce, 2014). The shortlist for interviews was done based on passengers that had traveled at least twice a year, either for leisure or business motives.

For data analysis, content analysis methods (Blumberg et al., 2005) were used and the relevant information was summarized. For this study the interviews were first listened after which the quotes that were related to an attribute were written under it. In total there were six interviews that were carried out which lasted between 45 min and 90 min. Content analysis should be followed in a stage-by-stage manner starting by segregating of categories as per Blumberg et al. (2005) which concurs with the method used for this study. Quotes which did not fit under an attribute of the framework based from literature review then were noted down under the dimension and close attention was paid to it during the course of the whole data analysis so that repetition could be noted. If there was a high mentioning from the participants for a new attribute then they were added as an attribute in the research model. As per this method from the qualitative data analysis two attributes were found to be not useful ('efficient guidance', and 'prompt service delivery'). While two attributes were found to be not clear as a result needed to be rephrased ('advance equipment' -initially 'equipment'- and 'consistent inspections' –initially 'announcements and inspections'). Finally, one new attribute was thought to be useful and was not a part of the old research model, as a result was added ('dependability to handle bureaucratic issues'). The overview of criteria is given in Table 1.

**Table 2**
Post pilot-test list of criteria and sub-criteria.

| Service quality dimensions | Attributes |
|---|---|
| Tangibles | Physical facilities |
| | Advance equipment |
| | Personnel |
| Reliability | Consistent service |
| | First time right |
| Responsiveness | Willingness to help |
| | Prompt handling of request |
| Assurance | Trustworthy crew |
| | Knowledge to answer questions |
| | Crew Courteousness |
| Empathy | Personal attention |
| | Your best interest at heart |
| | Understand needs |

### 3.3. Questionnaire for preference valuation

In total 31 questions were posed in the survey to acquire the data. The survey consists of three parts. Part 1 includes 21 questions aimed at acquiring the respondents' preferences of service quality. The answers to the questions were recorded on 9-point scale (see Section 3.4). Part 2 contains four questions to understand the expectations for service. The questions have five options to select from namely *very low, low, medium, high and very high.* Part 3 is focused on gathering demographic information about respondents like nationality, age, profession, times travelled by air per year and gender.

The survey was sent online to multiple people all over the world. The main focus was on getting maximum responses from China, India, Indonesia, USA and European countries as these are all among the countries that will shape the air travel markets in the coming 20 years (Pearce, 2014).

Based on the feedback from a pilot test it was found that the logic given to the survey was good. However, there were a few questions where intermediate values were found to be missing. The most important change was removal of two attributes from the reliability dimension. The participants of the pilot test found *consistent service* and *consistent inspection* similar, making it hard for them to select between the two attributes. To repair this problem the attribute of consistent inspection was removed. The other attribute that was removed was *dependability to handle bureaucratic issues*. Respondents related this attribute to *prompt*

**Table 1**
The relevant sub-criteria per dimension of SERVQUAL.

| Attributes | Number of quotes | Service quality dimensions |
|---|---|---|
| Physical facilities | 5 | Tangibles |
| Advance equipment | 8 | |
| Personnel | 2 | |
| Consistent service | 7 | Reliability |
| First time right | 8 | |
| Consistent inspections | 7 | |
| Dependability to handle bureaucratic issues | 7 | |
| Efficient Guidance | 1 | Responsiveness |
| Willingness to help | 11 | |
| Prompt service delivery | 0 | |
| Prompt handling of request | 13 | |
| Trustworthy crew | 5 | Assurance |
| Knowledge to answer questions | 4 | |
| Crew Courteousness | 7 | |
| Personal attention | 6 | Empathy |
| Your best interest at heart | 9 | |
| Understand needs | 7 | |

*handling of request* and *consistent service*. For a modified list of criteria per dimension see Table 2.

The final questionnaire was sent to 480 potential respondents via social media networks. After initial cleaning of the responses we ended up with 140 completed questionnaires[2] from four different regions and the EU.

### 3.4. Calculating attribute weights using best worst method

Multi-criteria decision-making (MCDM) methods have been devised to solve evaluation/decision problems in which we have a set of criteria. There exist several MCDM methods, each has advantages and disadvantages. A recently developed MCDM method called best worst method (BWM) is used in this study. The BWM uses a unique structure, and needs less comparison data, due to which there are less issues with the inconsistency that are experienced by pairwise comparison methods (Rezaei, 2015; Rezaei, 2016). The method has been applied for various decision-making and evaluation problems such as supplier selection (Gupta & Barua, 2017; Rezaei, Nispeling, Sarkis, & Tavasszy, 2016), supplier segmentation (Rezaei, Wang, & Tavasszy, 2015), sustainable supply-chain (Ahmadi, Kusi-Sarpong, & Rezaei, 2017; Wan Ahmad, Rezaei, Sadaghiani, & Tavasszy, 2017), water resource management (Chitsaz & Azarnivand, 2017), risk management (Torabi, Giahi, & Sahebjamnia, 2016), university-industry collaboration (Salimi & Rezaei, 2016), optimal bundling configurations in ground transport of air freights (Rezaei, Hemmes, & Tavasszy, 2017), evaluation of scientific outputs (Salimi, 2017), airports evaluation (Shojaei, Haeri, & Mohammadi, 2017), and firms' R&D evaluation (Salimi & Rezaei, 2018). Below we describe the stepwise approach of applying the BWM to estimate weights for the identified attributes.

#### 3.4.1. The steps of BWM

*Step 1* – A set of $n$ decision/assessment criteria are identified.
*Step 2* – As per the personal preference the decision-maker/assessor selects the best criterion (e.g. the most important one) and worst criterion (e.g. the least important one) among the available set of criteria identified in Step 1.
*Step 3* – The decision-maker/assessor then carries out pairwise comparisons between the best criterion and other criteria. This is done by determining preferences using a number between 1 and 9, where 1 is 'equally important' and 9 is 'extremely more important'. The resulting vector of this is as mentioned below:

$$A_B = (a_{B1},\ a_{B2}, \ldots,\ a_{Bn}) \tag{1}$$

*Step 4* - The decision-maker/assessor then carries out pairwise comparisons between the other criteria and the worst criterion. This is done by determining preferences using a number between 1 and 9, where 1 is 'equally important' and 9 is 'extremely more important'. The resulting vector of this is as mentioned below:

$$A_W = (a_{1W},\ a_{2W},\ \ldots,\ a_{nW}) \tag{2}$$

*Step 5* – In order to find the optimal weights of the criteria $(w_1^*,\ w_2^*,\ \ldots,\ w_n^*)$ and $\xi^{L*}$, the maximum absolute differences $\{|w_B - a_{Bj}w_j|$ and $|w_j - a_{jW}w_W|\}$ are minimized, or equivalently:

$$\min\max_j \{|w_B - a_{Bj}w_j|,\ |w_j - a_{jW}w_W|\}$$

s.t.

$$\sum_j w_j = 1,\quad w_j \geq 0,\ \text{for all } j. \tag{3}$$

This problem is transferred to the following linear programming problem:

$$\min\ \xi^L$$

s.t.

$$|w_B - a_{Bj}w_j| \leq \xi^L$$

$$|w_j - a_{jW}w_W| \leq \xi^L$$

$$\sum_j w_j = 1,\quad w_j \geq 0,\ \text{for all } j. \tag{4}$$

Solving this problem presents the optimal weights $(w_1^*,\ w_2^*,\ \ldots,\ w_n^*)$ and $\xi^{L*}$. $\xi^{L*} \in [0,\ 1]$ is considered as an indicator of consistency of the pairwise comparisons, i.e. there is high level of consistency for values close to zero.

For this study, there is a set of criteria each of which has some sub-criteria. Weighting is done for the 5 main criteria (tangibles, reliability, responsiveness, assurance and empathy) and also for the sub-criteria, which provides local weights. For obtaining the global weights for each sub-criterion the weights of the corresponding main criterion should be multiplied by the local weights of each sub-criterion.

## 4. Results

### 4.1. Sample demographics

Of the 140 respondents 29% respondents were Indian, 23% EU citizens, 20% Indonesians, 14% American, and 14% Chinese. As for the gender breakdown 55% respondents were male while 45% respondents were female. For the question 'purpose of travel' 57% respondents selected leisure as their preference, while 32% people selected work as their reason for flying. There were 11% respondents who said they travel for other reasons which are not known. As for the age group of the respondents the focus was to gather data from young respondents as these age groups in future will dominate the market needs. Of the responses useful for study 20% were from 18 to 24 age group and 63% were from 24 to 30 age group. Approximately 11% respondents were of 30–36 age group and 4.3% are of 36–42 age group while 1.4% were above 42 years age.

### 4.2. BWM results

By solving the BWM model, the weights of different criteria and sub-criteria were obtained. Table 3 and Fig. 1 show the weights of the main criteria on the basis of responses received from the

[2] In fact we received 241 complete questionnaires, 140 of which are from our target countries.

**Table 3**
Weights of the main criteria.

| | Tangibles | Reliability | Responsiveness | Assurance | Empathy | $\xi^{L*}$ |
|---|---|---|---|---|---|---|
| Mean | 0.140 | 0.346 | 0.230 | 0.152 | 0.134 | 0.168 |
| Median | 0.112 | 0.362 | 0.183 | 0.134 | 0.093 | 0.166 |
| Standard deviation | 0.112 | 0.177 | 0.154 | 0.098 | 0.103 | 0.078 |

respondents in the surveys.

Based on the mean values obtained, it can be seen that the respondents have selected 'reliability' as the most important main criterion of ground services by a big margin followed by 'responsiveness' criterion. The 'assurance' criterion is third closely followed by 'tangibles' and lastly 'empathy'. The standard deviation and the median values are also taken into consideration. Tangible elements are rated lowest in importance, together with empathy of service staff.

The standard deviation values show that reliability and responsiveness, the two most important criteria as per mean are the highest. However, the median value for reliability is higher which implies that majority of the respondents find this criterion more important. Based on the median value the order of importance does not change; however, it should be noted that the median value of the responsiveness criterion is considerably lower than its average value, which implies a positively skewed distribution of the weights for this criterion. That is to say, for this criterion, the frequency of weights below the median is higher than the frequency of the weights above the median.

Apart from the values of the main criteria, it is necessary to note the value of $\xi^{L*}$ as it indicates consistency of the comparisons. The $\xi^{L*}$ values are very close to zero showing a high consistency of the comparisons, and a high reliability of the results.

After the weights of the main criteria were obtained, the local weights for the sub-criteria were calculated. After the local weights for the sub-criteria were obtained, the weights were multiplied by their corresponding main criteria weights giving an output of global weights (see Table 4 and Fig. 2).

As can be seen from Table 4 and Fig. 2, the mean values of sub-criteria of 'reliability' are the highest. 'Consistent service' is seen as the most important sub-criterion by the respondents followed by 'first time right'. The sub-criteria of 'responsiveness' come next. On closer attention to sub-criteria of 'tangible' and 'assurance' it can be seen that sub-criteria of 'assurance' have higher values than those of 'tangible' sub-criteria.

As this is the first study that evaluates the criteria of baggage handling systems, we cannot compare our results with other studies. Nonetheless, we can show how our findings are similar or different from the findings of other airline-related problems. For instance, while Tsaur et al. (2002) found 'tangibility' as the most important dimension (weight = 0.245) in evaluating the overall service quality of an airline, in our study it is one of the least important dimensions. The 'tangibility' dimension has been found by some other studies on airline service quality as one of the least important (e.g. Chou et al., 2011). 'Reliability' is the most important criterion in our study. It is found to be the most important dimension in evaluating the overall service quality of an airline in the study of (Chou et al., 2011) and the second most important criterion in evaluating the overall service quality of an airline in the study of Tsaur et al. (2002). With respect to the position of the other main criteria, our findings are relatively close to the study of Tsaur et al. (2002). Pakdil and Aydın (2007) found responsiveness as the most important dimension airline service quality, which is close to our results as we found it the second most important dimension.

**Table 4**
Global weights of the sub-criteria.

| Sub-criteria | Mean | Median | Standard deviation |
|---|---|---|---|
| Advanced Equipment | 0.048 | 0.026 | 0.064 |
| Personnel | 0.028 | 0.014 | 0.038 |
| Physical Facility | 0.063 | 0.038 | 0.076 |
| Consistent service | 0.197 | 0.119 | 0.165 |
| First time right | 0.149 | 0.079 | 0.147 |
| Prompt handling of request | 0.094 | 0.067 | 0.095 |
| Willingness to help | 0.134 | 0.099 | 0.125 |
| Crew courteousness | 0.037 | 0.026 | 0.032 |
| Knowledge to answer questions | 0.045 | 0.022 | 0.063 |
| Trustworthy crew | 0.070 | 0.043 | 0.073 |
| Personal attention | 0.036 | 0.020 | 0.052 |
| Understand needs | 0.054 | 0.032 | 0.072 |
| Your best interest at heart | 0.044 | 0.026 | 0.049 |

The results can be used by airlines and airports to improve the baggage handling system. That is to say, the most important criteria show priorities for improvement. From among the main criteria, the decision-makers should mainly focus on 'reliability' dimension of the baggage handling system, while when it comes to the sub-criteria, the main priorities are 'consistent service', 'first time right', 'willingness to help', and prompt handling of request'.

It is interesting to explore to what degree the aggregate weights, as found here, are the results of scores of homogeneous sub-populations. If such subpopulations can be identified, and weights are different from the aggregate, this could improve our

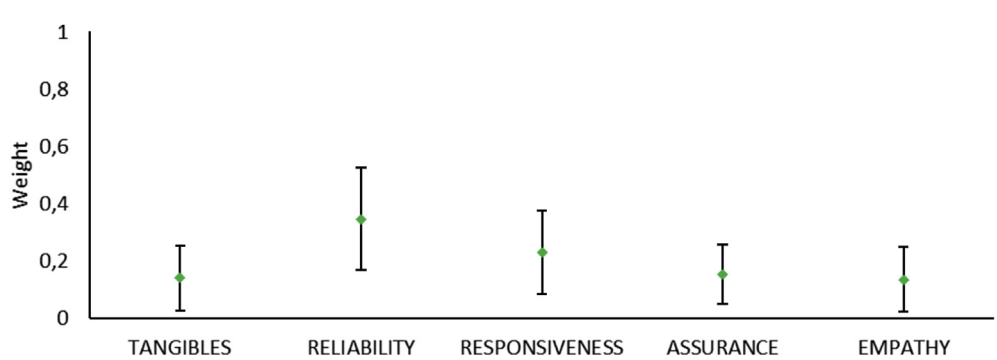


**Fig. 1.** Main criteria weights (mean and standard deviation).

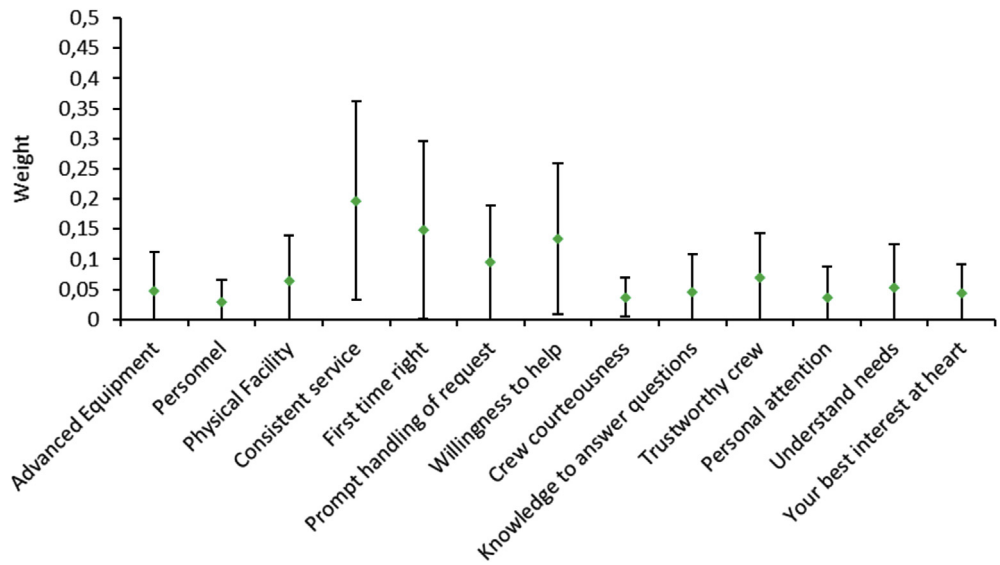


**Fig. 2.** Global weights of the sub-criteria (mean and standard deviation).

understanding of the reasons why travelers weigh the criteria differently. We describe this test in the next section.

### 4.3. Cluster analysis

To explore further the heterogeneity of the sample given the weights attached to the five criteria, a two-step cluster analysis was performed (Chiu, Fang, Chen, Wang, & Jeris, 2001). This analysis helps us to identify natural classes in the group of respondents with homogeneous preferences. Two-step cluster analysis has been shown to perform well in case of quantitative indicators (Bacher, Wenzig, & Vogler, 2004). An additional benefit of this method, compared to K-means cluster analysis, is that it provides statistical criteria to select the optimal number of clusters through BIC (Bayes Information Criterion) and AIC (Akaike Information Criterion) values. The software package SPSS version 22 was used to perform the analysis.

Application of the two-step clustering procedure revealed that the 3-cluster solution was optimal based on the BIC criterion (this solution had the lowest BIC value). Table 5 presents the results of this solution.

Overall, the three clusters can be interpreted clearly:

- Cluster 1, representing the majority of the sample (60%), is mostly concerned with reliability (mean weight of 0.46), followed by responsiveness and assurance, equally valued with a mean weight of 0.17, and finally by tangibility and empathy, which are equally valued with a mean weight of 0.10.
- Cluster 2, representing 24% of the sample, is mostly focused on responsiveness (mean weight of 0.49). In this cluster, the other criteria are valued more or less equally in the 12–15% range.
- Cluster 3, representing only 16% of the sample, attaches most importance to tangibility (mean weight of 0.28) and empathy (mean weight of 0.26). The reliability, responsiveness and assurance follow in decreasing order of importance.

Overall, the results indicate that most people (84%) mostly focus on a single criterion (being either reliability or responsiveness).

The cluster analysis reveals that the aggregate valuation of the two most important quality dimensions, reliability and responsiveness, rests mainly on a contrasting prioritization by two distinct user groups. In addition, the two dimensions that score lowest overall, tangibles and empathy, appear to have a group of users that do value these dimensions very highly. As the underlying pattern is different from the aggregate it is interesting for further exploration. The main implication for baggage handling system design is that a design targeted towards subgroups will probably be more effective than a design targeted to the aggregate average. In addition, it raises important questions for service package design and marketing, as the identification of these groups in terms of passenger characteristics would allow to target the design parameters to specific passenger groups.

To assess whether the clusters could be profiled on directly observable characteristics, relationships were explored between cluster membership and several background variables of respondents. For this, the following variables were considered:

**Table 5**
Results of the cluster analysis.

| Cluster size (%) | Cluster 1: Reliability | | Cluster 2: Responsiveness | | Cluster 3: Tangibility and empathy | |
|---|---|---|---|---|---|---|
| | 0.60 | | 0.24 | | 0.16 | |
| | Mean | SD | Mean | SD | Mean | SD |
| Tangibility | 0.10 | 0.05 | 0.10 | 0.04 | 0.28 | 0.17 |
| Reliability | 0.46 | 0.12 | 0.15 | 0.05 | 0.20 | 0.12 |
| Responsiveness | 0.17 | 0.07 | 0.49 | 0.08 | 0.15 | 0.04 |
| Assurance | 0.17 | 0.12 | 0.13 | 0.05 | 0.11 | 0.06 |
| Empathy | 0.10 | 0.06 | 0.12 | 0.07 | 0.26 | 0.21 |

gender, age, occupation, nationality and travel purpose. Interestingly, none of these variables were significantly associated with cluster membership, indicating that the revealed clusters manifest themselves equally across the categories of the variables considered. An interesting topic for new research would be the formation and testing of different hypotheses concerning cluster membership of passengers based on other observable characteristics than those recorded here (e.g. lifestyle). In any case, this classification will already help to predict responses to service improvement programs more accurately than when only using the aggregate valuations.

## 5. Conclusion

In this study, we operationalize and valuate service quality attributes for the airline baggage handling system. Through a multi-method approach, we identify a hierarchical set of service quality criteria and determine the weights for these criteria. The study is unique as it provides quantitative results for a comprehensive set of service attributes for the baggage handling system, formally supported by the SERVQUAL and the best worst method (BWM). The use of the SERVQUAL framework allows the inclusion of non-tangible performance dimensions. The application of BWM allows a robust measurement of weights of service quality attributes.

Starting from the initial SERVQUAL dimensions of tangibles, reliability, responsiveness, assurance and empathy, we arrive at 13 key underlying service attributes, based on a set of semi-structured interviews with passengers. To establish weights for these criteria we held a survey amongst air passengers from all key regions in the world, distributed by age, gender and travel motive. Weights obtained are based on 140 valid survey responses. The BWM scores suggest a strong numerical consistency amongst these responses. A secondary cluster analysis suggests 3 clusters of passengers with different attribute weights. We find that the clusters have different sets of weights, more focused towards individual dimensions. The differences between criteria are more pronounced, and even structured differently than the aggregate weights. This suggests that designs targeting the underlying clusters of customers may be more effective than designs targeting the aggregate.

Our findings are relevant for all stakeholders of the airline industry who have an interest in customer-centric design of baggage handling services. These include airlines, airports, handling service providers, public authorities such as customs and baggage handling equipment designers and operators. The results can be used to aid decisions in the (re-) design of baggage handling systems, with an aim to improve service quality.

The sample size of our study is rather small, so we recommend future studies gathering more data from the target countries in order to make the findings more generalizable. In doing so, it is suggested to use *proportional sampling* considering the population of those countries. Finally while in this study we found the importance (weights) of the dimensions of SERVQUAL for evaluating a baggage handling system, future studies could measure a full gap analysis to see the difference between the perception of passengers and their expectations considering the weights we have found in this study.

## Contribution statement

Jafar Rezaei has contributed to structuring the paper, writing the literature review and checking the methodology and calculations. Oshan Kothadiya has contributed to writing some parts of the paper, collecting data, and data analysis. Lori Tavasszy has contributed to strengthening the discussion part of the paper. Maarten Kroesen has contributed to the cluster analysis.

## References


Ahmadi, H. B., Kusi-Sarpong, S., & Rezaei, J. (2017). Assessing the social sustainability of supply chains using best worst method. *Resources, Conservation and Recycling, 126*, 99–106.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *The Journal of Marketing*, 53–66.

Augustyn, M. M., & Seakhoa-King, A. (2005). Is the Servqual scale an adequate measure of quality in leisure, tourism and hospitality?. In *Advances in hospitality and leisure* (pp. 3–24). Emerald Group Publishing Limited.

Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS TwoStep Cluster-a first evaluation*.

Barnett, A., Curtis, T., Goranson, J., & Patrick, A. (1992). Better than ever: Nonstop jet service in an era of hubs and spokes. *Sloan Management Review, 33*(2), 49–54.

Basfirinci, C., & Mitra, A. (2015). A cross cultural investigation of airlines service quality through integration of Servqual and the Kano model. *Journal of Air Transport Management, 42*, 239–248.

Becker, W. S., & Wellins, R. S. (1990). Customer-service perceptions and reality. *Training & Development Journal, 44*(3), 49–52.

Berry, L. L., Zeithaml, V. A., & Parasuraman, A. (1985). Quality counts in services, too. *Business Horizons, 28*(3), 44–52.

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2005). *Business research methods*. the McGraw-Hill: McGraw-Hill Education UK.

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research, 30*(1), 7–27.

Bowen, D. E., & Schneider, B. (1988). Services marketing and management-implications for organizational-behavior. *Research in Organizational Behavior, 10*, 43–80.

Brown, S. W., Gumesson, E., Edvardsson, B., & Gustavsson, B. (1991). *Service quality: Multidisciplinary and multinational perspectives*. New York, USA.

Buzzell, R. D., & Gale, B. T. (1987). *The PIMS principles: Linking strategy to performance*. New York, USA: Simon and Schuster.

Chen, I. S. (2016). A combined MCDM model based on DEMATEL and ANP for the selection of airline service quality improvement criteria: A study based on the Taiwanese airline industry. *Journal of Air Transport Management, 57*, 7–18.

Chitsaz, N., & Azarnivand, A. (2017). Water scarcity management in arid regions based on an extended multiple criteria technique. *Water Resources Management, 31*(1), 233–250.

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001, August). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 263–268). ACM.

Chou, C. C., Liu, L. J., Huang, S. F., Yih, J. M., & Han, T. C. (2011). An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. *Applied Soft Computing, 11*(2), 2117–2128.

Cohen, D., & Crabtree, B. (July 2006). *Qualitative research guidelines project*.

Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *The Journal of Marketing, 56*(3), 55–68.

Davis, G. F., Yoo, M., & Baker, W. E. (2003). The small world of the American corporate elite, 1982-2001. *Strategic Organization, 1*(3), 301–326.

Grönroos, C. (1982). *Strategic management and marketing in the service sector*. Florida, USA: Krieger Pub Co.

Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing, 18*(4), 36–44.

Grönroos, C. (1988). Service quality: The six criteria of good perceived service. *Review of Business, 9*(3), 10–13.

Gupta, H., & Barua, M. K. (2017). Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS. *Journal of Cleaner Production, 152*, 242–258.

Hadjetian, L. (2015). *The end of the airline price wars?*. Mercator.com. Retrieved 27 December 2015, from http://www.mercator.com/blog/the-end-of-the-airline-price-wars.

Haywood-Farmer, J. (1988). A conceptual model of service quality. *International Journal of Operations & Production Management, 8*(6), 19–29.

IATA (2015) Reterived in March 2017 from https://www.iata.org/about/…/iata-annual-review-2015.pdf.

International Airport Review. (2014). *Baggage handling: Achieving operational excellence*. Retrieved 8 December 2016, from http://www.internationalairportreview.com/16104/airport-extra/baggage-handling-achieving-operational-excellence/.

Joseph, M., Sekhon, Y., Stone, G., & Tinson, J. (2005). An exploratory study on the use of banking technology in the UK: A ranking of importance of selected technology on consumer perception of service delivery performance. *International Journal of Bank Marketing, 23*(5), 397–413.

Kasper, H., Van Helsdingen, P., & de Vries, W., Jr. (1999). *Services marketing management: An international perspective*. Wiley.

Khatib, F. S. (1998). *An investigation of airline service quality, passenger satisfaction and Loyalty: The case of royal jordanian airline*. Sheffield University. Doctoral dissertation.

Kuo, M. S. (2011). A novel interval-valued fuzzy MCDM method for improving airlines' service quality in Chinese cross-strait airlines. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1177–1193.

Lewis, B. R. (1989). Quality in the service sector: A review. *International Journal of Bank Marketing, 7*(5), 4–12.

Lin, H. Q. (1996). *Empirical study of the service quality of domestic airline companies*. Taipei: Master's Thesis of the Department of Commerce of National Taiwan University.
Liou, J. J., Hsu, C. C., Yeh, W. C., & Lin, R. H. (2011). Using a modified grey relation method for improving airline service quality. *Tourism Management, 32*(6), 1381–1388.
Liou, J. J., Tsai, C. Y., Lin, R. H., & Tzeng, G. H. (2011). A modified VIKOR multiple-criteria decision method for improving domestic airlines service quality. *Journal of Air Transport Management, 17*(2), 57–61.
Pakdil, F., & Aydın, Ö. (2007). Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores. *Journal of Air Transport Management, 13*(4), 229–237.
Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing, 67*(4), 420.
Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*(4), 41–50.
Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perc. *Journal of Retailing, 64*(1), 12–40.
Pearce, B. (2014). The shape of air travel markets over the next 20 years. In *Global airport development conference*. Athens.
Philip, G., & Hazlett, S. A. (1997). The measurement of service quality: A new PCP attributes model. *International Journal of Quality & Reliability Management, 14*(3), 260–286.
Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57.
Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega, 64*, 126–130.
Rezaei, J., Hemmes, A., & Tavasszy, L. (2017). Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight. *Journal of Air Transport Management, 61*, 95–105.
Rezaei, J., Nispeling, T., Sarkis, J., & Tavasszy, L. (2016). A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. *Journal of Cleaner Production, 135*, 577–588.
Rezaei, J., Wang, J., & Tavasszy, L. (2015). Linking supplier development to supplier segmentation using best worst method. *Expert Systems with Applications, 42*(23), 9152–9164.
Rust, R. T., & Oliver, R. L. (Eds.). (1994). *Service quality: New directions in theory and practice*. California, USA: Sage Publications.
Salimi, N. (2017). Quality assessment of scientific outputs using the BWM. *Scientometrics, 112*(1), 195–213.
Salimi, N., & Rezaei, J. (2016). Measuring efficiency of university-industry Ph.D. projects using best worst method. *Scientometrics, 109*(3), 1911–1938.
Salimi, N., & Rezaei, J. (2018). Evaluating firms' R&D performance using best worst method. *Evaluation and Program Planning, 66*, 147–155.
Seth, N., Deshmukh, S. G., & Vrat, P. (2005). Service quality models: A review. *International Journal of Quality & Reliability Management, 22*(9), 913–949.
Shojaei, P., Haeri, S. A. S., & Mohammadi, S. (2017). Airports evaluation and ranking model using Taguchi loss function, best-worst method and VIKOR technique. *Journal of Air Transport Management*. In press.
SITA. (2013). *Air transport industry insights*. Air Transport World. Retrieved from http://cdn.thejournal.ie/media/2013/04/baggage-report-2013-final.pdf.
Torabi, S. A., Giahi, R., & Sahebjamnia, N. (2016). An enhanced risk assessment framework for business continuity management systems. *Safety Science, 89*, 201–218.
Tsaur, S. H., Chang, T. Y., & Yen, C. H. (2002). The evaluation of airline service quality by fuzzy MCDM. *Tourism Management, 23*(2), 107–115.
UKEssays. (2015). *The history of baggage handling systems information Technology essay*. Retrieved 23 November 2015, from http://www.ukessays.com/essays/information-technology/the-history-of-baggage-handling-systems-information-technology-essay.php.
Wan Ahmad, W. N. K., Rezaei, J., Sadaghiani, S., & Tavasszy, L. A. (2017). Evaluation of the external forces affecting the sustainability of oil and gas supply chain using best worst method. *Journal of Cleaner Production, 153*, 242–252.
Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science, 21*(1), 1–12.
Zeithaml, V., Parasuraman, A., & Berry, L. (1990). *Delivering quality service*. New York: Free Press.
Zhang, Y., & Round, D. K. (2011). Price wars and price collusion in China's airline markets. *International Journal of Industrial Organization, 29*(4), 361–372.

**Jafar Rezaei** is an associate professor of operations and supply chain management at Delft University of Technology, The Netherlands, where he obtained his Ph.D. in 2012. His main research interests are in the areas of logistics and supply chain management, and multi-criteria decision-making (MCDM). He has presented his works in several international conferences, and has published in various academic journals, including International Journal of Production Economics, International Journal of Production Research, International Journal of Systems Science, Industrial Marketing Management, Transportation Research Record, Expert Systems with Applications, Applied Soft Computing, IEEE Transactions on Engineering Management, Journal of Air Transport Management, European Journal of Operational Research, International Journal of Logistics Research and Applications, Omega. He is the Editor-in-Chief of the open access Journal of Supply Chain Management Science, and serves on the editorial board of some other journals.

**Oshan Kothadiya** is an alumnus of Delft University of Technology from where he completed his Master in Management of Technology in 2016. The topic of his thesis was to find out 'Customer Preferences for Airline Ground Services' using multi-criteria decision-making (MCDM). He completed his Bachelor in Mechanical Engineering in 2012 from University of Pune, India. Currently he is working on his innovative concept to improve passenger service which will help aviation service organizations to save revenue. His research interests are related to aviation service, baggage handling and entrepreneurship.

**Lori Tavasszy** (1967) is full professor in Freight Transport & Logistics at the Delft University of Technology. He studied Civil Engineering and completed his PhD research at Delft University of Technology in 1996 on multimodal freight transport models for Europe. Until 2016, he worked with the Dutch national research institute TNO as researcher, manager and principal scientist and held part time chairs at the University of Nijmegen and TU Delft. His research involves freight transportation and logistics modelling at urban, national and global level. He has published over 50 articles in scientific journals and over 100 in books and conference proceedings. He received international awards for his work in 1998 and 2013. He is editorial board member for EJTIR, CSTP and JSMCS and acts as regular reviewer for several journals in transportation such as Transportation Research parts A-E, Transport Policy, Transportation and Transport Geography. He is now the chair of the scientific Committee of the World Conference for Transport Research Society; Co-chair of the Working Group on Hubs, Corridors and Synchromodality of the European Technology Platform on Logistics, ALICE; Core partner of the VREF Global Center of Excellence for Sustainable Urban Freight Systems; partner of the OECD/ITF Decarbonization project and member of various scientific committees of the Transportation Research Board.

**Maarten Kroesen** is an assistant professor at the Section of Transport and Logistics at Delft University of Technology, the Netherlands. His main research field is travel behavior research. He aims to contribute to this field by developing conceptual frameworks and related statistical models to analyze mobility panel data. Apart from this specific interest in the field of travel behavior, his research interests extend to the fields of acoustics, health, psychology, tourism and IT.