

# Creating a Short Scale for In-flight Experience Quality

Rian Mehta and Stephen Rice

**Abstract**—Customer service and satisfaction is of great importance to aviation industry as it has an economic impact on the airlines. Customer's in-flight experience, is tied into their satisfaction with that airline, and in turn has become an area of competition between airlines. This study seeks to develop a statistically valid and reliable scale to empirically measure the quality of the in-flight experience that a passenger would witness on a commercial airline flight within the United States. Actual consumers from the general public were used for each of the five stages of the process, along with input from aviation experts, to generate items for the scale, narrow down the list of items to those most relevant to in-flight experience quality, and test the final scale for validity, reliability and discriminability. A factor analysis using the principle components and varimax rotation loaded strongly on one factor, providing evidence for validity. Reliability was tested via Cronbach's Alpha and Guttman's Split-half tests, indicating high consistency and reliability. The final scale that was developed contained eight items, which were good condition, arrived on time, comfortable chairs, air conditioning was favorable, clean smelling air, comfortable cabin temperature, comfortable seat spacing, good customer service.

**Index Terms**—In-flight experience, scale, valid and reliable, and experience quality.

## I. INTRODUCTION

### A. Customer Service/Satisfaction

The airline industry is focused on delivering a pleasant customer experience for its passengers. There are several areas of competition between the commercial airlines. One of these main areas is the level of customer service they provide to their passengers, and the overall satisfaction of the passenger with their experience [1]. This study aims at addressing the in-flight aspect of this experience. The purpose of this research is to develop a valid and reliable scale measuring the quality of passengers' in-flight experience. One of the salient features of this scale is that the development involves items solicited from actual airline passengers and not just industry experts. Previous research in the general field of customer service identify five aspects to be represented, which are tangibility, reliability, responsiveness, assurance and empathy [2].

Customer service and customer satisfaction are of great importance to the airlines. As mentioned it is a source of competition between the airlines. While ticket prices remain a major factor in a customer's decision-making processes, their perceptions of the customer service offered does play a role as well [3]. Previous negative experiences with an airline may reduce the person's desire to fly with said

airline in the future [4]. This factor therefore has a direct economic impact on the airlines' bottom line, and therefore warrants serious attention. Ostrowski, O'Brien, and Gordon (1993) stated that there was a significant relationship between passengers' ratings of customer service and retained preferences, which is a measure of the passengers' decisions to fly with the same airline again (i.e. passenger loyalty) [5]. The study showed that as customer service quality decreased, so did customer loyalty to the brand (airline). Oftentimes, the airlines' customer service reputation is the deciding factor due to the fact that ticket prices are fairly consistent amongst competitors. Airlines have to use to provide superior customer service in order to set them apart from their competitors.

Interestingly, a study found that when analyzing the impact of minor incidents (i.e., failures that do not result in physical harm) and major incidents (i.e., failures that result in injury or death), minor incidents have a stronger negative relationship with future market share and customer satisfaction [6]. Similarly, a longitudinal study by Cunningham, Young, and Lee (2004) stated that there was no statistically significant decrease in passengers' overall satisfaction with the airline industry, and their loyalty to their airline from the time frame before the attacks of 9/11 to after the attacks, even though the number of flights decreased [7].

Customer service and satisfaction have even become important factors in airline ranking. This once again ties into marketing as tool for competition to increase profit margins. In 2015 Alaska Airlines and Jet Blue Airways were the top ranking airlines in the traditional and low cost categories respectively [8]. This study also shows that when passengers rank their satisfaction with the airline based on customer service and reputation, they are more likely to fly on board again in the future or recommend the airline to others as compared to those who rank the airlines based off of ticket prices. This shows that the impact of customer service can oftentimes outweigh higher ticket prices.

### B. Need for Scales

The most efficient method improvement is to identify areas of weakness. This is true for airline customer service quality as well. Surveying passengers to identify the airlines' strengths and weaknesses in terms of customer service and in-flight experience quality can be extremely beneficial. Managers, or even crewmembers can only do so much to identify areas that need improvement, but receiving feedback from actual passengers could highlight areas that industry experts may not have realized to be problem areas. While airlines may have attempted to receive feedback from passengers in the past, no universal scale was found in the scientific literature that measured the quality of passengers' in-flight experience, and was one developed using items solicited by passengers themselves.

Manuscript received February 6, 2016; revised April 30, 2016.

The authors are with the Florida Institute of Technology, College of Aeronautics, Melbourne, FL. 32901 USA (e-mail: rmehta2009@fit.edu).

Studies such as Elliott and Roach (1993), and Truitt and Haynes (1994), use metrics for measuring customer service and satisfaction in categories such as timelines, the luggage transportation, the quality of F&B (food and beverages), the comfort of seat, the check in process and onboard service, the convenience of transit, the clearness of seat, and the customer complaints handling [9], [10]. While many of these categories refer to a broad spectrum of items relation to overall customer satisfaction, they are oftentimes handled by a vast crew of different employees and take place over extended periods of time. The experience at the check-in counter may be vastly different than that of the on board experience. For this reason, this study proposes to develop a scale purely focused on the in-flight experience quality in order to gain a better understanding of the true perceptions of passengers and of items that may not be up to standard.

Developing a scientifically valid and reliable scale to measure the quality of the passengers' in-flight experience allows there to be one universal metric upon which all airlines may accurately gauge the same. Future use of the scale may even lead to possibilities of comparisons or rankings between airlines on the topic area.

### C. Previous Scales Developed Using Similar Methodology

Previous research studies have developed valid and reliable scales of measurement using the same methodology as this study. This line of research has potential to develop consumer-constructed instruments that can be of great practical use to the aviation industry. Based on these studies, valid and reliable scales may also be developed in other consumer-oriented fields or otherwise.

Some of the salient scales developed using this methodology of consumer-solicited items are as follows. Mehta *et al.* (2015) developed a valid and reliable scale to measure consumer perceptions towards intermodal rail networks at airports that aid in movement of passengers within the airport and the neighboring facilities [11]. Rice, *et al.* (2014) and Rice, *et al.* (2015) created two similar scales to measure trustworthiness of commercial airline pilots [12], [13]. The two versions were country specific. Rice *et al.* (2014) was developed for Indian consumers, while Rice *et al.* (2015) was created for American consumers [12], [13]. While the basic core concepts remain the same items generated were different, highlighting the aspect that perceptions of passengers and what is important to them varies as a function of country. It is unlikely that one scale can universally and accurately measures the same construct across the globe. This also highlights the need for future research to create other country specific scales to measure different topics.

While several other scientific scales do exist using the same and different methodologies, these highlight some of the broad range of topic areas that may be covered by scale development. The need for scientifically valid and reliable scales is definitely present and this current study seeks to fill a portion of that gap in the literature. No previous study has created a consumer based scale for measuring in-flight experiences, which sets this paper apart from existing research. As described later, the need to understand the quality of the passengers' experience is of tremendous value

to the airline industry, and further adds to the credence of this study. The issue is that no current study provides both valid and reliable scale and one that is developed by consumers for consumers.

## II. METHODOLOGY

### A. Stage 1: Item Generation

Generating items for the scale was the purpose of study 1. The aim of this scale was to be one that solicited words or phrases from current or future air travelers. This is different from some traditional scales that solely rely on industry or content experts for item generation. Since the scale will be ultimately be used on customers and passengers themselves it was deemed apt to include consumers in the scale development process to get the true mindset of the passenger. In addition, to supplement the items generated by consumers, items were also sourced from aviation industry experts.

#### 1) Participants

Study 1 utilized 70(32 females) participants from the United States. The mean age was 33.86 ( $SD = 10.82$ ). Participants were recruited using Amazon's <sup>®</sup> Mechanical Turk <sup>®</sup> (MTurk). This service provides participants compensation for their completion of human intelligence tasks. Buhrmester, *et. al* (2011) and Germine, *et. al.*, (2012) both suggested that MTurk data is reliable as laboratory data [14], [15]. In all the stages of this research, the studies ensured that the participants were airline consumers, i.e. had travelled on a commercial airline flight in the past.

#### 2) Materials and stimuli

After the participants gave electronic consent they were presented with the following scenario: "*In the context of traveling a commercial airline flight, please identify 6 words or short phrases that you associate with In-Flight experience quality (e.g. tasty meal, clean surroundings, friendly service, etc.).*" Once the participants had given their six words or phrases, they were debriefed and dismissed. This stage generated a total of 247 unique words or phrases. All items were reviewed for correct spelling and all words were de-capitalized to ensure uniform saliency in the next steps.

### B. Stage 2: Nominal Paring

The goal of stage 2 was to narrow down the initial list of items generated from stage 1 by eliminating words or phrases that were not perceived by participants as being relevant to in-flight experience quality.

#### 1) Participants

62 (32 females) participants from the United States took part in this stage. The mean age was 32.53 ( $SD = 9.12$ ). In line with the previous study, participants were recruited via a convenience sample using MTurk.

#### 2) Materials and stimuli

The 247 words generated in the first stage were presented to participants with the following statement, "*In the context of a commercial airline flight, please rate whether each word below is related to (similar to) In-Flight experience*"

quality, not related to (not similar to) In-Flight experience quality, or you don't know." 32 items were deemed to be related to in-flight experience quality by at least 85% of participants, and were included in the next stage of the research.

#### A. Stage 3: Likert-Scale Paring

The purpose of stage 3 was to continue to pare down the list of words that would result in the ones most related to in-flight quality. The items that remained after this stage would be the ones used to generate the final scale. A Likert-type scale was used in this stage instead of a nominal scale, in order to measure the relationship between the items and in-flight quality more sensitively.

##### 1) Participants

82(42 females) participants from the United States were recruited via a convenience sample using MTurk to participate in the study as in previous stages. The mean age was 34.44 ( $SD = 11.63$ ).

##### 2) Materials and stimuli

In this stage, the 32 items that were carried over from stage 2 were presented to participants with the following statement, "In the context of a commercial airline flight, please rate how strongly each word below is related to In-Flight experience quality". Participants responded based on a Likert-type scale from "Not at all related to In-Flight quality experience" (0) to "Extremely related to In-Flight quality experience" (+3). An average score for each item was calculated across each participant. Items that received an average score of 2.4 or higher (equivalent to the average participant saying that this item was at least "quite related") were retained for the next stage. This resulted in 8 words and phrases being carried over to stage 4.

#### B. Stage 4: Scenario-Based Testing

The aim of the first three stages was to generate and pare down words or items that are related to in-flight quality experience. Stage 4 was designed in order to collect initial validity and reliability evidence for the newly created scale.

##### 1) Participants

375 (151 females) participants from the United States took part in the study. The mean age was 32.25 ( $SD = 10.20$ ). Participants were recruited via a convenience sample using MTurk. In order to get a true representation of a travelling consumer, a logic rule in the survey was set-up such that participants were first asked if they had flown on a commercial airline flight in the past five years. Those who answered "yes" were forwarded to the survey. Those who answered "no" were debriefed and paid.

##### 2) Materials and stimuli

In this stage, participants were presented with the following scenario: "Please try to remember the last commercial airline flight that you were on. Think about your experience in flight. You may remember all of the specifics of that flight; however you are aware of how you perceived it. Please respond to the following statements below regarding that flight." Participants were then given the questionnaire (see Appendix A) and asked to provide statements of agreement or disagreement on a 5-point

Likert-type scale (coded from -2 to +2). Once the questionnaire was completed, the participants were debriefed and dismissed.

##### 3) Scale development

A factor analysis using the principle components and varimax rotation resulted in all items loading strongly on one factor. A Cronbach's Alpha test resulted in a value of .84, indicating high internal consistency. A Guttman split-half test generated a coefficient of .87, indicating good reliability.

#### C. Study 5: Scenario-Based Experiment

In the previous study, the validity and reliability of the newly created scale was tested. The current study was conducted in order to test the ability of the scale to discriminate between a good in-flight experience and a bad in-flight experience.

##### 1) Participants

140(58 females) participants from the United States took part in the study. The mean age was 33.27 ( $SD = 11.13$ ). Participants were recruited via a convenience sample using Amazon's ® Mechanical Turk ® (MTurk).

##### 2) Materials and stimuli

Participants were given the following instructions: "Please try to remember your BEST experience on a commercial airline flight. Think about your experience in flight. You may not remember all of the specifics of that flight; however you should have a general memory of the flight experience. Please respond to the following statements below regarding that flight". Participants were then given the questionnaire (see Appendix A) and asked to provide statements of agreement or disagreement on a 5-point Likert-type scale, scored from -2 to +2, with a neutral zero option. They were then presented with the second scenario: "Please try to remember your WORST experience on a commercial airline flight. Think about your experience in flight. You may not remember all of the specifics of that flight; however, you should have a general memory of the flight experience. Please respond to the following statements below regarding that flight". Following this, participants were given the same questionnaire once again (see Appendix A).

##### 3) Scale development

For the "BEST" condition, a Cronbach's Alpha test was conducted to measure internal consistency within the scale, resulting in a coefficient of .85, indicating very high internal consistency. A Guttman split-half test resulted in a coefficient of .92, indicating high reliability. For the "WORST" condition, a Cronbach's Alpha test was conducted to measure internal consistency within the scale, resulting in a coefficient of .86, indicating very high internal consistency. A Guttman split-half test resulted in a coefficient of .90, indicating high reliability. Separate factor analyses using the principle components and varimax rotation were conducted on both conditions, and the findings suggested that all items loaded strongly on one factor. These results support the findings of study 4.

A within participants comparison of the two conditions revealed a significant difference in scores on the in-flight

experience quality scale,  $t(139) = 15.49, p < .001, d = 1.51$ , revealing that the scale was easily able to discriminate effectively between the good experience ( $M = 0.97, SD = 0.56$ ) and the bad experience ( $M = -0.03, SD = 0.76$ ) conditions. We note the large effect size and small standard deviations, which is an indication of the sensitivity of the scale.

### III. DISCUSSION

The purpose of this study was to create a valid and reliable scale that could be used to measure passengers' in-flight experience quality. As mentioned earlier, in-flight experience quality refers to an all round experience during a flight including the comfort of the environment, the ambience, and the service quality of the crew. The items that were most relevant to the consumers themselves were identified through the multi-stage design of this study and consequently were the items used to develop this scale. The study began with the generation of 247 "satisfaction" items by the participants that they associated with positive experiences in relation to in-flight experience quality. Through the following stages, the 247 "satisfaction" items were narrowed down to the final 8 items. These items were good condition, arrived on time, comfortable chairs, air conditioning was favorable, clean smelling air, comfortable cabin temperature, comfortable seat spacing, good customer service. The final scale, located in Appendix A, using these items can be used to measure the quality of a passengers' in-flight experience. This scale will be useful to the commercial airline industry so as to be able to gauge the satisfaction of their customers with the quality of their in-flight experience.

It must be mentioned that the final scale that was developed, was done so only using positive words associated with in-flight experience quality. Previous research has suggested if positive and negative items are randomly placed within the scale, there is a risk of "detrimental effect on psychometric properties of a measure" [16]. Hinkin (1998) therefore suggested that it would be preferable if only one spectrum of the scale be used as a measure [17]. While having both sets of items does have its advantages, this study chose to only utilize positive terms in order to avoid cognitive confusion in passengers when actually using this scale in real world practical settings.

#### A. Validity

A scientifically developed scale or measure must be successfully tested for validity. For this study, construct validity was tested and ensured in two ways. Firstly, since this measure is to be filled out by passengers, it was important that actual aviation consumers themselves were involved in the item generation and development of the scale. Expert opinions are of value to the industry but they often times have a different mindset as the average consumer and may miss certain aspects that are important a traveller. In this study, passengers were not only responsible for the initial generation of items, but also the secondary narrowing down stages that lead to the final items. Secondly, through the data analysis, it was found that when a factor analysis was conducted on the final items, they all loaded

on a single factor. Both these two findings suggest strongly that the final scale has good validity.

#### B. Reliability, Consistency & Discriminability

A scientific and useful scale must not only have validity but also good reliability. Field (2009) stated that the reliability of an instrument is based on how consistently it measures what it is intended to be measured [18]. In order to test for reliability, the final scale underwent two different tests. Firstly, a Cronbach's Alpha test was conducted to measure internal consistency. The findings suggest high internal consistency as the values ranged from .84-.86. Additionally, a Guttman Split-half test was conducted and the resulting values ranged from .87-.92. Both these tests suggested the presence of high internal consistency and reliability for the final scale items. The last area of focus was the ability of the scale to discriminate and differentiate between a good and a bad in-flight experience, in order to prove useful in the aviation industry. The scale also showed low standard deviations and large effect sizes. Since the scale was able to discriminate well between good and bad in-flight experiences, it showed its versatility, and its effectiveness to be used appropriately by the industry.

#### C. Practical Implications

This scale has practical benefits not only to the commercial air operators, but the aviation industry as a whole. By implementing this scale commercial operators, may be able to gauge the passengers' satisfaction with their in-flight experience and the overall quality of the product they offer. The satisfaction of the customer with the product has a direct relationship to the economics of the company as it affects revenue and profit. If customers are dissatisfied with their experience, it may influence them to choose to fly a different operator in the future. In the same vane as competition comes the ability of this scale to be used as a marketing tool and in the fight for customers. This scale may be used as a method potentially ranking consumer satisfaction with the in-flight experience and therefore allow a quantifiable metric of comparison. Lastly, more in-depth use and analysis of the results from this scale will allow operators to also identify areas that passengers rate poorly. This can help them to focus on areas that need improvement and be able to gauge the priority of the same.

#### D. Limitations

As per any research endeavor, there are certain limitations associated with the study. Firstly, the final items identified and used for this scale are not defined. Different sets of passengers may define or interpret certain words differently, and therefore lead to ambiguity in what is being measured. Along those same lines, since this scale was designed, created and tested using American participants, it is only valid and reliable so long as it is used on passengers within the United States. Invalid results may be generated if this scale is used outside of this population. Lastly, the data collection source of MTurk has its own limitations. Data collected from this online source is deemed to be as reliable laboratory data [14], [15] however it may be that the views do not completely represent those of the population of the United States.

### E. Future Research

As mentioned as one of the limitations of the study, the items or terms used in the scale are not defined. Clear definitions of each term may be needed in order to reduce ambiguity or variations in results. By clearly defining each term, future studies can be certain as to what is actually being measured. Future research may seek to identify the depth and scope for each term. In order to solidify the applicability of the scale, future research should conduct field studies with actual airline passengers in order to fully test the scale. Lastly, it was mentioned that this scale might produce invalid results if used outside of the research population. Future studies may seek to develop similar valid and reliable scales on this topic using participants from different countries so that those scales will be usable by those populations. It may also be possible to conduct research in several different countries of the world and attempt to assemble a universally valid and reliable scale that can be used by the entire industry.

### IV. CONCLUSION

The primary goal of this research was to create a valid and reliable scale to measure passengers' in-flight experience quality. While other metrics and measurements of customer service and passenger satisfaction may exist, this scale is differentiated by the fact that actual airline consumers were the primary source of item generation and were used in every stage of the scale's development. The value of the scale lies in the creation of a universal metric that commercial airlines within the United States can use to gauge customer satisfaction with the quality of the in-flight experience. Additionally, this scale can provide airline managers with information on areas of the in-flight experience that may not be up to standard or unsatisfactory to passengers. This scale development procedure can lastly aid in providing a template for future scientific scale developments in several industries.

### Appendix

Please respond how strongly you agree or disagree with the following statements.

1. The airplane was in good condition				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
2. The airplane arrived on time				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
3. The airplane had comfortable chairs				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
4. The airplane air conditioning was favorable				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
5. The airplane had clean smelling air				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
6. The airplane had a comfortable cabin temperature				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
7. The airplane had comfortable seat spacing				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				
8. The airplane had good customer service				
Strongly Disagree	Disagree	Neutral		Agree
Strongly Agree				

### REFERENCES

- [1] E. Pels, P. Nijkamp, and P. Rietveld, "Airport and airline competition for passengers departing from a large metropolitan area," *Journal of Urban Economics*, vol. 48, no. 1, pp. 29-45, 2000.
- [2] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "A conceptual model of service quality and its implications for future research," *The Journal of Marketing*, pp. 41-50, 1985.
- [3] J. L. Heskett and L. A. Schlesinger, *Putting the 164-174*, 1994.
- [4] L. Moutinho, "Consumer behaviour in tourism," *European Journal of Marketing*, vol. 21, no. 10, pp. 5-44, 1987.
- [5] P. L. Ostrowski, T. V. O'Brien, and G. L. Gordon, "Service quality and customer loyalty in the commercial airline industry," *Journal of Travel Research*, vol. 32, no. 2, pp. 16-24, 1993.
- [6] T. L. Keiningham, F. V. Morgeson, L. Aksoy, and L. Williams, "Service failure severity, customer satisfaction, and market share an examination of the airline industry," *Journal of Service Research*, 2014.
- [7] L. F. Cunningham, C. E. Young, and M. Lee, "Perceptions of airline service quality pre and post 9/11," *Public Works Management and Policy*, vol. 9, no. 1, pp. 10-25, 2004.
- [8] J. D. Power. (2015). North America airline satisfaction study. [Online]. Available: <http://www.jdpower.com/press-releases/2015-north-america-airline-satisfaction-study>
- [9] K. M. Elliott and D. W. Roach, "Service quality in the airline industry: Are carriers getting an unbiased evaluation from consumers?" *Journal of Professional Services Marketing*, vol. 9, no. 2, pp. 71-82, 1993.
- [10] L. J. Truitt and R. Haynes, "Evaluating service quality and productivity in the regional airline industry," *Transportation Journal*, pp. 21-32, 1994.
- [11] R. Mehta, S. Rice, D. Carstens, I. Cremer, and K. Oyman, "A brief intermodal rail network (IRN) scale: Establishing validity and reliability," *Journal of Sustainable Development*, vol. 8, no. 6, 2015.
- [12] S. Rice, R. Mehta, L. Steelman, and S. R. Winter. (2014). A trustworthiness of commercial airline pilots (T-CAP) scale for Indian consumers. *International Journal of Aviation, Aeronautics, and Aerospace*. [Online]. Available: <http://commons.erau.edu/ijaaa/vol11/iss3/3>
- [13] S. Rice, R. Mehta, S. R. Winter, and K. Oyman. (2015). A trustworthiness of commercial airline pilots (T-CAP) scale for american consumers. *Journal of Aviation Technology and Engineering*. [Online]. Available: <http://docs.lib.purdue.edu/jate/vol4/iss2/6/>
- [14] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's mechanical turk: A new source of inexpensive, yet high-quality data?" *Perspectives on Psychological Science*, vol. 6, no. 3, pp. 3-5, 2011.
- [15] L. Germine, K. Nakayama, B. C. Duchaine, C. F. Chabris, G. Chatterjee, and J. B. Wilmer, "Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments," *Psychonomic Bulletin and Review*, vol. 19, no. 5, pp. 847-857, 2012.
- [16] D. A. Harrison and M. E. McLaughlin, "Exploring the cognitive processes underlying responses to self-report instruments: Effects of item content on work attitude measures," in *Proc. the 1991 Academy of Management Annual Meetings*, pp. 310-314, 1991.
- [17] T. R. Hinkin, "A brief tutorial on the development of measures for use in survey questionnaires," *Organizational Research Methods*, vol. 1, no. 1, pp. 104-121, 1998.
- [18] A. Field, *Discovering Statistics Using SPSS*, London, UK: Sage, 2009.

**Rian Mehta** is a graduate teaching assistant at the Florida Institute of Technology. He pursuing a Ph.D. in aviation science. He received his bachelor's degree in aviation management with flight in 2013 and his master's degree in applied aviation safety in 2015, both from the Florida Institute of Technology. His research interests include consumer perceptions, automation, pilot fatigue, and decision-making in the cockpit.

**Stephen Rice** is an associate professor at Florida Institute of Technology. He received his Ph.D. in human factors from the University of Illinois at Urbana-Champaign in 2006. His research interests include aviation psychology, automation, trust, stigmas, and human performance.