



Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Impact of Service Operations Failures on Customer Satisfaction: Evidence on How Failures and Their Source Affect What Matters to Customers

Shannon W. Anderson, L. Scott Baggett, Sally K. Widener,

To cite this article:

Shannon W. Anderson, L. Scott Baggett, Sally K. Widener, (2009) The Impact of Service Operations Failures on Customer Satisfaction: Evidence on How Failures and Their Source Affect What Matters to Customers. Manufacturing & Service Operations Management 11(1):52-69. <http://dx.doi.org/10.1287/msom.1070.0193>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2009, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Impact of Service Operations Failures on Customer Satisfaction: Evidence on How Failures and Their Source Affect What Matters to Customers

Shannon W. Anderson

Jesse H. Jones Graduate School of Management, Rice University, Houston, Texas 77005, and
Department of Accounting and Business Information Systems, University of Melbourne,
Melbourne, Australia, swa@rice.edu

L. Scott Baggett, Sally K. Widener

Jesse H. Jones Graduate School of Management, Rice University, Houston, Texas 77005
{baggett@rice.edu, widener@rice.edu}

Research in consumer psychology shows that customers seek reasons for service failures and that attributions of blame moderate the effects of failure on the *level* of customer satisfaction. This paper extends research on service operations failures by hypothesizing that attributions of blame also affect *what matters* to the customer during service failures. Specifically, we hypothesize that the relative weights that customers assign to key service elements in reaching an overall assessment of customer satisfaction are affected by customer attributions of blame for service failures. We use the U.S. airline industry as a quasi-experimental research setting to investigate the components of customer satisfaction for three samples of customers who experience (1) routine service, (2) flight delays of external (i.e., weather) origin, and (3) flight delays of internal origin. Although the level of customer satisfaction is lower for all service failures, we find that the key components of satisfaction differ between delayed and routine flights only when customers blame the service provider for the failure. Specifically, when delays are of external origin satisfaction is lower than for routine flights, but there is virtually no difference in the weight that customers assign to the components of customer satisfaction (including employee interactions). In contrast, when delays are of internal origin, satisfaction is lower than for either routine flights or flights delayed by external factors, and employee interactions have a significantly diminished role in customer satisfaction evaluations. Contrary to the popular view that employee interactions take on a greater role in determining customer satisfaction during service failures, we find that the opposite is true if the customer attributes blame to the service provider. Our findings highlight the important role of customer attributions during service failures and present more nuanced evidence on the role of employee-customer interactions in mitigating the effects of service failures on customer satisfaction.

Key words: service operations; customer satisfaction; attribution theory; failure recovery; airline industry

History: Received: November 7, 2006; accepted: July 11, 2007. Published online in *Articles in Advance* January 4, 2008.

1. Introduction

The service operations literature has focused on minimizing the impact of failures and designing robust operations that include recovery processes (e.g., Soteriou and Chase 1998, Cook et al. 2002, Craighead et al. 2004). A complicating factor is the customer, who is likely to perceive the failure differently from the service provider. Attribution theory posits that customers seek reasons for failures (Weiner 1985, 2000), and empirical research shows that attributions of responsibility for failures affect customer

satisfaction (e.g., Folkes 1984, Folkes et al. 1987). Studies typically focus on the change in the *level* of overall customer evaluations associated with service failure (e.g., Bejou and Palmer 1998, Tsiros et al. 2004). In this paper, we synthesize theory from service operations and service marketing to test the hypothesis that service failures affect *what matters* to the customer—or the *composition* of overall customer satisfaction.

Natural events in the U.S. airline industry make it a useful quasi-experimental setting for testing our propositions. Using a comprehensive database of

customer satisfaction surveys from an independent market research firm, we compare models relating key service elements to customer satisfaction for three comparable samples of coach class business passengers who experienced (1) routine service, (2) flight delays of external (i.e., weather) origin, and (3) flight delays of internal origin. In these models, the greater the influence of a particular service element on evaluations of the overall service, the more importance the customer assigns to it. We use a holdout sample to identify the key service components of customer satisfaction for passenger air travel. These include physical attributes of the service (e.g., food and seat space), employee-customer interactions (e.g., with flight attendants and gate agents), and core operating performance outcomes (e.g., flight timeliness). We then estimate the model of customer satisfaction for the research sample using a three-group, structural equations model (SEM). Following Donaldson (2001), we estimate a model that allows each treatment group to have a different structural model (i.e., different composition of satisfaction).

We find that for customers who experience routine service, overall satisfaction is well explained by evaluations of four service elements (in order of importance):¹ employee-customer interactions, personal space onboard the plane, the aircraft, and food. For customers who experience delays, the mean level of satisfaction is lower and a fifth service element—timely arrival—is significantly related to overall satisfaction. However, even with delays, timeliness ranks fifth in importance, compared with the other service elements.² Although customers value the same service elements, we find significant differences in the relative importance of these elements. The difference is

associated with attributions about the service failure. External sources of failure (low blame) are *not* associated with meaningful differences in what matters to the customer, compared with routine service. In contrast, internal sources of failure that are likely to be accompanied by the airline being blamed are associated with different weights being assigned to service elements.

Examining specific elements of the service concept, we find pronounced differences in the role of employee-customer interactions. The service operations literature posits that in times of failure, capable employees are critical to mitigating the effects of operations failures (e.g., Kellogg and Chase 1995, Miller et al. 2000, Goldstein 2003). However, employees may be hampered in this role if customers blame the firm (though not necessarily the specific employee) for the failure (e.g., Chung-Herrera et al. 2004). We find evidence supporting both perspectives. Customers who experience weather-related delays assign the greatest importance to employee interactions, and customers who experience delays for which the airline is likely to be responsible assign a *diminished* importance to employee interactions. Sensitivity analysis indicates that our conclusions are robust to the inclusion of air carrier effects.

In sum, we conclude that what matters to customers differs between customers who experience routine service and those who face service failures; however, this difference is contingent on customers' attributions about the source of failure. Although we find that employee interactions are the most critical determinant of customer satisfaction for coach class passengers traveling for business, we find no evidence that employees are *more* critical during operations failures of external origin. However, during failures for which the air carrier is likely to be held responsible, we find that employee interactions become markedly *less* important to customers. This suggests that customers may not trust employees to solve problems that originated with the firm.

The paper is organized in five sections. Section 2 develops hypotheses about the effects of service operations failures on customer satisfaction. Section 3 discusses the quasi-experimental research setting and describes the survey data, variable measures, and analysis methods. Section 4 presents the results of

¹ Chi-squared difference tests reveal that three of the four service elements are significantly different in order of importance. The importance of employee-customer INTERACTION ranks significantly higher than the importance of PERSONAL SPACE ($p < 0.01$), which ranks significantly higher than the importance of AIRCRAFT ($p < 0.01$). The importance of AIRCRAFT and FOOD are not significantly different.

² For delay groups, Chi-squared difference tests reveal that the importance of TIME (the fifth element in order of magnitude) is not significantly different from FOOD (the fourth element), but is significantly different from the AIRCRAFT, the third element ($p < 0.10$, $p < 0.05$, for the weather and other delay groups, respectively).

the analysis as well as sensitivity tests of the results. Section 5 summarizes key findings and managerial implications and discusses opportunities for future research.

2. Literature Review and Research Hypotheses

Sasser et al. (1978) coined the term “service concept” to describe the bundle of elements packaged for sale to the customer. Key service elements include core attributes that define the basic service provided; peripheral physical attributes, those features and amenities that are bundled with the core service attributes; and interactional attributes that define how employees interact with customers in service delivery (Booms and Bitner 1981). Overall customer satisfaction is a function of satisfaction with service attributes (Oliver 1997). The importance of each attribute is defined by the degree to which changes in attribute performance are accompanied by changes in overall satisfaction (Anderson and Mittal 2000).

Prior research posits a unique operational role for employees in creating customer satisfaction in service industries (e.g., Loveman 1998, Cook et al. 2002). Research showing that employee performance in the service encounter affects the level of customer satisfaction includes studies by Westbrook (1981) in retail settings; Loveman (1998) in banking; Bitner et al. (1990) in the hotel, restaurant, and airline industries; and Goldstein (2003) in hospitals. Although service failures negatively affect customer evaluations and repurchase intentions (e.g., Bejou and Palmer 1998, Tsiros et al. 2004), employees’ ability to diagnose and respond dynamically to problems is an opportunity for real-time service recovery. Bitner et al. (1990) find that a positive and appropriate response from an employee can overcome the negative effects of a service failure. Dewitt and Brady (2003) review conflicting theories and evidence on the impact of customer-employee rapport on a customer’s response to failure and conclude that most studies find a positive relation.

Service failures are an example of what the marketing literature terms a “critical incident.” Critical incidents are events in the service encounter that significantly influence the consumption process in a negative or positive manner (Bitner et al. 1990). Beaujean

et al. (2006, p. 64) describe these as “‘moments of truth:’ those few interactions...when customers invest a high amount of emotional energy in the outcome.” Attribution theory posits that customers evaluate a critical incident in an attempt to understand the underlying reasons for it. While visible failures affect customer evaluations of service, attribution theory posits that customers seek reasons for failures and that these reasons also influence evaluations of the service encounter (Weiner 1985, 2000).

Weiner (1985) posited that customers consider three things: (1) *locus of control*, the identification of the responsible party; (2) *controllability*, the influence the various parties have over the incident and its resolution; and (3) *stability*, whether the incident is likely to be repeated. Customers ascribe responsibility only after attributing causality. Folkes et al. (1987) interviewed passengers waiting at an airport for delayed flights and found that they attributed the cause to mechanical problems, airline personnel, delayed departure of a previous flight, other passengers, and weather. For each cause, the airline was ascribed somewhat different levels of responsibility.

The customer also considers whether the effect of the critical incident is within the provider’s control and whether actions could have been taken to mitigate that effect. For example, Folkes et al. (1987) found that passengers believed that the airline controlled personnel problems but had little control over weather conditions.

Finally, customers hold service providers responsible for routines to mitigate the effects of stable sources of critical incidents. Folkes et al. (1987) found that passengers believe that delays caused by airline personnel are the most stable and those caused by mechanical problems are least stable. In sum, different attributions are associated with different behavioral outcomes (i.e., complaints) and different affective outcomes (i.e., anger) (Oliver 1997, Weiner 1985) as well as with reductions in customer satisfaction (Bitner 1990).

In passenger air travel, our research setting, industry-level research has shown that performance failures and inconsistency in performance are associated with increased complaints (Behn and Riley 1999, Tsikriktsis and Heineke 2004). Studies using passenger-level data identify delays, cancelled flights,

overcrowding, and noisy children as significant negative critical incidents associated with different attributions of blame and customer reactions (Edvardsson 1992, Gilbert and Morris 1995). Folkes et al. (1987) found that passengers who judged the delay to be stable or controllable experienced more anger, complained more, and stated that they are less likely to repurchase a ticket on the airline. Using data from 287 delayed passengers, Taylor (1994) found that if customers perceived an airline delay to be controllable or stable, they felt more anger, and anger was negatively associated with evaluations of overall service. A limitation of these studies is that they focus on customer sentiments *prior to* the flight. Customers were not surveyed at the conclusion of the flight, so failures and related attributions are not linked to satisfaction with the flight.

Bitner (1990) is noteworthy for directly linking attributions about service failures to overall customer satisfaction in air travel. She hypothesized that before forming an opinion on satisfaction, the customer first attributes the results of the encounter along the dimensions of locus of control, stability, and controllability. Thus, the causal sequence of previous research is modified. An attribution leads to customer satisfaction, which in turn leads to behavioral and affective responses. Bitner tested her model experimentally with 145 subjects and found support for the hypothesis: In the event of service failure, customer perceptions of controllability and stability are negatively associated with satisfaction. However, Bitner's study focused on the *level* of overall satisfaction; thus, it provides no basis for assessing whether service operations failures and related attributions alter the *composition* of overall satisfaction.

Although the literature is fairly clear that service failures and customer attributions affect the *level* of overall customer satisfaction, prior studies have not considered whether failures and attributions affect the *composition* of customer satisfaction. The level of overall satisfaction may be altered by either a change in the level of satisfaction with constituent elements or a change in the importance of these elements. Other studies have shown that customer characteristics (Kekre et al. 1995), the competitive environment (Anderson and Mittal 2000), and temporal factors (Mittal et al. 2001) can also change what matters to

customers. Consequently, our first hypothesis examines the proposition that service failures alter what matters to customers.

HYPOTHESIS 1 (H1). *When service operations failures arise, the composition of the customer's evaluation of satisfaction will differ compared with the composition of satisfaction during routine service delivery.*

The service failure and recovery literature posits that an appropriate response by employees may compensate for the negative effects of the failure on customer satisfaction (Soteriou and Chase 1998, Miller et al. 2000, Cook et al. 2002, Dewitt and Brady 2003). This compensation may arise if customers "weight" interactions with employees more heavily in their overall satisfaction evaluation during service failures than during routine service. Alternatively, there may be no change in the importance of employee interactions, and any compensating differential may arise with a higher level of satisfaction with employee interactions. Because studies of service operations failures typically identify employees as especially critical to effective recovery, we hypothesize that employee interactions are more important to overall customer satisfaction during service failures:

HYPOTHESIS 2 (H2). *When service operations failures arise, the customer's evaluation of satisfaction will become more dependent on satisfaction with interactions with employees than during periods without service failures.*

Research has shown that overall customer satisfaction is lower for service failures of internal origin than for failures of external origin. This may arise if the level of satisfaction with some or all of the service elements is lower for internal attributions than for external attributions and/or if the structure of what matters to customers changes in such a way that even equivalent levels of satisfaction with constituent service elements produce different overall evaluations.

Using a simple example, if overall satisfaction is related to the level of satisfaction with two service elements, A and B, in the following way: overall satisfaction = 0.25 (satisfaction with A) + 0.75 (satisfaction with B), then differences in the level of overall satisfaction may be brought about by shifts in the level of satisfaction with A or B (i.e., shifting to a different point on the satisfaction measurement scale) or

by shifts in the importance attached (0.25 and 0.75) to each element. Attribution theory provides little guidance on the relative importance of these alternatives in explaining the pattern of overall satisfaction that accompanies failures of different sources. More generally, Oliver (1997) notes that there is little overlap between marketing research that focuses on the cognitive antecedents (i.e., attributions) of customer satisfaction and research that focuses on the constituent service elements of customer satisfaction.

Prior research provides some support for the hypothesis that attributions are accompanied by a shift in what matters to customers. A number of studies (e.g., Kamakura et al. 2002, Stafford 1994) have shown that employee interactions are more influential than other service elements in overall satisfaction evaluations. One theoretical reason, advanced by Bitner (1990), for expecting employee interactions to be more important is that services are inherently difficult to evaluate. Because they are high in credence or experience attributes, service evaluations depend heavily on peripheral cues. Employee actions and physical environments are particularly salient cues in settings in which performance is difficult to measure. When service failures arise, core operating performance (i.e., on-time arrival) ceases to be a sufficient statistic for evaluating service quality. In these situations, customers understandably must rely on other service elements to reach evaluation conclusions.

Together these factors suggest a structural difference between the weight that customers assign to service elements for routine versus delayed flights (H1) and the relative importance of employee interactions versus other service elements (H2). However, customers may view employees with suspicion if blame for the failure is attributed to the firm (Chung-Herrera et al. 2004). In particular, psychological approaches to service recovery (e.g., apologies, empathy) described by Miller et al. (2000) may prove less effective in these circumstances. So, although the theory is incomplete and thus the analysis is admittedly exploratory, we hypothesize that compared with failures of external origin, customers on flights with delays of internal origin place less weight on employee interactions and more weight on physical cues (e.g., aircraft, in-flight amenities) in evaluating overall satisfaction. In this manner, attributions of blame interact with the

weights that customers assign to elements of the service concept. Stated more generally:

HYPOTHESES 3A AND 3B (H3A AND H3B). *The relations hypothesized in H1 and H2 are moderated by the customer's attribution of blame for the service failure.*

Goldstein et al. (2002) point out that we know very little about the role of the service concept during service failure and recovery. In this paper we seek to address this shortfall by testing whether the service elements that matter most to customers change with service failures.

3. Research Design, Data, and Variable Measures

3.1. A Quasi-Experimental Research Setting

Service failures of internal and external origin are common in our sample. We exploit these common occurrences using a quasi-experimental research design in which we compare the relation between key service elements of customer satisfaction with delayed flights against a benchmark model for flights without delays. This model allows us to test whether delays are associated with a lower level of satisfaction, as previous studies have found, and whether delays are associated with a change in what matters to customers, as we hypothesize in H1 and H2. Partitioning our sample of delayed flights into delays associated with external sources (i.e., weather) and those associated with sources more likely to be attributed to the air carrier (i.e., equipment failure, crew shortages), we are able to test whether compositional differences in customer satisfaction are related to attributions (H3A and H3B).

3.2. Data and Sample Construction

A major vendor of customer satisfaction data to the airline industry provides data for this study. The vendor distributes a self-addressed, postage-paid one-page survey to a random sample of passengers on randomly selected flights. Passengers are asked to provide general demographic information, information on the particular flight (e.g., date, airline, flight number, class of service flown), and evaluations of many aspects of the air travel experience as well as of the overall flight experience. The ratings are on

Table 1 Sample Construction

	Sample <i>N</i>
(1) Number of customer surveys from flights in January, April, July, and October 2000	24,796
Less surveys completed by (*)	
first and business class passengers	3,411
frequent flyer award and upgrade travelers	4,619
travelers flying for leisure purposes	14,164
(2) Number of surveys in (1) completed by paying, economy-class, business travelers	7,848
(3) Number of surveys in (2) from flights to and from the 44 most trafficked U.S. airports	4,723
Number of these flights that are recorded as “delayed” by the BTS	1,260
Number of these delays that are predicted to be due to weather	703
Number of these delays that are predicted to be due to other causes	557
Number of these flights that are recorded as “on time” by the BTS	3,463

Notes. This table summarizes the steps taken to construct the three samples of delayed (weather and other sources) and on-time flights.

*Passengers may fit within more than one of these categories.

a seven-point scale with anchors “very poor” and “excellent.”

We first construct three otherwise comparable treatment groups of passengers: (1) those who experienced delays associated with external sources, (2) those who experienced delays associated with internal sources, and (3) those who experienced no delay. Because delays are more likely in some seasons than others, we start by selecting the year 2000 for study and sample from all four seasons, starting with customer surveys for flights completed in January, April, July, and October. We use data from a single calendar year because input price changes (e.g., labor negotiations) often follow an annual cycle and are accompanied by fare increases. We selected the year 2000 to avoid the aftermath of the 2001 terrorist attacks. As Table 1 indicates, for the four months we have 24,796 customer surveys.

The data do not include information on the ticket price. In an effort to control for the effect of price on customer satisfaction, we eliminate surveys completed by first-class passengers (i.e., very high fares), by those who indicate that they are traveling for leisure rather than business purposes (i.e., typically lower fares because of schedule flexibility and advance purchase), and by passengers who used frequent flyer awards or upgrades (i.e., free or discounted fares).³ By focusing on paying, coach class

business travelers, we attempt to limit our analysis to customers who paid similar fares.⁴ Removing passengers who are traveling in first class also ensures that our sample of passengers evaluated a similar service offering. After eliminating passengers according to these criteria, 7,848 surveys remain.

We use the Bureau of Transportation Statistics (BTS) aviation data library (http://www.transtats.bts.gov/Databases.asp?Mode_ID=1&Mode_Desc=Aviation&Subject_ID2=0) to sort passengers who experienced a flight delay into two groups. The BTS records a flight as “on time” if it arrives at the gate of the destination

troubling because even coach fare frequent flyers may receive special treatments. However, because we have no reason to suspect that frequent flyers are disproportionately present in any of the treatment groups, this omission is unlikely to jeopardize our tests.

⁴ In the absence of historical price data, we substantiate this claim by collecting current price data on the 160 routes flown by at least 10 respondents in our final sample (together, 72% of the full sample). We examine four scenarios. We first consider one week (fewer than 7-day) advance purchase of nonstop flights that (a) depart on Sunday and return on Wednesday, and (b) depart on Sunday and return on Friday. We then repeat the analysis using a 7- to 14-day advance purchase. In all four scenarios, the range of variation for each of the 160 routes, calculated as: [(max price – min price)/min price], ranged from 0% to no more than 5.71%, indicating little within-flight variation in price for common business purchase patterns. We compare these imputed prices for our three treatment groups and find no meaningful differences in either the mean price or the mean price per mile (an approximation of value) for any of the treatment groups. Thus, we feel confident that our sample construction steps have mitigated the potentially deleterious effects of omitting price from the analysis.

³ Unfortunately, our data do not indicate whether passengers are members of the airlines’ frequent flyer program. This is potentially

airport fewer than 15 minutes after the scheduled arrival time. Cancelled and diverted flights are considered late. The data are collected by all airlines with at least one percent of domestic passenger revenues and reported monthly to the Department of Transportation. In 2000, the BTS did not collect data on the cause of delay; consequently, we follow an approach used in Mazzeo (2003) to separate delayed flights into those most likely associated with weather and those associated with other causes. We identify 44 airports that together account for a majority of flights in the United States (http://www.bts.gov/programs/airline_information) and obtain data from the National Weather Service on weather conditions (e.g., temperature, amount of snow, and amount of rain) for these airports each day. As Table 1 indicates, 4,723 of the remaining sample originated and departed from these airports, of which 1,260 were on flights that the BTS recorded as delayed.

We match the weather data to our survey data based on flight number, date, and the airports of origin and destination and partition the sample based on extreme values of the weather variables. For each month we use a combination of three weather variables to develop a model that produces the highest accuracy rate for predicting delays.⁵ To the extent that weather at the origin or destination airport is not the source of a weather delay (e.g., en route weather conditions can also produce delays), the record will be misclassified as an internal source of delay, biasing against finding the hypothesized differences (H3A and H3B) between the two groups of delayed passengers. These procedures yield three groups of customer satisfaction surveys: (1) 703 surveys from customers who were likely delayed due to weather, (2) 557 surveys from customers who were likely delayed due to other causes, and (3) 3,463 surveys from customers who were not delayed.

As expected, air carriers are not represented in the three samples in proportion to their share of passenger traffic in the U.S. domestic air travel market. Nor are they equally represented in the three samples.

This is because carriers have different route structures that are not fully represented in the 44 most trafficked airports. Moreover, the 44 airports experience distinct local weather patterns that subject flights in particular airports to greater or lower weather delay risk than other airports. Finally, air carriers may have unique operational competencies that produce differences in the incidence of delays of internal source. Because air carriers are not randomly assigned to the delayed and on-time groups, we conduct sensitivity analysis to rule out the possibility that air carrier effects may explain any differences in the level or composition of customer satisfaction in the three treatment groups.

3.3. Variable Measures: Identifying Service Attributes of U.S. Air Travel

We empirically identify service elements that are important to customers by using a holdout sample of customer surveys from several months in 2000 that are not used in our hypothesis tests.⁶ We use exploratory factor analysis with an oblique rotation to reduce the large number of rated attributes of airline service to a much smaller number of service elements that are viewed as distinct by passengers (see online appendix for factor solution).

Some elements of the service are experienced by a small number of passengers (e.g., in-air telephone service). Listwise deletion of cases that omitted evaluation of unused services would produce a biased sample because evaluations are missing by design rather than at random. Consequently, we limit our analysis to service elements for which at least 50 percent of all surveyed customers provide an evaluation and use full information maximum likelihood estimation methods. The items load on six factors with a high degree of coherence. The factor solution is generally well behaved, with items typically loading in

⁵ For example, the rule for January was: Snow at origin (last 24 hours) ≥ 5 inches OR precipitation (last 24 hours) at origin ≥ 6 inches OR [minimum temperature (last 24 hours) at origin ≤ 32 and arrival delay ≥ 20 minutes].

⁶ It is important wherever possible to use a holdout sample to develop variable measures. Although it is not uncommon, using the estimation sample to first identify the latent constructs of interest and to then simultaneously estimate the measurement model and the structural model is problematic. The reported fit statistics of structural equations modeling are based on a joint test of the measurement model and the structural model. However, if the measurement model has been tested and modified outside of the SEM estimation using a common sample, then the data are overfit and the model fit statistics are overstated.

excess of 0.40 on one factor. Six items with significant cross loadings are deleted to improve the uniqueness of the factors (Verbeke and Bagozzi 2002). In an untabulated analysis we repeated the confirmatory analysis separately for the nine airlines that comprise our study and found the factor structure to be qualitatively similar for all airlines.⁷

The factor structure conforms nicely to prior research that considers the service concept of passenger airlines (Danaher 1998, Bejou and Palmer 1998). Key elements include employee interactions, the physical setting (i.e., seating, food, aircraft interior) and core flight performance. The first factor, labeled INTERACTION, loads on evaluations of the gate and flight personnel. Although in our study this is a single factor, consistent with Soteriou and Chase (1998), we find it is comprised of two critical aspects of employee interactions: communication (i.e., providing timely and accurate information on board and in the gate area) and intimacy (i.e., responsiveness to passenger needs). Aspects of the physical setting are evident in the third, fourth, and fifth factors. The third factor, labeled PERSONAL SPACE, loads on evaluations of arm and shoulder room, leg room, and seating comfort. The fourth factor FOOD, loads on evaluations of the amount and quality of food served on board. The fifth factor, labeled AIRCRAFT, loads on evaluations of the appearance, cleanliness, and condition of the airplane cabin.

Finally, factors two and six reflect the core operating performance of air travel. The second factor, labeled FLIGHT, loads on evaluations of the quietness of the aircraft during flight and take-off and flight smoothness. The sixth factor, labeled TIME, reflects customer evaluations of on-time arrival and departure. (Note that customers are not asked to provide objective measures of on-time arrival and departure; rather, they are asked to evaluate their *satisfaction* with the arrival and departure times.)

⁷ This does not mean that customers are equally satisfied with the air carriers or with each element of an air carrier service. For example, although customers of different airlines have a common notion of what constitutes personal space on board, they may be much less satisfied with the personal space provided by one airline than that provided by another. We are not permitted to report carrier-specific satisfaction for service elements or overall satisfaction.

We use confirmatory factor analysis to combine three measures of overall satisfaction: overall satisfaction with the flight experience, overall satisfaction with the aircraft, and overall satisfaction with on-board services. To establish construct validity, we collected quarterly customer satisfaction data on airlines included in the University of Michigan's American Customer Satisfaction Index (ACSI) from the same period. The customer satisfaction rankings for the ACSI correlated well ($r = 0.61$) with rankings produced using our survey data.

In Table 2 we present descriptive statistics for the raw responses to the survey items for our estimation sample.⁸ Survey items are grouped according to factors identified in the holdout sample analysis and presented separately for each of the three passenger groups. Untabulated statistics indicate that all survey item responses are within the tolerable range of skewness and kurtosis and that the full rating scale (1–7) is used for all questions for all groups. Confirmatory factor analysis of the proposed variables for the research sample indicates a single well-defined factor for each variable. The standardized Cronbach alphas range from 0.79 to 0.91 and indicate a high degree of coherence among items comprising each scale. Moreover, alpha exceeds the correlation of the latent construct with the remaining latent variables. Consequently, we conclude that our measures also have discriminant validity.

Because the data are drawn from responses by individual customers to a single survey, we conduct a Hartman's one-factor test to assess common method bias. The factor solution yields five factors with Eigenvalues greater than 1.0; thus, we conclude that there is little evidence of significant common method bias (Podsakoff and Organ 1986). Although one can never

⁸ We use Rubin's (1976, 1987) multiple imputation method to incorporate information contained in incomplete records (i.e., missing responses to a small number of items). Missing values are deemed missing at random and are replaced with an imputed value from a Monte Carlo Markov chain method that uses all remaining survey values. This imputation method is considered superior to other approaches because it relies on a random sample of missing values. Inferences from the imputed data reflect uncertainty in the missing values, and the multivariate distributions of the data are unchanged. The imputed item means differ from the reported raw item means (standard deviations) of Table 2, Panels A, B, and C, by less than 0.07 (0.09), 0.06 (0.07), and 0.04 (0.09), respectively.

Table 2 Descriptive Statistics for Raw Survey Responses and Estimated Common Measurement Model

Latent variables for service elements	Panel A: Weather delay			Panel B: Other delay			Panel C: No delay			Panel D: Measurement model	
	<i>N</i>	Item mean	Item std. dev.	<i>N</i>	Item mean	Item std. dev.	<i>N</i>	Item mean	Item std. dev.	Factor loading ^a	Std. error
Interaction (alpha = 0.79)											
Efficiency of boarding aircraft	679	4.92	1.63	540	4.77	1.67	3,393	5.49	1.30	0.931***	0.003
Wait to check in	680	5.18	1.78	544	5.05	1.82	3,387	5.48	1.55	0.936***	0.004
Helpfulness of check-in personnel	683	5.72	1.43	543	5.65	1.51	3,385	5.91	1.23	1.012***	0.003
Timely/accurate information on board	667	5.05	1.70	529	4.92	1.84	3,333	5.94	1.17	1.000	
Responsiveness of flight attendants	644	5.36	1.38	522	5.24	1.45	3,252	5.65	1.26	0.964***	0.003
Aircraft (alpha = 0.91)											
Cabin appearance	682	5.32	1.26	543	5.24	1.24	3,425	5.54	1.13	0.994***	0.002
Cleanliness of aircraft interior	684	5.32	1.21	545	5.28	1.24	3,420	5.59	1.15	1.003***	0.002
Condition of aircraft interior	687	5.33	1.20	545	5.28	1.20	3,418	5.56	1.14	1.000	
Time (alpha = 0.89)											
On-time arrival	649	3.20	2.14	500	3.30	2.17	3,149	6.09	1.25	1.000	
On-time departure	670	2.99	2.15	529	2.90	2.06	3,348	5.70	1.55	0.941***	0.003
Personal space (alpha = 0.88)											
Arm and shoulder room	685	3.93	1.74	545	3.84	1.71	3,415	4.19	1.70	0.994***	0.005
Leg room	684	3.99	1.87	541	3.90	1.83	3,407	4.18	1.85	1.000	
Seating comfort	684	4.18	1.71	542	4.08	1.69	3,389	4.50	1.69	1.059***	0.005
Food (alpha = 0.83)											
Amount of food	430	4.02	1.79	341	3.91	1.73	2,168	4.32	1.75	1.000	
Quality of food	436	3.69	1.63	343	3.49	1.67	2,185	4.10	1.62	0.942***	0.004
Flight (alpha = 0.85)											
Aircraft quietness during flight	677	4.99	1.33	527	4.94	1.33	3,368	5.26	1.26	1.023***	0.002
Aircraft quietness during take-off	680	4.91	1.35	530	4.81	1.42	3,373	5.13	1.29	1.000	
Smoothness of flight	673	5.07	1.38	524	5.19	1.35	3,327	5.55	1.16	1.056***	0.003
Customer satisfaction (alpha = 0.80)											
Overall experience of flight	688	4.47	1.76	542	4.21	1.86	3,363	5.65	1.25	1.000	
Overall rating of aircraft	679	4.83	1.36	535	4.69	1.57	3,353	5.22	1.27	0.944***	0.003
Overall on-board services	675	4.79	1.46	528	4.67	1.54	3,331	5.23	1.39	0.946***	0.003

^a As is required for structural equations modeling, the loading of one item for each latent variable is arbitrarily set equal to 1.0, and all other loadings are defined in relation to that item. As recommended, we select the item with the highest loading in our exploratory factor analysis of a holdout sample (see online appendix) for this treatment.

***, **, * corresponds to *p* values significant at the <0.01, 0.05, and 0.10, levels, respectively (two-tailed test).

eliminate the possibility of common method bias in a single-method study, the result of this test, as well as evidence that our measure of overall customer satisfaction correlates well with an independent measure of airline satisfaction (the ASCI), provides some assurance that our data are not inordinately influenced by this effect.

Several points about Table 2 deserve mention. First, casual inspection indicates that for every item, the mean evaluation rating of passengers who experienced no delay is greater than that for the passengers who experienced weather delays, which in turn is generally greater than that for passengers who experienced delays caused by other sources. These comparisons

are consistent with the expectation that regardless of source, delays reduce customer satisfaction, and with the hypothesis that delays with external attributions of blame are less prone to customer dissatisfaction than those with internal attributions of blame.⁹

⁹ Mean comparison tests reveal that, for every item, the mean evaluation of passengers who experienced no delay is greater than that of passengers who experienced weather delays (*p* value < 0.01). However, when comparing the weather delay group with the other delay group, with the exception of ratings of boarding (*p* value < 0.10), food quality (*p* value < 0.10), smooth flight (*p* value < 0.10), and overall flight (*p* value < 0.01), when one considers the standard deviation of responses, the other differences do not reach conventional levels of statistical significance.

A second observation is that in most cases the standard deviation of responses is similar across groups or is slightly greater for the two delayed groups. The largest differences emerge, again, in the ratings of timely performance. That delayed customers have a more dispersed opinion of timely performance is perhaps not surprising, because our categorization does not discriminate on the basis of the severity of the delay. What is more important for our analysis is that the distribution of delay severity does not differ greatly between the delayed groups. Specifically, responses from customers on weather-delayed flights are distributed across three levels of delay severity (15–24 minutes, 25–51 minutes, 52–356 minutes) as 29%, 35%, and 36%. The distribution for the other delay group is 40%, 30%, and 30%.

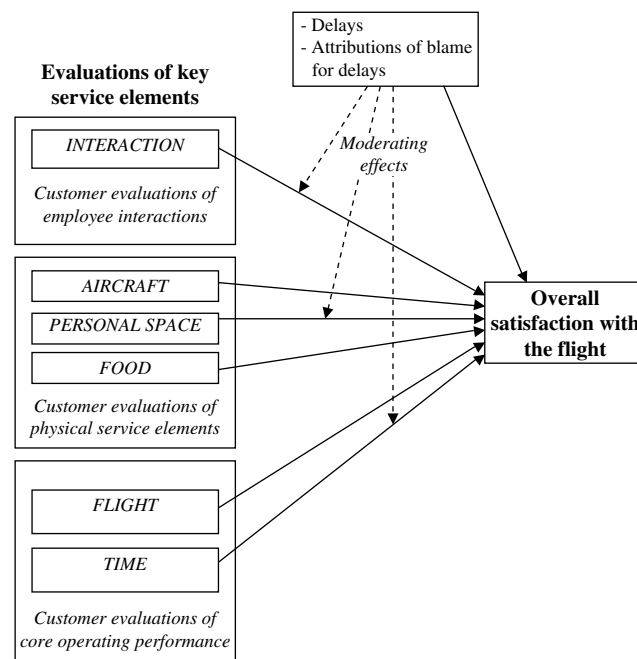
4. Results

4.1. Research Methods

Figure 1 depicts the conceptual relation between evaluations of key service elements and overall customer satisfaction with the flight. The coefficients that relate evaluations of each service element to overall satisfaction reflect the importance that customers assign to the service elements (e.g., Oliver 1997, Kamakura et al. 2002). Also depicted are the hypothesized direct and moderating effects of delays that are associated with different attributions of blame. Although developing the model relating attributes of the service concept to customer satisfaction is nontrivial, the objective of this research is testing the propositions that (1) when service failures arise, the coefficients that relate each element to overall satisfaction change (H1), particularly the coefficient related to INTERACTION (H2), and (2) these changes are associated with customer attributions of blame for the service failure (H3A and H3B).

The model relating service elements to overall customer satisfaction is estimated using SEM software, LISREL 8.5 (see online appendix). The measurement model and the structural model are jointly estimated using the full information maximum likelihood estimation method. SEM allows us to explicitly model measurement error of the latent dependent (i.e., customer satisfaction) and independent (i.e., service elements) variables as well as covariance among the

Figure 1 The Composition of Customer Satisfaction in Passenger Air Travel



Notes. This diagram depicts a conceptual model relating customer evaluations of key service elements to overall satisfaction. The coefficients that link service elements to overall satisfaction indicate the relative importance of the element to the customer. In addition to a direct effect of service failure, which has been previously reported, we hypothesize that service failures (delays) moderate the relation between evaluations of service elements and overall satisfaction (H1), in particular, that employee interactions become more important during delays (H2). We further hypothesize that delays associated with different attributions of blame for the air carrier are associated with differences in these relations (H3).

service elements. For larger samples such as ours, full information maximum likelihood methods provide efficient estimators when data are based on ordinal scales (Boomsma and Hoogland 2001).

To test the effects of service failures that are hypothesized to moderate the relation between satisfaction with attributes of the service concept and satisfaction with the overall flight experience, we use the multisample approach discussed in Rigdon et al. (1998). Specifically, the model of customer satisfaction is estimated *simultaneously* for the delayed and on-time groups, and the effect of the service failure is evident in group differences in the structural model parameters. This is conceptually similar to a pooled regression specification in which intercepts and coefficients are interacted with indicator variables for

group membership; a key difference is the modeling of covariance structure in the SEM approach. Ping (1995) argues that this SEM approach is preferred to alternatives when, as is the case for this study, the moderating variable is categorical (e.g., weather versus other delays) or when the functional form of the nonlinear relation is unknown.¹⁰

4.2. A Multigroup Structural Model of the Moderating Effects of Delays on the Relation between Key Service Elements and Overall Customer Satisfaction

Hypothesis H1 requires a test of whether allowing the structural parameters that relate service elements to customer satisfaction to differ for delayed customers improves model fit, compared with a model in which all customers are constrained to value the service elements identically. Similarly, H3A requires a test of whether further freeing the structural parameters for the groups that experience delays of internal versus external origin improves model fit. For the sake of parsimony, we present the full SEM model results and the tests of contrasts for the structural coefficients of the groups for only the three-group model (Tables 3 and 4). However, we conduct Chi-squared difference tests on the three nested models: the pooled (one-group) model, the two-group (delayed versus ontime) model, and the three-group (weather delay, other delay, on time) model. In support of H1, we find that model fit is significantly enhanced (Chi-squared difference = 1,347.37, with 13 degrees of freedom; p value < 0.000) when the structural parameters relating service elements and customer satisfaction are allowed to differ between delayed and on-time customers. In support of H3A, the model is further improved (Chi-squared difference = 30.05 with

¹⁰ In separate untabulated analyses, we also estimate the model using regression analysis with indicator variables to reflect group membership, substituting for each latent construct the average response to survey items that comprise the construct. The results are qualitatively similar to those reported for the SEM. We prefer to report and discuss the results of the SEM because of the benefits that it offers with respect to reduced measurement error, explicit estimation of covariance among key service elements, and strengths of modeling the moderating effects of delays, a categorical variable (Ping 1995).

Table 3 The Three-Group Structural Equation Model

	Weather delay	Other delay	No delay
Intercept	−0.43** (−2.27)	−0.48** (−1.99)	0.00 (0.02)
Interaction	0.64*** (9.81)	0.38*** (4.28)	0.54*** (14.66)
	0.60	0.33	0.45
	0.55	0.33	0.46
Aircraft	0.13** (2.40)	0.23*** (3.29)	0.12*** (5.58)
	0.13	0.19	0.12
	0.13	0.22	0.12
Time	0.04* (1.66)	0.12*** (3.44)	0.02 (1.31)
	0.07	0.17	0.03
	0.05	0.15	0.03
Personal space	0.16*** (5.34)	0.27*** (6.54)	0.27*** (20.09)
	0.22	0.30	0.38
	0.22	0.37	0.36
Food	0.08* (1.95)	0.19*** (4.63)	0.07*** (5.00)
	0.09	0.21	0.09
	0.10	0.24	0.09
Flight	0.04 (0.92)	−0.01 (−0.26)	0.03* (1.92)
	0.04	−0.01	0.04
	0.04	−0.01	0.04
R^2 of customer satisfaction	0.935	0.843	0.844
Model fit:			
DF		574	
χ^2		6,933.347***	
RMSEA		0.0843	
NFI		0.957	
CFI		0.960	
RFI		0.953	
IFI		0.960	

Notes. This table reports the structural model coefficients from maximum likelihood estimation of the structural equation model relating satisfaction with the service elements to overall customer satisfaction. In this model, after suitable tests (see text), the measurement model is constrained to be identical and the structural coefficients are allowed to vary for three groups: (1) travelers that experience weather delays, (2) other sources of delays, and (3) no delays. The cells contain the unstandardized coefficient, the t -statistic (in parentheses), the within-group completely standardized coefficient, and the completely standardized common metric coefficient. The within-group completely standardized coefficients permit comparisons within a group of the relative impact of each attribute of the service concept. The completely standardized common metric coefficients permit comparisons of the impact of an attribute across groups.

***, **, * corresponds to p values significant at the <0.01, 0.05, and 0.10, levels, respectively (two-tailed test).

13 degrees of freedom; p value < 0.005) when the relation between service elements and customer satisfaction is allowed to differ between customers delayed

Table 4 Tests of Contrasts: Comparisons Between Groups of the Relation Between Service Elements and Customer Satisfaction

Coefficient	Hypothesis	Weather delay versus other			Weather delay versus no delay			Other delay versus no delay		
		Chi-squared	d.f.	<i>p</i> value	Chi-squared	d.f.	<i>p</i> value	Chi-squared	d.f.	<i>p</i> value
Intercept	H0	7,013.83	575		7,019.97	575		7,015.91	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	0.00	1	1.000	6.14	1	0.013	2.08	1	0.149
Service elements Interaction	H0	7,019.15	575		7,014.83	575		7,017.33	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	5.32	1	0.021	1.00	1	0.317	3.50	1	0.061
Aircraft	H0	7,016.89	575		7,013.83	575		7,017.85	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	3.06	1	0.080	0.00	1	1.000	4.02	1	0.045
Time	H0	7,020.17	575		7,013.83	575		7,021.76	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	6.34	1	0.012	0.00	1	1.000	7.95	1	0.005
Personal space	H0	7,018.33	575		7,020.99	575		7,013.95	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	4.50	1	0.034	7.16	1	0.007	0.12	1	0.729
Food	H0	7,014.38	575		7,013.85	575		7,015.63	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	0.55	1	0.458	0.02	1	0.888	1.80	1	0.180
Flight	H0	7,016.57	575		7,014.09	575		7,015.28	575	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	2.74	1	0.098	0.26	1	0.610	1.45	1	0.229
Joint test of all service elements	H0	7,028.48	581		7,039.00	581		7,042.24	581	
	H1	7,013.83	574		7,013.83	574		7,013.83	574	
	Difference	14.65	7	0.041	25.17	7	0.001	28.41	7	0.000

Notes. This table presents summary statistics from 21 tests of whether the intercept or individual coefficients relating a service element to overall customer satisfaction differs between groups and from 3 joint tests of whether all coefficients relating service elements to customer satisfaction differs significantly between groups. Tests are conducted for each pair of groups by comparing the model of Table 3 with the nested model that constrains the coefficient(s) of interest to be identical for two groups. The reported *p* values from Chi-squared tests indicate whether allowing the coefficients to differ between groups yields a better fitting model (bold indicates values less than 0.10).

by weather and those delayed by other causes. We turn now to the substance of these differences.

Tables 2 and 3 present results of estimating the three-group structural equation model that relates evaluations of service attributes to overall customer satisfaction. Panel D of Table 2 provides the common measurement model that relates the survey item measures to the latent variables that comprise the structural model. To identify the scale of the latent variables, we fix the loading of one item for each latent construct to 1.0, as shown in the column of unstandardized factor loadings. Consistent with the exploratory factor analysis of the holdout sample, the confirmatory analysis yields large unstandardized loadings, all of which are statistically significant (*p* value < 0.01).¹¹

¹¹ To establish the appropriateness of a fixed measurement model, we tested whether there are meaningful differences between the

Table 3 reports the structural model parameters that indicate the strength of the association between

groups in the factor loadings and whether any such differences affect the estimated structural model. Measurement model invariance is a difficult condition to satisfy; however, if constructs that do not support invariance are not a significant part of the model, group comparisons of structural models can still be made without affecting the results (Cheung and Rensvold 2002). In a Chi-squared test comparing our model (Table 3) with one in which the measurement model is allowed to freely vary between the groups, we find that model fit is significantly improved; however, there are no meaningful differences in the factor loadings of the latent variables. Allowing the measurement models to differ also does not affect the size or significance of the structural model parameter estimates; so although freeing the measurement model produces statistical improvements in model fit, it does not affect inferences about the structural model. Thus, we conclude that the reported group differences in the structural models indicate how the groups adapt their preferences for service elements in response to different operational outcomes or different sources of delay.

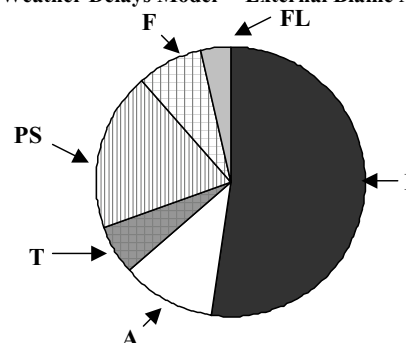
the service elements and overall satisfaction for each group. The model fit statistics reported in the lower portion of the table indicate very good fit of the data to the hypothesized model (Bagozzi and Yi 1988). (Although the structural equations modeling approach does not constrain the service elements to be unrelated, even in the untabulated regression analysis the variance inflation factors [all less than 5.0] indicated no serious multicollinearity among the service indicators.) Consistent with the proposition that these service elements are key components of overall satisfaction, approximately 85%–95% of the variation in overall customer satisfaction is explained by ratings of the service elements. The intercept in Table 3 is a measure of the direct effect of delays on satisfaction. It contrasts the mean effect on overall satisfaction of membership in the delay groups with the “no-delay” group after including all the service elements. Consistent with prior research, we see that overall satisfaction is significantly impaired by service operation failures, regardless of the source. With two exceptions (FLIGHT and TIME), satisfaction with all service attributes contributes to satisfaction with the overall flight experience for each group. FLIGHT has a modest statistical relation (p value < 0.10) to overall satisfaction for the “no-delay” group; however, the magnitude of the coefficient is quite small compared with other attributes. This suggests that the relation is not meaningful. TIME is significantly related to satisfaction for the delayed groups (more so for delays of internal than external origin) but is insignificantly related to satisfaction for passengers who experience no delays. This result may indicate a situation in which quality is unimportant to customers until it is absent (Taguchi et al. 1989). It is also consistent with marketing research that identifies diminishing returns in the relation between evaluations of service elements that customers take for granted and overall satisfaction (Anderson and Mittal 2000). In a separate sensitivity analysis we investigated whether the difference in the importance of time might be caused by differences in the severity of delay between the two delayed groups but found little support for this interpretation.

Before turning to statistical comparisons between the three groups of the relation between key service elements and customer satisfaction, we first consider the relative importance of each attribute for

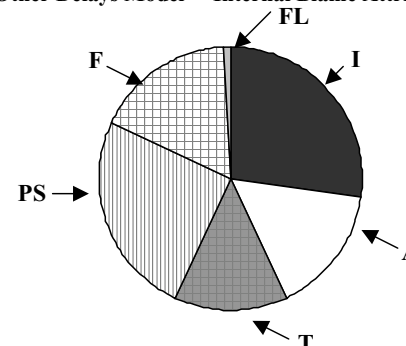
each group. The within-group completely standardized coefficients (the third number in each cell of Table 3) are used for this purpose and are depicted in the graphs of Figure 2 to facilitate comparison. In all groups, the most important attribute to overall customer satisfaction is employee interactions. The

Figure 2 Within-Group Comparisons of the Relative Influence of Service Attributes on Satisfaction

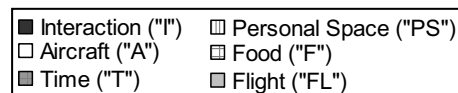
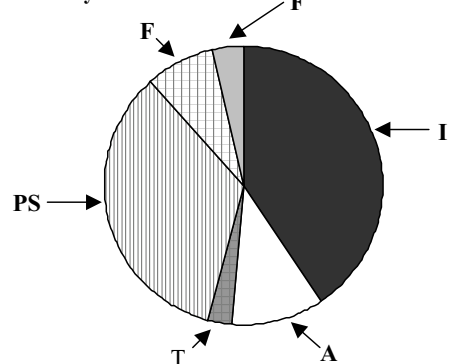
Panel A: Weather Delays Model—External Blame Attribution



Panel B: Other Delays Model—Internal Blame Attribution



Panel C: No Delays Model



second most important attribute is the personal space that the passenger has on board the plane. However, even the relative importance of these attributes, compared with others, differs dramatically between groups. Although these graphs are not formal tests of H2 and H3B, they are consistent with the increase of employee interactions in relative importance during service failures of external origin (i.e., weather delays) and the decrease in importance during service failures of internal origin (i.e., other delays). Although service interactions and personal space are the most important attributes of service by a wide margin for both the weather delay and no-delay groups, other attributes are almost equally important when the delays are more likely to be blamed on the air carrier.

In sum, the results indicate a well-specified model of the components of customer satisfaction and support previous research on the negative impact of service failures on satisfaction. We demonstrate that model fit is improved when customers in the three groups are allowed to value the service elements differently. Thus, we conclude that what matters to customers differs during delays, compared with routine service, and that the source of delays influences these relations. Having provided preliminary evidence that the model differences are associated with substantive differences in what customers value under differing circumstances, we turn to tests of our hypotheses about the impact of service failures on the structural model relating the weights that customers place on service attributes to overall satisfaction.

4.3. Tests of Contrasts: Comparisons Between Groups of the Relation Between Ratings of Service Elements and Overall Satisfaction

The common metric completely standardized coefficients found in the fourth row of each cell of Table 3 suggests differences across groups in the importance of each service attribute. Table 4 presents summary statistics from 21 separate tests (seven coefficients, each with three contrasts) on whether the intercept and the coefficients relating attributes to overall satisfaction differ between the three groups. Tests for each attribute and the model intercept are conducted for each pair of groups by comparing the model of Table 3 with a nested model that constrains the coefficient of interest to be identical for two groups. Table 4

reports the results of testing the significance of the improvement in model fit that is achieved by allowing the two structural parameters to vary by comparing the model χ^2 for the restricted and unrestricted models (Bagozzi and Yi 1988).

We first consider the general pattern of significance in the contrasts among the three groups. Several important observations emerge. Considering first the middle column, which contrasts weather-delayed passengers with passengers who experience no delay, with the exception of the intercept and the importance assigned to personal space on board, there is no statistically significant difference between the two groups. The one exception is personal space, where we see that weather-delayed passengers place less weight on this attribute than passengers who are not delayed. In sum, although weather-delayed passengers are less satisfied (i.e., smaller intercept) than those who are not delayed, the weights that they assign to different aspects of the service are virtually identical.

In contrast, the weights that customers in the other delay group attach to service attributes differ quite markedly from both the weather-delay group and the group without delays, although the mean level of satisfaction does not differ between the two delayed groups. With the exception of FOOD, which is valued similarly across all groups, and FLIGHT, which is valued differently but still at a low level of overall significance for all groups (Table 3), the group that experiences delays attributed to the air carrier assigns significantly less weight to employee interactions (H3B) and more weight to every other service attribute than the weather-delayed group does. (In separate unreported tests of a two-group model that does not discriminate based on the source of delay, H2 is also supported. That is, employee interactions are valued differently by delayed customers than by nondelayed customers).

The results in Table 4 show that service failures of internal origin, in which customers are likely to attribute blame to the service provider, cause customers to value attributes of the service concept differently than normal service operations and as compared with service failures of external origin. We see particularly pronounced differences in the role of employee-customer interactions in customer satisfaction, and these differences are related to attributions of blame for the service failure.

The service operations literature posits that in times of service failure, employees are a critical resource to respond to failures and restore customer satisfaction (e.g., Cook et al. 2002, DeWitt and Brady 2003). However, the marketing literature suggests that employees may be hampered in this role if customers blame the firm (though not necessarily the specific employee) for the failure (e.g., Chung-Herrera et al. 2004). We see evidence supporting both perspectives in our results. Figure 2, which illustrates the relative importance of employee interactions as compared with other attributes for each of the groups, highlights both views. The relative importance of employee interactions in Panel A for customers who experience delays that are unlikely to be blamed on the air carrier is consistent with the former view, and the diminished relative importance of employees in Panel B for customers who are most likely to attribute blame for delays to the air carrier supports the latter view. Thus it appears that when customers blame the firm for a service failure, they also cease to rely on employees to solve the problem. This is particularly interesting in the airline setting, where frontline service employees (gate and flight personal) are unlikely to cause internal delays (e.g., equipment failure, delayed incoming flights).

4.4. Sensitivity Analysis: Air Carrier Effects

As noted earlier, airlines are not represented in equal proportion in the three treatment groups. Consequently, an important concern is that differences in airline strategy and customer clientele may better explain our results than the treatments and related attributions. Although we have concluded using a large holdout sample that the measurement model is invariant to air carriers, we have not examined whether the structural model is invariant under different treatments. Ideally this would be resolved by estimating a 27-group (9 air carriers \times 3 treatments) model; however, the estimation demands of such a model exceed our data capacity. The current model contains 41 parameters (i.e., reported coefficients and covariance parameters); 14 are common to all groups and 27 vary by group, for a total of 95 unique estimated parameters. A common rule of thumb recommends five observations for each estimated parameter, so our total sample size of 4,723 and a minimum

treatment group size of 557 fall well within recommended limits. However, only two airlines have sufficient representation in the three treatment groups to support the estimation of an airline-specific model. Our agreement with the data provider does not permit us to report airline-specific results; however, when we estimate the three-group model for these two airlines, the within- and between-group differences are of approximately the same magnitude and significance as reported for the full sample. This suggests that at least for these two airlines, the previously reported effects of the three treatments hold.

A second avenue for examining airline effects is to estimate the relations using simple ordinary least squares (OLS) regression (i.e., for each airline, substituting for each latent variable the average value of the items that comprise the latent variable and interacting each coefficient with indicator variables for the treatment groups). As noted earlier, although there are serious incompatibilities between our data and the distributional assumption of OLS, we nonetheless find that the broad conclusions of the SEM analysis hold when the weights for each service element are allowed to vary by airline. Although we find some airline-specific differences in the relative weights that customers assign to the service elements, all the elements matter for each airline.¹² More important for our analysis, customer satisfaction for routine flights and flights delayed by weather have a similar composition, while external sources of delay are associated with differences and employee interactions are generally significantly less influential than in the routine or weather-delayed groups.

Finally, although data limitations preclude us from using SEM to test whether all coefficients in the customer satisfaction model differ by airline, we can include a fixed (i.e., intercept) air carrier effect in our model. When fixed effects of the nine air carriers are included, we reject the joint hypothesis of air carrier effects on customer satisfaction (p = value of 0.88). The indicator variables associated with delays continue to be significantly associated with diminished customer satisfaction in the regression analysis. Thus, although there are no incremental air carrier effects

¹² An exception is FOOD, which has no role in customers' flight satisfaction on airlines that do not serve meals.

on customer satisfaction *after including* the effect of delays, air carriers could nonetheless experience very different levels of satisfaction *because of* differing levels of delays. With the inclusion of air carrier effects, we note no meaningful changes in the other estimated coefficients, further suggesting that omission of the air carrier is not producing biased coefficients in the relation between service elements and overall customer satisfaction.

In summary, although we cannot definitively rule out air carrier effects as being confounded with our treatment groups, three separate approaches to examining the sensitivity of our findings fail to generate evidence in support of this alternative interpretation.

5. Conclusion

We find strong support for the proposition that in the airline industry, service failures affect what matters to customers and that attributions of blame are critical moderators of the composition of overall customer satisfaction during service failures. Specifically, when failures are not blamed on the service provider, average overall satisfaction is lower than in the case of routine service; however, there are virtually no differences in the importance that customers assign to the different service elements. In contrast, when failures are blamed on the service provider, customers rely far *less* on employee interactions to evaluate the service, depending instead on physical aspects of the service (e.g., aircraft cleanliness). Thus, even if a specific employee is not the source of the delay, it appears that the customer's distrust of the firm extends to all employees and overall satisfaction is both lower and associated with different service elements, compared with either delays of external origin or no delay.

Our findings are relevant for managers who design service operations to accommodate service failure and recovery. Our study demonstrates that on average, what matters to airline passengers changes when delays are associated with blame. Consequently, the first implication for managers is to be certain that customers draw correct conclusions about attributions of blame when the source is external. Although local weather may be readily observable at the originating airport, passengers need to be aware of weather problems that are less readily observable, including problems at the destination as well as along the flight

route. Clear, accurate communication with customers about the source of delay will ameliorate problems that might arise with incorrect attributions.

However, conditional on correct attributions, our analysis also suggests an approach for identifying an appropriate strategic response to failures of both internal and external origin. If the goal is to maintain satisfaction during service failures (subject to some budget constraint) and the marginal effects on overall satisfaction for elements of the service concept change during internal failures, then managers have two options.

The first is to provide *more* of the things that all customers value, during routine service. If the firm has identified things that matter more to customers during failures, as we have done in this paper, then this strategy can be refined to providing more of the things that customers value during internal failures. This option takes the elements of the service concept as given and, within the constraints of what is possible in the short term, adjusts the level of service provision to compensate for increasing or decreasing importance of the service element. In the airline industry, where seat configurations are not adjustable in the short run, the most adjustable elements may be food, aircraft cleanliness, and employee interactions. Thus, providing delayed passengers with free food or with upgraded food choices (e.g., airport restaurant vouchers), using delays as an opportunity to clean aircraft more thoroughly, augmenting gate staff with additional personnel, and communicating more with customers during the delay may be the only levers for mitigating the effects of failures using this approach.

The second option is to provide *different things* to customers who experience internal failures. This approach does not seek to narrowly optimize the basic service concept in the clearly second-best scenario of a service failure. Rather, this option designs the best service concept for a service failure and activates this alternative process in the event of failure. For example, providing hotel booking and rebooking services as well as free phone access may be critical service elements during an airline delay that are not included in the basic service concept. This approach is distinct from the previous option of providing *more* of the things that comprise the basic routine service.

Whether the objective of maintaining customer satisfaction is most effectively met by reoptimizing the basic service concept around emergent customer preferences or by designing a unique service concept for failures is an empirical question. Although our data cannot support such analysis, a firm that collected customer satisfaction data in conjunction with a quasi-experiment aimed at testing alternative service failure response could repeat our analysis to determine the best approach for its setting. In sum, although preventing service failures is clearly desirable, companies may need a different mix of service elements to maintain customer satisfaction in cases where failures arise.

Although this study provides a unique opportunity to examine the association between attributions of blame for service failures and customer satisfaction, it is not without limitations. One limitation of this study is that we do not directly measure the underlying attribution process, but rely instead on evidence from Folkes et al. (1987) about the attributions airline passengers typically make about various sources of delays. This was necessitated by the use of pre-existing survey data from a third-party data provider. A study that combined direct measurement of attributions about service failures with objectively identified sources of failure and direct measures of customer satisfaction with the completed service would permit a more powerful test of our hypotheses. This research design would be an ambitious and costly undertaking for an academic researcher; however, it might be feasible in partnership with a third-party market survey firm.

A second limitation concerns the issue of service recovery efforts. Our data do not include information on whether service recovery efforts (e.g., rescheduling delayed customers, offering cash or noncash compensation) were attempted or whether customers were satisfied with service recovery efforts. Consequently, although our results indicate that on average customer satisfaction is lower for all forms of service failure and that the elements of customer satisfaction differ as a function of different attributions for that failure, we cannot assess the role that service recovery might have played in these conclusions. An important direction for future research is relating the findings of this study to the presence of service recovery efforts

and to the success of those efforts in ameliorating service failures.

This study advances research at the interface between service operations and marketing. Other fields of business research have used behavioral theory to enrich the understanding of economic systems. In this study we find that combining attribution theory with theory about the design of service operations—in particular on the role of employees during service failures—produces a new perspective on what it means to satisfy the customer during service failure.

Data Availability: Data for replicating the results of this study are available online at: <http://msom.pubs.informs.org/ecompanion.html>. Included in an online appendix and as electronic data files are the LISREL program code for the basic model and data files containing the variable means, standard deviations, and covariance matrices for each of the three treatment groups. In addition to replicating the results of the study, the reader may explore any model that is nested within our model by making changes to the original program code to reflect constraints of a nested model. Confidentiality and nondisclosure agreements with the data provider preclude us from redistributing the raw survey data or reporting results that may be used to identify the customer satisfaction performance of any air carrier.

Electronic Companion

An electronic companion to this paper is available on the *Manufacturing & Service Operations Management* website (<http://msom.pubs.informs.org/ecompanion.html>).

Acknowledgments

The authors gratefully acknowledge the KPMG-University of Illinois at Urbana-Champaign (KPMG-UIUC) Performance Measurement Program for research funding, and an anonymous market research firm that provided data on customer satisfaction for the U.S. airline industry. The authors also thank Ginger Davis for helpful research assistance.

References

- Anderson, E. W., V. Mittal. 2000. Strengthening the satisfaction-profit chain. *J. Service Res.* 3(2) 107–120.
- Bagozzi, R. P., Y. Yi. 1988. On the evaluation of structural equation models. *J. Acad. Marketing Sci.* 16 74–94.
- Beaujean, M., J. Davidson, S. Madge. 2006. The “moment of truth” in customer service. *McKinsey Quart.* (1) 62–73.

- Behn, B. K., R. A. Riley. 1999. Using non-financial information to predict financial performance: The case of the U.S. airline industry. *J. Accounting, Auditing Finance* **14** 29–56.
- Bejou, D., A. Palmer. 1998. Service failure and loyalty: An exploratory empirical study of airline customers. *J. Service Marketing* **12**(1) 7–22.
- Bitner, M. J. 1990. Evaluating service encounters: The effects of physical surroundings and employee responses. *J. Marketing* **54** 69–82.
- Bitner, M. J., B. H. Booms, M. Stanfield-Tetreault. 1990. The service encounter: Diagnosing favorable and unfavorable incidents. *J. Marketing* **54** 71–84.
- Booms, B. H., M. J. Bitner. 1981. Marketing strategies and organization structures for service firms. J. H. Donnelly, W. R. George, eds. *Marketing of Services*. American Marketing Association, Chicago, IL, 47–52.
- Boomsma, A., J. J. Hoogland. 2001. The robustness of LISREL modeling revisited. R. Cudeck, S. du Toit, D. Sörbom, eds. *Structural Equation Models: Present and Future. A Festschrift in Honor of Karl Jöreskog*. Scientific Software International, Chicago, IL, 139–168.
- Cheung, G. W., R. B. Rensvold. 2002. Evaluating the goodness-of-fit indexes for testing measurement invariance. *Structural Equation Model* **9** 233–255.
- Chung-Herrera, B. G., N. Goldschmidt, K. D. Hoffman. 2004. Customer and employee views of critical service incidents. *J. Services Marketing* **18**(4) 241–254.
- Cook, L. S., D. E. Bowen, R. B. Chase, S. Dasu, D. M. Stewart, D. A. Tansik. 2002. Human issues in service design. *J. Oper. Management* **20** 159–174.
- Craighead, C. W., K. R. Kirwan, J. L. Miller. 2004. The effects of severity of failure and customer loyalty on service recovery strategies. *Production Oper. Management* **13**(4) 307–321.
- Danaher, P. J. 1998. Customer heterogeneity in service management. *J. Service Res.* **1**(2) 129–139.
- DeWitt, T., M. K. Brady. 2003. Rethinking service recovery strategies: The effect of rapport on consumer response to service failure. *J. Service Res.* **6**(2) 193–207.
- Donaldson, L. 2001. *The Contingency Theory of Organizations*. Sage Publications, London, UK.
- Edvardsson, B. 1992. Service breakdowns: A study of critical incidents in an airline. *Internat. J. Service Indust. Management* **7** 219–230.
- Folkes, V. 1984. Consumer reactions to product failure: An attributional approach. *J. Consumer Res.* **10** 398–409.
- Folkes, V. S., S. Koletsky, J. L. Graham. 1987. A field study of causal inferences and consumer reaction: The view from the airport. *J. Consumer Res.* **13** 534–539.
- Gilbert, D. D., L. Morris. 1995. The relative importance of hotels and airlines to the business traveler. *Internat. J. Contemporary Hospitality Management* **7** 19–23.
- Goldstein, S. M. 2003. Employee development: An examination of service strategy in a high-contact service environment. *Productions Oper. Management* **12**(2) 186–203.
- Goldstein, S. M., R. Johnston, J. Duffy, J. Rao. 2002. The service concept: The missing link in service design research? *J. Oper. Management* **20** 121–134.
- Kamakura, W. A., V. Mittal, F. de Rosa, J. A. Mazzon. 2002. Assessing the service-profit chain. *Marketing Sci.* **21**(3) 294–317.
- Kekre, S., M. S. Krishnan, K. Srinivasan. 1995. Drivers of customer satisfaction for software products: Implications for design and service support. *Management Sci.* **41**(9) 1456–1470.
- Kellogg, D. L., R. B. Chase. 1995. Constructing an empirically derived measure for customer contact. *Management Sci.* **41**(11) 1734–1749.
- Loveman, G. W. 1998. Employee satisfaction, customer loyalty, and financial performance: An empirical examination of the service profit chain in retail banking. *J. Service Res.* **1**(1) 18–31.
- Mazzeo, M. J. 2003. Competition and service quality in the U.S. airline industry. *Rev. Indust. Organ.* **22** 275–296.
- Miller, J. L., C. W. Craighead, K. R. Karwan. 2000. Service recovery: A framework and empirical investigation. *J. Oper. Management* **18** 387–400.
- Mittal, V., J. M. Katrichis, P. Kumar. 2001. Attribute performance and customer satisfaction over time: Evidence from two field studies. *J. Services Marketing* **15**(5) 343–356.
- Oliver, R. L. 1997. *Satisfaction: A Behavioral Perspective on the Consumer*. Irwin-McGraw Hill, Boston, MA.
- Ping, R. A. 1995. A parsimonious estimating technique for interaction and quadratic latent variables. *J. Marketing Res.* **32**(August) 336–347.
- Podsakoff, P. M., D. W. Organ. 1986. Self-reports in organizational research: Problems and prospects. *J. Management* **12** 531–544.
- Rigdon, E. E., R. E. Schumacker, W. Wothke. 1998. A comparative review of interaction and non-linear modeling. R. E. Schumacker, G. A. Marcoulides, eds. *Interaction and Nonlinear Effects in Structural Equation Modeling*. Lawrence Erlbaum Associates, Mahwah, NJ, 1–16.
- Rubin, D. B. 1976. Inference and missing data. *Biometrika* **63** 581–592.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.
- Sasser, W. E., R. P. Olsen, D. D. Wyckoff. 1978. *Management of Service Operations*. Allyn and Bacon, Boston, MA.
- Soteriou, A. C., R. B. Chase. 1998. Linking the customer contact model to service quality. *J. Oper. Management* **16** 495–508.
- Stafford, M. R. 1994. How customers perceive quality. *J. Retail Banking* **17**(2) 29–37.
- Taguchi, G., E. Elsayed, T. Hsiang. 1989. *Quality Engineering in Production Systems*. McGraw-Hill, New York.
- Taylor, S. 1994. Waiting for service: The relationship between delays and evaluations of service. *J. Marketing* **58** 56–69.
- Tsikriktsis, N., J. Heineke. 2004. The impact of process variation on customer dissatisfaction: Evidence from the U.S. domestic airline industry. *Decision Sci.* **35**(1) 129–142.
- Tsiros, M., V. Mittal, W. T. Ross. 2004. The role of attributions in customer satisfaction: A reexamination. *J. Consumer Res.* **31**(September) 476–483.
- Verbeke, W., R. Bagozzi. 2002. A situational analysis on how salespeople experience and cope with shame and embarrassment. *Psych. Marketing* **19**(9) 713–741.
- Weiner, B. 1985. An attributional theory of achievement motivation and emotion. *Psych. Rev.* **92** 548–573.
- Weiner, B. 2000. Attributional thoughts about consumer behavior. *J. Consumer Res.* **27**(December) 382–387.
- Westbrook, R. A. 1981. Sources of consumer satisfaction with retail outlets. *J. Retailing* **57** 68–85.