The Effects of Service Failures and Recovery on Customer Loyalty in E-Services: An Empirical Investigation

Rui Sousa  
Catholic University of Portugal (Porto)  
Rua Diogo Botelho, 1327  
4169-005 Porto, Portugal  
Tel: +351-22 6196 229  
Fax: +351-22 6196 291  
rsousa@porto.ucp.pt

Christopher A. Voss  
London Business School  
Sussex Place, Regent’s Park  
London NW1 4SA, United Kingdom  
Tel. +44 - (0)20 7000 7000  
Fax: +44 - (0)20 7000 7001  
cvoss@london.edu


---

© Rui Sousa and Christopher A. Voss, 2007. This paper should not be cited without authorization of the authors.
The Effects of Service Failures and Recovery on Customer Loyalty in E-Services: An Empirical Investigation

Structured Abstract

**Purpose:** Despite having been widely studied in traditional (bricks-and-mortar) services, the effect of service failures and recovery on customer loyalty has received only limited attention in the context of e-services. Taking the default stance that theory from traditional services will find replication in e-services, this study sets out to empirically test the following set of hypotheses in an e-service setting: H1: Service failures result in decreased customer loyalty intentions; H2: Achieving recovery from failures increases customer loyalty intentions; H3: Satisfaction with the way recovery is accomplished increases customer loyalty intentions; H4: Superior recovery results in loyalty intentions which are more favorable than they would be had no failure occurred (service recovery paradox).

**Methodology:** The study was based on an online survey of actual customers of a commercial e-banking service.

**Findings:** H1-H3 were supported, suggesting that: i) the detrimental effects of failures are also present online (H1); and ii) it is possible to achieve effective recovery in e-services leading to customer loyalty, despite the challenging nature of online failures and the reduced degree of human interaction (H2, H3). H4 was only partially supported in that it was found that unless recovery efforts “delight” the customer, they can, at best, restore loyalty to the levels existing prior to the failures.

**Research Limitations:** Future research should examine other types of e-services.

**Practical Implications:** E-service delivery systems should be designed which a strong failure-prevention mindset and include effective service recovery mechanisms. However, in general, e-service providers should not look at superior recovery as a substitute for error-free service.

**Originality/Value:** The study provides empirical evidence of the effects of service failure and recovery in the context of online service, an area which has received limited attention to date.

**Paper Category:** Research paper.

**Keywords:** E-Services, Service Operations, Service Quality, Service Design, Internet.
Introduction

There has been substantial research in traditional services on the effect of service failures and recovery on customer loyalty (Johnston, 2005). This work has produced three main findings. First, service failures have a negative effect on loyalty and have been found to be a driving factor in customer switching behavior (e.g., Hays and Hill, 1999; McCollough et al, 2000; Roos, 1999; Zeithaml et al, 1996). Second, in the event of a service failure, customers expect effective recoveries (Bitner et al, 1990) and their satisfaction with the recovery increases customer loyalty (Miller et al, 2000; Spreng, 1995; Tax and Brown, 1998; Zeithaml et al, 1996). Third, a “recovery paradox” (McCollough and Bharadwaj, 1992) has been proposed where customers who experience a service failure followed by superior recovery exhibit behavioral intentions towards the service provider which are more favorable than they would be had no failure occurred (service recovery seen as an opportunity). However, while several studies report results consistent with a recovery paradox effect (e.g., Hansen and Danaher, 1999; Smith and Bolton, 1998) and several textbook authors make assumptions about the existence of such phenomenon (e.g., Johnston and Clark, 2005; Kotler, 1997; Rust et al, 1996), some empirical studies have cast doubts over it (Andreassen, 2001; McCollough et al, 2000; Zeithaml et al, 1996). A recent meta-analysis of empirical research on the service recovery paradox concluded that its effect was significant on satisfaction, but non-significant on loyalty intentions (repurchase intentions and word-of-mouth) (Matos et al, 2007).

The ability to achieve effective recovery from failures is an important responsibility of the Operations function (Miller et al, 2000; Prajogo, 2006; Roth and Menor, 2003). It has been considered as a critical success factor for e-services (e.g., Collier and Bienstock, 2006b; Holloway and Beatty, 2003), for example, as a means to mitigate online trust issues. However, although service failure and recovery issues have received considerable attention in the literature, these topics have received only limited attention in the context of online services. Specifically, we lack an understanding of whether the knowledge gained in traditional services translates to online
environments, given the absence of human intervention in the service encounter and the fact that the reasons for dissatisfactory online service encounters have been found to be different from traditional offline service research (Forbes et al, 2005; Holloway and Beatty, 2003).

As such, the main objective of this paper is to empirically examine the impact of service failures and recovery on customer loyalty in e-services. The structure of the paper is as follows. First, we review the literature in traditional services and e-services and develop research hypotheses related to the study’s research question. Second, we describe the employed methodology, a survey study of customers of an e-banking service. Third, we present the data analysis and discuss the results. Finally, we put forward suggestions for future research and discuss the limitations of the study, followed by a summary of the main conclusions.

**Literature review and research hypotheses**

In developing the research hypotheses, our default stance is to apply theory from traditional services to e-service settings. A prominent study which has addressed most of the discussed failure/recovery relationships in traditional services has been Zeithaml et al (1996). The authors conducted a mail survey of actual customers of four different companies to find that customer loyalty was highest for customers experiencing no service problems; next highest for customers experiencing service problems that were resolved, and lowest for customers experiencing service problems that were not resolved. Consistent with our default stance, we adapt Zeithaml et al’s (1996) hypotheses by adding a variable representing the degree with which customers were satisfied with the way the problem was solved, thus putting forward the following hypotheses:

H1. Loyalty behavior will be higher for customers experiencing no service problem (A) than for those experiencing service problems (B).
H2. Within the customer group who have experienced service problems (B), loyalty behavior will be higher for those for whom the problem has been resolved (B2), than for those for whom the problem has not been resolved (B1).

H3. Within the customer group who have experienced service problems and have had the problems resolved (B2), loyalty behavior will be higher for customers who are more satisfied with the way the problem was resolved (B2.1) than for those who are less satisfied (B2.2).

In addition, we test the service recovery paradox in an e-service context by adding the following hypothesis:

H4. Loyalty behavior will be higher for customers who have experienced service problems and have had the problems resolved with a high degree of satisfaction (B2.1) than for those who have not experienced service problems (A).

Failure and recovery issues in e-services exhibit some differences relative to traditional services. It is largely an empirical question whether these differences will affect the verification of the proposed hypotheses. Concerning H1, it has been found that the types of service failures online are considerably different from traditional services (Forbes et al, 2005; Holloway and Beatty, 2003). Whether this hypothesis holds may depend on the degree of severity that customers attach to online service failures (Craighead et al, 2004) and the extent to which customers may be more or less forgiving towards online failures (McCollough et al, 2000).

The validity of H2-H4 largely hinges on the degree of effectiveness of the service recovery that can be achieved in e-service settings. If recovery effectiveness can reach high levels, then
problems may be speedily and adequately resolved to the extent that problem resolution per se as well as the quality of the resolution may have an impact on customer loyalty (support for H2 and H3). Moreover, if the recovery effort exceeds customer expectations, customer loyalty might even emerge as strengthened relative to the level existing prior to the occurrence of the service failure (McCollough et al, 2000) (support for H4).

Some studies suggest that the effectiveness of service recovery can be increased in e-services. This is due to two main reasons. First, the multi-channel nature of many of these services may facilitate customer complaints by offering new channels that reduce the time and effort required in the process (Holloway and Beatty, 2003; Tax and Brown, 1998) or remove the embarrassment of the face-to-face complaining process (Tax and Brown, 1998). Thus, as barriers to complaints are lowered, the complaining frequency is expected to increase, providing more opportunities for service providers to salvage dissatisfied customers. Second, the use of technological support to handle complaints (e.g., FAQs and troubleshooting engines that can automatically walk customers through problem-identification and resolution processes) may lead to more efficient service recovery systems (Tax and Brown, 1998).

However, there is some evidence of overall customer dissatisfaction with the service recoveries provided by online retailers (Holloway and Beatty, 2003; Voss, 2003) and other studies have suggested that it may be more difficult to perform effective recovery in e-service contexts. This is due to two types of reasons. The first is associated with the different nature of service failures online. Meuter et al (2000) identified three types of failures in e-services - technology failures, process failures and customer-driven failures - and acknowledged the difficulty in performing service recovery in such a context. Technology failures are those that prevent the customer from engaging with the e-service (e.g., web site is down or not working properly). In these cases, the customer may resort to interpersonal service either to complain or to receive the desired service (Meuter et al, 2000). Because of the wide reach of the Internet, there are a potentially large
number of users affected by the breakdown and it may be difficult for the service provider to enable convenient and rapid customer access to a support agent (Sousa and Voss, 2006). This may compromise the speed of response to a service failure, a key part of maintaining loyalty (Miller et al, 2000). *Process failures* are those that occur at some point after the customer’s interaction with the web site, but preventing correct service fulfillment from occurring. In these cases, because the initial interaction has taken place as expected, the customer expects the service to be provided successfully, and the failure only becomes apparent at a later stage (e.g., the items ordered through the Internet are never received). Because of this deferred nature of fulfillment, it is difficult to detect such failures early and proactively contacting the customer to overcome them (Sousa and Voss, 2006). *Customer-driven failures* are those that occur as a result of a customer mistake on the web site (e.g., not being able to understand and complete the steps required for placing an order online). Because of the absence of customer contact, it is difficult to detect customer difficulties or mistakes in a timely manner and perform on-the-spot customer support (Bitner et al, 2002).

The second type of reasons is the reduced degree of human interaction. Findings from traditional services have shown that the role of front-line employees is key for effective recovery. A customer’s anger is abated when employees act in a polite and empathetic manner and demonstrate a strong effort to solve the problem (Tax et al, 1998). An existing rapport between the customer and the service provider (achieved through the face-to-face interactions with service employees) also increases the likelihood of a satisfactory recovery from a service failure, acting as a switching barrier (DeWitt and Brady, 2003). On the contrary, e-services encounters are depersonalized and the technology creates a distance between customers and service personnel (Walker et al, 2002). Although e-service providers may offer access to human customer support agents, this typically happens via low richness media, such as e-mail and the phone.
Methodology

The study was based on the administration of an online questionnaire to a sample of customers at a major retail e-banking service, which was part of a broader multi-channel banking service. Our single-industry focus is in keeping with similar industry-specific research in e-services (e.g., Boyer and Hult, 2005; Verma et al, 2004). E-banking was chosen for several reasons. First, it is a mature and one of the most widespread types of e-services, with high adoption levels among both service providers (e.g., the majority of banks now offer such a service) and users (according to Forrester Research, it is predicted that 35% of US households with Internet connections will bank through the web by 2008). Second, in e-banking the web site plays a major role in service provision, given that it is an information service. Finally, the range of services offered at e-banking sites tends to be similar across different service providers and countries, enhancing the generalizability of our findings.

The chosen service, located in Portugal, had about 600,000 customers at the time of the study and is considered a “best practice” service, being ISO9001 certified and having won, among other awards, the “Best Consumer Internet Bank 2003” award by the Global Finance magazine. The next sections discuss the measurement of the research variables and data collection.

Measurement

There have been several conceptualizations of the customer loyalty construct (e.g., Dick and Basu, 1994; Oliver, 1997, 1999; Zeithaml et al, 1996). We have focused on loyalty in a conative sense, i.e., related to behavioral intentions (Oliver, 1997, 1999). In this connection, we drew on the definition of conative loyalty by Oliver (1997, 1999) to measure loyalty towards an e-service (web site) by the following items: L1) The intention to re-use the e-service (web site); and L2) Word of mouth recommendation. Similar scales have been commonly used in service quality research (e.g., Boulding et al, 1993; Spreng et al, 1995).
The service failure and recovery (SFR) issues associated with the research hypotheses H1-H4 were addressed in the questionnaire by three cascaded questions related to: SFR1) Whether a customer had experienced a service problem in the last 6 months; SFR2) If so, whether it had been resolved; and SFR3) If so, the customer was asked to rate the satisfaction with the way the problem had been resolved. The measures are presented in the Appendix.

We took several recommended steps to reduce the threat of common method bias (Podsakoff et al, 2003). First, we provided a psychological separation in our instrument with an introduction that said that we were examining quality issues, and not suggesting any link between our predictor variables (SFR1- SFR3) and loyalty. Second, there was an indication on the first and last pages of the questionnaire that responses were anonymous and confidential. Finally, we used short, specific questions to address each of the separate issues in the survey.

**Data collection**

The data were collected via the administration of an online questionnaire to a sample of the e-service’s customers. We drew on the overall customer database to exclusively target customers fulfilling the following criteria: active customers (customers having made at least two logins to the service in the previous three months), age over 18 years old, excluding bank employees. This screening resulted in a pool of 51,125 eligible customers. The actual target sample for our study consisted of a subset of 70% of this pool, generated through random extraction from the pool (35,781 customers).

The questionnaire was posted on the e-banking service web site, and placed immediately after the login stage. After the targeted customers logged in, they were asked whether they would like to fill in the questionnaire, in which case they were directed to the respective web page. The questionnaire software application kept track of the identification of all targeted customers, recorded respondents and non-respondents and, for the respondents, recorded their actual responses.
to the questionnaire. The software also ensured that customers who declined to fill in the questionnaire as well as those who did fill in the questionnaire were not asked again. The questionnaire was active on the site for one month, resulting in 5,942 valid responses, yielding a 16.6% response rate.

The final sample is characterized in Table I. The table shows that the predominant demographic profile in the sample was that of a male, young and educated customer. This pattern is in line with the patterns observed for general Internet users in the European Union and the US, as well as for e-banking users (e.g., SIBIS, 2003). We conducted a non-respondent bias analysis employing t-tests to compare the profiles of respondents and non-respondents in terms of age, gender and education level. The analysis showed no significant differences, indicating the absence of non-respondent bias.

<table>
<thead>
<tr>
<th>Analysis and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>The analysis comprised two stages: i) measurement analysis of the loyalty construct; and ii) testing of hypotheses H1-H4. Missing data relative to the loyalty items was handled via the EM method (Little and Rubin, 1987). Concerning the measurement analysis, the reliability of the loyalty construct was assessed by computing its composite reliability. The obtained value (0.798) was found to be well in excess of the suggested level of 0.70 (Hair et al, 1998). Convergent validity was assessed by computing the item-to-total correlations. The obtained value for both measurement items (0.659) was found to surpass the minimum suggested level of 0.4 (Kline, 1986). Collectively, the results support the reliability and convergent validity of the loyalty construct.</td>
</tr>
<tr>
<td>In order to test hypotheses H1-H4, the combined sample was classified into different customer groups obtained by splitting the sample according to the responses to variables SFR1</td>
</tr>
</tbody>
</table>
(Groups A and B), SFR2 (Groups B1 and B2) and SFR3 (Groups B2.1 and B2.2) – see Figure 1. The hypotheses tests consisted in comparing the means of the summated loyalty scales across the different groups. The summated scales were obtained by taking the average of the measurement items, which assumes that the individual item weights are equal (Hair et al, 1998, p.129). Analysis of variance revealed that loyalty scores differed across the mutually exclusive groups A, B1, B2.1 and B2.2 (F-value significant at p<0.000). The results of the individual group comparisons are shown in Table II. Because multiple comparisons were evaluated to test these hypotheses, the alpha level for testing the significance of individual comparisons was reduced by applying the Bonferroni correction to ensure that the overall probability of Type I error across all four comparisons did not exceed .05 (for details, see footnote b) in Table II).

Table II shows that close to 12% of customers experienced a problem in using the service. This customer group (B) exhibits a loyalty level which is significantly lower than that observed for the group who did not experience any problem (A) (p<0.000). Hence, there is support for H1. Within the customer group who experienced a problem, 56% reported that the problem had been solved. This group (B2) exhibits a significantly higher loyalty level than the group who reported that the problem had not been solved (B1) (p=0.005). Therefore, there is support for H2. Within the customer group who reported that the problem had been solved, about 37% reported high levels of satisfaction with recovery (B2.1). This group exhibited a significantly higher loyalty level than the group who reported low satisfaction with recovery (B2.2) (p<0.000). Consequently, there is support for H3. Although loyalty was higher for group B2.1 than for group A, the difference was not significant (p=0.147). Hence, the results do not provide definite support for H4.

Take in Figure 1

Take in Table II
The non-dichotomous nature of variable SFR3 allowed for additional more detailed analyses concerning H4. Specifically, we compared the means of the summated loyalty scales between each of five individual customer groups corresponding to the five possible reported levels of satisfaction with recovery (variable SFR3) and Group A (see Table III). Table III shows that whether loyalty levels after recovery are higher or lower relative to error-free service levels, depends on what the level of satisfaction with recovery is. In particular, loyalty levels after recovery only surpass error-free levels for the maximum possible level of satisfaction with recovery (5). For a satisfaction level of 4, loyalty is slightly lower than for Group A, although this difference is not statistically significant. For satisfaction levels of 3 or below, customer loyalty is significantly lower relative to Group A. Therefore, there is only partial support for H4; that is, there is only evidence of the “recovery paradox” for the maximum possible level of satisfaction with recovery, but not for other levels.

Take in Table III

Discussion

The results have several implications for e-service settings. Support for H1 suggests that the detrimental effects of service failures also appear to be present online. This may mean that customers do not attach a lower degree of severity to failures nor seem to be more forgiving when interacting with e-services when compared to traditional services. Because failures in e-services have the potential to affect a large number of customers, this highlights the importance of investing in the prevention of failures in these settings. Therefore, the application of quality management practices focusing on prevention, in particular when applied in the service delivery system design stages, should be an important requirement for e-services.
The support received by hypotheses H2 and H3 suggests that it is possible to achieve effective recovery in e-services. Such recovery efforts must deal not only with the actual service failure (H2), but also to do so in such a way that the customer is satisfied with the way in which the problem is resolved (H3). Therefore, it seems possible to overcome the previously discussed difficulties to achieving effective recovery created by the different nature of service failures and the reduced degree of interpersonal interaction in e-services.

The recovery paradox (H4) only manifested itself for the maximum level of satisfaction with the recovery effort. Hence, it seems that taking advantage of a service failure and associated recovery effort may only be possible if service providers can actually “delight” the customer in the process (we employ the term delight to mean a very high satisfaction with the recovery process, as perceived by the customer; hence, this notion is associated with customer perceptions rather than with an absolute assessment of the recovery effort). This is broadly consistent with the recovery paradox, but the empirical evidence in this study qualifies this notion by suggesting that only for very high levels of satisfaction with recovery (or possibly the maximum level of satisfaction) can service providers actually increase customer loyalty. This reinforces the detrimental effect of failures in e-services: for levels of recovery satisfaction below the very highest, at best, e-service providers may be able to restore loyalty to levels existing prior to the failures. Thus, there is only weak support for considering recovery in e-services as an “opportunity” when compared to the loyalty level resulting from error-free service. Therefore, in general, e-service providers should not look at superior service recovery as a substitute for error-free service.

The findings on the recovery paradox raise the issue of how likely it is to reach very high levels of customer satisfaction with recovery in a real e-service setting. In traditional services, there is support for the notion that service providers can only induce high levels of satisfaction with recovery for a small proportion of customers (e.g., Kelley et al, 1993). In fact, the concept of “delighting” itself may entail doing something out of the ordinary to positively surprise the
customers or going the “extra mile” (e.g., Johnston and Clark, 2005, p. 107; Oliver et al, 1997). In addition, delighting the customer may mean providing a personal touch in the recovery effort (e.g., Johnston and Clark, 2005, p. 107). Unfortunately for service providers, as customers get used to rising service levels and recovery standards both in the context of a given service and the broader marketplace, the potential to delight may be reduced over time (Rust and Oliver, 2000): something which delights the customer today may not do so in a few months’ time.

In e-services, as discussed earlier, the reduced degree of interpersonal interaction, the typically higher number of customers affected by failures and, to some extent, the fairly depersonalized nature of the customer interface may make delighting customers even more difficult to achieve than in traditional services. Therefore, it would seem reasonable to posit that generally any given e-service would not be able to achieve maximum ratings of satisfaction with recovery for a large majority of its customers. The findings for the specific e-service that we studied are consistent with this notion. This e-service was generally considered a best practice service and had no policy to deliberately vary the intensity of the recovery effort across different types of customers (although this effort might naturally vary across individual customers due to other reasons, such as the type of failure experienced, random factors, etc.). Despite this, only about 8% of its customers were "delighted" in the recovery process (see Table III). Although we must exert caution in generalizing this figure to other e-services, this result is consistent with findings in traditional services and suggests that in e-services it may also be difficult to “delight” the customers in the recovery process.

As a consequence, e-service providers may have to accept that they may not be able to delight all customers in the recovery process or, at least, that it may not be cost-effective to do so. In recognition of this, they might consider applying a Customer Relationship Management (CRM) logic (e.g., Greenberg, 2001), by varying the recovery effort (or investment) across different types of customers. For example, they might consider applying extraordinary recovery efforts only for
their most profitable customers. Because of the limited number of these customers, this would have the additional advantage of facilitating the use of personalized recovery efforts, thus maximizing the potential to delight.

Overall, the findings stress the importance of developing effective service recovery for customer loyalty. Inferior recovery performance can lead to what Bitner et al (1990) termed a double deviation from customer expectations: the service provider fails to deliver on the initial service and the recovery service.

**Future research directions and limitations**

Our findings open a number of directions for future research. Given the significant effect of satisfaction with recovery on loyalty (H3), we need to understand what differentiates recovery that gives high satisfaction from that that has low satisfaction. Parasuraman et al (2005) proposed that the quality of online recovery comprises the dimensions of responsiveness, compensation and ease of contact with human agents. However, future research needs to go into more detail and actually identify different types of recovery strategies (e.g., e-mail apologies, discounts on future transactions, personal phone calls and letters, etc.) and examine the factors that influence their effectiveness.

One relevant factor could be the type and degree of severity of the online service failures. Research in traditional services suggests that failures of different types and of different degrees of severity may require different recovery strategies (Smith et al, 1999). In this process, it would be important to classify into meaningful categories the several types of service failures that may occur, as a pre-requisite for developing appropriate recovery strategies. Meuter et al’s (2000) classification (technology, process and customer-driven failures) and Holloway and Beatty’s (2003) six categories of problems (delivery, web site design, customer service, payment, security and miscellaneous) are good starting points, but they need to be complemented with an assessment of the degree of severity
for customers of the different types of failures, along the lines of existing research in traditional services (Craighead et al, 2004).

Another important factor could be the profile of customers, given that research in traditional services has often made claims that service providers need to cater to the individual customer in dealing with service failures (e.g., Reichheld and Sasser, 1990). This issue may be especially relevant in e-services because, compared to traditional services, these services are typically exposed to a larger number of customers with different profiles and requirements (Boyer et al, 2002). Of particular interest would be to examine the effectiveness of different recovery strategies for customers exhibiting different degrees of preference and familiarity with the Internet channel, as opposed to offline channels. Some customers may prefer a less intrusive recovery approach (e.g., by e-mail), while others may value a personal phone call (Collier and Bienstock, 2006a). These differences are important because they influence the cost of the service recovery effort.

Our results are consistent with the notion that it is difficult to “delight” the customers in the recovery process. It would be important for future research to investigate recovery strategies that may lead to “customer delight” and to reaping the benefits of the recovery paradox in e-service settings. These strategies may well depend on the type of e-service, the degree of severity of the failure and the profile of the customers.

Another important area for future research is to provide guidance to service providers as to the level of service recovery that should be offered to customers. As discussed earlier, setting up recovery strategies in e-services may be a challenge (Meuter et al, 2000) and, if they require human intervention, may decrease the level of scalability of the service and increase costs (Hallowell, 2001). Our study does not address the issue of whether it is cost-effective to employ these strategies. Future research should investigate the costs and benefits associated with individual recovery strategies, as well as the overall recovery effort. Reichheld and Schefter (2000) have argued that loyalty may be more important in e-services than for comparable, traditional bricks-and-
mortar services because: i) attracting new customers has been found to be considerably more expensive in e-services; ii) the profitability of individual customers accelerates much faster on the web; and iii) the cost of serving a customer decreases much faster on the web. In addition, Hitt and Frei (2002) in a study of the banking sector found that online customers were more profitable than offline customers. Both these studies suggest that more overall investment in service recovery efforts per customer may be awarded in e-services than in traditional services, but this needs to be further examined. The concept of customer efficiency (Xue and Harker, 2002) may be a powerful tool to analyze cost-benefit issues in the context of service recovery activities. In particular, it would be important to identify recovery strategies that make the recovery process transaction-efficient, that is, saving the time expended by the e-service provider and the customer in these activities (Xue and Harker, 2002).

Another aspect of this question is the need to understand the extent to which different customers should be offered different levels of recovery and/or different recovery processes. As an example, earlier we have suggested that strategies aiming at customer delight might only be employed for the most profitable customers. As another example, customers who have a high intrinsic preference for the Internet channel may be more profitable (Hitt and Frei, 2002) and may require less intrusive recovery strategies. This discussion resonates with previous research in traditional services that highlights the need for service providers to segment customers for the purpose of designing recovery strategies (Craighead et al, 2004). A practical implication of this may be the need for e-service providers to put systems in place that, in the event of a failure, can provide quick access to a customer’s profile (e.g., profitability, degree of preference for the Internet channel, etc.) so as to quickly determine the type and level of service recovery that should be provided.

The study is not without its limitations, which themselves provide opportunities for future research. The investigation is based on one service industry, e-banking, a very important type of
service in today’s e-service landscape. We believe that the findings can be generalized to other task-oriented e-services, but caution must be exercised in extending the conclusions of this study to other services. It may be important for future research to test the developed hypotheses in other types of e-services, in particular, types of e-services which are more strongly associated with experiential/hedonic use (e.g., entertainment services).

Previously, we argued that many types of e-services in today’s business landscape are multi-channel (Sousa and Voss, 2006; Vishwanath and Mulvin, 2001) and we identified the existence of multiple channels as one possible facilitator of effective recovery in e-services. Our empirical investigation of a multi-channel e-service concluded that it was possible to achieve effective recovery and is thus consistent with this explanation. Despite the prevalence of multi-channel e-services, single-channel (internet-only) e-services exist in a limited number of sectors and it may be questionable whether effective service recovery can also be achieved in these settings. Future research should examine this.

Finally, our study has focused on loyalty behavioral intentions, what Oliver (1999) called conative loyalty. Although previous research has provided empirical support for the causal link between intentions and actual actions (Venkatesh and Davis, 2000), future research may examine if loyalty behavior is linked to loyalty actions (what Oliver (1999) calls action loyalty).

Conclusions

This study contributes to research in e-services in several ways. Overall, the results support the key role of service recovery in maintaining, although not necessarily driving, customer loyalty. This answers calls from a number of researchers for a better understanding of how to retain customers in e-services (e.g., Bolton et al, 2004; Reichheld and Schefter, 2000; Zeithaml et al, 2002). The findings suggest that the different nature of encounters in e-services relative to traditional services – notably, the reduced degree of interpersonal interaction and the different
nature of the service failures - does not reduce the key role of recovery in maintaining customer loyalty. This is an important result, given that loyalty has been considered harder to achieve in e-services than in traditional services.

With the exception of the service recovery paradox, the research hypotheses, largely based on theory from traditional services, were broadly supported. This further extends the generalizeability of research in traditional services into the context of online service. Thus, the article provides general support in defense of theories that draw parallels between online and offline services, answering calls for research examining the applicability of traditional notions of service management in e-service settings (Roth and Menor, 2003).

Concerning the recovery paradox hypothesis, this study goes beyond validation of the relationships found in traditional services to achieve a deeper understanding of this phenomenon in e-services. The empirical evidence qualifies this notion in e-service settings, suggesting that unless recovery efforts “delight” the customer, they can, at best, restore loyalty to the levels existing prior to the failures. Our analysis also indicated that it may be difficult to “delight” the customers in e-service settings and, as a consequence, e-service providers may have to accept that they may not be able to delight all customers in the recovery process. Therefore, in general, e-service providers should not look at superior recovery as a substitute for error-free service and might benefit from concentrating their efforts in applying extraordinary recovery only for their most profitable customers.

The study provides empirical evidence of the effects of service failure and recovery in the context of online service, an area which has been relatively unexplored to date despite the considerable attention it has received in traditional services (Holloway and Beatty, 2003). In doing so, the study investigated actual customers in a real e-service setting, increasing the external validity of the results. Such an approach has been lacking in existing e-service research which tends to
employ convenience samples (mainly students), rather than samples taken from the actual customer base of service providers.

Finally, the study opens a number of directions for future research with the goal of increasing our still limited understanding of service failure and recovery issues in e-service settings. This constitutes a research agenda that explicitly considers issues that are specific to e-services and that depart from traditional services issues.

The main responsibility for delivering error-free service and achieving effective recovery lies with the Operations function and the service delivery system (Miller et al, 2000; Roth and Menor, 2003). The main implications of this research for Operations managers are twofold.

First, operations should carefully design e-service delivery systems which a strong failure-prevention mindset. Because failures in e-services have the potential to affect a large number of customers, this design area may have to receive stronger attention than in traditional services. As discussed earlier, this could be achieved by the application of quality management practices focusing on prevention in the service delivery system design stages. The operations that support an e-service web site have a specific nature in that they mainly process information (Sousa and Voss, 2006). Therefore, this effort might require the development of increased knowledge relating to operations for which the main input is information (Hayes, 2002; Heim and Sinha, 2001).

Second, because delivery systems will never be completely failsafe, the design of these systems should explicitly include the devising of effective service recovery strategies and processes. In doing this, operations managers should take into account the previously discussed issues, such as determining the level of the recovery effort for different types of customers and devising different recovery strategies for different types of failures and different types of customers. This may require putting systems in place that, in the event of a failure, can provide quick access to a customer’s profile (e.g., profitability, degree of preference for the Internet channel, etc.) and that can quickly identify the failure type. These systems could then employ these two pieces of information to
determine the type and level of service recovery that should be provided. In general, the Internet
data exchange between providers and customers provides opportunities for the automated
generation and processing of a great deal of information about personal customer profiles at a low
cost (Wiedmann et al, 2002). As a result, the segmentation of customers for recovery alignment
decisions may be easier to accomplish in e-services than in traditional services.

By having conceptualized, empirically examined and extended the knowledge on the impact
of service failures and recovery on customer loyalty in e-service settings, we hope that our study
will contribute to fostering much needed prescriptive work on these two fronts.
Appendix. Measures of Research Variables.

Customer Loyalty

L1. Intention to re-use the [name of e-banking service] service (web site).

L2. Intention to recommend the [name of e-banking service] service (web site) to a friend or relative.

Both items used a 5-point Likert-type scale ranging from “1 - Strongly Disagree” to “5 - Strongly Agree”.

Service Failure/Recovery

SFR1. Have you experienced any problem in using the [name of e-banking service] service (web site) in the last 6 months? (Yes/No)

SFR2. Has the problem been resolved? (Yes/No) [only for customers replying Yes to SFR1]

SFR3. How satisfied are you with the way the problem was resolved? [only for customers replying Yes to SFR2] (1- Very Dissatisfied; 5 – Very Satisfied).
References


<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=5,942</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>[18-25]</td>
<td>9.4 %</td>
</tr>
<tr>
<td>[25-35]</td>
<td>44.8 %</td>
</tr>
<tr>
<td>[35-45]</td>
<td>22.1 %</td>
</tr>
<tr>
<td>[45-55]</td>
<td>14.2 %</td>
</tr>
<tr>
<td>55+</td>
<td>9.5 %</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>72.0 %</td>
</tr>
<tr>
<td>Female</td>
<td>28.0 %</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Primary education</td>
<td>3.4 %</td>
</tr>
<tr>
<td>Secondary education</td>
<td>37.9 %</td>
</tr>
<tr>
<td>Higher Education (Bachelor’s degree and above)</td>
<td>58.7 %</td>
</tr>
</tbody>
</table>
Table II. Comparison of the means of the summated loyalty scale across different customer groups (Hypotheses H1-H4).

<table>
<thead>
<tr>
<th>Customer Group</th>
<th>N</th>
<th>% Within Groups</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>P-Values (b)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample</td>
<td>5942</td>
<td>--</td>
<td>4.12</td>
<td>0.79</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SFR1</td>
<td>5942</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A – No problem experienced</td>
<td>5030</td>
<td>84.7%</td>
<td>4.16</td>
<td>0.77</td>
<td>A vs. B: &lt;0.000</td>
<td>H1 Supported</td>
</tr>
<tr>
<td>B – Problem experienced</td>
<td>695</td>
<td>11.7%</td>
<td>3.94</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Response to SFR1</td>
<td>217</td>
<td>3.6%</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFR2</td>
<td>695</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 – Problem not solved</td>
<td>267</td>
<td>38.4%</td>
<td>3.82</td>
<td>0.95</td>
<td>B1 vs. B2: 0.005</td>
<td>H2 Supported</td>
</tr>
<tr>
<td>B2 – Problem solved</td>
<td>389</td>
<td>56.0%</td>
<td>4.01</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Response to SFR2</td>
<td>39</td>
<td>5.6%</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFR3</td>
<td>389</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2.1 High satisfaction with recovery</td>
<td>145</td>
<td>37.3%</td>
<td>4.21</td>
<td>0.64</td>
<td>B2.1 vs. B2.2: &lt;0.000</td>
<td>H3 Supported</td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B2.1 vs. A: 0.147</td>
<td></td>
</tr>
<tr>
<td>B2.2 Low satisfaction with recovery</td>
<td>219</td>
<td>56.3%</td>
<td>3.85</td>
<td>0.85</td>
<td></td>
<td>H4 Not Supported</td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Response to SFR3</td>
<td>25</td>
<td>6.4%</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Groups B2.1 and B2.2 were obtained by splitting group B2 by the mean (3.28 or, for practical purposes, level 4 of satisfaction).

(b) P-values correspond to one-tailed t-tests for group means. Because multiple comparisons were evaluated to test the hypotheses, the Bonferroni correction was applied to the customary alpha level of .05 to control the Type I error rate. Specifically, the alpha level was lowered by a factor of four (the total number of comparisons) to yield a critical alpha level of .0125 for testing the significance of each comparison (Hair et al, p. 328).
Table III. Differences in the means of the summated loyalty scale between groups corresponding to different levels of satisfaction with recovery and Group A (Hypothesis H4).

<table>
<thead>
<tr>
<th>Level of Satisfaction (SFR3)</th>
<th>Number of Customers</th>
<th>% Within Group B2</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Difference in Means Relative to Group A</th>
<th>P-Value (a)</th>
<th>Loyalty Level vs. Group A (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>29</td>
<td>8.0%</td>
<td>4.48</td>
<td>0.54</td>
<td>+0.32</td>
<td>0.001</td>
<td>Significantly Higher</td>
</tr>
<tr>
<td>4</td>
<td>116</td>
<td>31.9%</td>
<td>4.14</td>
<td>0.65</td>
<td>-0.02</td>
<td>0.433</td>
<td>Non-Significantly Lower</td>
</tr>
<tr>
<td>3</td>
<td>162</td>
<td>44.5%</td>
<td>3.91</td>
<td>0.85</td>
<td>-0.24</td>
<td>&lt;0.000</td>
<td>Significantly Lower</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>11.8%</td>
<td>3.82</td>
<td>0.74</td>
<td>-0.33</td>
<td>0.003</td>
<td>Significantly Lower</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>3.8%</td>
<td>3.29</td>
<td>0.99</td>
<td>-0.87</td>
<td>0.003</td>
<td>Significantly Lower</td>
</tr>
</tbody>
</table>

(a) One-tailed t-tests.
(b) The Bonferroni correction was applied by lowering the alpha level by a factor of five (the total number of comparisons) to yield a critical alpha level of .01 for testing the significance of each comparison (Hair et al, p. 328).
Figure 1. Classification of total sample into different customer groups.