

# Journal of Service Research

<http://jsr.sagepub.com>

---

## **Service Recovery Paradox: A Meta-Analysis**

Celso Augusto de Matos, Jorge Luiz Henrique and Carlos Alberto Vargas Rossi

*Journal of Service Research* 2007; 10; 60

DOI: 10.1177/1094670507303012

The online version of this article can be found at:  
<http://jsr.sagepub.com/cgi/content/abstract/10/1/60>

---

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:

Center for Excellence in Service, University of Maryland

**Additional services and information for *Journal of Service Research* can be found at:**

**Email Alerts:** <http://jsr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jsr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** (this article cites 56 articles hosted on the SAGE Journals Online and HighWire Press platforms):  
<http://jsr.sagepub.com/cgi/content/refs/10/1/60>

# Service Recovery Paradox: A Meta-Analysis

**Celso Augusto de Matos**

**Jorge Luiz Henrique**

**Carlos Alberto Vargas Rossi**

*School of Management, Federal University of Rio Grande do Sul (PPGA-EA-UFRGS)*

---

*The Service Recovery Paradox (SRP) has emerged as an important effect in the marketing literature. However, empirical research testing the SRP has produced mixed results, with only some studies supporting this paradox. Because of these inconsistencies, a meta-analysis was conducted to integrate the studies dealing with the SRP and to test whether studies' characteristics influence the results. The analyses show that the cumulative mean effect of the SRP is significant and positive on satisfaction, supporting the SRP, but nonsignificant on repurchase intentions, word-of-mouth, and corporate image, suggesting that there is no effect of the SRP on these variables. Additional analyses of moderator variables find that design (cross-sectional versus longitudinal), subject (student versus nonstudent), and service category (hotel, restaurant, and others) influence the effect of SRP on satisfaction. Finally, implications for managers and directions for future research are presented.*

**Keywords:** *service recovery; Service Recovery Paradox; meta-analysis; moderation analysis*

The Service Recovery Paradox (SRP) is a peculiar effect in the services marketing literature and has been conceptually defined as a situation in which a customer's postfailure satisfaction exceeds prefailure satisfaction (McCullough and Bharadwaj 1992). When reviewing this literature, the theoretical paper by Hart, Heskett, and Sasser (1990) is one of those frequently cited, especially the statement that "a good recovery can turn angry, frustrated customers into loyal ones. It can, in fact, create more goodwill than if things had gone smoothly in the first place" (p. 148). In this way, recovery encounters would mean an opportunity for service providers to increase customer retention (Hart, Heskett, and Sasser 1990).

The topic of SRP has been of great importance for managers and researchers. Given that failure is one of the main reasons that drive customer-switching behavior, understanding recovery is relevant because a successful recovery may lead to customer retention, which will affect company profitability (McCullough, Berry, and Yadav 2000). On the other hand, there has always been a question in the service literature as to whether high recovery efforts can really create greater satisfaction

---

The authors are thankful for the support provided by the Brazilian Funding Council for Research (CNPq and CAPES) and the Graduate School of Management. The authors also would like to thank various authors who sent their recent articles on the topic of service recovery, including Chihyung Ok, David A. Cranage, Stefan Michel, Steven H. Seggie, and Vincent P. Magnini. The authors are also grateful to Professor Frank L. Schmidt and Professor David B. Wilson for their support in answering questions about methods of meta-analysis and to the editor and three anonymous reviewers of *JSR* for their insightful comments.

Journal of Service Research, Volume 10, No. 1, August 2007 60-77

DOI: 10.1177/1094670507303012

© 2007 Sage Publications

when compared to the situation of no failure (Etzel and Silverman 1981).

However, empirical studies investigating the SRP have produced results that vary considerably in terms of statistical significance, direction, and magnitude. Although some studies provide support for the SRP (Hocutt, Bowers, and Donovan 2006; Hocutt and Stone 1998; Magnini et al. 2007; Maxham and Netemeyer 2002; McCollough 2000; Michel 2001; Michel and Meuter 2006; Smith and Bolton 1998), others have found no support (Andreassen 2001; Halstead and Page 1992; Hocutt, Chakraborty, and Mowen 1997; Maxham 2001; McCollough, Berry, and Yadav 2000; Ok, Back, and Shanklin 2006; Zeithaml, Berry, and Parasuraman 1996).

These conflicting results might be a consequence of a number of factors, from different methodological aspects in the studies to certain conditions moderating the paradox. In this respect, some variables have been proposed in the service recovery literature as potential moderators of the paradox, including severity of the failure, prior failure with the firm, stability of the cause of the failure, and perceived control (Magnini et al. 2007).

Along with these conflicting results, "there is a considerable body of conjecture and intuition pertaining to the existence of the service recovery paradox" (Andreassen 2001, p. 40). Thus, this inconsistency with regard to the effect of the SRP suggests the need for a meta-analysis to provide both a systematic review and a quantitative integration of all the available SRP research. A meta-analysis can provide insights into these inconsistencies by accumulating effects across studies after adjusting for the studies' main artifacts (i.e., measurement and sampling error), identifying measurement and sample characteristics that affect the support/nonsupport of the SRP, and also testing the generalizability of the results (Farley, Lehmann, and Sawyer 1995).

Through meta-analysis we aim to (a) reflect on the different methodological approaches used to test the SRP, (b) map the dependent variables that have been considered when the SRP is tested in the literature, (c) reveal which of these dependent variables support the SRP, (d) investigate which methodological differences across the studies moderate the results for the effects of the SRP, and (e) identify research questions worthy of future empirical investigations regarding the SRP.

First, we present a theoretical background about the SRP to guide the meta-analysis. Second, we discuss the procedures for building the database, computing, and integrating the effect sizes. Third, we present a quantitative summary that includes the adjusted cumulative mean values of the effect of service recovery on dependent variables and test whether the paradox is supported or not. Fourth, we conduct a more detailed examination,

using subgroup meta-analysis and hierarchical moderator analysis (Hunter and Schmidt 2004) to provide insights regarding studies' characteristics that might moderate the effects of the SRP. Finally, we present a discussion with theoretical and managerial implications, limitations, and future research directions.

## THE SERVICE RECOVERY PARADOX

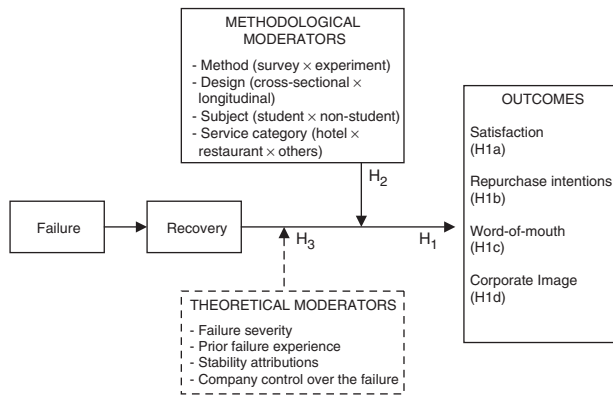
The SRP is defined as the situation in which postrecovery satisfaction is greater than that prior to the service failure when customers receive high recovery performance (Maxham 2001; McCollough 1995; McCollough and Bharadwaj 1992; Smith and Bolton 1998). In this context, effective service recovery may lead to higher satisfaction compared to the service that was correctly performed the first time, and recovery encounters would mean an opportunity for service providers to increase customer retention (Hart, Heskett, and Sasser 1990).

Based on the disconfirmation framework (McCollough, Berry, and Yadav 2000; Oliver 1997), the paradox is related to a secondary satisfaction following a service failure in which customers compare their expectations for recovery to their perceptions of the service recovery performance. If there is a positive disconfirmation, that is, if perceptions of service recovery performance are greater than expectations, a paradox might emerge (secondary satisfaction becomes greater than prefailure satisfaction). Otherwise, in the case of a negative disconfirmation, there is a double negative effect, as service failure is followed by a flawed recovery (Bitner, Booms, and Tetreault 1990; McCollough, Berry, and Yadav 2000; Smith and Bolton 1998).

The paradox can also be justified by the script theory and the commitment-trust theory for relationship marketing (Magnini et al. 2007). Script theory proposes that there is a common sequence of acts in a service delivery, in such a way that employees and customers share similar beliefs regarding the expected order of events and their respective roles in the process (Bitner, Booms, and Mohr 1994). If a service failure occurs, it works as a deviation from the predicted script and produces an increased sensitivity in the customer regarding the failure and the redress process. Because of this, satisfaction with the recovery process becomes more relevant than satisfaction with the initial attributes in influencing the final cumulative satisfaction (Bitner, Booms, and Tetreault 1990; Magnini et al. 2007).

Because an excellent service recovery has a direct impact on how much consumers trust the firm (Kau and Loh 2006; Tax, Brown, and Chandrashekar 1998), there is also a foundation for the Service Recovery Paradox in Morgan and Hunt's (1994) commitment-trust

**FIGURE 1**  
**Meta-Analytic Framework of the SRP**



NOTE: dashed line ( - - ) indicates path not tested in the meta-analysis.

theory for relationship marketing (Magnini et al. 2007). In this view, both service recovery efforts and relationship marketing focus on customer satisfaction, trust, and commitment. Trust exists when one party has confidence in another's reliability and integrity (Moorman, Zaltman, and Deshpande 1992; Morgan and Hunt 1994). As failures contribute to create insecurity in the customers and affect trust in the firm, an effective service recovery can be an opportunity to make customers feel that the firm is able and willing to correct the problem. As a result, a fair conflict resolution may have a positive impact on consumer trust (Achrol 1991).

Although most studies test the Service Recovery Paradox for satisfaction and repurchase intentions,<sup>1</sup> there are also studies considering the paradox for word-of-mouth (Hocutt, Bowers, and Donovan 2006; Kau and Loh 2006; Maxham 2001; Maxham and Netemeyer 2002; Ok, Back, and Shanklin 2006), corporate image (Andreassen 2001; Kwortnik 2006), trust (Kau and Loh 2006), quality (McCullough 1995), complaint intentions (Hocutt, Chakraborty, and Mowen 1997), switching intentions, pay-more intentions, and external response (Zeithaml, Berry, and Parasuraman 1996). Figure 1 provides an overarching conceptual framework for our meta-analysis and also synthesizes key insights from previous studies and discussions about the SRP in the extant literature. If the SRP exists, a service failure that is followed by a high recovery effort should produce outcome variables that are higher when compared to a situation in which no failure occurred. Based on this framework, we expect the following:

*Hypothesis 1a:* There is a significant positive SRP effect for satisfaction.

*Hypothesis 1b:* There is a significant positive SRP effect for repurchase intentions.

*Hypothesis 1c:* There is a significant positive SRP effect for word-of-mouth.

*Hypothesis 1d:* There is a significant positive SRP effect for corporate image.

Conflicting results have been found in the literature, with some studies supporting the SRP (Hocutt, Bowers, and Donovan 2006; Hocutt and Stone 1998; Magnini et al. 2007; Maxham and Netemeyer 2002; McCullough 2000; Michel 2001; Michel and Meuter 2006; Smith and Bolton 1998) and others not supporting this effect (Andreassen 2001; Halstead and Page 1992; Hocutt, Chakraborty, and Mowen 1997; Mattila 1999; Maxham 2001; McCullough, Berry, and Yadav 2000; Ok, Back, and Shanklin 2006; Zeithaml, Berry, and Parasuraman 1996). Also, there is support for the notion that the SRP is more likely when service failure causes low harm, indicating that recovery strategies may be more effective when the failure is perceived by the customers as less severe (Magnini et al. 2007; Mattila 1999; Smith and Bolton 1998). These contingencies are discussed later in the Moderators section.

Another possible explanation for the mixed findings might be related to the nature of the paradox (Michel and Meuter 2006). In this view, it is considered that the SRP is a very rare event (Boshoff 1997), that only a minority of dissatisfied customers complains (Singh 1990), and that only few recoveries lead to customer satisfaction (Kelley, Hoffman, and Davis 1993). As a result, it becomes very difficult to achieve a large sample of customers who have received a very satisfactory recovery,<sup>2</sup> and this requirement may have an influence on nonsignificant results presented in the literature (Michel and Meuter 2006).

## Moderators

The mixed findings can also be caused by certain conditions moderating the paradox. For example, although the theoretical definition of the SRP seems to be convergent in the literature, the same is not true for the operationalizations of the concept. Although some authors use a between-subjects approach, comparing a recovery group with a no-failure control group (Hocutt, Bowers, and Donovan 2006; Kau and Loh 2006; McCullough 1995; McCullough, Berry, and Yadav 2000; Michel and Meuter 2006; Ok, Back, and Shanklin 2006), others use a within-subjects approach, comparing different measures from the same subject before and after a failure and/or recovery (Magnini et al. 2007; Maxham 2001; Maxham and Netemeyer 2002; Smith and Bolton 1998). These differences also extend to the type of research design used in terms of experiment or survey approach, cross-sectional or longitudinal measures, student or nonstudent subjects,

single or multiple items measuring dependent variables, scenario- or non-scenario-based research, and the different manipulated factors (in case of experiments).

Assmus, Farley, and Lehmann (1984) suggested that four categories of characteristics might help identify systematic patterns in a meta-analysis: research context, model specification, measurement methods, and estimation procedure. However, because our unit of analysis was bivariate correlations, we seek systematic differences in the study characteristics. This procedure is common in meta-analysis using correlations (e.g., Pan and Zinkhan 2006). In our investigation, we examine four potential moderators: method (survey versus experiment), design (cross-sectional versus longitudinal), subject (student versus nonstudent), and service category (hotel versus restaurant versus others).

Based on the methodological differences across the studies, we propose the following (see Figure 1):

*Hypothesis 2a:* The SRP effects differ in studies using survey methods versus those using experimental methods.

*Hypothesis 2b:* The SRP effects differ in studies using cross-sectional designs versus those using longitudinal designs.

*Hypothesis 2c:* The SRP effects differ in studies using student subjects versus those using non-student subjects.

*Hypothesis 2d:* The SRP effects differ across studies using different service categories.

Regarding the boundary conditions for the SRP effects, some theoretical variables have also been proposed as potential moderators in the service recovery literature, including severity of the failure, prior failure with the firm, stability of the cause of the failure, and perceived control (Magnini et al. 2007). Most of these contingencies, however, have been tested only in the more recent literature (with severity of the failure<sup>3</sup> being an exception), which precluded their empirical assessment as moderators in our meta-analysis. Nevertheless, they are included in Figure 1 as propositions to be investigated further in future research. Their rationale for proposing the various theoretical moderators is discussed next.

Studies support the notion that it is harder to recover from high-magnitude failures (Magnini et al. 2007; Mattila 1999; McCollough 1995; Smith and Bolton 1998) or that the perceived harm caused by the failure interacts with the recovery effort to influence customer satisfaction (McCollough, Berry, and Yadav 2000). It has been found that the higher the magnitude or severity of the failure, the lower the overall customer satisfaction (Mattila 1999; Weun, Beatty, and Jones 2004), just as less favorable recoveries tend to be more memorable (Kelley,

Hoffman, and Davis 1993). For instance, a recovery action (e.g., apology or compensation) might increase customer satisfaction after a delay in waiting in a line. But what if this delay has caused a serious consequence for the customer (e.g., he missed his flight after a delay at the hotel desk)? It is unlikely that a recovery action would be able to either bring the customer to the original level of satisfaction or, even more improbable, increase his satisfaction. Thus, we propose:

*Hypothesis 3a:* The SRP is more (less) likely to occur when the customer perceives the failure as less (more) severe.

Given that customers usually have a history of interactions with the firm, their cumulative satisfaction, as opposed to a transaction-specific satisfaction, is based on their evaluations of multiple experiences with the firm over time (Bolton and Drew 1991). In this way, satisfactory recoveries may yield paradoxical gains only in the short run, and customers will likely infer that multiple failures are because of problems inherent to the firm (Maxham and Netemeyer 2002). Hence, when a customer experiences a second failure, he or she is more likely to attribute the cause of that problem to the firm than when the customer experienced failure for the first time (Magnini et al. 2007; Maxham and Netemeyer 2002). Thus, we propose:

*Hypothesis 3b:* The SRP is more likely to occur when the customer experiences the failure for the first time when compared to the situation in which there has already been a previous service failure.

Another influencing factor is the stability of the cause of the failure. Stability attributions refer to customers' inferences about whether similar failures are likely to occur in the future, given the customers' dissatisfaction with a product or service (Blodgett, Granbois, and Walters 1993; Folkes 1984). When customers experience a service failure, they ask themselves whether the failure has temporary (i.e., unstable) or permanent (i.e., stable) causes, and if they think that the problem has stable causes (i.e., it is likely to occur again), then they will try to avoid this service provider in the future (Folkes 1984). Studies have indicated that consumers who perceive a service failure as more stable present lower repatronage intentions (Folkes 1984, 1988). Smith and Bolton (1998) found similar results. In their findings, if a customer believed that the unavailability of the requested food item was because of a consistent omission of the restaurant, he or she would be less satisfied and less likely to repatronize this restaurant. Hence, customers are more likely to forgive failures with unstable (temporary) causes (Kelley, Hoffman, and Davis 1993; Magnini et al.

2007) and to express a situation of recovery paradox in this context (Magnini et al. 2007). Thus, we expect:

*Hypothesis 3c:* The SRP is more (less) likely when customers perceive that the failure is less (more) likely to repeat in the future.

Finally, attributions related to whether the firm had much or little control over the occurrence of the failure also influence the recovery paradox. When customers perceive that the firm had little control over the service failure, they are more likely to comprehend and forgive the problem (Maxham and Netemeyer 2002). This is in agreement with studies showing that the perceived reason for a product's failure is an important predictor of how consumers react (Folkes 1984). For instance, complainants who believed that firms were responsible for the failure were more likely to expect redress (e.g., apologies, refunds). Also, customers who attribute failures to controllable factors will probably be more dissatisfied with the failure and less forgiving in their evaluations. Indeed, it has been found that an SRP is more (less) likely to occur when the customer perceives that the firm had little (much) control over the cause of the failure (Magnini et al. 2007). Based on these findings, we propose:

*Hypothesis 3d:* The SRP is more (less) likely when customers perceive that the firm had little (much) control over the cause of the failure.

## METHOD

### Search Process and Sampling Frame

Studies were identified by a computerized bibliographic search. Databases included Blackwell Synergy, Elsevier Science Direct, Ebsco, Emerald Insight, Infotrac College, Proquest, Scopus, Thompson Gale, Wilson Web, and Google Scholar. First, we searched for the terms "service failure" and "service recovery" in keywords and abstracts. Then we narrowed our search by "service recovery paradox" in abstracts, keywords, or full text. Using this procedure, we found a total of 319 articles ranging from 1987 to 2006. By searching Proquest and Google, we could also access 14 dissertations on the research topic, leading to a total of 333 studies. Of this total, 42 (12.6%) were theoretical papers and the remaining 291 investigated service failure and/or recovery empirically. Among these studies, 21 were identified as testing the SRP empirically and were chosen for the analysis, producing a total of 24 observations (independent samples) in our data set (see Table 1).

All identified studies were then examined in terms of the following relevant variables: authors, year, journal, service category (hotel, restaurant, and others), method (survey versus experiment), subjects (students versus nonstudents), number of compared groups, number of factors manipulated or measured, dependent variables (satisfaction, repurchase intentions, word-of-mouth, trust, image, quality, intentions to complain, switching intentions, pay-more intentions, and external response), reliabilities for the dependent variables, and effect sizes.

### Effect Size Computation

Our meta-analytic procedure followed common guidelines for meta-analysis of experimental studies (Lipsey and Wilson 2001), in which standardized mean differences (Cohen's  $d$ ) are computed first and then converted to correlation coefficients ( $r$ ). This procedure is the same as that employed by other meta-analyses in the marketing literature (e.g., Brown and Stayman 1992; Eisend 2006). We selected the correlation coefficient,  $r$ , as the effect-size metric because it is easier to interpret and a scale-free measure. Also, the correlation coefficient is the mostly used effect size in meta-analyses in the marketing literature (e.g., DelVecchio, Henard, and Freling 2006; Eisend 2004, 2006; Franke and Park 2006; Janiszewski, Noel, and Sawyer 2003; Palmatier et al. 2006; Pan and Zinkhan 2006). As we included in our data set both surveys and experiments, we could integrate them by using  $r$  as the common effect size. Positive (negative) values of the correlation coefficient indicate the presence (absence) of the SRP. This procedure followed recommendations by Lipsey and Wilson (2001, pp. 14, 173) for conducting meta-analysis with group contrasts (in experiments or surveys) and is based on the following rationale:

1. The SRP refers to the effect that an outcome variable (e.g., satisfaction) is greater for a customer that has experienced a failure and a high recovery effort when compared to a customer that has experienced no failure.
2. The recovery group is considered as the treatment group and the no failure group as the control group.
3. "The contrast between the experimental and control group on the values of an outcome variable is interpreted as the effect of treatment" (Lipsey and Wilson 2001, p. 14). Treatment in our case is the high recovery effort.
4. Effect size (ES) =  $(\text{mean}_{\text{Sat experimental group}} - \text{mean}_{\text{Sat control group}}) / \text{standard deviation}$ .
5. Then, if satisfaction has a significantly higher mean value in the experimental group (high recovery) when compared to the control group

**TABLE 1**  
**Studies Included in the Meta-Analysis**

No.	Study	Effect Sizes for the Relationship of SRP and ...									
		Sat	Rep	Wom	Ima	Tru	Qua	IntC	Swi	Pay	Ext
1	Hocutt, Bowers, and Donovan (2006)	X		X							
2	Kau and Loh (2006)		X	X		X					
3	Ok, Back, and Shanklin (2006)	X	X	X							
4	Magnini et al. (2007)	X									
5	Michel and Meuter (2006)	X	X								
6	Kwortnik (2006)		X		X						
7	Oh (2003) <sup>a</sup>	X									
8	Maxham and Netemeyer (2002)	X	X	X							
9	Maxham (2001), Study 1	X	X	X							
10	Maxham (2001), Study 2	X	X	X							
11	Andreassen (2001)		X		X						
12	Michel (2001)	X									
13	McCullough, Berry, and Yadav (2000), Study 1	X									
14	McCullough, Berry, and Yadav (2000), Study 2	X									
15	McCullough (2000)	X									
16	Mattila (1999)	X									
17	Smith and Bolton (1998), Study 1	X	X								
18	Smith and Bolton (1998), Study 2	X	X								
19	Hocutt and Stone (1998)	X									
20	Boshoff (1997)	X									
21	Hocutt, Chakraborty, and Mowen (1997)	X						X			
22	Zeithaml, Berry, and Parasuraman (1996)		X						X	X	X
23	McCullough (1995) <sup>b</sup>	X					X				
24	Halstead and Page (1992)		X								
	Total	19	12	6	2	1	1	1	1	1	1

NOTE: Total number of effect sizes: 45. Sat = satisfaction; Rep = repurchase intentions; Wom = word-of-mouth; Tru = trust; Ima = image; Qua = quality; IntC = intentions to complain; Swi = switching intentions; Pay = pay more intentions; Ext = external response.

a. Classified as outlier.

b. Experiment 3 was used.

(no failure), the SRP is present and ES is positive. Otherwise, if satisfaction is higher in the condition of no failure, then ES is negative and reflects a situation of an inverse SRP. Finally, if there are no significant differences between satisfaction in the conditions of high recovery and no failure, then ES is close to zero and the SRP is null.

6. In conclusion, a positive effect size reflects a positive effect of the treatment (high recovery effort) and support for the SRP.

However, direct calculation of effect size (as present above in Item 4) is uncommon because rarely is there enough information. Under those circumstances, effects sizes are obtained through a range of statistical data (e.g., Student's *t*, *F* ratios with one *df* in the numerator,  $\chi^2$ ) by means of the formulas given by Lipsey and Wilson (2001, p. 198).

All together, using this approach, 45 effect sizes were available for the purpose of our meta-analysis. As presented in Table 1, most studies reported multiple effect sizes, particularly if we consider the first two outcome variables (satisfaction and repurchase intentions), for which there are more frequencies of effects (31 of the total 45). However, there were no significant mean differences for satisfaction when comparing studies reporting single or multiple effect sizes ( $t = .383$ ,  $p < .706$ ). When considering repurchase intentions, only one study presented a single effect size ( $ES = -.52$ ), and for all others, there were multiple effect sizes ( $n = 11$ ; mean  $ES = .025$ ). Mean differences in this case were not significant at the .05 significance level ( $t = 2.042$ ,  $p < .068$ ).

For most cases (35 of the 45 effect sizes), articles presented ANOVA results, including mean values of the outcome variable (e.g., satisfaction) for the treatment group

**TABLE 2**  
**Descriptive Statistics for Effect Size Integration**

Dependent Variable	k <sup>a</sup>	O <sup>b</sup>	N <sup>c</sup>	Min.	Max.	Sample-Weighted		Q Statistic for Homogeneity Test	Sig.	LCI	UCI	File Drawer N		
						Simple Average r	Adjusted Average r							
Sat	15	18	7,502	-.450	.595	.031	.154	.125	.017	.032	.217	45.992	.000	27
Rep	10	12	7,788	-.520	.435	-.020	-.061	-.072	.068	-.143	-.002	32.965	.001	5
Wom	5	6	672	-.292	.152	-.120	-.078	-.080	.330	-.227	.066	5.875	.319	NC
Image	2	2	762	-.084	.435	.175	.020	-.084	.556	-.282	.113	.000	1.000	NC

NOTE: LCI = lower confidence interval; UCI = upper confidence interval; Sig. = significance; Min. = minimum; Max. = maximum; Sat = satisfaction; Rep = repurchase intentions; Wom = word-of-mouth; Image = corporate image. Fail-safe number attenuated at .05. NC means the file drawer *N* was not calculated because the 95% confidence interval contained zero (effect size with significance  $p > .05$ ).

a. Number of studies.

b. Number of observations.

c. Combined *N* over all independent samples.

(high recovery effort) and for the control group (no failure), sample size for each one, and the *t* or *F* statistic. This was enough for computing the standardized difference and transforming it into *r*. In some cases (3 of 45), nonparametric tests were presented (e.g., chi-square) and the appropriate formula was used for conversion. In the remaining cases (7 of 45), authors provided either complete information for the direct calculation (mean, standard deviation, and sample size of each group) or *F* value and group sizes (without group means).

### Effect Size Integration

Effect size integration (mean, significance, and confidence intervals) was performed following common guidelines (Lipsey and Wilson 2001). Because the true relationship between variables is mainly influenced by sampling and measurement error, correlations were first weighted by the inverse variance and then by the inverse variance corrected for measurement error (cf. Lipsey and Wilson 2001, p. 110).<sup>4</sup> Thus, our effect size integration is presented in three stages, based first on observed correlations, then on correlations corrected for sampling error, and finally on correlations corrected for both sampling and measurement error. A confidence interval is presented for each effect size, and it is significant when it does not include zero. Significance for the mean effect size can also be tested by *z* statistic ( $p < .05$  if  $z > 1.96$ ). When the mean effect size is significant, a fail-safe *N* (also known as file drawer *N*) is calculated, estimating the number of non-significant and unavailable studies that would be necessary to bring the cumulated effect size to a nonsignificant value (known as the "file drawer problem"; Rosenthal 1979). This statistic is an indication of how robust results are. Table 2 presents a summary of this information.

Homogeneity of the effect size distribution was tested by the *Q* statistic, which is distributed as a chi-square (Hedges and Olkin 1985).<sup>5</sup> If the null hypothesis of homogeneity is rejected, it indicates that the variability in effect sizes is larger than it would be expected from sampling error, or in other words, differences in effect sizes may be attributed to factors other than sampling error alone, maybe moderating variables related to studies characteristics (Lipsey and Wilson 2001, p. 115).

For each relationship in which homogeneity of effect size was rejected (i.e., heterogeneity evidence), an analysis of moderating effects was performed, considering the study characteristics that were coded based on information provided in the articles. These variables included service category (hotels, restaurants, and others), method (survey versus experiment), design (cross-sectional versus longitudinal), subjects (students versus nonstudents), scenario (used or not), sample size (total, control group, and treatment group),<sup>6</sup> number of items used to measure the dependent variable,<sup>7</sup> and reliability of the dependent variable. These data were available for most studies.

When information about the number of subjects in each experimental group was not available, we used the mean group size by dividing the total sample size by the number of groups in the experiment (in 6 of the 24 observations). In some studies, authors mentioned in tables' notes the sizes of minimum and maximum groups. In these cases, we used the minimum as the sample size for both the treatment and the control group (in 7 of the 24 observations).

When studies measured dependent variables using single items<sup>8</sup> or when reliability values were unavailable,<sup>9</sup> these reliabilities were estimated using the Spearman-Brown procedure suggested by Hunter and Schmidt (2004, p. 311), a common approach in meta-analyses in marketing (e.g., Grewal et al. 1997). Among the 19 studies



providing effect size for satisfaction, 5 were based on single items and had an average reliability estimated as .701. In the same way, among the 12 studies providing effect size for repurchase intentions, 5 studies did not provide information on reliability (2 were based on single items and had reliability estimated as .66, 2 were based on two items and had reliability estimated as .795, and 1 used four items and had reliability estimated as .886).

## ANALYSIS AND RESULTS

Table 2 presents the results for the integration of effect sizes of the SRP on satisfaction, repurchase intentions, word-of-mouth, and image. Note that only those dependent variables with 2 or more observations are presented (trust, quality, intentions to complain, switching intentions, pay-more intentions, and external response are not submitted to further analyses because they were based on a single observation).

In this stage of the analysis, 1 of the 24 studies presented in Table 1 (Oh 2003) had to be excluded because it was classified as an outlier. Its sample size was 30,905 for the experimental group and 17,278 for the control group (the mean sample size for the remaining studies was 90 for the experimental group and 418 for the control group). If included, this study would produce an inverse variance weight of 11,053, whereas the mean value for the other studies was only 23. In summary, the study was excluded because it would bias the cumulated effect size.<sup>10</sup> All of the subsequent analyses were performed after the exclusion of this study from the data set.

The integrated effect size (corrected for sampling and measurement error) shows a positive and significant value for satisfaction ( $r = .125, p < .017$ ), supporting the existence of the SRP for satisfaction (Hypothesis 1a), with a medium effect size (Lipsey and Wilson [2001] consider  $r \leq .10$  as small effect;  $.10 < r < .40$  as medium effect; and  $r \geq .40$  as large effect). This indicates that satisfaction increases after a high service-recovery effort. The confidence interval ranged from .032 to .217, suggesting a small to medium effect. These results were based on 18 independent observations and 7,502 subjects. Fail-safe  $N$  suggests that 27 studies with non-significant effect size would be needed to reduce the cumulated effect size to a level of just significant (a level of .05 was used as "just significant," similar to Grewal et al. 1997).<sup>11</sup> In other words, to bring the significant effect of the SRP on satisfaction down to the level of just significant at  $\alpha = .05$ , it would be necessary to find 27 studies with null results to be included in our analysis. This is rather a small number, but it is somewhat unlikely that with only 18 studies identified for satisfaction, 27

studies remain in the "file drawer," especially because this is a recent research topic in the services marketing literature.

Conversely, the cumulated effect of the SRP on repurchase intentions is negative and not significant at the .05 significant level ( $r = -.07, p < .068$ ), not supporting Hypothesis 1b and suggesting that there is no SRP effect on repurchase intentions. Confidence intervals ranged from  $-.143$  to  $-.002$ , indicating a small effect. These findings were based on 12 independent observations and 7,788 subjects. Because of this small number of observations and the low magnitude of the average effect size, a relatively small number was found for the fail-safe  $N$ , indicating that 5 unpublished studies with null results would be needed to reduce the average effect size to the level of .05. Compared with satisfaction, it is more likely that this number of unpublished studies exists. Therefore, these results do not support Hypothesis 1b and indicate that repurchase intentions are not increased by a high service-recovery performance.

The integrated effect sizes of word-of-mouth and image were also negative but not significant, and they were based on a smaller number of observations (six and two, respectively). These results do not support Hypotheses 1c and 1d. Because of these null results, the file drawer number is not calculated for word-of-mouth and image.

A heterogeneous subset of effect sizes was obtained both for satisfaction and for repurchase intentions (see  $Q$  test of homogeneity in Table 2), indicating that moderating variables might help explain the variance in the effect sizes. Hence, we tested whether there were differences in effect sizes across study characteristics or, in other words, if the coded study characteristics were significant moderators. These analyses are presented in the next section.

### Moderating Effects

A common procedure for testing whether studies' characteristics can explain variability in the effect sizes is regression analysis, in which effect sizes are entered as dependent variables and moderators as independent variables (e.g., Eisend 2006; Szymanski and Henard 2001). However, this procedure may be limited if there are only few observations for each level of the moderators and/or there is small number of effect sizes. In this case, confidence in the results is threatened by low statistical power and capitalization on sampling error (Hunter and Schmidt 2004, p. 70). This was the case in our data set, as there were only 18 effect sizes for satisfaction and 12 for repurchase intentions.<sup>12</sup>

Because of this, we conducted a subgroup meta-analysis, comparing the mean effect size between the levels of each moderator, a common procedure in meta-analyses in

**TABLE 3**  
**Effects of Moderator Variables**

Moderator	Level	Satisfaction				Repurchase Intentions			
		No. of Studies	Effect Size	Test Statistic (Z-Value)	Sig.	No. of Studies	Effect Size	Test Statistic (Z-Value)	Sig.
Method	Survey	4	.124	.005	.498	8	-.073	.049	.481
	Experiment	14	.125	—	—	4	-.068	—	—
Design	Cross-sectional	11	.042	1.590	.056	7	-.090	.962	.168
	Longitudinal	7	.192	—	—	5	-.004	—	—
Subject	Student	8	.184	1.311	.095	2	-.218	1.198	.115
	Nonstudent	10	.060	—	—	10	-.060	—	—
Service category	Hotel (a)	4	.495	3.388 (a versus b)	.000	2	.324	-.746 (a versus b)	.772
	Restaurant (b)	5	.044	5.007 (a versus c)	.000	2	.160	2.953 (a versus c)	.002
	Others (c)	9	-.057	.840 (b versus c)	.200	8	-.112	1.593 (b versus c)	.056

marketing (see, e.g., Geyskens, Steenkamp, and Kumar 1998; Grewal et al. 1997; Palmatier et al. 2006; Pan and Zinkhan 2006). Also, an additional moderator analysis was conducted following the hierarchical method proposed by Hunter and Schmidt (2004). In this method, moderator variables are considered in combination to avoid confounding of correlated moderators.

In Table 3, we present the results of the moderator analyses, based on the subgroup meta-analyses. These results show the mean corrected effect size (both for measurement and sampling error) in each level of the moderators, together with the number of studies and the test of mean differences. Our moderator variables included service category (hotels, restaurants, and others), method (Survey  $\times$  Experiment), design (Cross-Sectional  $\times$  Longitudinal), subjects (Students  $\times$  Nonstudents), and scenario (used or not).<sup>13</sup>

As indicated in Table 3, the use of surveys or experiments did not change the direction of effect sizes, either in satisfaction or in repurchase intentions, not supporting Hypothesis 2a. On the other hand, longitudinal studies tended to present relatively higher means for satisfaction when compared to cross-sectional studies (.192 versus .042,  $p < .056$ ). This factor did not have influence on the effect sizes of repurchase intentions, not supporting Hypothesis 2b for this variable. Also, studies using student samples presented higher mean-effect sizes for satisfaction when compared to those using nonstudent samples (.184 versus .060,  $p < .095$ ). Thus, there was support for Hypothesis 2c for satisfaction only at the .10 significance level. Findings suggested no difference between Students  $\times$  Nonstudents in the repurchase intentions, not supporting Hypothesis 2c for this variable.

Considering service categories, studies conducted in hotels presented relatively higher effect sizes for satisfaction (.495) when compared to those in restaurants (.044)

or in the other categories (–.057), with significant differences ( $p < .000$  in both cases), which support Hypothesis 2d for satisfaction. In repurchase intentions, the difference between hotels versus others was replicated ( $p < .002$ ), and the studies conducted in restaurants presented higher effect sizes when compared to studies in other categories ( $p < .056$ ), giving partial support for Hypothesis 2d for repurchase intentions. These findings indicated that the context in which the SRP is investigated might also influence the results.<sup>14</sup>

Following a recommendation by Hunter and Schmidt (2004, p. 424) that subgroup meta-analysis may yield confounded results if moderators are correlated, we conducted an additional test for moderators, using hierarchical analysis. In this analysis, moderators are considered together. We provide results from this analysis in Table 4, considering the three moderators dealing with the methodological differences. One difficulty we encountered in this analysis was the small number of studies in each cell when considering the eight groups (a combination of  $2 \times 2 \times 2$ ). Despite this limitation, we could make five comparisons for the satisfaction effect sizes and one for the repurchase intentions. In these analyses, two factors are fixed and the levels of the third factor are compared.

In Table 4, we first compare cross-sectional versus longitudinal studies between experiments conducted with students (.086 versus .239). Results did not suggest a significant difference ( $p < .129$ ). In the sequence, the same comparison (cross-sectional versus longitudinal) was made between experiments conducted with nonstudents and a relatively higher effect size of satisfaction was found for longitudinal studies (–.149 versus .156,  $p < .085$ ). When we compared cross-sectional versus longitudinal between surveys conducted with nonstudents (.147 versus .103), no significant difference was found ( $p < .399$ ). This is an indication that the *difference* between cross-sectional versus longitudinal (suggested in Table 3)

**TABLE 4**  
**Hierarchical Moderator Analysis: Cross-Tabulation of Effect Sizes Among Moderators**

	Cross-Sectional		Longitudinal		Total
	Student	Nonstudent	Student	Nonstudent	
Survey	—	2 (6)	—	2 (2)	4 (8)
	—	.147 <sup>c</sup> (–.092) <sup>f</sup>	—	.103 <sup>c</sup> (.10) <sup>f</sup>	
	—	.453 (.072)	—	.554 (.56)	
Experiment	5	3 (1)	4 (2)	2 (1)	14 (4)
	.086 <sup>a,d</sup>	–.149 <sup>b,d</sup> (.013)	.239 <sup>a,e</sup> (–.218)	.156 <sup>b,e</sup> (.233)	
	.470	.353 (NC)	.099 (.335)	.537 (NC)	
Total	5 (0)	5 (7)	4 (2)	4 (3)	18 (12)

NOTE: Values in each cell represent number of effect sizes, mean effect size, and significance; em-dashes indicate no data. Values of repurchase intentions are in parentheses and values of satisfaction are outside parentheses. For example, there were two effect sizes of satisfaction in the category Survey/Nonstudent/Cross-sectional, with a mean effect size of .147 and significance of .453. NC = significance not calculated when only one effect size was available.

a. Contrast 1: .086 versus .239,  $p < .129$ .

b. Contrast 2: –.149 versus .156,  $p < .085$ .

c. Contrast 3: .147 versus .103,  $p < .399$ .

d. Contrast 4: .086 versus –.149,  $p < .088$ .

e. Contrast 5: .239 versus .156,  $p < .334$ .

f. Contrast 6: –.092 versus .10,  $p < .137$ .

is likely to be higher when experiments are conducted with nonstudents rather than with students.

Another comparison was made between students versus nonstudents, considering experiments using cross-sectional design. In this case, a relatively higher effect size of satisfaction was found for the studies using students (.086 versus –.149,  $p < .088$ ). This difference was not statistically significant when we made this same comparison in experiments using longitudinal approach. Therefore, the *difference* in satisfaction effect sizes between students versus nonstudents (suggested in Table 3) seems to be stronger for experiments using a cross-sectional (rather than longitudinal) approach.

A similar analysis was conducted for the effect sizes of repurchase intentions. However, the limitation of a small sample size in each cell was more severe in this case, as the total number of effect sizes was smaller for this variable ( $n = 12$ ). As a consequence, only one comparison was possible: cross-sectional versus longitudinal for surveys using nonstudents (–.092 versus .100,  $p < .137$ ). However, no significant difference was found.

## DISCUSSION

The meta-analysis presented in this study provides a systematic review and a quantitative integration of the effects of high recovery efforts on the dependent variables (satisfaction, repurchase intentions, word-of-mouth, and corporate image), revealing the cumulative effect of the SRP on these variables. Because there are

mixed results in the SRP literature, a meta-analysis can help in understanding these inconsistencies by accumulating results after adjusting for measurement and sampling error and by identifying study characteristics that may account for the variability in effect sizes. Although the SRP is a relatively recent topic in the services marketing literature (first empirical studies begin in the 1990s), a total of 21 studies (24 independent samples) could be included in the meta-analysis and used for the effect size integration.

Our primary results reveal support for the SRP only in the case of satisfaction, with a mean adjusted effect size of .125, which was significant at the 5% level. This is interpreted as a medium effect size (Lipsey and Wilson 2001), with a confidence interval ranging from .032 to .217. Thus, from the cumulative studies reviewed by our meta-analysis, findings indicate that satisfaction increases after a high service-recovery effort, suggesting the existence of the SRP for this variable. This finding supports the notion that a customer's postfailure satisfaction exceeds prefailure satisfaction (McCullough and Bharadwaj 1992).

Based on this result, are recovery encounters really good opportunities for service providers to increase customer *retention*, as recommended by Hart, Heskett, and Sasser (1990)? The empirical integration provided by our meta-analysis suggests a negative answer, in that the SRP effect was not evident for the repurchase intentions variable at the 5% significance level. Even if we consider a 10% significance level, the small mean effect size with a negative sign (–.072) suggests that the SRP effect on

repurchase intentions runs counter to what the SRP predicts. In other words, customers' postfailure repurchase intentions are likely to be lower than or equal to their pre-failure intentions.

These results are interesting because they show that the SRP works for satisfaction but not for repurchase intentions. Customers are willing to make a positive evaluation of a firm providing a high recovery effort, but they are not likely to repatronize this firm. Why would this happen?

A possible explanation is that satisfied customers are not necessarily loyal (Reichheld 1994). A meta-analysis of customer satisfaction found, for instance, that it explains less than 25% of the variance in repurchase intentions (Szymanski and Henard 2001). In agreement with this rationale, a recent study in e-retailing by Forbes, Kelley, and Hoffman (2006) found that customers are not likely to repurchase once a failure has been experienced, even if they are satisfied with the recovery effort. These findings might be influenced by the low switching costs that customers experience in online shopping. Nevertheless, they are further evidence of the importance of considering satisfaction levels and switching levels in combination (Jones and Sasser 1995).

Another possible explanation is the following line of reasoning: When evaluating postrecovery satisfaction, customers can be more influenced by the recovery process itself and the positive rewards that it may provide (e.g., a free service, a compensation for the failure), but when evaluating their likelihood of repurchasing from the same firm, customers might think that their original desired result was not accomplished in the purchasing process (the company did not provide a correct service in the first time), and therefore, it is not worth repurchasing from this firm. In agreement with this, a recent study by Magnini et al. (2007) supports the notion that customers who have experienced previous failure do not experience the SRP effect (i.e., the SRP is more likely to occur when it is the firm's first failure with the customer).

The accumulated effect size was not significant for word-of-mouth and corporate image. A limitation in this case was the reduced number of available studies for testing the SRP effect for these variables. As a consequence of the nonsignificant result and the homogeneous effect sizes, they were not included in the subsequent analysis of moderators.

Further analyses of homogeneity of effect sizes for satisfaction and repurchase intentions suggested possible moderators, as heterogeneous effect sizes were found for both variables. By using subgroup meta-analysis and hierarchical moderating analysis, we first identified three factors that moderated the SRP effect for satisfaction (design, subject, and service category) and one factor influencing repurchase intentions (service category).

Results suggested that effect sizes were not significantly different across studies using experiments or surveys, either for satisfaction or for repurchase intentions. This is an indication that this methodological decision did not affect the support/nonsupport of the SRP in the related studies, in agreement with what is predicted by Michel and Meuter (2006).

When comparing longitudinal versus cross-sectional studies, it was found that longitudinal studies provided stronger evidence for the SRP in satisfaction. This difference was further investigated in the hierarchical moderator analysis. It was found not only that longitudinal studies provided stronger support for the SRP effect for satisfaction (contrary to the prediction of Michel and Meuter 2006), but also that the *difference* between cross-sectional versus longitudinal was likely to be higher when experiments were conducted with nonstudents, rather than with students.<sup>15</sup> This finding is interesting because it suggests a possible interaction between moderators.

Another influencing factor in the effect of SRP on satisfaction was the use or nonuse of students as research subjects (significance was found only at the 10% level). It was found that students were more likely to support the SRP than nonstudents were.

A possible explanation may be that nonstudents are usually more experienced customers, and because of this, they are less likely to experience a positive disconfirmation if they have higher expectations. Furthermore, students are usually researched out of the purchasing environment and, therefore, their satisfaction-evaluation process may not be influenced by their past experiences, as in the case of "real" customers. Hence, satisfaction evaluations of nonstudents will likely be lower than those of students, and in the case of a service failure, this pattern may be maintained because students out of their purchasing context might not be as severe in their evaluations as the nonstudents, who will also be closer to their previous experiences and encounters with the firm.

These differences between students versus nonstudents were more pronounced in cross-sectional experiments when compared to longitudinal experiments, suggesting again a possible interaction between moderators. In summary, these results should make us reflect on the following: If researchers conduct *cross-sectional* experiments, then a difference in support for the SRP might exist when using students or nonstudents; but when conducting *longitudinal* experiments, no significant difference exists between students or nonstudents. As longitudinal studies provided stronger effect sizes for the SRP, this is an indication that longitudinal experiments are more likely to provide support for the SRP, whether having students as respondents or not. This finding contributes to the SRP literature by shedding light on

the question “Are methodological aspects of the studies influencing their support/nonsupport for the SRP?” Our meta-analysis seems to offer an affirmative answer and suggests that future studies should focus on these boundary conditions of the SRP.

Finally, service category was also a significant moderator, with influence both on satisfaction and on repurchase intentions. It was found that studies in hotels provided higher support for the SRP effect on satisfaction when compared to all other categories. This should also be a point of future debate: “Are there differences across service categories that facilitate the support for the SRP?” For example, when a service failure occurs in a context of the hospitality industry, most customers will seek a solution to their problem, and therefore, there may be a greater tendency toward redress-seeking behavior in this context (McCullough 2000), and they may also have a higher likelihood of receiving recovery service. In this case, customers cannot easily look for another service provider or interrupt their travel plans. As mentioned above, switching costs may be an influencing factor in this context. The same factor may explain why hotels and restaurants scored higher in effect sizes for repurchase intentions. Future studies could investigate this proposition further and test, for instance, the differences in the SRP effects for customers with high versus low switching costs within a given industry.

An additional analysis of possible moderators indicated that studies using more reliable scales did not provide support (or provided weaker support) for the SRP (a significant negative correlation was found) both for satisfaction and repurchase intentions. As expected, the same influence was found for the number of items used to measure the dependent variables. On the other hand, sample size was not significantly related (at the 5% level) to the effect sizes in these two dependent variables, although a correlation of .43 (significant at the 10% level) was found between treatment group and satisfaction, indicating that studies using larger samples for the service recovery group were more likely to provide support for the SRP. This is in agreement with Michel and Meuter’s (2006) argument that, once the SRP is a very rare event (Boshoff 1997), it becomes very difficult to achieve a large sample of customers who have received a very satisfactory recovery, and this limitation may be responsible for the nonsignificant results presented in the literature. Indeed, as we found, studies with a larger number of respondents in the recovery group tend to present greater support for the SRP and have higher statistical power, as discussed in the next section.

### Statistical Power<sup>16</sup>

Statistical power is related to the probability of not rejecting a false null hypothesis (Type II error, defined

by  $\beta$ ). The power of an experiment is thus defined as  $(1 - \beta)$  and interpreted as the probability that a statistical test will correctly reject a false null hypothesis (Cohen 1988). Although Cohen (1988) recommends using .80 as the threshold for statistical power, there is also a suggestion for using .50 for the social sciences, in which errors are less likely to have life-threatening consequences (Muncer, Craigie, and Holmes 2003).

It has been argued that including the statistical power discussion in the context of a meta-analysis can contribute to enhance the reliability of the meta-analysis (Muncer, Craigie, and Holmes 2003). Following the guidelines presented by these authors, we (a) used the mean effect size computed for the included studies to estimate the average statistical power of the combined studies and (b) estimated the statistical power of each study, indicating the ability to detect an effect size of the magnitude of the mean effect size obtained in the meta-analysis (estimated as population effect size), given the sample size and the significance level of .05. We used G\*Power 3.0<sup>®</sup> (Faul et al. in press) for these analyses and the results are presented and discussed below.

The mean effect size of satisfaction ( $r = .125$ ;  $d = .251$ ), in conjunction with the mean sample size of the groups ( $n_{\text{treatment}} = 57$ ;  $n_{\text{control}} = 401$ )<sup>17</sup> and the significance level of .05, produced a statistical power of .42. This value is below the recommended .80 threshold level but relatively close to the level of .50 (Muncer, Craigie, and Holmes 2003). For the studies of repurchase intentions ( $r = -.072$ ;  $d = -.145$ ;  $n_{\text{treatment}} = 134$ ;  $n_{\text{control}} = 558$ ), statistical power was estimated as .32. This analysis indicates that the average statistical power was smaller in the 12 studies of repurchase intentions when compared to the 18 studies of satisfaction.

A deeper investigation of the statistical power of the individual studies showed that power varied between .11 and .98 (mean = .30;  $SD = .23$ ;  $n = 18$ ) for the satisfaction variable and between .07 and .80 (mean = .22;  $SD = .20$ ;  $n = 12$ ) for repurchase intentions. There was no significant difference between these two means ( $t = .91$ , sig. = .37), indicating that statistical power was not statistically different between the observations of satisfaction and repurchase intentions. Note that these means are not weighted by any factor, justifying why they are different from the previous mean statistical power presented for satisfaction (.42) and repurchase intentions (.32), when average statistical power was computed directly from the final average effect size weighted by sample size and reliability.

We also recomputed statistical power after excluding from the database those studies with power lower than .5 and checked whether there were major changes in the estimates of the mean effect size (weighted by sample size and reliability), significance and confidence intervals. For satisfaction, four studies produced a mean effect

size of  $r = .334$ , with  $\text{sig.} = .018$ , and confidence intervals between  $.20$  and  $.47$ . We can note that these studies provide stronger support for the SRP in the case of satisfaction ( $r$  changed from  $.125$  to  $.334$ ). Considering only these four studies with power greater than  $.50$ , we should conclude that there is a stronger SRP effect for satisfaction, although it is still a medium effect.

For repurchase intentions, only one study presented power higher than  $.50$  (power =  $.80$ ). This study presented an effect size of  $r = -.10$ , lower bound =  $-.20$ , upper bound =  $.00$ . These values are similar to the ones obtained from the 12 integrated studies (see Table 2:  $r = -.072$ , lower bound =  $-.143$ , upper bound =  $.002$ ), indicating that including the studies with lower power did not change the interpretation of the results in the cumulated effect size of repurchase intentions.<sup>18</sup>

Overall, these power analyses indicate that the studies conducted to test the SRP and included in our meta-analysis have relatively low statistical power, which can be one additional reason why "conflicting results" are found regarding the existence or nonexistence of the SRP. Our analyses showed, for example, that the mean effect size for satisfaction was greater ( $r = .33$ ) when only studies with acceptable power (higher than  $.50$ ) were retained in the analysis. This suggests a stronger effect ( $r = .33$ ) than when studies with lower power are also included in the meta-analysis ( $r = .125$ ). Because of the limited number of studies, however, we could not include the power estimates as one of the variables in our moderation analysis.

## Implications and Further Research

The main implication of this meta-analysis is to show that the SRP effect is more likely to occur for satisfaction than for repurchase intentions. This result challenges us to understand why satisfied customers are not necessarily loyal. In a recent investigation of this topic, Chandrashekar et al. (2007), by decomposing satisfaction in two factors (satisfaction level and satisfaction strength), have theorized that weakly held satisfaction does not translate into loyalty and that only strongly held satisfaction is able to translate into loyalty. Based on this rationale, future SRP studies could check the influence that satisfaction strength might exert as a possible moderator (i.e., the SRP effect may occur for repurchase intentions when it occurs for satisfaction and this satisfaction is strongly held).

Also, the lack of support for the SRP effect on repurchase intentions may suggest differences of the SRP on the diverse stages of the loyalty pyramid (cognitive  $\rightarrow$  affective  $\rightarrow$  conative  $\rightarrow$  action; Oliver 1997). Because the studies included in our meta-analysis did not take this factor into account empirically, future studies should be conducted to investigate whether the SRP can exist for

the different stages of loyalty (e.g., the SRP may exist during the cognitive stage but not the action stage).

Future studies are needed to provide further investigation of the reasons behind this difference of support of the SRP in satisfaction but not in repurchase intentions. In this sense, switching costs may be one of the relevant factors accounting for this difference, as suggested by the findings of Forbes, Kelley, and Hoffman (2006). The difference between "satisfaction level" and "satisfaction strength" (Chandrashekar et al. 2007) can also contribute to this discussion.

Another possibility is that studies investigating repurchase intentions may have used approaches that tend to give no support for the SRP. For example, 6 out of 12 of the effect sizes for repurchase intentions came from cross-sectional surveys with nonstudents. It was suggested in the moderator analysis of satisfaction that (a) studies using cross-sectional design produce relatively smaller effect sizes when compared to longitudinal approach, and (b) studies with nonstudents produced lower effect sizes when compared to students use. Given that 50% of the effect sizes for repurchase intentions used this design, it may be possible that a SRP was not supported for repurchase intentions because of these methodological differences across the studies.

Our moderation analyses were intended to be only an exploratory investigation of the possible methodological aspects that might have an influence on the effects of SRP, especially because of the limited number of studies in each cell in our hierarchical moderation procedure (Hunter and Schmidt 2004). Because of this, future research is necessary to provide further investigation of these moderating effects.

Also, although theoretical moderators have been suggested in the SRP literature (see Magnini et al. 2007), because only few studies have tested them empirically, they could not be included as variables in our data set,<sup>19</sup> as illustrated in Figure 1. Examples of these moderators include severity of the failure, prior failure with the firm, stability of the cause of the failure, and perceived company control. Hence, we suggest that these moderators be further investigated in future SRP studies.

Our revision of the mean effect sizes when taking statistical power into account, as suggested by Muncer, Craigie, and Holmes (2003), indicates the relevance of considering statistical power in the context of the studies testing the SRP, especially for the satisfaction variable, whose findings suggested a stronger effect of the SRP. Thus, future studies of the SRP should consider statistical power a priori so as to be able to achieve greater confidence in the results of the significance tests.

Our statistical power analysis can also be used to suggest sample sizes for future studies. Based on the integrated effect size as an estimate of the effect size of the

considered population, we can calculate the required sample size that future studies should use to be able to detect the desired effect. For example, for satisfaction, considering  $\text{sig.} = .05$ ,  $\text{power} = .8$ , it would be necessary to have a sample size of 251 in each group (experimental and control) to be able to identify an effect size of  $r = .125$  ( $d = .251$ ) as significant. This value would be reduced to 123 if we considered the statistical power in the suggested level of .50 (Muncer, Craigie, and Holmes 2003). On the other hand, for the repurchase intentions, it would be necessary to have a sample size of 748 in each of the two groups to identify an effect size of  $r = -.072$  ( $d = -.145$ ) with a statistical power of .8. If we lowered the power to .5, the sample size would be reduced to 367. As expected, for a given power, it is necessary to have a larger sample size to identify smaller effect sizes.

Another relevant area for further research includes the cognitive/affective mechanisms behind the SRP. Because there is support for the SRP effect on satisfaction, as suggested by the integrated results of the meta-analysis, and satisfaction involves cognitive and affective dimensions in prepurchase, purchase, and postpurchase phases of consumption (e.g., Westbrook 1980), future studies should investigate further the cognitive and/or affective mechanisms driving the SRP effect. Although previous studies have investigated the influence of the cognitive and/or affective factors on satisfaction with service recovery (e.g., Andreassen 2000), there is a lack of studies examining the cognitive and affective dimensions in the context of the SRP.

As discussed by Parasuraman (2007), future research should also investigate whether there is an optimal mix of reliability versus recovery investments or, in other words, how much should be invested in delivering reliable service rather than in superior recovery when problems occur. What are the main variables in this context and what are their influences? These questions are also relevant for future research on the service-recovery context.

Finally, we would suggest that more studies should be conducted to investigate not only satisfaction and repurchase intentions as main recovery outcomes but also important constructs like word-of-mouth and corporate image, for which only a limited number of studies were found in the literature. Other relevant variables include trust, quality, intentions to complain, and switching behavior. Testing the SRP effect on the customers' actual behavior rather than on their behavioral intentions alone would also contribute to the current state of the knowledge about the SRP.

## Managerial Implications

Our meta-analysis has a number of implications for service managers. First, the reviewed studies of the SRP

indicated that customer satisfaction after a high recovery effort is greater when compared to that satisfaction prior to the service failure. However, the same is not true for the customer repurchase intentions. Because of this, service managers should make every effort to provide services correctly on the first time, rather than permitting failures and then trying to respond with superior recovery. This view has already been advocated by single studies (e.g., Andreassen 2001; McCollough, Berry, and Yadav 2000), but our meta-analysis makes this argument stronger, given that the meta-analysis provides an integrative review and a quantitative integration of the conflicting results about the SRP.

Second, trust is considered a key variable when managing customer relationships (see Morgan and Hunt 1994). In the context of service failure and recovery, it is expected that satisfaction with service recovery would lead to the building of trust. However, because results have demonstrated that customers who were initially satisfied with the service expressed greater trust when compared to the satisfied complainants, not supporting the SRP in trust (Kau and Loh 2006), a service failure seems to be a serious threat to trust. From the manager's point of view, it is critical to manage customer trust in the service provider. Trust can be built and/or enhanced with a company providing a reliable service over time. Hence, service failure should be avoided also because of its negative impact on customer trust. Indeed, research investigating why customers stay, given a switching dilemma, has suggested that the most important reason (out of all the 28 revealed reasons) was "lack of a critical incident," or in other words, customers stayed because a negative critical event had not occurred (Colgate et al. 2007). Thus, the service provider should perform as promised if customers' perceived confidence is expected to be strengthened. This confidence can be increased by investing in customers' feeling of comfort, trust in the service provider, satisfaction with the service provider, familiarity with the service provider, history with the current service provider, and lack of negative critical incidents (Colgate et al. 2007). Service managers should invest in these factors to earn the customer's confidence.

Nevertheless, achieving 100% service reliability can be impossible or cost prohibitive in most settings. Thus, in case of a failure, companies should strive to provide a service recovery of high performance anyway because a delight with the recovery can contribute in moving the customer up in the loyalty hierarchy (cognitive  $\rightarrow$  affective  $\rightarrow$  conative  $\rightarrow$  action), as suggested by Andreassen (2001). In other words, the effect of the SRP on repurchase intentions may be influenced by the loyalty stage in which the customer is found. This prediction needs to be confirmed by future studies, and the findings from such studies will

provide service managers a deeper understanding of this process. In the interim, we suggest that if a service failure occurs, the first step should be to provide a recovery of high performance in such a way to restore customer satisfaction and achieve customer delight with the recovery process. The second step should be to consider whether this customer's manifested satisfaction would be translated into future loyalty with the company. In this stage, managers could estimate through surveys the customer's satisfaction strength (i.e., the strength with which the satisfaction judgment is held; Chandrashekar et al. 2007) in such a way as to target those customers with weakly held satisfaction, as they are less likely to turn their satisfaction into loyalty (Chandrashekar et al. 2007). For instance, these customers could receive long-term benefits (e.g., discounts based on history with the company), which would increase their switching costs, and the company would use the future service encounters as opportunities to increase the customer satisfaction and make it more strongly held.

Third, service managers should be cognizant of the differences of the SRP across service settings. Findings from the meta-analysis indicated a stronger effect size of satisfaction in hotels, compared to all other categories, suggesting that SRP is more likely in this setting. This may be because of the high-contact characteristic of the hospitality industry. Furthermore, this is a service of relatively longer duration because even customers who stay for a short time in a hotel (e.g., 1 day) may engage in a series of service encounters. Also, given a service failure, the customer of a hotel may be more prone to engage in redress seeking, as he or she would not like (or be able) to change his or her schedule (e.g., cancel a business meeting or a vacation plan) because of this failure. Interestingly, there was also a higher likelihood of the SRP effect in repurchase intentions in the hotel setting. Thus, managers dealing with failures in a hotel context should also be able to provide recoveries of high performance so as to boost customer satisfaction and repatronage intentions.

In addition, service managers should also monitor customers' word-of-mouth, which can be very negative if a failure occurs and the company is not able to provide a satisfactory recovery (i.e., a "double deviation," as termed by Bitner, Booms, and Tetreault 1990). Indeed, the findings from the meta-analysis showed that there was a negative average effect size for word-of-mouth resultant from the six independent observations reviewed. As recommended by Andreassen (2001), positive word-of-mouth from existing customers can make the company more attractive in the eyes of the new customers. Negative word-of-mouth derived from unsatisfactory recoveries, on the other hand, could push to competitors not only potential new customers but also existing customers.

## Limitations

Meta-analyses offer several benefits, but they also have intrinsic limitations, which are common in most meta-analytic studies in the marketing literature. We discuss the main limitations of our study below.

First, our analyses are based on secondary data, and therefore, we cannot use information other than those presented in the articles. For example, we could test the methodological moderators presented in Figure 1 but not the theoretical moderators because very few studies tested the SRP considering these categories. Also, we had to consider a mean group size when studies did not inform the exact size of the groups being compared. Moreover, only satisfaction and repurchase intentions presented relatively high frequency of studies and could be entered in further analysis of moderators. For example, even though there was reference for other recovery outcomes (e.g., trust, quality, intentions to complain, and switching intentions), they could not be included in our effect-size integration.

Second, although there were a large number of studies investigating service failure and recovery (about 300 were identified), a relatively small number of them tested the SRP empirically and could be included in the analysis. Even with the recommendation that 10 or more studies should be an acceptable minimum number (see note 12), we should be careful in interpreting the results based on a small number of studies, especially regarding the moderator analysis. We recognize that the presented moderator analysis has a more exploratory perspective. Nevertheless, results from the moderator analysis can help researchers to design new studies that address the boundary conditions for the SRP effect.

Finally, studies included in the meta-analysis presented relatively limited statistical power in general. Because of this, the number of studies limited the conservative procedure of calculating more robust effect sizes only from studies with acceptable statistical power. When we implemented this procedure for satisfaction, only 4 studies remained, with the remaining 14 studies having low power. This is a clear indication that future studies in this context should consider statistical power a priori and determine the minimum sample size required to detect the effect size.

## CONCLUSION

Notwithstanding the presented limitations, the findings from this meta-analysis contribute to a greater understanding of the SRP by (a) estimating its cumulated mean effect for the key dependent variables (satisfaction



and repurchase intentions), (b) testing how studies characteristics might influence these results, and (c) suggesting further research directions.

Meta-analyses should not be viewed as conclusive or as a substitute of new primary research but only as a methodological tool that makes a temporary “balance sheet” of the current state of the knowledge in a given area. Its main contribution to science is to help researchers to direct their next wave of research toward still-unexplored questions and boundary conditions.

In this spirit, we hope that the results reported by our meta-analysis provide managers and researchers with inspirations for designing new in-depth and extensive investigations that will keep advancing the services marketing literature.

## NOTES

1. Among our 24 studies, 10 used the term “repurchase intentions” as dependent variables and only 2 considered “loyalty” (Kau and Loh 2006; Zeithaml, Berry, and Parasuraman 1996). Because these two studies worked with loyalty intentions, we considered these variables together, named as “repurchase intentions.”

2. Indeed, we noticed in our data analysis process that the treatment group was relatively smaller (mean = 90) when compared to the control group (mean = 418).

3. Even severity of the failure could not be empirically evaluated in our moderation analysis because only 4 studies tested SRP considering Low  $\times$  High severity. With missing values in the other 20 studies, this variable could not be integrated in our moderation analyses.

4. The adjustments for unreliability can be applied directly to the inverse variance weights using  $w' = w \times (r_{yy})$ . In this approach, the effect size is corrected both for sampling and measurement error (cf. Lipsey and Wilson 2001, p. 110).

$$5. \text{ The formula used is: } Q = \left( \frac{\sum w_i ES_i^2 - \left( \frac{\sum w_i ES_i}{\sum w_i} \right)^2}{\sum w_i} \right)$$

6. For most of the studies the total sample was greater than the sum of the treatment and the control group. This was because each study tested relationships other than the paradox (e.g., a given study used a  $2 \times 2 \times 2$  experiment with a control group, and the paradox was tested comparing one of the eight experimental groups with the control group). In this case, we coded three variables: (a) total sample size from the nine groups, (b) size of treatment group, and (c) size of the control group.

7. It is expected that measures with more items produce greater reliability (Nunnally and Bernstein 1994) and then stronger effect sizes.

8. Seven out of the 24 studies used single items for any of the measured variables; a total of nine observations of the 45 values in our data set.

9. There were only four missing values of reliability for studies using multiple items in dependent variables.

10. The cumulated effect size for satisfaction would become  $-.135$  if Oh (2003) were included in the data set. The observed effect size for this study was  $-.149$ .

11. The formula for the fail-safe number is  $k \times (r - r_c) / r_c$ , where  $k$  is the accumulated number of studies,  $r$  is the mean effect size, and  $r_c$  is the critical effect size, or the “just significant” level (Hunter and Schmidt 2004, p. 501; Lipsey and Wilson 2001, p.166).

12. We asked experts in meta-analysis what would be the minimum number of studies needed to conduct a meta-analysis. Professor Frank L. Schmidt (author of the book *Methods of Meta-Analysis*; see Hunter and Schmidt 2004) told us, “The really bare minimum is 2 studies, but

most journals will not publish a meta-analysis unless it contains at least 5 to 7 studies. So with 10 or more, you are OK” (personal communication, October 16, 2006). Professor David B. Wilson (author of the book *Practical Meta-Analysis*; see Lipsey and Wilson 2001) also said, “Minimum number of studies: 2. Of course, this limits the analyses that you can do” (personal communication, October 14, 2006).

13. Method and scenario are analyzed together because all studies using experiments are based on scenarios, and none of the surveys used scenarios.

14. To achieve a better understanding of the variability of the effect sizes, we also checked if effect sizes were influenced by scale reliability, number of items measuring dependent variables, and sample size. We found a significant correlation between effect size and scale reliability ( $-.62$  in satisfaction and  $-.57$  in repurchase intentions), with negative values indicating that studies using more reliable scales did not provide support (or provided weaker support) for the SRP. As expected, significant positive correlations were found between number of items measuring the dependent variable and scale reliability ( $.59$  in satisfaction and  $.81$  in repurchase intentions). As a result, the number of items measuring satisfaction was also negatively related to the satisfaction effect size ( $r = -.47$ ) and the same was true for repurchase intentions ( $r = -.57$ ). Sample size (total, treatment, and control) was not significantly correlated with effect sizes either in satisfaction or in repurchase intentions (the highest correlation was found in the pair treatment group–satisfaction,  $r = .43$ ,  $p < .075$ ). A regression analysis of these variables on the effect sizes produced no significant results for either satisfaction or repurchase intentions. However, these regression results might not be reliable because they may be influenced by low statistical power and capitalization on chance because of the small number of observations (see Hunter and Schmidt 2004, p. 70).

15. We could not test if surveys with nonstudents (rather than with students) produce greater difference between cross-sectional versus longitudinal because there were no observation for the cells “survey with students.”

16. We are very thankful to Reviewer C for suggesting the inclusion of this topic in our discussion.

17. These are mean values based on 18 observations with total sample size of 7,218 in the control group ( $7,218 / 18 = 401$ ) and 1,015 in the experimental group ( $1,015 / 18 = 57$ ). For repurchase intentions, 12 observations had a total sample size of 6,686 in the control group ( $6,686 / 12 = 558$ ) and 1,601 in the experimental group ( $1,601 / 12 = 134$ ).

18. We also checked if there was a correlation between power and effect sizes of the individual studies. This correlation was not significant for either satisfaction or repurchase intentions.

19. Most of the studies did not test these moderators or provide information that could allow the authors to classify the study in one category or another. Because of this, we chose to tabulate only information available in the studies.

## REFERENCES

\*Asterisks denote studies used in meta-analysis.

- Achrol, Ravi S. (1991), “Evolution of the Marketing Organization: New Forms for Turbulent Environments,” *Journal of Marketing*, 55 (4), 77-93.
- Andreassen, Tor Wallin (2000), “Antecedents to Satisfaction with Service Recovery,” *European Journal of Marketing*, 34 (1/2), 156-75.
- \*——— (2001), “From Disgust to Delight—Do Customers Hold a Grudge?” *Journal of Service Research*, 4 (1), 39-49.
- Assmus, Gert, John U. Farley, and Donald R. Lehmann (1984), “How Advertising Affects Sales: Meta-Analysis of Econometric Results,” *Journal of Marketing Research*, 21 (1), 65-74.
- Bitner, Mary Jo, Bernard H. Booms, and Mary S. Tetreault (1990), “The Service Encounter: Diagnosing Favorable and Unfavorable Incidents,” *Journal of Marketing*, 54 (1), 71-84.
- , ———, and L. A. Mohr (1994), “Critical Service Encounters: The Employee’s Viewpoint,” *Journal of Marketing*, 58 (3), 95-106.

- Blodgett, Jeffrey G., Donald H. Granbois, and Rockney G. Walters (1993), "The Effects of Perceived Justice on Complainants' Negative Word-of-Mouth Behavior and Repatronage Intentions," *Journal of Retailing*, 69 (4), 399-428.
- Bolton, Ruth N. and James H. Drew (1991), "A Multistage Model of Customers' Assessments of Service Quality and Value," *Journal of Consumer Research*, 17 (March), 375-84.
- \*Boshoff, Christo (1997), "An Experimental Study of Service Recovery Options," *International Journal of Service Industry Management*, 8 (2), 110-30.
- Brown, Steven and Douglas M. Stayman (1992), "Antecedents and Consequences of Attitude toward the Ad: A Meta-Analysis," *Journal of Consumer Research*, 19 (1), 34-51.
- Chandrashekar, Murali, Kristin Rotte, Stephen S. Tax, and Rajdeep Grewal (2007), "Satisfaction Strength and Customer Loyalty," *Journal of Marketing Research*, 44 (1), 153-63.
- Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
- Colgate, Mark, Vicky T. U. Tong, Christina K. C. Lee, and John U. Farley (2007), "Back from the Brink: Why Customers Stay," *Journal of Service Research*, 9 (3), 211-28.
- DelVecchio, Devon, David H. Henard, and Traci H. Freling (2006), "The Effect of Sales Promotion on Post-Promotion Brand Preference: A Meta-Analysis," *Journal of Retailing*, 82 (3), 203-13.
- Eisend, Martin (2004), "Is It Still Worth to be Credible? A Meta-Analysis of Temporal Patterns of Source Credibility Effects in Marketing," in *Advances in Consumer Research*, Vol. 31, Barbara Kahn and Mary Frances Luce, eds. Urbana, IL: Association of Consumer Research, 352-57.
- (2006), "Two-Sided Advertising: A Meta-Analysis," *International Journal of Research in Marketing*, 23 (2), 187-98.
- Etzel, Michael J. and Bernard I. Silverman (1981), "A Managerial Perspective on Directions for Retail Customer Dissatisfaction Research," *Journal of Retailing*, 57 (3), 124-36.
- Farley, John U., Donald R. Lehmann, and Alan Sawyer (1995), "Empirical Marketing Generalization Using Meta-Analysis," *Marketing Science*, 14 (3), 36-46.
- Faul, Franz, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner (in press), "G\*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences," *Behavior Research Methods*, <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>.
- Folkes, Valerie (1984), "Consumer Reactions to Product Failure: An Attributional Approach," *Journal of Consumer Research*, 10 (March), 398-409.
- (1988), "Recent Attribution Research in Consumer Behavior: A Review and New Directions," *Journal of Consumer Research*, 14 (March), 548-60.
- Forbes, Lukas P., Scott W. Kelley, and K. Douglas Hoffman (2006), "Typologies of E-Commerce Retail Failures and Recovery Strategies," *Journal of Services Marketing*, 19 (5), 280-92.
- Franke, George R. and Jeong-Eun Park (2006), "Salesperson Adaptive Selling Behavior and Customer Orientation: A Meta-Analysis," *Journal of Marketing Research*, 43 (4), 693-702.
- Geyskens, Inge, Jan-Benedict E. M. Steenkamp, and Nirmalya Kumar (1998), "Generalizations about Trust in Marketing Channel Relationships Using Meta-Analysis," *International Journal of Research in Marketing*, 15 (3), 223-49.
- Grewal, Dhruv, Sukumar Kavanoor, Edward F. Fern, Carolyn Costley, and James Barnes (1997), "Comparative versus Noncomparative Advertising: A Meta-Analysis," *Journal of Marketing*, 61 (4), 1-15.
- \*Halstead, Diane and Thomas J. Page (1992), "The Effects of Satisfaction and Complaining Behavior on Consumer Repurchase Intentions," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 5, 1-11.
- Hart, Christopher W., James L. Heskett, and W. Earl Sasser, Jr. (1990), "The Profitable Art of Service Recovery," *Harvard Business Review*, 68 (August), 148-56.
- Hedges, Larry V. and Ingram Olkin (1985), *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- \*Hocutt, Mary Ann, Goutam Chakraborty, and John Mowen (1997), "The Impact of Perceived Justice on Customer Satisfaction and Intention to Complain in a Service Recovery," *Advances in Consumer Research*, 24, 457-63.
- \*——— and Thomas H. Stone (1998), "The Impact of Employee Empowerment on the Quality of a Service Recovery Effort," *Journal of Quality Management*, 3 (1), 117-32.
- \*———, Michael R. Bowers, and D. Todd Donovan (2006), "The Art of Service Recovery: Fact or Fiction?" *Journal of Services Marketing*, 20 (3), 199-207.
- Hunter, John E. and Frank L. Schmidt (2004), *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd ed. Beverly Hills, CA: Sage.
- Janiszewski, Chris, Hayden Noel, and Alan G. Sawyer (2003), "A Meta-Analysis of the Spacing Effect in Verbal Learning: Implications for Research on Advertising Repetition and Consumer Memory," *Journal of Consumer Research*, 30 (1), 138-49.
- Jones, Thomas O. and W. Earl Sasser, Jr. (1995), "Why Satisfied Customers Defect," *Harvard Business Review*, 73 (November-December), 88-99.
- \*Kau, Ah-Keng and Elizabeth Wan-Yium Loh (2006), "The Effects of Service Recovery on Consumer Satisfaction: A Comparison between Complainants and Non-Complainants," *Journal of Services Marketing*, 20 (2), 101-11.
- Kelley, Scott W., K. Douglas Hoffman, and Mark A. Davis (1993), "A Typology of Retail Failures and Recoveries," *Journal of Retailing*, 69 (4), 429-52.
- \*Kwortnik, Robert J. (2006), "Shining Examples of Service when the Lights Went Out: Hotel Employees and Service Recovery during the Blackout of 2003," *Journal of Hospitality and Leisure Marketing*, 14 (2), 23-45.
- Lipsey, Mark W. and David B. Wilson (2001), *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- \*Magnini, Vincent P., John B. Ford, Edward P. Markowski, and Earl D. Honeycutt (2007), "The Service Recovery Paradox: Justifiable Theory or Smoldering Myth?" *Journal of Services Marketing*, 21 (3), 213-225.
- \*Mattila, Anna S. (1999), "An Examination of Factors Affecting Service Recovery in a Restaurant Setting," *Journal of Hospitality & Tourism Research*, 23 (3), 284-98.
- \*Maxham, James G., III (2001), "Service Recovery's Influence on Consumer Satisfaction, Positive Word-of-Mouth, and Purchase Intentions," *Journal of Business Research*, 54 (1), 11-24.
- \*——— and Richard G. Netemeyer (2002), "A Longitudinal Study of Complaining Customer's Evaluations of Multiple Service Failures and Recovery Efforts," *Journal of Marketing*, 66 (4), 57-71.
- \*McCullough, Michael A. (1995), "The Recovery Paradox: A Conceptual Model and Empirical Investigation of Customer Satisfaction and Service Quality Attitudes after Service Failure and Recovery," doctoral dissertation, Texas A&M University, College Station, United States.
- \*——— (2000), "The Effect of Perceived Justice and Attributions regarding Service Failure and Recovery on Post-Recovery Customer Satisfaction and Service Quality Attitudes," *Journal of Hospitality & Tourism Research*, 24 (4), 423-47.
- and Sundar G. Bharadwaj (1992), "The Recovery Paradox: An Examination of Consumer Satisfaction in Relation to Disconfirmation, Service Quality, and Attribution Based Theories," in *Marketing Theory and Applications*, Chris T. Allen et al., eds. Chicago: American Marketing Association, 119.
- \*———, Leonard L. Berry, and Manjit S. Yadav (2000), "An Empirical Investigation of Customer Satisfaction after Service Failure and Recovery," *Journal of Service Research*, 3 (2), 121-37.
- \*Michel, Stefan (2001), "Analyzing Service Failures and Recoveries: A Process Approach," *International Journal of Service Industry Management*, 12 (1), 20-33.

- \*—— and Matthew L. Meuter (2006), "The Service Recovery Paradox: True but Overrated?" working paper, [www.dienstleistungsmarketing.ch/documents/IJSIMParadoxMichelMeuter.pdf](http://www.dienstleistungsmarketing.ch/documents/IJSIMParadoxMichelMeuter.pdf).
- Moorman, Christine, Gerald Zaltman, and Rohit Deshpande (1992), "Relationships between Providers and Users of Market Research: The Dynamics of Trust within and between Organizations," *Journal of Marketing Research*, 29 (3), 314-39.
- Morgan, Robert and Shelby Hunt (1994), "The Commitment-Trust Theory of Marketing Relationships," *Journal of Marketing*, 58 (3), 20-38.
- Muncer, Steven, Mark Craigie, and Joni Holmes (2003), "Meta-Analysis and Power: Some Suggestions for the Use of Power in Research Synthesis," *Understanding Statistics*, 21 (1), 1-12.
- Nunnally, Jum and Ira Bernstein (1994), *Psychometric Theory*. New York: McGraw-Hill.
- \*Oh, Haemoon (2003), "Reexamining Recovery Paradox Effects and Impact Ranges of Service Failure and Recovery," *Journal of Hospitality & Tourism Research*, 27 (4), 402-18.
- \*Ok, Chihyung, Ki-Joon Back, and Carol W. Shanklin (2006), "Service Recovery Paradox: Implications from an Experimental Study in a Restaurant Setting," *Journal of Hospitality & Leisure Marketing*, 14 (3), 17-33.
- Oliver, Richard L. (1997), *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill.
- Palmatier, Robert W., Rajiv P. Dant, Dhruv Grewal, and Kenneth R. Evans (2006), "Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis," *Journal of Marketing*, 70 (4), 136-53.
- Pan, Yue and George M. Zinkhan (2006), "Determinants of Retail Patronage: A Meta-Analytical Perspective," *Journal of Retailing*, 82 (3), 229-43.
- Parasuraman, A. (2007), "Modeling Opportunities in Service Recovery and Customer-Managed Interactions," *Marketing Science*, 25 (6), 590-93.
- Reichheld, Frederick F. (1994), "Loyalty and the Renaissance of Marketing," *Marketing Management*, 2 (4), 10-21.
- Rosenthal, Robert (1979), "The 'File Drawer Problem' and Tolerance for Null Results," *Psychological Bulletin*, 86, 638-41.
- Singh, J. (1990), "Voice, Exit and Negative Word-of-Mouth Behaviors: An Investigation across Three Service Categories," *Journal of the Academy of Marketing Science*, 18 (1), 1-15.
- \*Smith, Amy K. and Ruth N. Bolton (1998), "An Experimental Investigation of Service Failure and Recovery: Paradox or Peril?" *Journal of Service Research*, 1 (1), 65-81.
- Szymanski, David M. and David H. Henard (2001), "Customer Satisfaction: A Meta-Analysis of the Empirical Evidence," *Journal of the Academy of Marketing Science*, 29 (1), 16-35.
- Tax, Stephen S., Stephen W. Brown, and Murali Chandrashekar (1998), "Customer Evaluations of Experiences: Implications for Relationship Marketing," *Journal of Marketing*, 62 (2), 60-76.
- Weun, Seungoo, Sharon E. Beatty, and Michael A. Jones (2004), "The Impact of Service Failure Severity on Service Recovery Evaluations and Post-Recovery Relationships," *Journal of Services Marketing*, 18 (2/3), 133-46.
- Westbrook, Robert A. (1980), "Intrapersonal Affective Influences upon Consumer Satisfaction with Products," *Journal of Consumer Research*, 7 (1), 49-54.
- \*Zeithaml, Valerie A., Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (2), 31-47.

**Celso Augusto de Matos** is a marketing doctoral candidate at the School of Management, Federal University of Rio Grande do Sul (PPGA-EA-UFRGS), Brazil. His main research interests lie in the areas of consumer behavior in services, attitude formation and change, and marketing research. His research has been published in the *Journal of Consumer Marketing*, *International Journal of Consumer Studies*, *ACR Conference*, and in a number of Brazilian journals and proceedings. He can be contacted at: [celsmatos@yahoo.com.br](mailto:celsmatos@yahoo.com.br).

**Jorge Luiz Henrique** is a marketing doctoral candidate at the School of Management, Federal University of Rio Grande do Sul (PPGA-EA-UFRGS), Brazil. His research focuses on consumer behavior in services and relationship marketing. He is a marketing manager at Banco do Brasil (Bank of Brazil). His research has been published in the *Journal of Internet Banking and Commerce*, *Global Information Technology Management (GITM)*, *The Business Association of Latin American Studies (BALAS)*, *International Association for Management of Technology (IAMOT)* conferences, and in a number of Brazilian journals and proceedings.

**Carlos Alberto Vargas Rossi** is a professor of marketing at the School of Management, Federal University of Rio Grande do Sul (PPGA-EA-UFRGS), Brazil. His research interests are consumer behavior and marketing theory. His research has been published in the *Journal of Consumer Marketing*, *International Journal of Consumer Studies*, *ACR*, *AMA* and *EMAC* conferences, and in a number of Brazilian journals and proceedings.