

# **Service employees' willingness to report complaints scale: Cross-country application and replication**

## **Abstract**

Although effective customer complaint management can be a key success factor for international service firms, relatively little is known about employees' proclivity to report complaints. The present study examines the meaningfulness and cross-national validity of the Israel-developed willingness to report complaints (WRC) scale. Recognized validation procedures, with samples of service employees from Germany and the U.S., demonstrate the reliability, validity and cross-national invariance of the WRC scale. This article offers implications for both research and service managers.

**Keywords:** complaint management, scale validation, service employees, willingness to report complaints

## 1. Introduction

Customer retention is a strategic imperative for service firms and service managers (Hart, Heskett, & Sasser, 1990). Therefore, high levels of service quality, a proper understanding of customer complaint behavior, and effective service recovery processes are critical to success (Liao & Chuang, 2004; Tax, Brown, & Chandrashekar, 1998). Inappropriate responses to customers' complaints and mismanaged service failure recoveries can foster customer dissatisfaction and result in the double deviation effect that ultimately drives customers away from the firm (Bitner, Booms, & Tetreault, 1990; Maxham & Netemeyer, 2002). Instead, complaint handling processes that reflect a comprehensive understanding of customer complaint behavior might bolster customers' perceptions of service quality, even after a service failure, and serve as important influences on customer satisfaction and loyalty (Homburg & Fürst, 2005; Liao & Chuang, 2004).

As service companies increasingly market their services internationally and enter global markets (La, Patterson, & Styles, 2005; Zhang, Beatty, & Walsh, 2008), service managers and scholars experience a greater need for sensitivity towards cultural differences and their effects on customers' and employees' behaviors and expectations toward complaining and service recovery (Lowe & Corkindale, 1998). Given that customers from different countries voice complaints and respond to firms' recovery activities in different ways (e.g., Chelminski & Coulter, 2007; Mattila & Patterson, 2004a), it is vitally important for international service firms to understand how service employees from varying countries or cultures deal with customer complaints, which is a key aspect of services (e.g., Orsingher, Valentini, & De Angelis, 2010).

Complaining behavior represents a specific type of customer behavior that is of interest to international marketers. In line with the demand for theoretical insights, a growing body of research investigates cultural differences in customer complaint behavior (e.g., Chan & Wan,

2008; Chelminski & Coulter, 2007; Hernandez, Strahle, Garcia, & Sorensen, 1991; Liu, Watkins, & Yi, 1997; Sharma, Marshall, Reday, & Woonbong, 2010; Wan, 2013). Cross-cultural studies also aim to detect effective service recovery and complaint handling designs that respect cultural differences (Hui & Au, 2001; Mattila & Patterson, 2004a, 2004b; Patterson, Cowley, & Prasongsukarn, 2006; Wong, 2004).

Because effective complaint handling can affect a service organization's bottom-line performance (Tax et al., 1998), firms seek out processes to effectively deal with complaints (e.g., Hart et al., 1990). An effective complaint handling process relies mainly on those closest to the customer—the service employees who receive and file complaints. Consequently, research reveals a growing interest in service employees' behavior in relation to customer complaints (Harris & Ogbonna, 2009). This focus is not particularly surprising, however, as service employees serve as gatekeepers in the complaint handling process (Liao, 2007; Liao & Chuang, 2004; Schneider & Bowen, 1985) and are usually the first contact for dissatisfied customers. For example, 45% of customers complain informally to service employees, but only 1–5% go directly to management (Goodman, 1999). Because most complaints are informal and submitted to service employees, service employees must report complaints to management for the complaint management process to begin (Schneider & Bowen, 1985; Zeithaml, Berry, & Parasuraman, 1988). However, empirical evidence suggests that service employees actually thwart organizational complaint management goals by distorting or failing to forward customer complaints (Homburg & Fürst, 2007).

To address this employee phenomenon, Luria, Gal, and Yagil (2009) introduce a new construct: service employees' willingness to report service complaints (WRC). WRC is defined as service employees' "discretion in terms of reporting or sharing information about clients' informal complaints" (Luria et al., 2009: 156). Despite garnering interest among researchers and

practitioners in services marketing, the WRC construct suffers from slow uptake. This slow diffusion might occur because the original scale development took place in an Israeli services context, which differs from most western services contexts with regard to how service employees and customers interact with each other (Rafaeli & Sutton, 1987). Gallois and Callan (1997: 86) suggest that customer-employee interactions differ across cultures by stating that all interactions between people are “governed by culture-specific social rules”. Thus, scholars might be reluctant to adopt the scale to measure WRC in other countries. The field needs a thorough reexamination of the scale and its measurement properties in another cultural context. Therefore, the present study assesses the robustness and validity of the WRC scale in the largest service economy in the world and the largest in the European Union; the U.S. and Germany. Hofstede (2001) advances a typology of principal cross-cultural values—power distance, individualism, masculinity, and uncertainty avoidance. Hofstede (2001) ranks Israel, Germany and the U.S. differently across these cultural dimensions. In addition, these three countries are geographically diverse and therefore, offer interesting comparisons.

By assessing the WRC scale in new cultural contexts, the present study responds to calls for more replication and extension studies in marketing and management research. For example, Albers (2012) asserts that articles in the marketing field often lack reproducible results and Evanschitzky, Baumgarth, Hubbard, and Armstrong (2007) advise academics to reserve judgment about published results in marketing and management journals and treat the findings with caution unless other studies replicate them. For measurement scales in particular, replication studies can assess psychometric properties and validity across different countries (Hassan, Shiu, & Walsh, 2011). Bruner (2003) even calls for action against the proliferation of scales that arises when replication studies do not confirm published scales.

Furthermore, this research is important from conceptual and practical perspectives. Conceptually, demonstrating the applicability of the WRC scale to different countries should prompt further research into the WRC construct and its correlates. Insight into the WRC scale's cross-cultural robustness is valuable because research shows that culture, the unstated standard operating procedures or ways of doing things (Triandis, 1994), is associated with work outcomes such as job satisfaction and organizational commitment (Diener, Oishi, & Lucas, 2003; Wasti, 2003). Therefore, a scale measuring organizational behavior does not necessarily work equally well across cultures. Practically, a valid WRC scale is useful for survey practitioners and enables service marketing researchers to use employee surveys as sources for empirical investigations of employees' WRC.

## **2. Background**

### *2.1. The WRC scale*

Luria et al. (2009) conducted three exploratory studies to examine the nature of WRC and identify organizational correlates. The first two studies relied on qualitative explorations and showed that employees can choose to report customer complaints to management or not. With a critical incidents technique, these authors asked service employees about customer complaints and expressions of dissatisfaction, as well as their own reporting behavior. More detailed interviews with 30 service employees and three service managers provided precise information about the key factors affecting service employees' WRC. Both studies showed that service employees consider several interacting factors when deciding whether and how to report customer complaints to management. Building on the results of their qualitative studies, Luria et al. (2009) developed a four-item WRC scale and examined the quantitative relationship of this WRC with service employees' organizational citizenship behavior (OCB), perceptions of the

service climate (SC), and perceived empowerment (EP). The results showed that the differences in service employees' willingness to report complains related to their OCB, SC, and EP.

The WRC scale achieved a Cronbach's alpha of .73, indicating adequate internal consistency. In a principle components factor analysis with Varimax rotation on the four WRC items, Luria et al. (2009) extracted a single factor. The factor loadings for the WRC items ranged from .33 to .85. To establish discriminant validity, Luria et al. (2009) also performed another principle component factor analysis with Varimax rotation including items from both the WRC and OCB scales. Two discrete factors emerged: one with all the OCB items and another with all the WRC items. These results strongly indicated that the WRC scale measured a one-dimensional, internally consistent construct.

## *2.2. Validation procedures*

Luria et al.'s (2009) results only include samples of Israeli employees, which may limit the generalizability of the WRC scale. A measurement instrument may have good psychometric properties in one cultural context but not in another. For example, in Israel, impersonal interactions differ from those in other countries based on Hofstede's (2001) dimensions. Frankness, for instance, is a unique characteristic of Israeli behavior such that Israelis generally are open and direct and freely disclose opinions and emotions (e.g., Mayselless & Salomon, 2003; Shamir & Melnik, 2002).

However, no studies confirm the cross-cultural validity of the WRC scale with Luria et al. (2009) calling for further research along these lines. To address this gap, the present study investigates whether the validity of the WRC scale, originally tested in Israel, applies to a German and U.S. context. For the cross-cultural validation, this study follows the procedure and criteria (internal, external, and judgmental) suggested by Nenkov, Morrin, Ward, Schwartz, and Hulland (2008). Internal criteria refer to the internal consistency and dimensionality of WRC.

External criteria are construct validity measures such as predictive and nomological validity. Judgmental validity relates to readability (Nenkov et al., 2008).

To address nomological validity, this study uses a structural model with WRC as the dependent variable and supervisor knowledge and role ambiguity as independent variables. Supervisors' knowledge indicates awareness of employees' behavior and outputs (Ramaswami, 1996). Role ambiguity describes a state without clear information or certainty about job responsibilities and expectations (Peterson et al., 1995; Zeithaml et al., 1988). The assessment of predictive validity involves considering a key employee outcome, job satisfaction, regressed on WRC. Spector (1985: 695) defines job satisfaction as an "emotional affective response to a job or specific aspects of a job," such as gratification. All characteristics of the job and work environment can be relevant for creating this emotion in an employee (Brown & Peterson, 1993). If the WRC measure has predictive and nomological validity, the correlation between the measure and other related constructs should behave as expected in theory (Churchill, 1995).

### **3. Method**

#### *3.1. Samples and questionnaire*

Among the Israeli respondents, the median age was 26 years, ranging from 19 to 61, and 56% of the respondents were women. The median job tenure was 1.9 years (see Table 1). Respondents completed a printed questionnaire. The WRC scale used a seven-point Likert agreement scale, from 1 ("strongly disagree") to 7 ("strongly agree").

An English version of the WRC scale was available from Luria et al.'s (2009) original article. The development of the German version of the WRC scale entailed translating and back-translating the four items of the original WRC scale (Luria et al., 2009). A native English-speaker, raised bilingually in Germany, took primary responsibility for the translation process.

Thirty-two German participants then judged the comprehensibility of the four items in the WRC scale following Nenkov et al.'s (2008) rules for establishing judgmental validity.

The German and U.S. questionnaires also contained multi-item measures for external validation. The measure of willingness to report complaints used four items (Luria et al., 2009), role ambiguity featured three items (adapted from Kohli & Jaworski, 1994), and supervisor knowledge involved a five-item measure (adapted from Ramaswami, 1996). The job satisfaction measure aggregated two items (adapted from Rich, 1997) and served as the dependent variable in the regression model (see Appendix A). All variables used seven-point Likert scales, ranging from 1 ("strongly agree") to 7 ("strongly disagree").

Both samples consisted of service employees who engaged in face-to-face customer contacts. In both countries, research assistants recruited service employees by distributing links to an online questionnaire. Each assistant received a brief description of the study's purpose and cover letters to give to respondents, which explained the nature of the study and provided instructions for the data collection procedures. The recruitment process relied on a snowballing technique, such that the research assistants recruited employees or relatives who worked in services (Salganik & Heckathorn, 2004), then enlisted the help of these service employees to recruit at least two other personally known service employees. A key advantage of online data collection is the avoidance of interviewer effects (Duffy, Smith, Terhanian, & Bremer, 2005).

Finally, participants provided demographic information, such as age, gender, and job tenure. The data collection in Germany lasted one month and, in the U.S., five weeks. In total, 232 German and 321 U.S. participants completed the questionnaire (see Table 1). In Germany, service employees' average age was 29 years ( $SD = 9.80$ ), and 60% ( $n = 139$ ) were women. The majority of respondents had very frequent face-to-face interactions with customers. The median job tenure was four years. To assess the potential for non-response bias, this study compared



early (weeks 1 and 2) and late (week 3 and 4) respondents on the primary variable, WRC (Armstrong & Overton, 1977). The *t*-test of the group means for the aggregated WRC measure revealed no significant difference, so non-response bias was not a notable problem for this study.

The median age of the American service employees was 24 years ( $SD = 12.97$ ), and 53% ( $n = 168$ ) were women. Overall, the majority of respondents had frequent (70%) contact or face-to-face interactions with customers as part of their daily routine. The median job tenure was 5.45 years.

Table 1 contains descriptions of the Israeli, German and U.S. samples, the frequency of personal contact for a service employee during a regular day, and median years of job tenure. Compared to the Israeli sample, the German sample contains more female and older participants, as well as employees with longer job tenures. The U.S. sample contains more male and younger participants than the Israeli or German sample, and employees with longer job tenures than the other samples.

*-- Table 1 here --*

## **4. Results**

### *4.1. Scale reliability and validity – Germany*

To assess the internal consistency of the WRC scale, the present study used a confirmatory factor analysis (CFA). The results of the measurement model indicated acceptable model fit (root mean square error of approximation [RMSEA] = .09; comparative fit index [CFI] = .98; goodness-of-fit index [GFI] = .98;  $\chi^2(2) = 5.8$  [ $\chi^2/df = 2.9$ ;  $p < .05$ ]). The average variance extracted (AVE) was .5, which is in line with the recommended threshold (Bagozzi & Yi, 1988; Fornell & Larcker, 1981), and composite reliability reached .78. The item-to-scale correlations showed that the first two items were responsible for the relatively low AVE (see Appendix A),

with CFA loadings of .57 for item 2 and .65 for item 1. The other two items showed good CFA loadings, namely, .79 for item 3 and .80 for item 4.

To embed WRC in a nomological network, this study also assessed the relationship with two predictor variables, supervisor knowledge and role ambiguity. The relevant indices pertaining to the structural model showed good model fit (RMSEA = .04; CFI = .99; GFI = .95;  $\chi^2(50) = 69.76$  [ $\chi^2/df = 1.4$ ;  $p < .05$ ]). The standardized path coefficients revealed that supervisor knowledge (.30) and role ambiguity (-.46) both significantly predicted WRC.

The assessment of predictive validity, by regressing job satisfaction on WRC, when job satisfaction represented an aggregation of two items ( $r = .69$ ,  $p < .001$ ), indicated that WRC significantly predicted job satisfaction ( $\beta = .37$ ;  $R^2 = .13$ ,  $p < .001$ ). That is, the findings offered support for the predictive validity of the WRC scale.

To further assess construct validity, a measurement model featured WRC and the related organizational variables: supervisor knowledge, role ambiguity, and job satisfaction. The CFA on the four-factor solution revealed a good fit (RMSEA = .051; CFI = .97; GFI = .94;  $\chi^2(70)$  of 113.39 [ $\chi^2/df = 1.62$ ;  $p < .001$ ]). The correlations of the latent variables, which appear in Table 2, indicated convergent validity. Furthermore, all of the constructs evidenced discriminant validity, according to Fornell and Larcker's (1981) test, because all squared correlations between WRC and the other constructs were lower than .5, the AVE of WRC.

-- Table 2 here --

The judgmental measure of item quality was readability (Nenkov et al., 2008). Thus, a sample of 32 service employees (65% women) indicated the ease of understanding of each item from the WRC scale (seven-point scale, anchored by "not very easy to understand" [1] and "very easy to understand" [7]). Table 3 gives an overview of these readability ratings. Three items (1, 3, and 4) achieved means greater than 6, indicating very good readability. The second WRC item

showed poorer readability, though the mean close to 5 indicated that this item still appeared easy to understand.

Luria et al. (2009) did not conduct a CFA, so a comparison with the CFA results across Germany and Israel was not possible. Instead, to attain comparability with the original results, the present study included an exploratory factor analysis (EFA). The Kaiser-Meyer-Olkin value of .76 indicated the good factorability of the data. Principal component analysis revealed a single factor that explained 60% of the variance. Table 3 contains a comparison of the EFA results between Luria et al.'s (2009) and the present study. Luria et al. (2009) reported a Cronbach's alpha of .73, similar to the alpha calculated herein (.77). However, the factor loadings differed. Notably, item 1 shows a high factor loading in the present study but a very poor loading in the reference study by Luria et al. (2009).

To further explore the WRC scale, groups with very strong (extreme) or very low occurrences of certain characteristics underwent a comparison (Tian, Bearden, & Hunter, 2001). Specifically, the group splits depended on the demographic variables that enabled respondents to self-assign to different groups (see Table 1). A one-way between-subjects analysis of variance compared the effects of age, gender, frequency of personal contact, and job tenure on service employees' willingness to report complaints, but no significant main effects emerged. Younger ( $M = 2.1$ ,  $SD = .11$ ) and older ( $M = 2.13$ ,  $SD = .12$ ) employees did not differ in their willingness to report complaints ( $F(1, 235) = .034$ ,  $p > .05$ ), nor did gender have any significant effect (men  $M = 2.14$ ,  $SD = .13$ ; women  $M = 2.1$ ,  $SD = .09$ ;  $F(1, 235) = .075$ ,  $p > .05$ ). Service employees with moderate ( $M = 2.07$ ,  $SD = .09$ ) or high ( $M = 2.17$ ,  $SD = .13$ ) customer contact levels also did not exhibit significant differences ( $F(1, 235) = .382$ ,  $p > .05$ ). Finally, the results revealed no significant difference between short-tenured ( $M = 1.99$ ,  $SD = .11$ ) and long-tenured

( $M = 2.25$ ,  $SD = 0.12$ ) service employees in willingness to report complaints ( $F(1, 235) = 2.604$ ,  $p > .05$ ). None of the interaction effects was significant (Appendix B).

#### 4.2. Scale reliability and validity – U.S.

As with the German data, a CFA was used to assess the internal consistency of the WRC scale for the U.S. data. The results of the model indicate good model fit (RMSEA = .02; CFI = 1.00; GFI = .99;  $\chi^2(2) = 2.28$  [ $\chi^2/df = 1.14$ ;  $p = .3$ ]). The AVE was .66 which is in line with the recommended threshold (Bagozzi & Yi, 1998; Fornell & Larcker, 1981). Composite reliability was .88. All items demonstrated good standardized loadings: .79 for item 1, .74 for item 2, .86 for item 3, and .85 for item 4.

Nomological validity was assessed by relating WRC to two predictor variables, supervisor knowledge and role ambiguity. The structural model showed good fit (RMSEA = .05; CFI = .99; GFI = .95;  $\chi^2(49) = 92.76$  [ $\chi^2/df = 1.89$ ;  $p < .05$ ]). The standardized path coefficients further revealed that supervisor knowledge (.25) and role ambiguity (-.56) both significantly predicted WRC.

Predictive validity of the WRC scale for the U.S data was assessed by regressing job satisfaction (represented by aggregation of two items) on WRC ( $r = .46$ ,  $p < .001$ ). This indicated that WRC significantly predicted job satisfaction ( $\beta = .46$ ;  $R^2 = .21$ ,  $p < .001$ ) which offered support for the predictive validity of the WRC scale.

Construct validity was assessed by analyzing a measurement model that featured WRC and the other three related variables: supervisor knowledge, role ambiguity, and job satisfaction. The CFA on the four-factor solution indicated good fit (RMSEA = .05; CFI = .99; GFI = .95;  $\chi^2(69) = 129.05$  [ $\chi^2/df = 1.87$ ;  $p < .05$ ]). The correlations of the latent variables (see Table 2) indicated convergent validity. In addition, all of the constructs demonstrated discriminant validity using the

Fornell and Larcker (1981) test because all the squared correlations between WRC and the three other constructs were lower than the AVE of WRC, .66.

Finally to compare the U.S. to the Israeli data used in the Luria et al. (2009) study, an exploratory factor analysis (EFA) was conducted. The Kaiser-Meyer-Olkin value of .83 indicated good factorability of the data. Furthermore, a principal component analysis demonstrated a one-factor solution that explained 75% of the variance (see Table 3). Cronbach's alpha for the U.S. data was .88, which is higher than the .73 reported by Luria et al. (2009). Further, the loadings differ as well with item 1's factor loading being higher in the U.S. data.

Overall, the results of the WRC scale assessment in a German and U.S. setting provide a robust defense of the four-item WRC scale. The scale exhibits good psychometric properties, with high levels of reliability and validity in both Germany and the U.S. Thus, the WRC scale has good potential for use in service settings across different cultures.

-- Table 3 here --

#### *4.3. Cross-cultural measurement invariance*

A primary goal of this research was to develop a scale that was applicable across countries. Therefore, testing the measurement invariance of the scale was a necessary final step. Measurement invariance was assessed for the two data sets (Germany and the U.S.) collected in this study by examining the data for similar patterns of factor loadings (configural invariance), for equality of factor loadings (metric invariance), and for equality of intercept terms (scalar invariance).

Using Lisrel 8.8 and following the measurement invariance testing procedure of Milfont and Fisher (2010), a multi-group CFA with the two countries representing the groups and WRC as a reflective, one-dimensional construct was conducted. Overall, the model fit was good (RMSEA = .06; CFI = 1.00; NNFI = .99;  $\chi^2(4) = 7.76$ ; see Table 4: Baseline model A). The

analysis supported configural invariance indicating that the same basic factor structure held for both groups and all four items loaded substantially on the target factor. To test for metric invariance, a constraint was added that forced the factor loadings to be equal across the two groups. Again the model showed good fit (RMSEA = .03; CFI = .99; NNFI = .99;  $\chi^2(7) = 8.96$ ; see Table 4: Model B). The increase in  $\chi^2$  was insignificant ( $p = .75$ ) indicating equality of the factor loadings and full measurement invariance. Finally, to test for scalar invariance, a constraint was added that forced the intercepts to be the same across the two groups. The model showed good fit (RMSEA = .04; CFI = .99, NNFI = .99;  $\chi^2(10) = 14.53$ ; see Table 4: Model C). The increase in  $\chi^2$  from the baseline model is insignificant ( $p = .34$ ) indicating that equality of the intercepts across the two groups. Thus scalar invariance for this scale was supported.

-- Table 4 here --

## 5. Discussion

This study aimed to test the reliability and validity of the WRC scale in cultural settings other than Israel. Using three criteria, this study demonstrates the robustness of the WRC scale. The findings provide support for the usefulness of the WRC scale while also highlighting important differences that must be considered when using the scale.

Following Nenkov et al. (2008), this study relied on three criteria (internal, external, and judgmental) to evaluate the WRC scale. No single criterion can separately judge the quality of scales; the three criteria should appear in conjunction, and items ideally should meet all three criteria. In terms of internal criteria, the composite reliability of .78 in Germany and .92 in the U.S. exceeds the recommended threshold level, and the AVE (.5) is in line with the recommended threshold (Bagozzi & Yi, 1988; Fornell & Larcker, 1981). All CFA loadings are above .50, ranging from .57 to .80 (Germany) and .74 to .86 (U.S.). However, one item appears

less effective in the German context, namely, the second WRC item: “I report to my direct manager about problems customers encounter even if a customer has not told me that he or she wants to complain about service.” This item implies employee proactivity, which is an important work-culture dimension (e.g., Thomas, Whitman, & Viswesvaran, 2010). The hierarchical structure of many German organizations and the compartmentalization of jobs may suppress such proactive behaviors, rendering the second item less effective.

To assess construct validity, this study embedded WRC into a nomological network of variables that are relevant for service processes. Supervisor knowledge and role ambiguity affected WRC significantly. In support of predictive validity, the path coefficient for job satisfaction remained significant when regressed on the WRC scale. The assessment of judgmental quality relied on an examination of the readability of the WRC scale items.

Although a direct comparison of the Israeli, German and U.S. samples was difficult because of the lack of reported CFA loadings in Luria et al.’s (2009) article, this study offers an assessment of the applicability of the scale in Germany and the U.S. Finally, the differences in factor loadings across scales (Table 3) may reflect sample specificities, different service contexts, or cultural variance, as reflected in Hofstede’s (2001) cultural dimensions. Overall, differences may emerge in the application and replication of the WRC scale, but the findings, in agreement with Nenkov et al.’s (2008) suggested criteria, show that the WRC scale is a viable instrument to measure service employees’ willingness to report customer complaints.

### *5.1. Theoretical implications, limitations, and further research*

Complaint management research continues to develop steadily, and support for the link between effective complaint handling and firm performance remains clear and robust. The present study assesses a scale with the potential to improve the effectiveness of complaint handling processes. Despite offering deeper insights into the cross-cultural validation of WRC,

this study is not free of limitations, which suggests avenues for further research. As Luria et al. (2009) note, WRC scale generalizability and application demand cross-cultural validation, which also could increase uptake of the scale outside Israel.

Because scale validation is a continuous process (Baumgartner & Homburg, 1996; Churchill, 1979), additional research should assess the cross-cultural validity of the WRC scale further. One limitation of this study is the convenience samples used. Future research should thus be based on more national samples versus convenience samples. Furthermore, studies could test the scale with larger samples and data from different contexts, collecting non-cross-sectional data. Cross-sectional data ensure representations of different types of service occupations but also prevent measures of true causality. Research also could consider additional variables in the nomological network, such as manager-related (i.e., leadership) variables and the effects on WRC, or additional correlates, such as OCB, job motivation, or workplace deviance. Weaving WRC into a wider nomological net could improve understanding of the mechanism by which WRC affects employee- and customer-related outcomes.

From a practitioner perspective, considering organizational antecedents is relevant as a means to foster employees' inclination to deal with complaints effectively. For example, organizational incentives influence employee behavior, including increasing employees' willingness to share knowledge with coworkers and supervisors, and thus, may drive WRC. Several studies focus on increased knowledge sharing prompted by incentives (e.g., Foss, Laursen, & Pedersen, 2011; Hansen, Mors, & Lovas, 2005). Future research could examine organizational incentives in relation to WRC.

Further research may also be warranted to investigate the relationship between WRC and customer outcomes. High levels of WRC should increase customer complaint satisfaction, although perhaps not in all countries. Laroche et al. (2004) report that East Asian customers



express lower ratings of quality perceptions under a superior service condition than their North Americans counterparts. This may suggest that in some Asian cultures even high WRC levels would not necessarily lead to high customer complaint satisfaction. Using dyadic data (e.g., Homburg & Fürst, 2005), cross-cultural studies could consider the complaint process from the customer perspective and determine if low levels of WRC in a service organization bring about negative customer outcomes. For example, service failures coupled with poor recovery efforts can prompt customer retaliation behavior, such as negative word of mouth and reduced patronage (Grégoire & Fisher, 2006, 2008). An added benefit of using dyadic data would be the elimination of potential effects of response contamination, which can be introduced through the use of same-source data.

Further research avenues should develop WRC-related insights, particularly related to the mediators and moderators of antecedent–WRC and WRC–outcome relationships. For example, employee age might moderate the relationship between organizational incentives and WRC, because older service employees tend to maximize positive and minimize negative emotional work experiences (Dahling & Perez, 2010), which implies a lower likelihood of reporting complaints that detrimentally affect emotional states. Replication studies could incorporate mediators and moderators to address these issues, as well as focus on various cultures to confirm the WRC scale’s cross-cultural suitability. A good starting point for further validation would be countries experiencing significant growth in service sectors, such as India and China (Javalgi, Gross, Joseph, & Granot, 2011). Measuring WRC might provide firms in these countries with a means to detect service failures and increase service recovery effectiveness. To extend the present study in Germany or the U.S., researchers also might apply the WRC scale to different service contexts.

## *5.2. Managerial implications*

The findings show that the WRC scale, as originally developed in Israel, is highly relevant for the management of service firms. The scale helps measure service employees' willingness to report or not report complaints to supervisors. In addition, the WRC scale relates to several important performance-related correlates.

Considering the emphasis that customers place on complaint handling (e.g., Gruber, Szmigin, & Voss, 2009), managers must recognize the reasons for and effects of suboptimal service complaint handling. In this sense, WRC is vital to complaint management. German and U.S. managers should make more use of the WRC scale to identify potential shortcomings in the service recovery process. If that process does not work efficiently, a service firm might lose customers and suffer revenue losses. Accordingly, this study shows that using the WRC scale makes good managerial sense. Service firms might embed the scale in employee performance surveys to measure WRC across departments, branches, or international subsidiaries.

Service firms with operations in multiple countries may benefit most from a cross-culturally robust WRC scale. For such firms, the WRC scale appears well-suited to gather benchmark data regarding levels of employees' willingness to report complaints, as well as to conduct periodic comparisons between countries and across company branches. Application of the WRC scale may reveal necessary adaptations of complaint management and service recovery processes to different national conditions. The present study advocates further research on the cross-cultural applicability of the scale.

Finally, a positive relationship appears between WRC and job satisfaction; a positive atmosphere in a service organization likely results in higher job satisfaction among service employees and thus better performance in complaint handling, which in turn enhances the firm's revenues. In addition, WRC relates positively to supervisor knowledge but negatively to service employees' role ambiguity. Service firms should tackle role ambiguity by investing in employee

training or providing clear service scripts, to help guide employee behavior, minimize variance in service outputs, and increase WRC. Because supervisor knowledge also relates positively to WRC, service organizations should invest in supervisor training and gain returns in terms of more reported (and better handled) complaints.

## References

- Albers, S. (2012). Optimizable and implementable aggregate response modeling for marketing decision support. *International Journal of Research in Marketing*, 29(2), 111–122.
- Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(3), 396–402.
- Bagozzi, R. P., & Yi, T. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research. *International Journal of Research in Marketing*, 13(2), 139–161.
- Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing*, 54(1), 71–84.
- Brown, S. P., & Peterson, R. A. (1993). Antecedents and consequences of salesperson job satisfaction: Meta-analysis and assessment of causal effects. *Journal of Marketing Research*, 30(1), 63–77.
- Bruner, G. C. II. (2003). Combating scale proliferation. *Journal of Targeting, Measurement and Analysis for Marketing*, 11(4), 362–372.
- Chan, H., & Wan, L. C. (2008). Consumer responses to service failures: A resource preference model of cultural influences. *Journal of International Marketing*, 16(1), 72–97.
- Chelminski, P., & Coulter, R. A. (2007). The effects of cultural individualism and self-confidence on propensity to voice: From theory to measurement to practice. *Journal of International Marketing*, 15(4), 94–118.
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64–73.
- Churchill, G. A. (1995). *Marketing research: Methodological foundations* (6<sup>th</sup> ed.). Chicago: Dryden.
- Dahling, J. J., & Perez, L. A. (2010). Older worker, different actor? Linking age and emotional labor strategies. *Personality and Individual Differences*, 48(5), 574–578.
- Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, 54, 403–425.

- Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*, 47(6), 615–639.
- Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60(4), 411–415.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Foss, N. J., Laursen, K., & Pedersen, T. (2011). Linking customer interaction and innovation: The mediating role of new organizational practices. *Organization Science*, 22(4), 980–999.
- Gallois, C., & Callan, V. J. (1997). *Communication and culture: A guide for practice*. Wiley series in culture and professional practice. Hoboken: John Wiley & Sons Inc.
- Goodman, J. (1999). Basic facts on customer complaint behavior and the impact of service on the bottom line. *Competitive Advantage*, 8(1), 1–5.
- Grégoire, Y., & Fisher, R. J. (2006). The effects of relationship quality on customer retaliation. *Marketing Letters*, 17(1), 31–46.
- Grégoire, Y., & Fisher, R. J. (2008). Customer betrayal and retaliation: When your best customers become your worst enemies. *Journal of the Academy of Marketing Science*, 36(2), 247–261.
- Gruber, T., Szmigin, I., & Voss, R. (2009). Handling customer complaints effectively: A comparison of the value maps of female and male complainants. *Managing Service Quality*, 19(6), 636–656.
- Hansen, M. T., Mors, M. L., & Lovas, B. (2005). Knowledge sharing in organizations: Multiple networks, multiple phases. *Academy of Management Journal*, 48(5), 776–793.
- Harris, L.C., & Ogbonna, E. (2009). Hiding customer complaints: Studying the motivations and forms of service employees' complaint concealment behaviours. *British Journal of Management*, 21(2), 262–279.
- Hassan, L., Shiu, E., & Walsh, G. (2011). An assessment of the long-term orientation scale: A replication and extension based on findings across 10 European countries. *International Marketing Review*, 28(1), 81–101.
- Hart, C. W. L., Heskett, J. L., & Sasser, W. E. (1990). The profitable art of service recovery. *Harvard Business Review*, 68(4), 148–156.

- Hernandez, S. A., Strahle, W., Garcia, H. L., & Sorensen, R. C. (1991). A cross-cultural study of consumer complaining behavior: VCR owners in the U.S. and Puerto Rico. *Journal of Consumer Policy, 14*(1), 35–62.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organization across nations*. Thousand Oaks: Sage.
- Homburg, C., & Fürst, A. (2005). How organizational complaint handling drives customer loyalty: An analysis of the mechanistic and the organic approach. *Journal of Marketing, 69*(3), 95–114.
- Hui, M. K., & Au, K. (2001). Justice perceptions of complaint-handling: A cross-cultural comparison between PRC and Canadian customers. *Journal of Business Research, 52*(2), 161–173.
- Javalgi, R. G., Gross, A. A., Joseph, W. B., & Granot, E. (2011). Assessing competitive advantage of emerging markets in knowledge intensive business services. *Journal of Business & Industrial Marketing, 26*(3), 171–180.
- Kohli, A. K., & Jaworski, B. J. (1994). The influence of coworker feedback on salespeople. *Journal of Marketing, 58*(4), 82–94.
- La, V. Q., Patterson, P. G., & Styles, C. W. (2005). Determinants of export performance across service types: A conceptual model. *Journal of Services Marketing, 19*(6), 379–391.
- Laroche, M., Ueltschy, L. C., Abe, S., Cleveland, M., & Yannopoulos, P. P. (2004). Service quality perceptions and customer satisfaction: Evaluating the role of culture. *Journal of International Marketing, 12*(3), 58–85.
- Liao, H. (2007). Do it right this time: The role of employee service recovery performance in customer-perceived justice and customer loyalty after service failures. *Journal of Applied Psychology, 92*(2), 475–489.
- Liao, H., & Chuang, A. (2004). A multilevel investigation of factors influencing employee service performance and customer outcomes. *Academy of Management Journal, 47*(1), 41–58.
- Liu, R. R., Watkins, H. S., & Yi, Y. (1997). Taxonomy of consumer complaint behavior: Replication and extension. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 10*(1), 91–103.

- Lowe, A. C.-T., & Corkindale, D. R. (1998). Differences in “cultural values” and their effects on responses to marketing stimuli: A cross-cultural study between Australians and Chinese from the People’s Republic of China. *European Journal of Marketing*, 32(9), 843–867.
- Luria, G., Gal, I., & Yagil, D. (2009). Employees’ willingness to report service complaints. *Journal of Service Research*, 12(2), 156–174.
- Mattila, A. S., & Patterson, P. G. (2004a). The impact of culture on consumers’ perceptions of service recovery efforts. *Journal of Retailing*, 80(3), 196–206.
- Mattila, A. S., & Patterson, P. G. (2004b). Service recovery and fairness perceptions in collectivist and individualist contexts. *Journal of Service Research*, 6(4), 336–346.
- Maxham, J. G., & Netemeyer, R. G. (2002). A longitudinal study of complaining customers’ evaluations of multiple service failures and recovery efforts. *Journal of Marketing*, 66(4), 57–71.
- Mayseless, O., & Solomon, G. (2003). Dialectic contradictions in the experiences of Israeli Jewish adolescents. Efficacy and stress, closeness and friction, and conformity and noncompliance. In F. Pajares (Ed.), *International Perspective on Adolescents* (pp.151–173). Charlotte: Information Age Publishing.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. Supplementary material. *International Journal of Psychological Research*, 3(1), 122–130.
- Nenkov, G. Y., Morrin, M., Ward, A., Schwartz, B., & Hulland, J. (2008). A short form of the maximization scale: Factor structure, reliability and validity studies. *Judgment and Decision Making*, 3(5), 371–388.
- Orsingher, C., Valentini, S., & De Angelis, M. (2010). A meta-analysis of satisfaction with complaint handling in services. *Journal of the Academy of Marketing Science*, 38(2), 169–186.
- Patterson, P. G., Cowley, E., & Prasongsukarn, K. (2006). Service failure recovery: The moderating impact of individual-level cultural value orientation on perceptions of justice. *International Journal of Research in Marketing*, 23(3), 263–277.
- Peterson, M. F., Smith, P. B., Akande, A., Ayestaran, S., Bochner, S., & Callan, V. et al. (1995). Role conflict, ambiguity, and overload: A 21-nation study. *Academy of Management Journal*, 38(2), 429–445.

- Rafaeli, A., & Robert, S. (1987). Expression of emotion as part of the work role. *Academy of Management Review*, 12(1), 23–37.
- Ramaswami, S. N. (1996). Marketing controls and dysfunctional employee behaviors: A test of traditional and contingency theory postulates. *Journal of Marketing*, 60(2), 105–120.
- Rich, G. A. (1997). The sales manager as a role model: Effects on trust, job satisfaction, and performance on salespeople. *Journal of the Academy of Marketing Science*, 25(4), 319–328.
- Salganik, M. J., & Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-drive sampling. *Sociological Methodology*, 34(1), 193–239.
- Shamir, B., & Melnik, Y. (2002). Boundary permeability as a cultural dimension. A study of cross-cultural working relations between Americans and Israelis in high-tech organizations. *International Journal of Cross Cultural Management*, 2(2), 219–238.
- Sharma, P., Marshall, R., Reday, P. A., & Woonbong, N. (2010). Complainers versus non-complainers: A multi-national investigation of individual and situational influences on customer complaint behaviour. *Journal of Marketing Management*, 26(1-2), 163–180.
- Schneider, B., & Bowen, D. E. (1985). Employee and customer perceptions of service in banks replication and extension. *Journal of Applied Psychology*, 70(3), 423–433.
- Spector, P. E. (1985). Measurement of human service staff satisfaction: Development of the job satisfaction survey. *American Journal of Community Psychology*, 13(6), 693–713.
- Tax, S. S., Brown, S. W., & Chandrashekar, M. (1998). Customer evaluations of service complaint experiences: Implications for relationship marketing. *Journal of Marketing*, 62(2), 60–76.
- Thomas, J. P., Whitman, D. S., & Viswesvaran, C. (2010). Employee proactivity in organizations: A comparative meta-analysis of emergent proactive constructs. *Journal of Occupational and Organizational Psychology*, 83(2), 275–300.
- Tian, K. T., Bearden, W. O., & Hunter, G. L. (2001). Consumer's need for uniqueness: Scale development and validation. *Journal of Consumer Research*, 28(1), 50–66.
- Triandis, H. C. (1994). *Culture and social behavior*. New York: McGraw-Hill.
- U.S. Central Intelligence Agency (2012). The World Factbook 2001 [WWW page]. URL. <http://www.cia.gov/cia/publications/factbook/index.html>.
- Wan, L. C. (2013). Culture's impact on consumer complaining responses to embarrassing service failure. *Journal of Business Research*, 66(3), 298–305.



- Wasti, S. A. (2003). The influence of cultural values on antecedents of organizational commitment: An individual-level analysis. *Applied Psychology, 52*(4), 533–554.
- Witkowski, T. H., & Wolfinbarger, M. F. (2002). Comparative service quality: German and American ratings across service settings. *Journal of Business Research, 55*(11), 875–881.
- Wong, N. Y. (2004). The role of culture in the perception of service recovery. *Journal of Business Research, 57*(9), 957–963.
- Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1988). Communication and control processes in the delivery of service quality. *Journal of Marketing, 52*(2), 35–48.
- Zhang, J., Beatty, S. E., & Walsh, G. (2008). Review and future directions of cross-cultural consumer services research. *Journal of Business Research, 61*(3), 211–224.

**Table 1**

## Sample descriptions

Characteristic		Israeli sample (n = 180)	German sample (n = 232)	U.S. sample (n = 321)
Age	Median	26	29	24
Gender	Male	44%	40%	47%
	Female	56%	60%	53%
Frequency of personal contact* (median percentage)		n.a.	80%	70%
Job tenure in years (median)		1.9	4.1	5.45

\*Respondents responded to the following question: “How often do you interact face-to-face with customers?” (0% to 100% of work time).

**Table 2**

## Correlations

		1	2	3	4
1	WRC	<i>.78/.88</i>	.45	-.62**	.46**
2	Supervisor knowledge	.51**	<i>.91/.92</i>	-.46**	.64**
3	Role ambiguity	-.59**	-.46**	<i>.74/.95</i>	-.46**
4	Job satisfaction	.46**	.47**	-.37**	<i>.75/.94</i>

\*\*  $p < .001$ .

Note: The italicized values on the diagonal are composite reliabilities (values below the diagonal and before the slash pertain to the German sample, values above the diagonal and after the slash to the US sample).

**Table 3**

EFA factor loadings from Luria et al. (2009) and current studies with readability ratings

<i>Items of willingness to report service complaints</i>	Luria et al. (2009)*	German study (n = 232)	Readability ratings (SD) (n = 32)**	U.S. study (n = 321)
Cronbach's alpha	.73	.77		.88
Item 1: I report to management about incidents in which customers complain about serious problems.	.33	.76	6.06 (1.22)	.85
Item 2: I report to my direct manager about problems customers encounter even if a customer has not told me that he or she wants to complain about service.	.68	.71	4.97 (1.2)	.83
Item 3: I feel comfortable discussing problems encountered with a dissatisfied customer with my direct manager.	.82	.79	6.34 (.97)	.89
Item 4: I am willing to tell my direct manager about difficulties I had when serving customers.	.85	.83	6.13 (1)	.88

\*Dr. Gil Luria kindly provided the EFA factor loadings: \*\*based on sample of German consumers.

**Table 4**

Cross-cultural measurement invariance

	<b>Model <math>\chi^2</math> (df)</b>	<b><math>\Delta\chi^2</math> (<math>\Delta df</math>)</b>	<b>RMSEA</b>	<b>CFI</b>	<b>NNFI</b>	<b>Equality supported</b>
(A) Baseline multigroup model (Configural Invariance)	7.76 (4)	-	.06	1.00	.99	-
(B) Metric Invariance	8.96 (7)	1.20 (3)	.03	.99	.99	yes
(C) Scalar Invariance	14.53 (10)	6.77 (6)	.04	.99	.99	yes

## Appendix A

### Items and results of confirmatory factor analysis

	Factor loadings Germany (CFA)	Factor loadings U.S. (CFA)	AVE*
Willingness to report complaints (Luria et al., 2009)			.50/.66
I report to management about incidents in which customers complain about serious problems.	.65	.79	
I report to my direct manager about problems customers encounter even if a customer has not told me that he or she wants to complain about service.	.57	.74	
I feel comfortable discussing problems encountered with a dissatisfied customer with my direct manager.	.79	.86	
I am willing to tell my direct manager about difficulties I had when serving customers.	.80	.85	
Role ambiguity (adapted from Kohli and Jaworski, 1994)			.51/.87
I know exactly what is expected of me (R).	.37	.89	
I know what my responsibilities are (R).	.83	.97	
I know the scope of my job (R).	.84	.93	
Supervisor knowledge (adapted from Ramaswami, 1996)			.66/.69
My supervisor knows how to accomplish the work I normally encounter.	.81	.70	
My supervisor is intimately familiar with the day-to-day decisions related to my work.	.77	.87	
My supervisor has developed an excellent working knowledge of my job.	.86	.86	
I am confident that my supervisor can assess my job performance.	.77	.87	
My supervisor can specify the most important variables to monitor my work.	.84	.84	
Job satisfaction (adapted from Rich, 1997)			.59/.81
All in all, I am satisfied with my job.	.79	.87	
In general, I like working for my company.	.76	.93	

Notes: N = 236. All factor loadings are significant at  $p < .001$ . R = reverse item; CFA = confirmatory factor analysis; AVE = average variance extracted.

\*(values before the slash pertain to the German sample, values after the slash to the US sample).

**Appendix B**

Analysis of variance results (based on German data)

Dependent variable: WRC	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10.848*	15	.723	.761	.720
Intercept	658.776	1	658.776	693.330	.000
Age	.032	1	.032	.034	.854
Gender	.071	1	.071	.075	.785
Frequency of personal contact	.363	1	.363	.382	.537
Job tenure in years	2.474	1	2.474	2.604	.108
Error	209.036	220	.950		
Total	1258.125	236			
Corrected Total	219.883	235			

\*R-square = .049