Vlerick Leuven Gent Management School
the Autonomous Management School of Ghent University and Katholieke Universiteit Leuven

Vlerick Leuven Gent Working Paper Series 2006/08

(R)E-TAIL SATISFACTION: RETAIL CUSTOMER SATISFACTION IN ONLINE AND OFFLINE CONTEXTS

BERT WEIJTERS

Bert.Weijters@vlerick.be

NIELS SCHILLEWAERT

Niels.Schillewaert@vlerick.be

D/2006/6482/8

(R)E-TAIL SATISFACTION: RETAIL CUSTOMER SATISFACTION IN ONLINE AND OFFLINE CONTEXTS

BERT WEIJTERS

Vlerick Leuven Gent Management School

NIELS SCHILLEWAERT

Vlerick Leuven Gent Management School

Acknowledgements: The authors would like to thank Michael Brady for his helpful feedback on a previous version of this paper. Bert Weijters also would like to thank the Intercollegiate Centre for Management Studies (ICM Belgium) for funding his research.

Contact:

Bert Weijters

Vlerick Leuven Gent Management School

Tel: +32 09 210 98 76

Fax: +32 09 210 98 75

Email: Bert.Weijters@vlerick.be

ABSTRACT

Building on the e-Satisfaction model proposed by Szymanski and Hise (2000) and further validated by Evanschitzky, Iyer, Hesse, and Ahlert (2004), we develop an instrument to measure shopper satisfaction in online and offline retail contexts: the (R)E-Tail Satisfaction scale. Using data from an online (N=202) and an offline (N=441) grocery shopper sample, the instrument is shown to be fit for cross-channel evaluation of levels of satisfaction and its antecedents. We find full metric invariance (identical factor loadings), sufficient partial scalar invariance (identical item intercepts for at least two items per construct), as well as some interesting structural differences. Most notably, online shoppers evaluate the facets of retail satisfaction generally lower than do offline shoppers.

INTRODUCTION

Some retailers use the Internet and their brick-and-mortar facilities as complementary and/or alternate channels (Wallace, Giese and Johnson 2004). For such retailers, it is of strategic importance to monitor customer satisfaction in all of its aspects in both the online and the offline channel (Shankar, Smith and Rangaswamy 2003). However, to be able to compare online and offline levels of satisfaction and its dimensions, it is a prerequisite to have measurement instruments that are invariant across the online and offline contexts (Meredith 1993; Little 1997). The contribution of the current paper is to offer a multidimensional satisfaction measure that may identify deficiencies and point to relevant strategies across retail channels. The scale we propose can be used to evaluate shoppers' satisfaction levels both online and offline. More specifically, we start from an instrument designed by Szymanski and Hise (2000) and cross-validated by Evanschitzky et al. (2004) measuring dimensions of e-tail satisfaction, and extend and adapt the instrument to make it applicable and equivalent across online and offline retail contexts. To this end, measurement invariance across both conditions is studied. Additionally, we use invariance tests to assess cross-channel differences in the levels and weights of formative facets of satisfaction. The current study focuses on grocery shopping.

THEORETICAL FRAMEWORK

Satisfaction in Retailing and E-Tailing

Customer satisfaction is a strategically important outcome in a service context (Gómez, McLaughlin, and Wittink 2004; Seiders et al. 2005) and this is equally true when the transactions are executed online (Shankar, Smith and Rangaswamy 2003; Wolfinbarger and Gilly 2003; Parasuraman, Zeithaml and Malhotra 2005). Szymanski and Hise (2003) propose a scale to measure dimensions of satisfaction specifically for an e-tailing context. Our focus is on building a scale from their model that measures relevant satisfaction dimensions across online and offline channels, thus enabling the formulation of actionable cross-channel strategies.

Szymanski and Hise (2000) propose five dimensions of satisfaction in an e-tailing context, based on qualitative and quantitative research among online shoppers. In their model, satisfaction is evaluated relative to offline shopping. (1) Shopping Convenience refers to the extent to which the e-tail environment allows for easy browsing and does not take too much time. Merchandising refers to the offerings available to the shoppers. In an exploratory factor analysis, Szymanski and Hise (2000) find that this factor consists of two facets: (2) Product Offerings, indicating whether the available product range is sufficiently diverse and varied and (3) Information, indicating whether the shopper finds the information s/he wants. (4) Site Design refers to the organization of the site and how it supports the search process. (5) Financial Security concerns the extent to which payment is safe. All these five dimensions are antecedents of overall satisfaction, defined as the overarching evaluation of the online shopping experience. Using multiple regression analysis, Szymanski and Hise (2000) find that the antecedents explain 28% of the variance in Satisfaction.

While such amount of explained variance is generally considered large in the behavioral sciences (Green 1991), in the case of formative facets this can be considered to be low (Diamantopoulos and Winklhofer 2001). Presumably, measuring satisfaction with online shopping in general operates at quite a high level of abstraction, which is a common reason for low levels of explained variance (Ajzen and Fishbein 1980). Evanschitzky et al. (2004) validate the e-satisfaction scale by Szymanski and Hise (2000) in an online shopping retail and an online financial services context in Germany. A confirmatory factor analysis shows only moderate fit for the measurement model (e.g., RMSEA is .08 for e-shopping and .09 for e-finance). More specifically, as shown by the modification indices, the Convenience item ‘Ease of browsing’ and the Site Design item ‘Presenting information fast’ lead to empirical misfit. Apart from this, Evanschitzky et al.’s (2004) confirmatory factor analysis provides support for the trans-national validity of the model. Additionally, using multiple regression analysis, the authors find that in a retail shopping context Convenience is by far the most important determinant of Satisfaction. Other significant predictors of Satisfaction are Site Design, Product Offerings and Financial Security. Product Information does not show a significant effect on online retail Satisfaction. In a retail shopping context, Evanschitzky et al. (2004) find that the antecedents in their model explain 18% of the variance in Satisfaction. Although such amount of explained variance is generally considered moderate to large in the behavioral sciences (Green 1991), again in the case of formative facets this can be considered to be low (Diamantopoulos and Winklhofer 2001).

To be able to simultaneously study online and offline Customer Satisfaction, we need to develop measures that are applicable in both environments. This is the objective of the current study. We want the resulting scale to meet the following requirements, listed in order of importance. First, the instrument should be relevant and show good psychometric properties simultaneously in online and offline retail contexts. Specifically, the scale needs a sufficient level of measurement invariance to be fit for cross-channel comparisons and benchmarking. Second, it should be parsimonious, thus guaranteeing time efficiency of data collection. As pointed out by Szymanski and Hise (2000) and Evanschitzky et al. (2004), this is a prerequisite especially in online marketing surveys. Third, within the limits set by the former two criteria, we want the scale to be as exhaustive as possible, i.e. covering all relevant formative facets of satisfaction in a retail and e-tail context.

Measurement Invariance

A growing body of literature asserts that measures of constructs are comparable across groups and contexts only if they show certain levels of measurement invariance. Measurement invariance refers to the condition where the relation between indicators and latent variables is the same across groups (Meredith 1993). More specifically, to compare mean scores of latent constructs as well as structural relations between the same, at least two of their indicators need to have equal factor loadings and equal item intercepts across groups (Vandenberg and Lance 2000; Little 1997; Ployhart and Oswald 2004; Steenkamp and Baumgartner 1998). The former criterion, equality of factor loadings, is labeled metric invariance; the latter, equality of item intercepts, is labeled scalar invariance (Steenkamp and Baumgartner 1998). As Cheung and Rensvold (2002) point out, metric and scalar non-invariance may lead to biased conclusions. However, as indicated by Little (1997), in stead of merely being a bothersome requirement for between-group (e.g. cross-cultural) research, non-invariance in itself should be an object of study, and can be instrumental in better understanding between-group differences. When comparing online and offline shoppers' customer satisfaction and its dimensions, it can be very instructive to investigate what differences emerge in measurement and structural models across these two groups. For example, online shoppers might systematically rate freshness of products lower than do offline shoppers, regardless of overall levels of satisfaction with Product Quality. Equally so, it is plausible to imagine offline shoppers being more price sensitive than online shoppers.

CONCEPTUAL MODEL

Conceptual Model and Hypotheses

To develop a conceptual model, we particularly draw upon the work of Szymanski and Hise (2000) and Evanschitzky et al. (2004). However, we make adaptations in several regards. First and foremost, the constructs in our model have to be equivalent across online and offline retail contexts. To this end, we evaluate the usefulness of each construct based on the findings presented in both studies. Additionally, we validate the equivalence of the resulting model in a qualitative study. Second, because measurement quality is a prerequisite for cross-channel validity of the measurement instrument, we need the measurement model to show acceptable model fit to start with. To this end, in scrutinizing the work of Szymanski and Hise (2000) and Evanschitzky et al. (2004), we pay special attention to indications of misspecification (such as cross-loadings) and omitted variables (resulting in relatively low explained variance of the Satisfaction variable).

In the conceptual model resulting from our literature review, overall retail satisfaction is the outcome of five formative facets. We discuss each dimension in turn, after which we list the resulting hypotheses. In the terms of Jarvis, Mackenzie and Podsakoff (2003), the model that we propose is a type II formative model, i.e. a model where the first order factors are measured by means of reflective indicators, and where these first order factors are the formative indicators of the second order factor (which additionally has some reflective indicators for reasons of identification).

Based on Szymanski and Hise (2000), as the first dimension of Satisfaction we propose Convenience. Convenience refers to consumers' perceptions of the time and effort they invest in buying or using a service, in this case online or offline shopping. As Berry, Seiders and Grewal (2002) argue, Convenience is an important antecedent of Satisfaction and captures effects of - among others - Design and Information. Therefore, in line with our goal of parsimony and optimal model specification and fit, but deviating from Szymanski and Hise (2000), we choose not to include Design or Information as separate constructs, since their impact is assumed to be captured by Convenience. Moreover, Evanschitzky et al. (2004) indicate that the Information dimension is not considered relevant by retail shoppers. It seems plausible that this is true especially for grocery shoppers. Further, in the confirmatory factor analysis by Evanschitzky et al. (2004), the authors note that the mediocre fit of their model is partly due to two specific items: 'Ease of browsing', which is modeled as a Convenience item, and 'Presenting information fast', which is modeled as a Site Design item. We believe the problems encountered with these items may reflect the conceptual status of Design and Information as actually being part of Convenience. Both refer to means to an end, the end being Convenience, or economizing on consumers' time and cognitive resources (Szymanski and Hise 2000).

Also in keeping with Szymanski and Hise (2000), Product Offerings is proposed as the second formative dimension of Satisfaction. This dimension refers to the extent to which the assortment meets the needs of the customer. As these authors argue, the right assortment of products makes it more probable that consumer needs can be met. As a second argument, they suggest that the wider assortment can include products of higher quality, this way providing a benefit to customers. We feel this is indicative of a separate construct, and therefore propose as a third dimension Product Quality. Product Quality is an important tangible outcome of the retail experience and is an important factor in evaluating services (Lee, Lee and Yoo 2000). Product Quality here is defined as the non-price related evaluation of the products in the assortment.

Fair Price is added as a fourth dimension. This dimension refers to the extent to which the customer evaluates a retailer's prices as being within the customer's acceptable range. Customers do not only take into account time and cognitive resources they invest when evaluating services, but also financial costs (Brady et al. 2005). When discussing their results, Evanschitzky et al. (2004) note that Price might be important to German retail customers, lending further support to the idea that Price may be an omitted variable in the model.

Financial Security is the final formative dimension in Szymanski and Hise's (2000) model. Financial Security refers to the customer's evaluation of the reliability and safety of payments to the retailer. This dimension is an ever recurrent theme in discussions on online retailing and can hence be expected to affect satisfaction. Szymanski and Hise's (2000) results indicate that this dimension indeed has a significant and substantial impact on satisfaction in an e-tail context. It is conceivable that the growing adoption of electronic payment in the offline world has made this factor relevant in an offline context as well. For example, customers could be concerned about such things as entering their ATM pin numbers in a public space, credit card theft by employees, etc.

Based on the above discussion, we now formulate the hypotheses concerning the relation between Satisfaction and its facets. The question of whether these effects apply equally to both contexts is an issue that will be addressed later (see Hypotheses 10 through 13). The conceptual and measurement model is depicted in Figure 1.

Insert Figure 1 About Here

H1: Convenience is positively related to Satisfaction

H2: Product Offerings is positively related to Satisfaction

H3: Product Quality is positively related to Satisfaction

H4: Fair Price is positively related to Satisfaction

H5: Financial Security is positively related to Satisfaction

Levels of Invariance

A fundamental contribution of this study lies in the cross-channel comparison of the satisfaction model. We are interested in studying three sets of invariance questions in this regard: (1) measurement invariance (Hypotheses 6 and 7); (2) invariance of the structural intercept and structural residual variance of the dependent variable, Satisfaction (Hypotheses 8 and 9); (3) invariance of structural means, variances, covariances and regression weights of the independent variables, i.e. the formative facets of Satisfaction (Hypotheses 10 through 13). For the specific items to which Hypotheses 6 and 7 apply, we refer to the items in Table 1 and the discussion below.

Measurement invariance refers to invariance in the relation between the items and the latent variables they tap into. More specifically, metric and scalar invariance will be tested for. While doing so, we are open to the possibility that specific items do have different relations to the underlying latent variable in these different groups, and consider this a worthwhile topic of investigation (Little 1997). For example, online shoppers might have different definitions of Product Quality than offline shoppers.

> H6 - Metric invariance: The factor loadings of the indicators of Satisfaction and its antecedents are invariant across online and offline shoppers.

> H7 - Scalar invariance: The intercepts of the indicators of Satisfaction and its antecedents are invariant across online and offline shoppers.

In a next major stage, we study structural invariance, which will be broken down into two sub-steps. First, we study invariance of the intercept and residual variance of overall Satisfaction. If invariance of the intercept is established, this indicates there is no effect of online versus offline shopping that has not been accounted for by the antecedents of satisfaction in our model. If the intercept of satisfaction is not invariant, this implies there is either a direct, non-mediated effect of online/offline shopping, or we omitted an antecedent of satisfaction which has different levels in the on- or offline condition. Invariance of the residual variance of satisfaction would indicate homoscedasticity across both groups. Heteroscedasticity would be diagnostic of omission of one or more variables having equal means across both groups, but different variances or weights. As is apparent from the above discussion, the evaluation of invariance of the dependent variable's intercept and residual variance is relevant for a general evaluation of the quality of our model.

> H8 – Structural intercept invariance: After controlling for Convenience, Product Offerings, Product Quality, Fair Price and Financial Security, the level of Satisfaction is equal among online and offline shoppers.

> H9 – Structural residual variance invariance: After controlling for Convenience, Product Offerings, Product Quality, Fair Price and Financial Security, the variance of Satisfaction is equal among online and offline shoppers.

In a second sub-stage of the structural invariance evaluation, we study more idiosyncratic differences between the online and offline shoppers. More specifically, we compare the means, variances, covariances and regression weights of the antecedents of Satisfaction. Means correspond to the average evaluation of the given dimensions. Variances correspond to the heterogeneity in these evaluations, either flowing forth from heterogeneity in the shoppers' needs or in the delivered service quality. Covariances are linked to how dimensions of Satisfaction relate to one another due to common sources of variance. Regression weights indicate the importance shoppers attach to a given dimension in forming an overall evaluation of the retail experience. Since the investigation into the differences between online and offline shoppers in this regard is exploratory in nature, we formulate the null hypotheses that the levels, variances, covariances and weights of the Satisfaction facets are equal across channels.

> H10 - Structural Means invariance: Evaluations of formative facets of Satisfaction are equal for online and offline shoppers.

H11 - Structural Variances Invariance: Variances of the evaluations of formative facets of Satisfaction are equal for online and offline shoppers.

H12 - Structural Covariances Invariance: The relations between formative facets of Satisfaction are equal for online and offline shoppers.

H13 - Structural Regression Weights Invariance: Relations between formative facets and Satisfaction are equal for online and offline shoppers.

METHODOLOGY

Scale Design

To adapt the instrument to be relevant and identical in meaning across online and offline settings, we conduct a qualitative phase of research. Twelve face-to-face interviews are conducted, six with offline shoppers, six with shoppers who have had experience with both off- and online shopping. The interviews are designed to elicit information on what drives satisfaction with shopping via either or both of the channels. In a second stage of the interviews, respondents are asked to evaluate to what extent the items in the scale by Szymanski and Hise (2000) and the constructs in our conceptual model are similar in meaning across on- and offline shopping experiences and whether specific items are either relevant or redundant. The latter questions also are the subject of two additional expert interviews.

We rephrase all items to be relevant in both an online and an offline retail context. Quality and Fair Price items are formulated based on input from our respondents. Moreover, we include an extra item to measure Financial Security in order to have multiple indicators per latent construct (Churchill 1979). This will allow us to assess reliability and use Means and Covariance Structures (MACS) to test for measurement invariance. The scale items are listed in Table 1; the measurement model is depicted in Figure 1.

The facets of satisfaction are evaluated on five point scales, under the heading "Please evaluate *[retailer A]* on the following dimensions", where 1 = bad and 5 = excellent. Overall satisfaction is evaluated on two five point scales anchored "very dissatisfied – very satisfied" and "very displeased – very pleased".

Insert Table 1 About Here

Data Collection

The scale is used in a survey among a sample of offline shoppers and a sample of online shoppers, both customers respectively of an offline grocery retail chain and its online daughter company in a European country. For this reason, all items are back translated and checked for functional and semantic equivalence (Kumar 2000). The offline retailer positions itself as a quality grocery retailer, with an extensive assortment of fresh vegetables, specialties and wine. Prices are slightly above average. The online retailer offers a selected subset of the products available in the offline supermarket at similar prices, with an additional small fee for home delivery.

Samples

In the offline sample, data are collected via personal interviews using the questionnaire. Respondents are recruited at the exit of four supermarkets of the same chain on four different days of the week, including a Saturday. Of 900 shoppers that are addressed by the interviewers, 441 provide us with complete and valid responses (participation rate = 49%). The average age in this sample is 39.4 years. On average people have 15.1 years of formal education. 42.2% of the respondents are male, 57.8% are female.

In the online sample, data are collected by means of an online survey. Respondents are recruited by means of a personalized e-mail linking to the survey. We send 913 e-mails to customers who made a purchase at the online retailer in the last four weeks. 290 people click through to the questionnaire (response rate =31.8%). 202 of these respondents fill out the questionnaire completely and correctly (net response rate = 22.1%; some respondents have merely surfed to the questionnaire without responding to it). The average age in this sample is 40.9 years. On average people have 14.8 years of formal education. 35.5% of respondents are male, 64.5% are female. The online sample is not significantly different from the offline sample in terms of age and educational level (respectively $t\ (571) = -1.108$, $p=.268$ and $t\ (639) = 1.082$, $p=.280$). The proportion of women is significantly higher though ($\chi^2\ (1) = 23.69$, $p<.001$). However, we believe it is improbable that this would bias our results.

Test Procedure

To study invariance across the online and the offline context, we specify nested models (Steenkamp and Baumgartner 1998; Little 1997; Cheung and Rensvold 2002). In each model, an additional set of parameters is constrained to equality across both groups. The validity of this restriction is then tested by comparing the fit of the nested model to the fit of the model in which it is nested. If the deterioration in fit is significant and substantial, the constraints are not accepted, indicating that the parameters in question are different ('non-invariant') across groups. We work in two major stages, respectively investigating measurement invariance and structural invariance, the former being a prerequisite for the latter (Steenkamp and Baumgartner 1998). The specifications of the nested models are detailed in Appendix A.

To test whether a model shows statistically worse fit than the model in which it is nested, we evaluate the chi square difference test (Jöreskog 1971). If statistical significance is found, we evaluate practical significance based on the recommendations by Little (1997), Cheung and Rensvold (2002) and Baumgartner and Steenkamp (1998). Little (1997) suggests that a nested model should be accepted as not being substantially worse in fit than the model in which it is nested if the following criteria are met: (1) the overall fit of the nested model is acceptable; (2) the difference in the Tucker-Lewis Index (TLI) is less than or equal to .05; (3) indices of local misfit are uniformly and unsystematically distributed with respect to the constrained parameters; and (4) the constrained model is substantively more meaningful and parsimonious than the unconstrained model. Criterion (2) has been empirically put to test by Cheung and Rensvold (2002). In line with their findings we rephrase this condition to: (2) the difference in the Comparative Fit Index (CFI) should be less than or equal to .01 (Cheung and Rensvold 2002). As for criterion (3), we take into consideration the recommendations by Steenkamp and Baumgartner (1998), and evaluate the modification indices (M.I.'s) and standardized residuals (s.r.'s) of constrained parameters.

FINDINGS

We specify several Means and Covariance Structures (MACS) corresponding to the model in Figure 1. For reasons of identification, per factor, one loading is fixed to one and one intercept is fixed to zero. The models are simultaneously tested in the online and offline sample. The fit indices for all models are summarized in Table 3.

Insert Table 3 About Here

Base Model

The unconstrained model allows for different measurement and structural parameters in the online and offline groups. The fit indices are given in Table 3, under model A. The chi square model fit test is significant on the .01- but not the .001-level. The alternative fit indices compare favorably against common cut-off criteria (Hu and Bentler 1999) and there is no reason to believe further model adaptations would result in better parameter estimates. Therefore, we accept this model as the reference model for the first round of nested model tests, which aim to evaluate measurement invariance. The model estimates show a clear factor structure, with good levels of discriminant validity. All standardized factor loadings are at least .60 (see Table 1). The Average Variance Extracted for each factor, as well as the Shared Variance between factors is displayed in Table 2 (Fornell and Larcker 1981), providing further evidence of discriminant validity.

Measurement Invariance

We test for measurement invariance in two steps. First, we specify a metric invariant model in which the factor loadings are held equal across groups. Next, we add scalar invariance by constraining the intercepts of the items to equality across groups. Again, the resulting fit indices are listed in Table 3.

Imposing metric invariance (model B in Table 3) does not induce a significant change in chi square. Hence, metric invariance (and hypothesis 6) is accepted. On the contrary, scalar invariance (model C; hypothesis 7) does lead to a significant decrease in model fit. Moreover, while the decrease in CFI is just below .01 (.009), the general deterioration in alternative fit indices is substantial. Consequently, in line with Steenkamp and Baumgartner (1998), we one-by-one release the intercepts showing the highest M.I.'s (respectively 10.0, 6.65 and 4.29): CONV3 ("effortlessly finding what I'm looking for"; .33 points lower online than offline), PRICE2 ("interesting reductions"; .23 points lower online than offline) and QUAL2 ("freshness of products"; .18 points lower online than offline). The relationship between these items and the factors they reflect are shown respectively in figure 2a, 2b and 2c.

Insert Figure 2a, 2b & 2c About Here

Compared to model B, the resulting partial scalar invariance model (model D) has a chi square difference test that is just significant on the .05-level. We therefore again evaluate the other evaluation criteria for nested models outlined above to evaluate whether hypothesis 7 is partially supported. First, the alternative fit indices are very good and not substantially different from the reference model (Steenkamp and Baumgartner 1998; Little 1997). Second, none of the intercepts in the model has a significant M.I. (i.e. above 3.84, the value corresponding to a χ^2 (1) p-value of .05) or a significant standardized residual mean (above 1.96, corresponding to a p-value of .05). Finally, the model is parsimonious and makes theoretical sense: from a substantive point of view, the released intercepts indicate that in an online context some evaluative ratings are lower regardless of the evaluation of the broader factor (evaluative dimension) to which they belong (Raju, Lafitte and Byrne 2002). For example, keeping overall ratings of product quality equal, on average products are less fresh in an online context (QUAL2). A similar reasoning applies to the other scalar non-invariant items: for equal levels of overall convenience, CONV3, ‘effortlessly finding what you are looking for’, scores lower in an online context; for equal levels of overall price evaluation, PRICE2, ‘there are interesting price reductions’ scores lower in an online context. These differences plausibly reflect customer perceptions in line with real differences between the online and offline retail experience. Based on the above, we accept partial scalar invariance and use model D as the reference model for testing aspects of structural invariance.

Structural Invariance

Taking full metric and partial scalar measurement invariance as given, we further investigate structural invariance. More specifically, we study the between-group (online/offline) differences in the structural intercept and residual variance of Satisfaction, as well as structural means, variances, covariances and weights of the Satisfaction facets. For each of these aspects, we specify a model that is nested in the partial scalar invariance model (D). Since the order in which the different aspects are tested for is described as rather arbitrary in the literature (Steenkamp and Baumgartner 1998), and to avoid confusion in the meaning of the tests, we do not nest different structural equality restrictions in one another.

Invariance of intercept and residual variance of satisfaction

Based on the chi square difference test comparing model E to model D (see Table 3), we accept invariance of the structural intercept of satisfaction. The same applies to the invariance of the structural residual variance of satisfaction (model F in Table 3). These results indicate that our model of (R)E-Tail satisfaction captures the most relevant sources of variance in overall satisfaction with both on- and offline retail (Hypotheses 8 and 9).

Invariance of means, variances, covariances and weights of antecedents

The chi square difference test comparing model G to model D is significant on the .001-level. Moreover, the drop in CFI substantially exceeds .01, and the overall model fit of model G is rather poor (Hu and Bentler 1999). The standardized residuals indicate that all indicator means in the offline group, except for those of Fair Price, are systematically underestimated in model G. All this indicates that the assumption of equal means of the satisfaction antecedents is not tenable. The means of the satisfaction antecedents in the online and offline samples are reported in Table 4. Since we set one intercept per latent variable to zero, these means are expressed in the same scale as the indicators, i.e. a scale with neutral point 3. Apart from the evaluation of Fair Price, all means are lower in the online sample, most notably so for Offerings.

Insert Table 4 About Here

Model H imposes equal variances of the satisfaction antecedents across groups. The resulting chi square test is significant on the .001 level. Additionally, the CFI for this model is more than .01 below the CFI of reference model D. While the overall model fit for H is acceptable, it is less acceptable than that of the reference model. Moreover, the standardized residuals indicate that all indicators' variances in the offline group, except for those of the Fair Price factor, are overestimated in model H. Based on this range of signals, we reject invariance of the structural variances. We report the estimates based on model D in Table 4. All variances, except for that of Fair Price, are notably larger in the online group than in the offline group.

Model fit of model I is not significantly worse than that of model D, indicating that the covariances between the antecedents of satisfaction are similar in the on- and offline samples. Note that correlations not involving Fair Price nevertheless tend to be bigger in the offline sample due to the smaller variances in this group. The correlations for both samples are given in Table 6.

Insert Table 6 About Here

The chi square difference test comparing model J to model D is significant on the .01 but not the .001 level. The decrease in CFI is below .01 and the overall model fit is acceptable, indicating that the difference in regression weights across groups is statistically significant, but that the practical significance of this effect is small. Inspection of the standardized residuals suggests, however, that in model J the regression weight of Fair Price is slightly (but not significantly) smaller in the online sample, while the regression weight of Financial Security and Product Offerings are slightly (also not significantly) smaller in the offline group. Closer inspection of the regression weights (see Table 5) shows that the effect of Fair Price is insignificant online. Financial Security is insignificant offline, and only marginally significant online ($.05 < p < .10$). Also, Offerings only has a marginally significant impact offline ($.05 < p < .10$), while being more significant online ($p < .001$).

Insert Table 5 About Here

DISCUSSION

Based on data collected from a sample of online and offline grocery shoppers, we show that our model of (R)E-Tail Satisfaction is applicable for cross-channel evaluations of shopper satisfaction. More specifically, we propose five formative facets of overall customer retail satisfaction that have equivalent meaning to both online and offline customers. The observation that the concepts' meanings are equivalent does not imply that their antecedents need to be identical. For example, irrespective of retail channel, convenience refers to consumers' perceptions of the time and effort they invest in shopping. How convenience is achieved may differ across channels though. Online, site design probably plays a role, while offline, lay-out of the shelves may be of importance. To resume our line of reasoning, while the meaning of these dimensions is similar to both groups, their ratings vary (as reflected in means differences), as does the impact each dimension has on overall satisfaction (as reflected in differences in regression weights).

This allows for cross-channel comparisons of satisfaction using a unified measurement model.

Three specific items are not invariant in their relation to the latent construct they reflect, but a sufficient number of items have metric and scalar invariance. This means the proposed scale and measurement model can be used in future studies aiming to measure retail satisfaction in both online and offline contexts. The non-invariant items can be either included and tested for invariance in the new context, or can be excluded ex ante. It should be noted that including the items does not impair measurement in any way, but the information captured by the non-invariant items is not taken into account when comparing means and variances of the latent constructs. We also note that one of the non-invariant items, QUAL2 ("freshness of the products") is specific to a grocery retail context and should not be included in future studies in contexts other than grocery retailing. All in all, the tools provided here level the ground for further cross-channel satisfaction research. In addition to establishing this measurement tool, we derived some preliminary findings from our data, which may not be generalizable to other countries and retailers though.

We find that Convenience, Product Quality, Product Offerings and Fair Price are significant antecedents of overall shopper satisfaction in either or both online and offline retail contexts. In our sample, Financial Security is only marginally significant for online shoppers and not significant for offline shoppers. The above dimensions explain a good proportion of variance in the overall shopper Satisfaction measure. The explained variance in satisfaction amounts to 67% online, 58% offline. The cross-channel difference in R^2 is due to differences in the variances of the antecedents of satisfaction. We see two reasons for the improvement as compared to previous studies (Szymanski and Hise 2000; Evanschitzky et al. 2004). First, the use of Structural Equation Modeling corrects for measurement error and thus disattenuates the regression weights and factor variances. Second, we focus on a more specific level of satisfaction, which tends to lead to higher explanatory power.

The data enable us to compare the levels of satisfaction with several facets as well as their impact on overall satisfaction. In our sample, the online shoppers evaluated Convenience, Quality, Offerings and Financial Security lower than did the offline shoppers. Remarkably, while online shoppers have to pay an additional fee to have the goods delivered at home, they did not rate 'Fair Price' any different than did the offline shoppers. Apparently, online shoppers' price perceptions are not negatively affected by this extra cost. This might be related to their lower price sensitivity (see below). Alternatively, the fee may be perceived as fair. In the case studied here, among the online shoppers the ranking of the antecedents of satisfaction in order of importance is (1) Convenience, (2) Quality, (3) Product Offerings, and (4) Financial Security. Offline, the ranking is (1) Convenience, (2) Quality, (3) Fair Price, and (4) Product Offerings. Thus, Convenience and Product Quality are the main driver of satisfaction in both contexts. The online shoppers seem to care somewhat more about Product Offerings and Financial Security. The higher focus on Offerings is most probably related to the more limited product range offered by this particular online retailer, due to smaller volumes and the required high rotation of fresh products. Financial Security seems to be an issue only for online shoppers. This is in line with expectations (Szymanski and Hise 2000). In our sample, the offline shoppers are more price sensitive than are the online shoppers. It is plausible that online shoppers are prepared to pay a price for higher efficiency, as opposed to the offline shoppers. As a final note of caution, we want to point out that the satisfaction levels and weights need not generalize to online and offline customers of other retailers. The specific evaluations that are obtained from customers are idiosyncratically related to the retail context, and that is exactly what is intended. The main objective of this paper is to establish a measurement scale and model that makes it possible to equivalently measure (differences in) retail satisfaction among online and offline shoppers.

LIMITATIONS AND FUTURE RESEARCH

Some limitations of the current study provide interesting opportunities for future research. Most importantly, this study was limited to grocery retailing in one European country. It would be interesting to investigate how our findings generalize to other countries, as Evanschitzky et al. (2004) studied with regard to the instrument proposed by Szymanski and Hise (2000). Also, applying the measurement instrument and model in sectors other than grocery retailing would shed light on the robustness of the model across domains.

Additionally, we limited the scope of this paper to satisfaction. As Brady et al. (2005) show, for a full understanding of how customers evaluate services and how this relates to outcomes for the company, satisfaction should be seen in a broader conceptual framework, including among others behavioral intentions. Future research should find ways of building such broader nomological networks using constructs and measures that are equivalent across channels. This is the only way to fully assess the differences in how customers operate across retail channels.

MANAGERIAL IMPLICATIONS

First and foremost, our study shows that retail customer satisfaction can be measured in a similar way in the online and offline context. This offers managers the possibility to evaluate performance in both channels using the same evaluative dimensions. Thus, with this instrument, the online and offline channels can be used as a benchmark for one another.

Second, our data show that in some cases, the online and offline channels need to deliver a different mix of benefits. In particular the levels of price sensitivity may differ. The online and offline channel of one retailer need therefore not be similar in positioning, since they can be used to cater to different segments of customers.

Finally, we observe that variability in ratings is higher in the online context. Seemingly, the automation of large part of the shopping experience does not necessarily even out the variability in shopping satisfaction from the customer point of view. In this specific setting, the reason may lie in the lower levels of competition in the online channel, i.e. customers have to choose from a smaller number of alternatives online than offline. Consequently, in such circumstances a more diverse set of customers - even including less satisfied customers - may come back to the same online retailer repeatedly.

REFERENCES

Ajzen, Icek, and Martin Fishbein (1980), *Understanding attitudes and predicting social behavior*, Englewood Cliffs, NJ: Prentice-Hall

Berry, Leonard L., Kathleen Seiders, and Dhruv Grewal (2002), "Understanding Service Convenience," *Journal of Marketing*, 66(3), 1-17.

Brady, Michael K., Gary A. Knight, J. Joseph Cronin, G. Tomas M. Hult, and Bruce D. Keillor (2005), "Removing the contextual lens: A multinational, multi-setting comparison of service evaluation models," *Journal of Retailing*, 81(3), 215-230.

Cheung, Gordon W., and Roger B. Rensvold (2002), "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling*, 9(2), 233-255.

Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* 16(1), 64-73.

Diamantopoulos, Adamantios, and Heidi M. Winklhofer (2001), "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research*, 38(2), 269-277.

Evanschitzky, Heiner, Gopalkrishnan R. Iyer, Josef Hesse, and Dieter Ahlert (2004), "E-Satisfaction: a Re-examination," *Journal of Retailing*, 80, 239-247.

Fornell and Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", *Journal of Marketing Research*, 18, 39-50.

Gómez, Miguel I., Edward W. McLaughlin, and Dick R. Wittink (2004), "Customer Satisfaction and Retail Sales Performance: an Empirical Investigation," *Journal of Retailing*, 80(4), 265-278.

Green, Samuel B. (1991), "How Many Subjects Does It Take To Do A Regression Analysis?" *Multivariate Behavioral Research*, 26(3), 499-510.

Hu, L., and P. M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives," *Structural Equation Modeling*, 6 (1), 1-55.

Jarvis, Cheryl Burke, Scott B. Mackenzie, and Philip M. Podsakoff (2003), “A critical review of construct indicators and measurement model misspecification in marketing and consumer research,” *Journal of Consumer Research*, 30, 199-218.

Jöreskog, Karl G. (1971), “Simultaneous Factor Analysis in Several Populations,” *Psychometrika*, 36 (4), 409-426.

Kumar, V. (2000), *International Market Research*, NJ: Prentice-Hall: Upper Saddle River.

Lee, Haksik, Yongki Lee, and Dongkeun Yoo (2000), “The determinants of perceived service quality and its relationship with satisfaction,” *Journal of Services Marketing*, 14(2), 217-231.

Little, Todd D. (1997), “Means and Covariance Structures Analyses of Cross-Cultural Data: Practical and Theoretical Issues”, *Multivariate Behavioral Research*, 32(1), 53-76.

Meredith, W. (1993), “Measurement invariance, factor analysis and factorial invariance”, *Psychometrika*, 58, 525-543.

Parasuraman, A., Valarie A. Zeithaml, and Arvind Malhotra (2005), “E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality,” *Journal of Service Research*, 7(3), 213-233.

Ployhart, Robert E., and Frederick L. Oswald (2004), “Applications of Means and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches,” *Organizational Research Methods*, 7(1), 27-65.

Raju, Nambury S., Larry J. Laffitte, and Barbara M. Byrne (2002), “Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory,” *Journal of Applied Psychology*, 87(3), 517-529.

Seiders, Kathleen, Glenn B. Voss, Dhruv Grewal, and Andrea Godfrey (2005), "Do Satisfied Customers Buy More? Examining Moderating Influences in a Retailing Context," *Journal of Marketing*, 69(4), 26-43.

Shankar, Venkatesh, Amy K. Smith, and Arvind Rangaswamy (2003), "Customer satisfaction and loyalty in online and offline environments", *International Journal of Research in Marketing*, 20, 153-175.

Steenkamp, J.B. & Baumgartner, H. (1998), "Assessing measurement invariance in cross-national consumer", *Journal of Consumer Research*, 25, 78-90.

Szymanski, David M., and Richard T. Hise (2000), "E-Satisfaction: An Initial Examination", *Journal of Retailing*, 76(3), 309-322.

Vandenberg, Robert J., and Charles E. Lance (2000), "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research," *Organizational Research Methods*, 3(1), 4-70.

Wallace, David W., Joan L. Giese, and Jean L. Johnson (2004), "Customer Retailer Loyalty in the Context of Multiple Channel Strategies," *Journal of Retailing*, 80(4), 249-263.

Wolfinbarger, Mary, and Mary C. Gilly (2003), "eTailQ: Dimensionalizing, Measuring and Predicting eTail Quality," *Journal of Retailing*, 79(3), 183-198.

TABLE 1

Items and factor loadings

Factor	Item Number & Label	Item	Offline loadings[c]	Online loadings[c]	Metric invariance	Scalar invariance
Convenience	1. CONV1	Total shopping time (how fast it goes to shop here) [b]	.67	.73	yes	yes
	2. CONV2	Convenience (how easy it is to shop here)[b]	.81	.83	yes	yes
	3. CONV3	Being able to effortlessly find what you are looking for[b]	.62	.70	yes	no
Quality	4. QUAL1	I know what I can expect of the products I buy here	.77	.77	yes	yes
	5. QUAL2	Freshness of the products	.73	.73	yes	no
	6. QUAL3	The products I buy here are completely ok and undamaged	.68	.78	yes	yes
Fair price	7. PRICE1	Shopping at X is affordable	.72	.83	yes	yes
	8. PRICE2	There are interesting price reductions	.60	.60	yes	no
	9. PRICE3	The price of products is fair	.77	.86	yes	yes
Offerings	10. OFFER1	Number of offerings[a]	.83	.85	yes	yes
	11. OFFER2	Variety of offerings[a]	.84	.89	yes	yes
Financial Security	12. PAY1	Financial Security of the transaction[a]	.88	.91	yes	yes
	13. PAY2	Reliable method of payment	.99	.96	yes	yes
Satisfaction	14. SAT1	Very dissatisfied – very satisfied[a]	.87	.91	yes	yes
	15. SAT2	Very displeased – very pleased [a]	.82	.84	yes	yes

[a] taken from Szymanski and Hise (2000); [b] rephrased based on Szymanski and Hise (2000); The estimates are based on the partial scalar invariance model (model D in Table 3). [c]Offline loadings: Standardized factor loadings in the offline sample; Online loadings: Standardized factor loadings in the online sample. Differences in loadings are non-significant. Note that the reported differences are due differences in factor variances.

TABLE 2

Shared Variance and Average Variance Extracted

	Online					Offline				
SV/AVE	Price	Offer	Quality	Con-venience	Financial Security	Price	Offer	Quality	Con-venience	Financial Security
Price	0.59					0.49				
Offer	0.20	0.75				0.11	0.69			
Quality	0.14	0.18	0.57			0.12	0.29	0.53		
Convenience	0.15	0.22	0.19	0.57		0.14	0.18	0.24	0.50	
Financial Security	0.09	0.06	0.16	0.13	0.88	0.10	0.12	0.22	0.13	0.89

On the diagonal, Average Variance Extracted of each factor is displayed; the other values display Shared Variance (i.e. r^2) between two factors.

TABLE 3

Fit indices for the nested models

	Model	χ^2	df	p	Ref. model	$\Delta\chi^2$	Δdf	Δp	TLI	CFI	RMSEA	SRMR
A.	Unconstrained	208.6	150	0.001					0.981	0.987	0.025	0.043
B.	Metric invariance	218.2	159	0.001	A	9.5	9	0.389	0.982	0.986	0.024	0.041
C.	Full scalar invariance	266.8	168	0.000	B	48.6	9	0.000	0.972	0.977	0.030	0.039
D.	Partial scalar invariance	230.9	165	0.001	B	12.7	6	0.048	0.981	0.985	0.025	0.040
E.	Structural intercept invariance	233.4	166	0.000	D	2.6	1	0.109	0.980	0.985	0.025	0.041
F.	Structural residual invariance	234.3	166	0.000	D	3.4	1	0.064	0.980	0.984	0.025	0.041
G.	Means invariance	445.1	170	0.000	D	214.3	5	0.000	0.922	0.937	0.050	0.093
H.	Structural variance invariance	296.9	170	0.000	D	66.1	5	0.000	0.964	0.971	0.034	0.073
I.	Structural covariance invariance	243.8	175	0.000	D	12.9	10	0.229	0.981	0.984	0.025	0.069
J.	Structural regression invariance	247.2	170	0.000	D	16.3	5	0.006	0.978	0.982	0.027	0.046

Ref model: reference model in which the model on this line is nested (Jöreskog 1971). $\Delta\chi^2$: chi square difference test for nested models; Δdf: degrees of freedom of the chi square difference test; Δp: p-value of the chi square difference test; TLI: Tucker-Lewis Index; CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Residual.

TABLE 4

Means and variances of satisfaction facets in the online and offline samples

	Online				Offline			
	Mean	s.e.	Variance	s.e.	Mean	s.e.	Variance	s.e.
Fair price	3.55	0.230	0.43	0.059	3.57	0.229	0.49	0.057
Offerings	3.36	0.151	0.77	0.095	4.63	0.164	0.34	0.031
Quality	4.08	0.253	0.41	0.057	4.43	0.259	0.26	0.028
Financial Security	3.97	0.170	0.45	0.051	4.25	0.174	0.24	0.021
Convenience	3.54	0.227	0.48	0.071	3.89	0.236	0.27	0.033

Estimates taken from partial scalar invariance model (D in Table 3).

TABLE 5

Regression weights of satisfaction facets in the online and offline samples

	Online (R^2 = .67)				Offline (R^2 = .58)			
Independent variable	Stdd weight	Unstdd weight	s.e.	C.R.	Stdd weight	Unstdd weight	s.e.	C.R.
Fair price	-0.04	-0.04	0.077	-0.55	0.29	0.22	0.042	5.19
Offerings	0.26	0.21	0.061	3.50	0.11	0.10	0.054	1.94
Quality	0.35	0.40	0.087	4.60	0.31	0.33	0.071	4.59
Financial Security	0.11	0.12	0.068	1.76	-0.05	-0.05	0.052	-0.93
Convenience	0.38	0.40	0.082	4.95	0.32	0.32	0.062	5.15

Dependent variable = Satisfaction. Estimates based on partial scalar invariance model (model D in Table 3). Stdd weight: Standardized weight; Unstdd weight: Unstandardized weight; s.e.: standard error of the estimate; C.R.: Critical Ratio = estimate / s.e.

TABLE 6

Correlations between the antecedents of satisfaction in the on- and offline groups

	Fair Price	Offerings	Quality	Financial Security	Convenience
					Offline
Fair Price		0.34	0.35	0.32	0.37
Offerings	0.44		0.54	0.34	0.43
Quality	0.37	0.43		0.47	0.49
Financial Security	0.31	0.25	0.40		0.35
Convenience	0.39	0.47	0.43	0.36	
	Online				

Correlations in the offline sample are given in the upper right triangle; correlations in the online sample are given in the lower left triangle.

FIGURE 1

Means and Covariance Structure Model

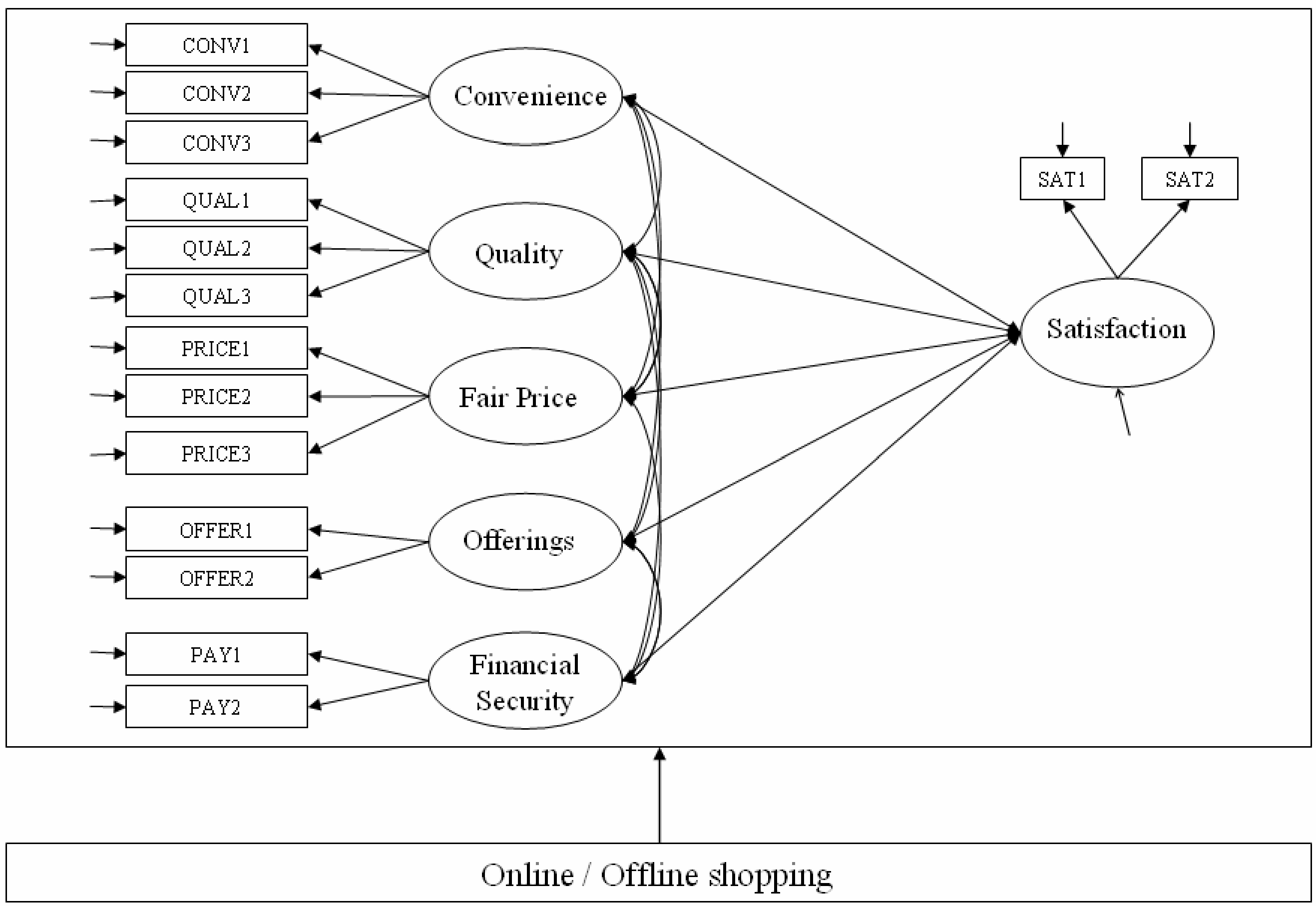


FIGURE 2

Relation between Factors and non-invariant items

FIGURE 2A

PRICE2, "There are interesting price reductions"

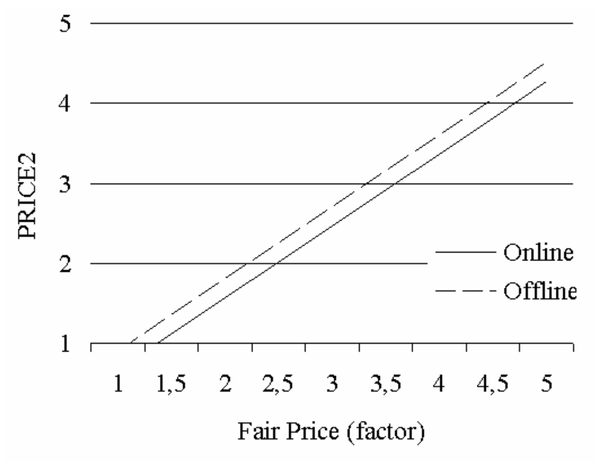


FIGURE 2B

CONV3, "Being able to effortlessly find what you are looking for"

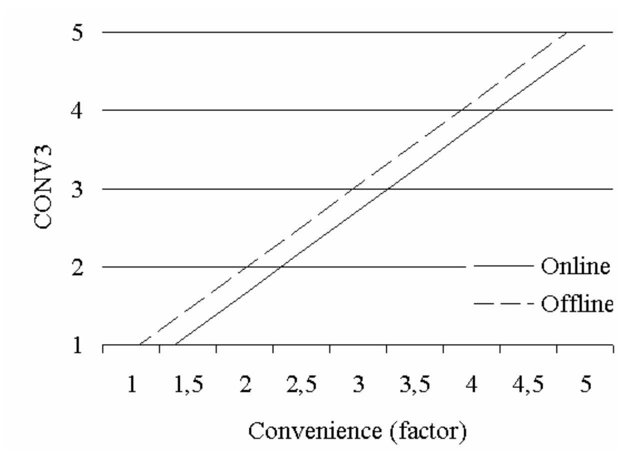
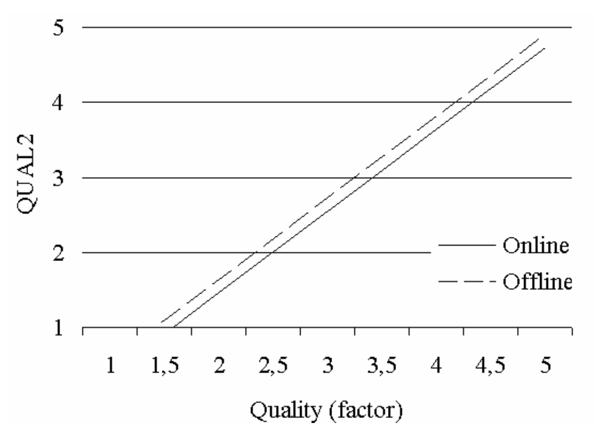


FIGURE 2C

QUAL2, "Freshness of the products"



APPENDIX A

Nested Means And Covariance Structures

In this appendix, we specify the nested models used to test our hypotheses. These models correspond to those reported in Table 3 and to the model depicted in Figure 1.

We model the observed item scores x and y as functions of latent variables. Satisfaction is modeled as a function of its formative facets:

$$x^{(g)} = \tau_x^{(g)} + \Lambda_x^{(g)} \xi^{(g)} + \delta^{(g)}$$
$$y^{(g)} = \tau_y^{(g)} + \Lambda_y^{(g)} \eta^{(g)} + \varepsilon^{(g)}$$
$$\eta^{(g)} = \alpha^{(g)} + \Gamma^{(g)} \xi^{(g)} + \zeta^{(g)}$$

Where g refers to groups (1) offline and (2) online; $x^{(g)}$ is a 13*1 vector with observed independent scores on the items reflecting the formative facets of Satisfaction (Convenience, Offerings, Quality, Transaction, Fair Price; see Table 1); $\tau_x^{(g)}$ is a 13*1 vector with measurement intercepts for the same items; $\Lambda_x^{(g)}$ is a 13*5 matrix with factor loadings; $\xi^{(g)}$ is a 5*1 vector with independent latent scores for the formative facets of satisfaction; $\delta^{(g)}$ is a 13*1 vector with residuals; $y^{(g)}$ is a 2*1 vector with observed scores for the indicators of overall Satisfaction; $\tau_y^{(g)}$ is a 2*1 vector with measurement intercepts; $\Lambda_y^{(g)}$ is a 2*1 vector with factor loadings; $\eta^{(g)}$ is a 1*1 vector with latent dependent scores; $\varepsilon^{(g)}$ is a 2*1 vector with residuals; $\alpha^{(g)}$ is a 1*1 vector with the structural intercept of Satisfaction; $\Gamma^{(g)}$ is a 1*5 vector with regression weights; $\zeta^{(g)}$ is a 1*1 vector with the disturbance term for Satisfaction. Moreover, $\kappa^{(g)}$ is a 5*1 vector with the means of $\xi^{(g)}$; $\phi^{(g)}$ is a 5*1 vector representing the variances of $\xi^{(g)}$; $\Psi^{(g)}$ is a 1*1 vector representing the variance of $\zeta^{(g)}$.

In all models and both groups, one loading per factor is fixed to one. Also, in all models and both groups the intercepts of one indicator per factor are fixed to zero. Below, $\tau_{x\text{ - }3,5,8}^{(1)}$ refers to vector τ_x not including its 3rd, 5th and 8th element.

Model	Constrained parameters
A. Unconstrained model:	No additional constraints
B. Metric invariance:	$\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$
C. Full scalar invariance:	$\tau_x^{(1)} = \tau_x^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$
D. Partial scalar invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$
E. Structural intercept invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\alpha^{(1)} = \alpha^{(2)}$
F. Structural residual invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\Psi^{(1)} = \Psi^{(2)}$
G. Means invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\kappa^{(1)} = \kappa^{(2)}$
H. Structural variance invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\phi_{ii}^{(g)} = \phi_{ii}^{(g)}$
I. Structural covariance invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\phi_{ij}^{(g)} = \phi_{ij}^{(g)}$ for i =/= j
J. Structural regression invariance	$\tau_{x - 3,5,8}^{(1)} = \tau_{x - 3,5,8}^{(2)}$; $\tau_y^{(1)} = \tau_y^{(2)}$; $\Lambda_x^{(1)} = \Lambda_x^{(2)}$; $\Lambda_y^{(1)} = \Lambda_y^{(2)}$ $\Gamma^{(1)} = \Gamma^{(2)}$