

An Assessment of Measurement Invariance between Online and Mail Surveys

Elisabeth Deutskens¹

Department of Marketing, University of Maastricht, The Netherlands

Ko de Ruyter

Department of Marketing, University of Maastricht, The Netherlands

Martin Wetzels

Department of Organization Science and Marketing, ECIS, Eindhoven University of
Technology, The Netherlands

¹ **Corresponding Author: Elisabeth Deutskens, ec.deutskens@mw.unimaas.nl**
Address: University of Maastricht, Faculty of Economics and Business Administration,
Department of Marketing, P.O. Box 616, 6200 MD Maastricht (The Netherlands), Tel.:
+31-43-3883716 Fax: +31-43-3884918

An Assessment of Equivalence between Online and Mail Surveys

Abstract

One of the latest trends in marketing research is the increasing use of online surveys, which offer lower costs and faster responses. Yet, critics question whether data collected via online surveys are equivalent to data collected via traditional mail surveys. Since existing evidence from the comparison of Web-based and paper-and-pencil surveys is inconclusive, we empirically examine the equivalence of online and traditional mail surveys in a marketing context.

Keywords: Online Marketing Research, Measurement Invariance, Response Quality

One of the latest trends in marketing research is the increasing use of Internet-based surveys (Couper 2000; Green, Johnson, and Neal 2003; Kiesler and Sproull 1986; Schaefer and Dillman 1998). Green et al. (2003, p.6) even state that “By the mid-1990s, [...] Web-administered surveys had become one of the most widely used ways to interview respondents”. Whereas in 1996 \$3 million were spent on commercial online research efforts in the US, this number has grown to \$0.5 billion at the end of 2001 (Hogg 2002). Also, academics increasingly use the Internet to collect data (e.g., Iqbal, Verma, and Baran 2003; Lynch and Ariely 2000; Mandel and Johnson 2002; Meuter et al. 2000; Novak, Hoffman, and Yung 2000; Toubia, Hauser, and Simester 2004).

Internet-based surveys offer great advantages over traditional mail surveys, such as lower costs, faster response, and higher quality data (e.g., Green, Johnson, and Neal 2003; Illieva, Baron, and Healey 2002; Schuldt and Totten 1994). However, one of the core concerns with respect to Internet-based research has not yet been addressed adequately (Epstein et al. 2001): are data collected via online surveys equivalent to data collected via traditional mail surveys? Evidence from the comparison of telephone and online surveys for example verifies that online and telephone surveys exhibit a different underlying factor structure (Roster et al. 2004). Existing evidence on this topic from the comparison of Web-based and paper-and-pencil surveys is scarce and inconclusive. Some studies indicate more socially desirable answers and more extreme responses (Kiesler and Sproull 1986), higher item completion (Klassen and Jacobs 2001), higher item response (Shermis and Lombard 1999), higher item variability, fewer missing values, or higher measurement error (Stanton 1998) in online surveys. In contrast, other studies ascertain equivalence between the two methods (Epstein et al. 2001; Knapp and Kirk 2003). Extrapolating equivalence from the previously mentioned studies is risky as most of these studies have methodological limitations and conduct only limited statistical comparisons. Since the success of Internet-based research will depend on

the soundness of psychometric properties in online surveys, it is imperative that the equivalence of online and traditional methods is examined critically. In fact, we will continue to conduct online surveys ‘in the dark’ until we know whether on- and offline data are indeed comparable. Mail surveys were chosen as a point of reference since they are the most common research method nowadays (Dillman 2000).

Therefore, the main purpose of this study is to empirically examine the equivalence of Internet-based and traditional mail surveys in a marketing context. In addition, we would like to call on researchers conducting multi-mode surveys to test the assumption of measurement invariance of on- and offline surveys. To address this critical issue and examine equivalence, we use a service quality survey from a large office equipment manufacturer, which has been administered via a mail and online questionnaire. Equivalence is assessed with a multigroup confirmatory factor analysis model (Jöreskog 1971; Vandenberg and Lance 2000).

The remainder of the article is structured as follows: firstly, we review existing empirical evidence on the equivalence of online and mail surveys. Next, we describe the study that was conducted to illustrate how the equivalence of on- and offline surveys is assessed. Finally, we conclude this paper with a discussion and the theoretical and practical implications of our findings.

1. Literature Review

Persuasive benefits of online surveys over traditional methods include lower costs, faster response, and a wide geographic reach (e.g., Green, Johnson, and Neal 2003; Illieva, Baron, and Healey 2002; Schuldt and Totten 1994). Furthermore, the Internet allows simpler directions (e.g., through automatic routing), as well as richer and more interesting question formats (Klassen and Jacobs 2001; Simsek and Veiga 2001). Yet, online and mail surveys might produce different results. Kiesler and Sproull (1986) for example find more socially

desirable answers online, which could be explained by privacy concerns as well as attitudes and beliefs about confidentiality of data on the Internet. Contradictory, they also find more extreme responses, which would indicate that respondents feel higher levels of perceived anonymity and thus are more open and outspoken in an online environment (Alonzo and Aiken 2004). Positive findings of higher item completion (Klassen and Jacobs 2001), higher item response, and fewer missing values (Shermis and Lombard 1999) in online surveys could be caused by the difference in the processes of responding to online and mail surveys. Online respondents are mostly not able to scan, preview, review, skip or change items. Higher item variability, higher measurement error (Stanton 1998), or a potential influence of self-presentation bias in online surveys (Epstein et al. 2001) could be explained by a variety of other reasons. For example, computer anxiety might affect participants' responses (Buchanan and Smith 1999) or biases could occur in the way people perceive questions on a screen or on paper. Different screen formats and otherwise inconsistent computer administration as well as technical or interface problems can elicit different responses to an online survey. Also, evidence from the comparison of telephone and online surveys actually find differences in the factor analysis, where the online survey has a much simpler underlying structure (Roster et al. 2004). In contrast to the previously mentioned evidence, some studies find that online and mail surveys are equivalent (Epstein et al. 2001; Knapp and Kirk 2003). In addition to inconclusive results, most previous studies that compare online and mail surveys have methodological limitations. For example, surveys are based on small samples, target populations with strong familiarity with the technology (e.g., student or academic samples) or self-selected convenience samples. Knapp and Kirk (2003) also use a different recruitment method, e.g., instead of sending an e-mail with the hyperlink to the web questionnaire, an envelope with the URL was handed out to respondents. This requires more effort from the respondent and is more time consuming and complicated. They also had

technical problems since the questionnaire was offline for 27 hours (Knapp and Kirk 2003). Epstein et al. (2001) conducted their study in a highly controlled environment. Furthermore, most studies are conducted in research areas different from marketing, such as psychology or public opinion research (e.g., Buchanan and Smith 1999; Kiesler and Sproull 1986; Stanton 1998; Wright, Aquilino, and Supple 1998). Finally, some studies provide contradictory results (e.g., Kiesler and Sproull 1986) and are limited both in the quantity and methodological quality. Most articles only compare the means of the online and offline groups. Since both means are just an approximation of the underlying true population mean, it is impossible to say which one is better and how much importance should be attached to small but significant differences.

Despite the prominent evidence for non-equivalence, we hypothesize that data collected via an online and mail survey are equivalent. First, later studies find equivalence between the two methods (Epstein et al. 2001; Knapp and Kirk 2003), indicating that factors such as computer anxiety or privacy concerns are reduced as people become more familiar with the Internet. Respondents also become more experienced in filling out online surveys, and it is thus unlikely that the response process or the way people perceive questions on a screen or on paper cause differences. Even though online communication in general was found to be more open, there should be no difference in perceived anonymity between online and mail questionnaires as both are filled out in the absence of an interviewer. Furthermore, the number of online panel members or e-mail addresses available in other databases increases steadily, leading to lower coverage problems. As there is a growing body of literature on the best design of online surveys (e.g., Couper, Traugott, and Lamias 2001; Schaefer and Dillman 1998), problems due to different screen formats and other technical or interface problems are also reduced. Therefore, we hypothesize that online and mail surveys are equivalent.

2. An Empirical Study

In order to test the equivalence between mail and online surveys, we conducted a service quality survey with a major multinational office equipment manufacturer. The data were collected in the US, which has the highest Internet acceptance rate worldwide (NUA Internet Surveys 2002). Customers in the traditional paper-and-pencil group received a mailing containing an introduction letter, the questionnaire, and a pre-paid return envelope.

Customers in the online group received an e-mail invitation for this research, including a short introduction to the study with a request to participate and the hyperlink to the web questionnaire. With one click on this link, respondents were directed towards the questionnaire. By using a unique 8-digit ID for each respondent, double entries could be avoided.

The items used to assess equivalence measured the service call quality, service visit quality, and the intentions to use the services of this provider again (see Table 1). The items were strongly driven by the SERVQUAL dimensions developed by Parasuraman et al. (1988) and have been used in earlier studies on service contact modes (Van Birgelen et al. 2002). All items were measured on a 9-point Likert-scale, ranging from '1 - much worse than expected' to '9 - much better than expected' for the service call quality and service visit quality, and from '1 - very unlikely' to '9 - very likely' for the intentions.

3. Results

Participants for both the online and offline group are recruited from the company's customer database. A stratified sampling procedure (business units, regions, product type) is used, followed by a random sample that is drawn from each group in order to make sure that a valid and representative sample of customers is obtained. 694 (16.58%) customers have responded

to the mail survey, while 255 (28.47%) customers have participated in the online questionnaire. The smaller sample for the online survey reflects the common problem that customer databases do not yet contain all customer e-mail addresses.

Measurement invariance is assessed following the procedure recommended by Vandenberg and Lance (2000). Increasingly restrictive hypotheses are tested to examine invariance, starting with the analysis of configural, then metric, scalar, factor covariance, and finally factor variance invariance. In line with Steenkamp and Baumgartner (1998), we also test for error variance invariance. The majority of SEM applications in the behavioral sciences employ the maximum likelihood (ML) estimation procedure to provide parameter estimates for the hypothesized models (Bollen 1989; Breckler 1990). However, the ML estimator exhibits desirable statistical properties (being unbiased, consistent, asymptotically efficient and approximating a χ^2 distribution) only if several important assumptions are met (Bollen 1989; West, Finch, and Curran 1995). Chief among these is the assumption that the manifest variables follow a multivariate normal distribution. However, this assumption frequently does not hold in behavioral research (Micceri 1989), as for example in most customer satisfaction and service quality research (Brown, Churchill Jr., and Peter 1993; Peterson and Wilson 1992). This is also the case in our study, where we find significant deviations from multivariate skewness and kurtosis in both the online and mail version (DeCarlo 1997). As far as univariate normality is concerned, we find no severe deviations from skewness and kurtosis. Both for the online sample and the offline sample none of the fourteen variables show significant deviation from univariate normality with respect to kurtosis (using a Bonferroni correction for the number of variables). For skewness, we find that for the offline sample twelve out of fourteen variables show excessive, negative skewness; for the online sample four out of fourteen variables show excessive, negative skewness. To assess multivariate normality, univariate normality is a necessary but not sufficient condition. We

tested for multivariate normality using Srivastava's multivariate test of skewness and kurtosis (Srivastava 1984) and Mardia's test of multivariate kurtosis (Mardia 1970). For multivariate kurtosis both Mardia's and Srivastava's test indicate a significant deviation from multivariate normality (for both tests $p < 0.001$). Similarly, Srivastava's test of multivariate skewness shows a significant deviation from multivariate normality ($p < 0.001$). As a consequence of the deviation from multivariate normality, the χ^2 statistic does no longer provide an adequate assessment of model fit and leads to an underestimation of the standard error of the estimates in confirmatory factor analysis models (Curran, West, and Finch 1996; Hu, Bentler, and Kano 1992; West, Finch, and Curran 1995). Several alternatives to ML in case of nonnormality have been proposed (cf. West, Finch, and Curran 1995). Basically, two alternative estimation procedures have received widespread attention in the extant literature and have been employed in several simulation studies: (1) the asymptotically distribution free (ADF) estimator (Browne 1984) and the Satorra-Bentler -scaled χ^2 statistic (χ^2_{SB}) with robust standard errors (Satorra and Bentler 1994). As far as the ADF estimation procedure is concerned, two major shortcomings for practical applications can be identified. First, the ADF estimator requires a relatively large sample size; Curran et al. (1996) report that a sample size of 1000 might be required for relatively simple models under typical conditions of nonnormality, while Hu et al. (1992) indicate that a sample size of 5000 might be needed with more complex models and under conditions of severe nonnormality (or both). Second, the ADF estimator is computationally quite intensive, even with the current generation of PC's (Bentler 1990; West, Finch, and Curran 1995). Several studies using Monte Carlo simulations found that the SB-scaled χ^2 statistic with robust standard errors outperformed ADF, especially at smaller sample sizes under nonnormality, and even under normality conditions (Chou, Bentler, and Satorra 1991; Curran, West, and Finch 1996; Hu, Bentler, and Kano 1992).

The SB-scaled χ^2 statistic cannot be straightforwardly employed as a difference test for nested models as compared to ML χ^2 statistic. Satorra & Bentler (2001) suggest that a difference test for the SB-scaled χ^2 statistic can be implemented as follows:

$$\Delta\chi_{SB}^2 = \frac{\Delta\chi^2}{c_d}, \text{ where}$$

$$\Delta\chi^2 = \chi_1^2 - \chi_2^2, \text{ and}$$

$$c_d = \frac{df_1 c_1 - df_2 c_2}{df_1 - df_2}, \text{ where}$$

$$c_1 = \frac{\chi_1^2}{\chi_{SB,1}^2}, \text{ and}$$

$$c_2 = \frac{\chi_2^2}{\chi_{SB,2}^2}$$

We have employed the procedure outlined above to test for the difference test for the SB-scaled χ^2 statistic.

We use EQS 6.1 to analyze the data, and first specify a baseline model for both the online and the offline sample, starting with 14 items. Yet, the fit indices are modest for both the offline ($\chi_{SB}^2(74) = 283.700$; NFI = 0.917; CFI = 0.937; RMSEA = 0.099) and the online data ($\chi_{SB}^2(74) = 180.448$; NFI = 0.881; CFI = 0.925; RMSEA = 0.122). Therefore, we retain 10 items for the subsequent analysis. This resulted in a good fit for both groups: offline $\chi_{SB}^2(32) = 77.637$; NFI = 0.984; CFI = 0.991; RMSEA = 0.045 and online $\chi_{SB}^2(32) = 84.921$; NFI = 0.959; CFI = 0.974; RMSEA = 0.081. Reliability analysis provides evidence for internal consistency in terms of composite reliability and average variance extracted (see Table 1). Composite scale reliability ranges from 0.96 to 0.98. All values exceed the cut-off value of 0.7 as suggested by Nunnally and Bernstein (1994). Average variance extracted ranges from

0.89 to 0.93, exceeding the 0.5 cut-off value proposed by Fornell and Larcker (1981).

Additionally, discriminant validity is assessed with a Satorra-Bentler χ^2_{SB} difference test (Satorra and Bentler 2001), showing discriminant validity for all constructs at $p < 0.001$.

[PLEASE INSERT TABLE 1 ABOUT HERE]

Next, we test whether the means for the online and offline group are significantly different.

With $\chi^2_{SB}(10) = 10.405$, $p=0.406$, there are no significant differences in means between the online and offline group. Testing for the differences in the variance-covariance matrix

($\chi^2_{SB}(55) = 52.858$, $p=0.557$) and simultaneously for the means and variance-covariance

matrix ($\chi^2_{SB}(65) = 62.453$, $p=0.567$) also results in a good model fit. Even though the means

are comparable across modes, it can be that the underlying factors structures are still different

(Byrne 1994). The simple analysis of mean score differences is only meaningful if

measurement equivalence is present (King Jr. and Miles 1995; Vandenberg 2002). Therefore,

we use the more rigorous, powerful and versatile multigroup confirmatory factor analysis to

assess equivalence (Steenkamp and Baumgartner 1998). We test for measurement invariance

using a hierarchical ordering of six nested models. The first model tests configural

invariance, which compares whether the pattern of salient and nonsalient factor loadings are

equal across modes. Our results indicate that the data fit well with the hypothesized model

($\chi^2_{SB}(64) = 162.092$; NFI = 0.977; CFI = 0.986; RMSEA = 0.040), which means that full

configural invariance can be established. Then, we test a second model for metric invariance,

which assesses whether factor loadings are equal across groups. Our results indicate again

that the data fit well ($\chi^2_{SB}(71) = 173.966$; NFI = 0.975; CFI = 0.985; RMSEA = 0.039) and

also the Satorra-Bentler χ^2 difference test (Satorra and Bentler 2001) is not significant

($p=0.222$). Next, we test for scalar invariance, resulting in a good model fit ($\chi^2_{SB}(81) =$

192.631; NFI = 0.974; CFI = 0.984; RMSEA = 0.038) and an insignificant χ^2_{SB} difference

test ($p=0.124$). To confirm the signs of equivalence, we test a fourth model with invariant factor covariances across groups. Our results support this with a good model fit ($\chi^2_{SB} (84) = 194.801$; NFI = 0.974; CFI = 0.984; RMSEA = 0.037) and a p-value of 0.909 for the χ^2_{SB} difference test. Furthermore, the results for the test of full factor variance invariance indicate that the data fit well with the hypothesized model ($\chi^2_{SB} (87) = 194.733$; NFI = 0.974; CFI = 0.985; RMSEA = 0.036; $\Delta \chi^2_{SB} (3) = 0.809$, $p=0.847$). In the sixth and final model, we test for error variance invariance across modes (Steenkamp and Baumgartner 1998). Our results show that the data fit well ($\chi^2_{SB} (97) = 186.148$; NFI = 0.975; CFI = 0.987; RMSEA = 0.031) and that the χ^2_{SB} difference test is insignificant ($p=0.707$). Hence, we can conclude that the data from the online and mail survey are equivalent.

[PLEASE INSERT TABLE 2 ABOUT HERE]

4. Discussion

This study aims to extend the existing literature on the comparison of online and mail surveys, using a more appropriate data analysis strategy. Specifically, we use “the most powerful and versatile approach to testing for [...] measurement invariance” (Steenkamp and Baumgartner 1998, p. 78), namely a multigroup confirmatory factor analysis model (Jöreskog 1971). Factor loadings, means, factor covariances, factor variances, and error variances of both the online and mail survey are compared and found to be invariant, and also the means and variance-covariance matrix are equal across modes. This implies that scales are equally reliable across surveys and that data collected via the online survey are indeed equivalent to data collected via the traditional mail survey. Previous findings of more extreme responses, higher item variability, or higher measurement error cannot be supported.

The selection of a data collection technique is generally based on four criteria: cost, completion time, response rate and response bias (Wiseman 1972). Since the online survey is cheaper, faster, has a higher response rate and comparable psychometric properties as the mail survey, our findings provide encouraging evidence for the quality and usefulness of data collected via the Internet. In fact, the evidence for the comparability of online and mail surveys has the potential to fundamentally modify the market research industry. The cost structure of research projects will change due to the low variable costs of online surveys. This implies that more companies can afford to conduct research. The shorter response time for online surveys gives companies faster access to information, allowing them to accelerate their project time. For some years already, mail and telephone surveys have suffered from low and even declining response rates, particularly for b-2-b surveys. The higher response rate for the online sample in our study suggests that online surveys are preferable to contact busy, hard-to-reach professionals.

Our findings also support the use of multi-mode surveys. Since Internet users are still not totally representative or relevant information such as e-mail addresses are missing in customer databases, it is often impossible to reach the whole population through an online survey, which could lead to coverage error. Our findings suggest that combining online and mail surveys is possible. Internet users can participate via the cheaper and faster online survey, while the rest of the population can be questioned with a mail survey. Like this, response rates can be maximized, coverage error reduced, and survey costs optimized. Since our study verifies the equivalence of online and mail surveys, the results of this mixed mode study can be accumulated. Yet, researchers should be aware that if they are combining on- and offline data, they make assumptions about measurement invariance in testing their hypotheses. Since the assumption of equivalence can easily be tested “as extension to the basic CFA framework” (Vandenberg and Lance 2000, p. 6), we strongly suggest researchers to do so. If

not, “ violations of measurement equivalence assumptions are as threatening to substantive interpretations as is an inability to demonstrate reliability and validity” (Vandenberg and Lance 2000, p.6).

5. Limitations and Future Research Guidelines

Our study was conducted in one country, the US, only. Future studies should examine whether our findings also hold for other countries, especially if they have a lower rate of Internet adoption than the US. In this study, we analyzed measurement invariance of online and mail surveys. We decided to focus on those two modes, since the type of research for which mail surveys are used is most likely to be replaced by online surveys. Yet, it could also be interesting to examine how alternative modes such as telephone surveys compare to online surveys. Insights into these issues will advance the knowledge about the quality of online surveys and thereby help to empirically assess the potential of the Internet –based research.

Table 1: Measurement items and scale reliabilities

Variable	Item					offline	online
						offline	online
Service Call Quality	Competence of the telephone support team.	CR ^a :	0.96	0.96	M	5.636	5.541
		AVE ^b :	0.89	0.89	SD	2.049	2.207
	Feedback on when your software-related problem is being resolved.				M	5.436	5.316
					SD	2.186	2.282
		Understanding of your needs by the call-handling agent.			M	6.268	6.389
					SD	1.878	1.932
Speed of response by the call-handling agent.			M	6.368	6.2551		
			SD	1.871	1.879		
Service Visit	Ability of the service technician to solve your problem in one visit.	CR:	0.97	0.98	M	6.425	6.480
		AVE:	0.93	0.93	SD	2.1311	2.136
	The degree to which the service technician provides feedback on the progress of the service visit.				M	6.615	6.398
					SD	2.070	2.114
	Competence of the service technician.				M	6.852	6.908
					SD	2.038	1.895
	Understanding of your needs by the service technician.				M	6.773	6.888
					SD	2.028	1.838
Amount of time it takes the service technician to repair your equipment.				M	6.433	6.408	
				SD	1.98	2.050	
Intentions	I recommend to someone who seeks my advise about [...].	CR:	0.97	0.98	M	6.302	6.327
		AVE:	0.91	0.91	SD	2.156	2.200
	I encourage associates, friends, and relatives to do business with [...]				M	6.131	6.235
					SD	2.179	2.224
	I intent to do more business with [...] in the next few years.				M	6.113	6.153
					SD	2.293	2.217
I consider [...] to be my first choice for [...].				M	5.986	6.133	
				SD	2.339	2.176	

^a = Composite reliability^b = Average variance extracted

Table 2: Model Fits

	Satorra-Bentler Chi-Square	Degrees of Freedom	NFI	CFI	RMSEA	ΔX^2_{SB} (d.f.)	P- value
Offline, 14 items	283.700	74	0.917	0.937	0.099		
Offline, 10 items	77.637	32	0.984	0.991	0.045		
Online, 14 items	180.448	74	0.881	0.925	0.122		
Online, 10 items	84.921	32	0.959	0.974	0.081		
Means	10.405	10					
Variance-covariance matrix	52.858	55					
Variance covariance matrix + Means	62.453	65					
Configural invariance	162.092	64	0.977	0.986	0.040		
Metric invariance	173.966	71	0.975	0.985	0.039	9.446 (7)	0.222
Scalar invariance	192.631	81	0.974	0.984	0.038	15.222 (10)	0.124
Factor covariance invariance	194.801	84	0.974	0.984	0.037	0.545 (3)	0.909
Factor variance invariance	194.733	87	0.974	0.985	0.036	0.809 (3)	0.847
Error variance invariance	186.148	97	0.975	0.987	0.031	7.191 (10)	0.707

References

- Alonzo, Mei and Milam Aiken. 2004. "Flaming in Electronic Communication," *Decision Support Systems* 36 (3): 205-13.
- Bentler, Peter M. 1990. "Comparative Fit Indices in Structural Models," *Psychological Bulletin* 107 (2): 238-46.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Breckler, Steven J. 1990. "Application of Covariance Structure Modeling in Psychology: Cause for Concern?," *Psychological Bulletin* 107 (2): 260-73.
- Brown, Tom J., Gilbert A. Churchill Jr., and J. Paul Peter. 1993. "Improving the Measurement of Service Quality," *Journal of Retailing* 69 (1): 127-39.
- Browne, Michael W. 1984. "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology* 37: 62-83.
- Buchanan, Tom and John L. Smith. 1999. "Using the Internet for Psychological research: Personality Testing on the World Wide Web," *British Journal of Psychology* 90 (1): 125-44.
- Byrne, Barbara M. 1994. *Structural Equation Modeling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*. Thousand Oaks, California: Sage.
- Chou, Chih-Ping, Peter M. Bentler, and Albert Satorra. 1991. "Scaled Test Statistics and Robust Standard Errors for Nonnormal Data in Covariance Structure Analysis: A Monte Carlo Study," *British Journal of Mathematical and Statistical Psychology* 44: 347-57.
- Couper, Mick P. 2000. "Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly* 64 (4): 464-94.

- Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web survey design and administration," *Public Opinion Quarterly* 65 (2): 230-53.
- Curran, Patrick J., Stephan G. West, and John F. Finch. 1996. "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis," *Psychological Methods* 1: 16-29.
- DeCarlo, Lawrence T. 1997. "On the Meaning and Use of Kurtosis," *Psychological Methods* 2 (3): 292-307.
- Dillman, Don A. 2000. *Mail and Internet Surveys, The Tailored Design Method* (Second ed.). New York: Wiley.
- Epstein, Joel, W. Dean Klinkenberg, D. Wiley, and L. McKinley. 2001. "Insuring Sample Equivalence across Internet and Paper-and-Pencil Assessments," *Computers in Human Behavior* 17 (3): 339-46.
- Fornell, Claes and David F. Larcker. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* 18 (3): 39-50.
- Green, Paul E., Richard M. Johnson, and William D. Neal. 2003. "The *Journal of Marketing Research*: Its Initiation, Growth, and Knowledge Dissemination," *Journal of Marketing Research* 40 (1): 1-9.
- Hogg, Allen (2002), "Conducting Online Research." 3 (2) ed. Vol. 2003: White Paper Series, Burke Interactive.
- Hu, Li-tze, Peter M. Bentler, and Yutaka Kano. 1992. "Can Test Statistics in Covariance Structure Analysis Be Trusted?," *Psychological Bulletin* 112: 351-62.
- Illieva, Janet, Steve Baron, and Nigel M. Healey. 2002. "Online Surveys in Marketing Research: Pros and Cons," *International Journal of Market Research* 44 (3): 361-82.

Iqbal, Zafar, Rohit Verma, and Roger Baran. 2003. "Understanding Consumer Choices and Preferences in Transaction-Based e-Services," *Journal of Service Research* 6 (1): 51-65.

Jöreskog, Karl G. 1971. "Simultaneous Factor Analysis in Several Populations," *Psychometrika* 36 (4): 409-26.

Kiesler, Sara and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey," *Public Opinion Quarterly* 50 (3): 402-13.

King Jr., Wesley C. and Edward W. Miles. 1995. "A Quasi-Experimental Assessment of the Effects of Computerizing Noncognitive Paper-and-Pencil Measurements: A Test of Measurement Equivalence," *Journal of Applied Psychology* 80 (6): 643-51.

Klassen, Robert D. and Jennifer Jacobs. 2001. "Experimental Comparison of Web, Electronic and Mail Survey Technologies in Operations Management," *Journal of Operations Management* 19 (6): 713-28.

Knapp, Herschel and Stuart A. Kirk. 2003. "Using Pencil and Paper, Internet and Touch-Tone Phones for Self-Administered Surveys: Does Methodology Matter?," *Computers in Human Behavior* 19 (1): 117-34.

Lynch, Jr., John G. and Dan. Ariely. 2000. "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution," *Marketing Science* 19 (1): 83-103.

Mandel, Naomi and Eric J. Johnson. 2002. "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," *Journal of Consumer Research* 29 (2): 235-45.

Mardia, K.V. 1970. "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika* 57: 519-30.

Meuter, Matthew L., Amy L. Ostrom, Robert I. Roundtree, and Mary Jo Bitner. 2000. "Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters," *Journal of Marketing* 64 (3): 50-64.

Micceri, Theodore. 1989. "The Unicorn, the Normal Curve, and other Improbable Creatures," *Psychological Bulletin* 105: 156-66.

Novak, Thomas P., Donna L. Hoffman, and Yiu-Fai Yung. 2000. "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach," *Marketing Science* 19 (1): 22-42.

NUA Internet Surveys (2002), "Ipsos-Reid: Internet Use Climbing in Most Markets," Vol. 2004.

Nunnally, J.C. and I.H. Bernstein. 1994. *Psychometric Theory*. New York: McGraw-Hill.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry. 1988. "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing* 64 (1): 12-40.

Peterson, Robert A. and William R. Wilson. 1992. "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the Academy of Marketing Science* 20 (1): 61-71.

Roster, Catherine A., Robert D. Rogers, Gerald Albaum, and Darin Klein. 2004. "A Comparison of Response Characteristic from Web and Telephone Surveys," *International Journal of Market Research* 46 (3): 359-73.

Satorra, Albert and P. M. Bentler. 2001. "A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis," *Psychometrika* 66 (4): 507-14.

Satorra, Albert and Peter M. Bentler. 1994. "Corrections to Test Statistic and Standard Errors in Covariance Structure Analysis," in *Analysis of Latent Variables in Developmental Research*, A. von Eye, et al., Eds. Newbury Park, CA: Sage Publications.

Schaefer, David R. and Don A. Dillman. 1998. "Development of a Standard E-Mail Methodology," *Public Opinion Quarterly* 62 (3): 378-97.

Schuldt, Barbara A. and Jeff W. Totten. 1994. "Electronic Mail vs. Mail Survey Response Rates," *Marketing Research* 6 (1): 36-39.

Shermis, Mark D. and Danielle Lombard. 1999. "A Comparison of Survey Data Collected by Regular Mail and Electronic Mail Questionnaires," *Journal of Business & Psychology* 14 (2): 341-54.

Simsek, Zeki and John F. Veiga. 2001. "A Primer on Internet Organizational Surveys," *Organizational Research Methods* 4 (3): 218-35.

Srivastava, M.S. 1984. "A Measure of Skewness and Kurtosis and a Graphical Method for Assessing Multivariate Normality," *Statistics and Probability Letters* 2: 263-67.

Stanton, Jeffrey M. 1998. "An Empirical Assessment of Data Collection Using the Internet," *Personnel Psychology* 51 (3): 709-25.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research* 25 (1): 78-90.

Toubia, Oliver, John R. Hauser, and Duncan I. Simester. 2004. "Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis," *Journal of Marketing Research* 41 (1): 116-31.

Van Birgelen, Marcel, Ko de Ruyter, Ad de Jong, and Martin G.M. Wetzels. 2002. "Customer Evaluations of After-Sales Service Contact Models: An Empirical Analysis of national Culture's Consequences," *International Journal of Research in Marketing* 19 (1): 43-64.

Vandenberg, Robert J. 2002. "Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures," *Organizational Research Methods* 5 (2): 139-58.

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.

West, Stephan G., John F. Finch, and Patrick J. Curran. 1995. "Structural Equation Models with Nonnormal Variables," in *Structural Equation Modeling: Concepts, Issues and Applications*, R.H. Hoyle, Ed. Thousand Oaks: Sage Publications.

Wiseman, Frederick. 1972. "Methodological Bias in Public Opinion Surveys," *Public Opinion Quarterly* 36 (1): 105-08.

Wright, Debra L., William S. Aquilino, and Andrew J. Supple. 1998. "A Comparison of Computer-Assisted and Paper-and-Pencil Self-Administered Questionnaires in a Survey on Smoking, Alcohol, and Drug Use," *Public Opinion Quarterly* 62 (3): 331-53.