The Measurement Invariance of Customer Loyalty and Customer Experience across Firms, Industries, and Countries

Gravelle, Timothy B.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Gravelle, T. B. (2021). The Measurement Invariance of Customer Loyalty and Customer Experience across Firms, Industries, and Countries. *Methods, data, analyses : a journal for quantitative methods and survey methodology (mda)*, *15*(2), 191-214. https://doi.org/10.12758/mda.2021.01

methods, data, analyses | 2021, Vol. 15(2), pp. 191-214 DOI: 10.12758/mda.2021.01

The Measurement Invariance of Customer Loyalty and Customer Experience across Firms, Industries, and Countries

Timothy B. Gravelle
SurveyMonkey

Abstract

Research on cross-national (and cross-group) measurement invariance is now well developed in the social and behavioural sciences, but this research has yet to engage research practitioners whose focus is measuring and modelling customer loyalty and customer experience. This is a notable gap in existing research on cross-group comparisons, especially considering the reliance of business decision-makers on customer feedback. Though standard measures of customer experience and loyalty are used in every industry, their measurement invariance across industries has not been subject to extensive testing. This article brings current thinking about cross-group comparisons and modern tools of multi-group confirmatory factor analysis (MGCFA) to the measurement of customer loyalty and customer experience across firms, industries, and countries, drawing on original large-scale survey data from the United States, United Kingdom, and Canada.

Keywords: Customer research; measurement invariance; confirmatory factor analysis; United States; United Kingdom; Canada

Methods for collecting and analysing cross-national and cross-group survey data have advanced considerably in the past two decades (Harkness et al., 2010; Harkness, van de Vijver, & Mohler, 2003; Johnson, Pennell, Stoop, Ineke, & Dorer, 2019). A critical component of comparative survey research is the assessment of measurement invariance, also called measurement equivalence. Measurement invariance refers to the notion that survey-based measures capture the same underlying constructs in different groups, and thus survey estimates for these groups offer a valid basis for comparison (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014). Indeed, comparative research in the social and behavioural sciences has been seized by the question of measurement invariance. There is now a wide-ranging, multi-disciplinary literature on the cross-national measurement invariance (or non-invariance) of core values (Cieciuch, Davidov, Vecchione, Beierlein, & Schwartz, 2014; Zercher, Schmidt, Cieciuch, & Davidov, 2015), personality traits (Marsh, Nagengast, & Morin, 2013), and attitudes toward a wide range of political concepts and policy issues such as support for democracy (Ariely & Davidov, 2011), the welfare state (Stegmueller, 2011) and foreign policy attitudes (Gravelle, Reifler, & Scotto, 2017, 2020).

Existing research on measurement invariance has thus tended to focus on concepts of interest to academic sociologists, political scientists, and psychologists. Consequently, the substantive focus of this literature has little engaged researchers and practitioners focused on measuring, modelling, and comparing customer feedback, which as a seminal text on survey research methods observes, is a core application of survey research (Dillman, Smyth, & Christian, 2014, pp. 462–463). This is not to say that customer research practitioners have failed to consider the issue of measurement invariance. Existing customer research nevertheless has marked limitations, having examined only single industries (Ueltschy, Laroche, Tamilia, & Yannopoulos, 2004), single countries (Fornell, Johnson, Anderson, Cha, & Everitt Bryant, 1996; Klaus & Maklan, 2013), or considered concepts with a narrow remit such as consumer ethnocentrism (Steenkamp & Baumgartner, 1998) instead of central constructs like customer loyalty and customer experience. Indeed, the premise of long-running, cross-industry measures of customer sentiment such as the American Customer Satisfaction Index (Fornell et al., 1996) is that such inter-firm and

Acknowledgements

The author thanks Jack Chen, Sarah Cho, Jon Cohen, Austin Pettis, Caroline Queny, and Brett Silverman at SurveyMonkey for advice and assistance with survey design, translation, and data collection. They power the curious every day. The statistical analysis and discussion also greatly benefitted from the constructive comments and suggestions provided by the two anonymous reviewers.

Direct correspondence to

Timothy B. Gravelle, Senior Manager, Research Science, SurveyMonkey, Canada
E-mail: tgravelle@surveymonkey.com

inter-industry comparisons yield comparable scores and serve as valid benchmarks. Still, the validity of such measures across industries and firms is assumed rather than tested. This working assumption persists even though there is a prima facie argument for the *in*comparability of measures of customer sentiment: durable consumer goods, travel, hospitality, retail shopping and financial services imply qualitatively different (and potentially incommensurate) customer experiences. Despite this, existing research has not presented simultaneously cross-national *and* cross-industry evidence of the measurement invariance of customer sentiment using modern confirmatory factor analysis tools – the preferred approach for testing measurement invariance – to validate this working assumption (cf. Yu & Yang, 2015).

To advance the current state of customer survey research, this article brings current thinking about cross-group comparisons and modern tools of multi-group confirmatory factor analysis (MGCFA) to the measurement of customer loyalty and customer experience across countries and industries. It draws on large-scale survey data from the United States, United Kingdom, and Canada with measures of customer loyalty and experience for firms in multiple industries to assess the cross-group measurement invariance of customer sentiment, with "group" defined here as, alternately, firms, industries, and countries. In brief, it finds support for strict measurement invariance of customer loyalty and customer experience across firms (or brands) and industries, as well as across the countries studied.

Measuring Customer Loyalty and Customer Experience

The existing research literature on customer sentiment and customer behaviour is vast, offering up a veritable cacophony of competing theoretical models and empirical measures. Indeed, marketing research and management consulting firms are the same, with every firm advancing its own perspective on the optimal questions to gauge customer sentiment, and that are meant to serve as antecedents of customer behaviours of interest: customer retention, repeat purchasing, and share of wallet (e.g., Reichheld, 2003; Yu & Yang, 2015).

Customer loyalty has been described as a favourable attitude toward a brand that differentiates it from competing brands (Dick & Basu, 1994), and as the composite of beliefs, affect, and intentions toward a brand (Oliver, 1999). Defined in this way, customer loyalty refers to *attitudinal* loyalty, and is distinguishable from *behavioural* loyalty, which refers to repeat patronage or repeat purchasing (Watson, Beck, Henderson, & Palmatier, 2015). Only the former is strictly a survey-based measurement, while the latter may be measured using business operational data or as a self-reported behaviour in a survey setting (allowing for some measure-

ment error). A closely related concept, customer satisfaction, is understood as the alignment between initial customer expectations and firm performance (Fornell et al., 1996). Different approaches to customer sentiment conceive the relationship between customer satisfaction and customer loyalty differently. In some models, customer satisfaction and customer loyalty are conceived as distinct concepts (Dick & Basu, 1994; Fornell et al., 1996). Still other models – especially those current in applied customer research – subsume customer satisfaction under customer loyalty, treating satisfaction as an indicator of loyalty, along with self-reported measures of one's likelihood to recommend a brand (word-of-mouth intention) and likelihood to repurchase a brand (repurchase intention). Empirical analyses indicate that measures of customer satisfaction, likelihood to recommend, and likelihood to repurchase measure the same underlying concept (Yu & Yang, 2015).

Customer experience can be defined as customers' internal affective, cognitive, emotional, and sensorial responses to engagement with a brand (Brakus, Schmitt, & Zarantonello, 2009; Lemon & Verhoef, 2016). A variety of measures have been proposed as capturing different facets of customer experience. These include perceptions of the quality of service interactions (across touchpoints, and at different points in the customer journey), perceptions of product design and quality, value in comparison to other brands, feeling of confidence in the brand, and feeling that the brand or firm cares about one as a customer (Klaus & Maklan, 2013; Lemon & Verhoef, 2016; Schwager & Meyer, 2007; Yu & Yang, 2015). Customer experience is thus theorised as distinct from (and an antecedent of) customer loyalty (Klaus & Maklan, 2013; Lemon & Verhoef, 2016).

These theoretical considerations informed the selection of survey items intended to measure customer loyalty and customer experience. Customer loyalty is operationalised using survey items capturing customers' overall satisfaction, likelihood to repurchase, and likelihood to recommend a brand, which aligns with industry-standard measures also used by Yu and Yang (2015). Adapting existing survey items designed to measure customer experience, it is operationalised here as customers' perceptions of a brand meeting their expectations, its value for money, comparisons to other brands, the brand delivering what it promises, and how much the brand cares about them as a customer (see, e.g., Klaus & Maklan, 2013; Yu & Yang, 2015). The measurement of customer loyalty and customer experience can thus be depicted in the two-factor model shown in Figure 1.

Data

To assess the cross-firm, cross-industry, and cross-national measurement invariance of the model of customer loyalty and customer experience depicted above,

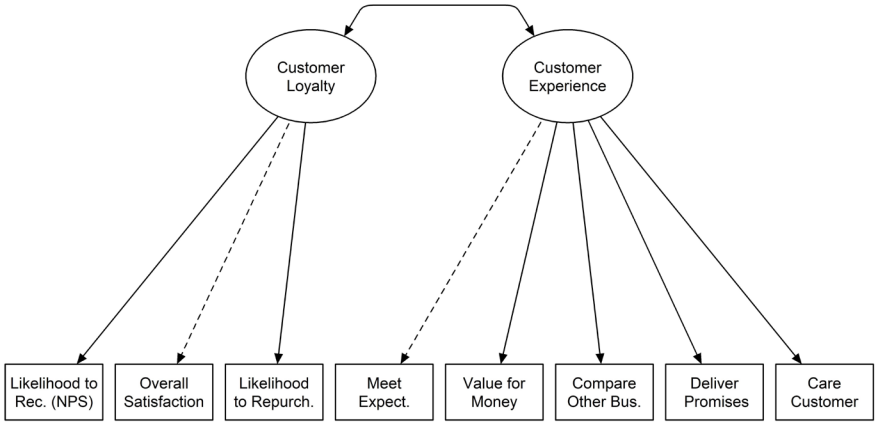

Figure 1 MGCFA Model of Customer Loyalty and Customer Experience

online surveys were conducted in the United States, United Kingdom, and Canada on 31 October–25 November 2019 using SurveyMonkey's endpage recruitment methodology. More than 2 million people around the world complete surveys designed by individuals, community organisations, and businesses using the SurveyMonkey online survey platform every day. This stream of survey respondents serves as an opportune recruitment pool for additional surveys (see, e.g., Chen, Valliant, & Elliott, 2019; Clinton, Cohen, Lapinski, & Trussler, 2021). After completing a survey on the SurveyMonkey platform, randomly selected respondents from the targeted countries (identified using their internet protocol (IP) addresses) were presented with a survey completion web page (endpage) inviting them to then complete another survey. (At the time of survey data collection, the endpage recruitment methodology was only available in these three countries.)

Five different surveys were administered in each country, each focusing on a specific industry: passenger airlines, hotels, consumer electronics, retail, and banking. These five industries represent major consumer-facing industries offering a selection of brands (or competing firms) in each national market. They also represent five of the seven industry groups represented in the ACSI: transportation, services, consumer durables, retail and finance (Fornell et al., 1996). Still, the industries and firms included in the surveys are not intended to be an exhaustive set. Respondents were first asked about their experiences (whether airline travel, hotel stays, purchases, or banking) with specific nationally leading or global companies and brands in the past 12 months. These included well-known brands such as American Airlines, British Airways, Air Canada, Hilton, Marriott, Apple, Sam-

sung, Best Buy, Wal-Mart, Bank of America, HSBC, and Royal Bank of Canada (the complete list of brands is reported in Tables A1 and A2 in the appendix). Those who reported engaging with a particular brand in the past 12 months were then asked a series of industry-standard questions measuring attitudinal customer loyalty: likelihood to recommend the brand – the widely-used "Net Promoter" question (Reichheld, 2003) – along with overall satisfaction with the brand, and likelihood to repurchase the brand. The surveys also asked about five key elements of customer experience: meeting expectations, value for money, comparisons to other brands, delivering what the brand promises, and how much the brand cares about you as a customer. All questions were asked using five-point, fully-labelled survey scales with the exception of likelihood to recommend which used the prescribed 0–10 scale, which was then recoded into the three categories specified by Reichheld (2003): "detractor" (0–6), "neutral" (7–8), and "promoter" (9–10). These eight questions are summarised in Table 1; full survey item wording appears in the appendix. The surveys were administered in English in the US and UK, and in both English and French in Canada. These samples were weighted (using weight raking) to be demographically representative of the national adult (18 years and older) populations across age, sex, region, and educational attainment categories (raking on race and ethnicity was also done in the US).

Though SurveyMonkey's endpage recruitment methodology differs from online opt-in panels, and has more in common river sampling methods, we can nevertheless calculate equivalent survey participation rates (AAPOR 2016), also called completion rates (Callegaro & Disogra, 2008), since the number of SurveyMonkey endpage views (functionally the survey invitation), click-throughs, and the number of completed surveys are all known quantities. The overall completion rate is thus calculated as 3.4 percent.

In total, 25,953 out of 41,581 respondents (or 62.4 percent) provided customer ratings, with 44,677 customer ratings collected for 60 brands across the five industries (airlines: n = 5,756; hotels: n = 6,796; consumer electronics: n = 8,347; retail: n = 16,638; banking: n = 7,140) and three countries (US: n = 12,392; UK: n = 14,974; Canada: n = 17,311). The number of ratings per brand range between 109 and 3,675, and the mean number of ratings per respondent (providing at least one brand rating) is 1.72.

It is important to acknowledge that these samples were recruited in a non-probabilistic manner. While some studies comparing probability and non-probability samples have concluded that they yield different sample point estimates (Malhotra & Krosnick, 2007; Yeager et al., 2011), other studies find few substantively meaningful differences (Ansolabehere & Schaffner, 2014; Sanders, Clarke, Stewart, & Whiteley, 2007). More germane to the present study, though, is the assessment of the measurement characteristics of a set of confirmatory factor models (more on this below) than sample point estimates for individual survey items. It is also worth

Table 1 Summary of Customer Loyalty/Experience Survey Items

Customer Loyalty
Likelihood to recommend brand (0–10, recoded into 0–6 [1], 7–8 [2], 9–10 [3])
Overall satisfaction with brand (1–5)
Likelihood to repurchase brand (industry-specific wording) (1–5)
Customer Experience
Brand met expectations (1–5)
Value for money provided by brand (1–5)
How does brand compare to other brands in industry (1–5)
How often does brand deliver what they promise (1–5)
How much does brand care about you as a customer (1–5)

noting the prevalence of non-probability samples in many marketing research and customer research applications. Given this focus on the factor structure in a customer research context, the samples collected by the endpage methodology are deemed to be fit for purpose (Baker et al., 2013).

Methods

Following from the conceptualisation of customer loyalty and customer experience presented in Figure 1 above, the models tested comprise two latent variables (or factors) corresponding to these two overarching concepts. In line with prevailing practice for testing measurement invariance, this two-factor model is analysed in a multi-group confirmatory factor analysis (MGCFA) framework (Davidov et al., 2014). As MGCFA is part of a broader structural equation modelling (SEM) framework, confirmatory factor analysis (CFA) differs from the more widely-employed exploratory factor analysis (EFA) by requiring the modeller to specify a factor model and the items measuring (i.e., that "load on") a given factor (Kline, 2016).

The typical practice for testing measurement invariance involves moving through a sequence of nested, increasingly constrained model specifications reflecting higher degrees of invariance while assessing overall model fit. *Configural invariance* is achieved when all groups have the same salient (non-zero) and non-salient (near-zero) factor loadings; no cross-group equality constraints are imposed. Configural invariance allows us to conclude that the same latent constructs exist

in all groups, but formal cross-group comparisons (e.g., of mean scores on those constructs) cannot be made. *Metric invariance* (or weak measurement invariance) requires good model fit while constraining factor loadings to be equal across groups. This allows for regression coefficients (e.g., structural relations between latent constructs) to be meaningfully compared between groups. *Scalar invariance* (or strong measurement invariance) requires good model fit while constraining factor loadings as well as item intercepts (or thresholds) to be equal across groups. This allows latent variable means meaningfully compared across groups (Davidov et al., 2014; Steenkamp & Baumgartner, 1998). Error variances (also called measurement residuals or residual variances) can be constrained to equality to test error variance invariance. *Error variance invariance* (or strict measurement invariance) allows us to conclude that a set of items serve as equally reliable indicators of the latent constructs in all groups (Steenkamp & Baumgartner, 1998). The forms of measurement equivalence described above can be subsumed under the heading of *exact* measurement invariance. Recent extensions of measurement invariance testing have investigated more flexible alternatives to exact measurement invariance, including *approximate* measurement invariance (Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014) and the *alignment method* (Asparouhov & Muthén, 2014). Given the focus here on making clear, direct comparisons of customer sentiment across countries, industries, and brands, the analysis here remains focused on exact measurement invariance.

The modelling approach pursued here thus entails testing in turn configural, metric, scalar, and scalar plus error variance invariance. Separate sets of MGCFA models are also fit where the groups comprise the 60 brands (or firms), five industries, and three countries, thus testing different levels of measurement invariance across different dimensions: brands, industries, and national contexts. Several of the ordinal survey items exhibit significant skew, with a preponderance of high scores reflecting positive brand experiences. Given the skew exhibited by the data, treating the survey data as continuous and estimating the MGCFA models by maximum likelihood is not advised (Rhemtulla, Brosseau-Liard, & Savalei, 2012). The survey data are therefore treated as categorical, and all models are estimated by robust weighted least squares (Finney & DiStefano, 2013) using M*plus* version 8.1 (Muthén & Muthén, 2017).

Concerning the assessment of overall model fit, the SEM and CFA literatures distinguish between measures of absolute fit – in particular, the model chi-square statistic – and approximate fit indices. Given the very large sample employed here, experienced structural equation modellers would expect model chi-squares to have little utility in practice: with large sample sizes, chi-square statistics routinely indicate model misfit for otherwise acceptable models. Approximate fit indices – particularly the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR) – are therefore

more useful in assessing model fit (Kline, 2016). Widely-used cut-off values for these fit indices were proposed by Hu and Bentler (1999); these are CFI $\geq$ 0.95, RMSEA $\leq$ 0.05 (or $\leq$ 0.06), and SRMR $\leq$ 0.08. An earlier proposal of Browne and Cudeck (1993) proposed RMSEA $\leq$ 0.08 as indicating acceptable model fit. Several authors, however, have argued against rigid cut-off values (regardless of the values chosen) or reliance on any single model fit index (e.g., F. Chen, Curran, Bollen, Kirby, & Paxton, 2008).

In addition to considering overall model fit, the methodological literature on measurement invariance also provides guidelines on *incremental* (or relative) model fit – that is, the change in model fit from a less constrained to a more constrained model (e.g., from configural invariance to metric invariance, or from metric invariance to scalar invariance). Initial work by Chen (2007) based on models comprising a single factor, two groups, and estimation by maximum likelihood proposed the following guidelines for the permissible change in the model fit indices: ΔCFI $\geq$ -0.01, ΔRMSEA $\leq$ 0.015, and ΔSRMR $\leq$ 0.03. Extending this work to consider multiple factors, several groups, and categorical indicators (as in the present study), Rutkowski and Svetina (2017) have proposed ΔRMSEA $\leq$ 0.05 for metric equivalence and ΔRMSEA $\leq$ 0.01 for scalar equivalence, while advising against the use of ΔCFI on account of its poor performance (their study did not examine ΔSRMR).

Accordingly, the approach to assessing model fit employed here entails examining the CFI, RMSEA, and SRMR model fit indices together to assess overall fit while also examining incremental fit, focusing on ΔRMSEA and ΔSRMR using current guidelines. In all cases, proposed cut-off values (for both overall and incremental model fit) are used as guides as opposed to rigid rules.

Results

Examining the MGCFA results for the analyses by brand, industry, and country, the model chi-square statistics (as expected) would lead one to conclude that none of the models achieve good fit (see Table 2). By contrast, the CFI, RMSEA, and SRMR model fit indices suggest good (or at a minimum, acceptable) fit.

For the MGCFA conducted with the 60 measured brands comprising the groups, the configural and metric invariant models both achieve good overall fit based on CFI ($\geq$ 0.95) and SRMR ($\leq$ 0.08); RMSEA indicates acceptable fit ($\leq$ 0.08). At a minimum, then, valid comparisons of the structural relationship between customer loyalty and customer experience can be made across firms. Still, benchmarking (i.e., comparing mean scores) on customer loyalty and customer experience metrics across firms is more typically the aim of customer research practitioners. This requires scalar invariance. The scalar invariant model again

Table 2 MGCFA Model Summary Statistics

	Model χ^2 (SB)	DF	CFI	RMSEA	(90 % c.i.)	*p* close fit	SRMR
By brand (60)							
Configural invariance	4,338.409	1,140	0.995	0.061	(0.059–0.063)	0.000	0.018
Metric invariance	4,845.274	1,494	0.995	0.055	(0.053–0.057)	0.000	0.019
Scalar invariance	10,840.100	2,674	0.988	0.064	(0.063–0.065)	0.000	0.030
Scalar + error variance invariance	12,809.688	3,138	0.986	0.064	(0.063–0.065)	0.000	0.037
By industry (5)							
Configural invariance	2,199.800	95	0.995	0.050	(0.048–0.052)	0.570	0.013
Metric invariance	2,096.521	119	0.996	0.043	(0.042–0.045)	1.000	0.013
Scalar invariance	4,139.169	199	0.991	0.047	(0.046–0.048)	1.000	0.019
Scalar + error variance invariance	4,831.688	223	0.990	0.048	(0.047–0.049)	1.000	0.023
By country (3)							
Configural invariance	1,871.708	57	0.996	0.046	(0.044–0.048)	1.000	0.011
Metric invariance	1,603.513	69	0.996	0.039	(0.037–0.040)	1.000	0.011
Scalar invariance	1,396.957	109	0.997	0.028	(0.027–0.029)	1.000	0.012
Scalar + error variance invariance	1,421.922	117	0.997	0.027	(0.026–0.029)	1.000	0.013

Note: Models are fit by robust weighted least squares. Chi-square statistics are Sartorra-Bentler scaled (mean-adjusted) chi-squares (Sartorra & Bentler, 1994).

indicates good overall fit according to the CFI (0.988) and SRMR (0.030); the RMSEA (0.064) points to increased misfit, though still yielding acceptable overall fit. The error variance invariant model is still further constrained, but still yields acceptable model fit (CFI = 0.986, RMSEA = 0.064, SRMR = 0.037). In terms of incremental model fit, the results for the metric invariance model are ΔRMSEA = -0.060 and ΔSRMR = 0.001 – well below the ΔRMSEA ≤ 0.05 and ΔSRMR ≤ 0.03 cut-offs. Change in model fit for the scalar invariance model is ΔRMSEA = 0.009 and ΔSRMR = 0.021, while change in model fit for the scalar plus error variance invariance model is ΔRMSEA = 0 and ΔSRMR = 0.007, thus meeting the ΔRMSEA ≤ 0.01 and ΔSRMR ≤ 0.03 cut-offs. Taking in the results of the overall and incremental measures of model fit, then, one can conclude that latent variable scores for customer loyalty and customer experience can be meaningfully compared across very different products and services – whether Apple or American Airlines, Best Buy or Barclays, Marriott or Marks and Spencer. Further, the three indicators of customer loyalty and five indicators of customer experience used here exhibit the same measurement properties across brands.

Not only does the MGCFA model fit across a wide variety of brands, it also fits across the five industries under study. With the groups comprised of the five industries, all model specifications – configural, metric, scalar, and scalar plus error variance invariance – achieve good overall model fit based on the guidelines for CFI (≥ 0.95), RMSEA (≤ 0.05), and SRMR (≤ 0.08) fit indices. Despite the highly constrained model specification, the error variance invariant model still indicates good model fit, with CFI = 0.990, RMSEA = 0.048, and SRMR = 0.023. The metric invariance model easily meets the guidelines for good incremental model fit with ΔRMSEA = -0.070 and ΔSRMR = 0. Change in model fit for the scalar invariance model (ΔRMSEA = 0.004, ΔSRMR = 0.006) and for the scalar plus error variance invariance model (ΔRMSEA = 0.001, ΔSRMR = 0.004) also indicate good incremental model fit. Substantively, then, latent variable scores for customer loyalty and customer experience can be compared directly across airline travel, hotel stays, consumers electronics brands, retailers, and banks.

Customer loyalty and customer experience can be similarly compared across the US, UK, and Canada. Each of the configural, metric, scalar, and scalar plus error variance invariance model exhibit good overall model fit. As with the industry groups, the highly constrained error variance invariant model achieves good model fit across countries, with CFI = 0.997, RMSEA = 0.027, and SRMR = 0.013. The metric invariance model exhibits good incremental model fit with ΔRMSEA = -0.070 and ΔSRMR = 0, as do the scalar invariant model with ΔRMSEA = -0.011 and ΔSRMR = 0.001, and the scalar plus error variance invariant model with ΔRMSEA = -0.001 and ΔSRMR = 0.001. These are important findings for large firms with global footprints and global customer bases. It implies that core customer metrics travel across the different national contexts studied. For customer

research practitioners, it similarly implies that there is little need for industry-specific customer metrics.

Looking at the standardised factor loadings for the scalar plus error variance invariance (strict measurement invariance) MGCFA models, it is worth pointing out that they are consistently high (see Table 3). Across groups by brand, industry, and country, and across all items, factor loadings range between 0.757 and 0.954. The items thus serve as good measures of customer loyalty and customer experience, respectively. It is also worth noting the standardised correlations between customer loyalty and customer experience factors, which range between 0.831 and 0.992 for specific brands; slightly narrower ranges are seen for the industry- and country-grouped models. These high correlations might suggest a lack of discriminant validity to readers accustomed to lower factor correlations, though such high correlations are common in customer research (e.g., Fornell et al., 1996). More to the point, these data pass the conventional CFA test of discriminant validity where the factor correlation is constrained to be equal to 1, implying a one-factor model. This test is highly significant ($\chi^2 = 2{,}069.515$, d.f. $= 1$, $p < 0.001$), indicating that a one-factor model has significantly worse fit than a two-factor model. A two-factor model thus remains preferable despite the high standardised correlation between the customer loyalty and customer experience factors.

Conclusion

The question motivating this article was whether widely used measures of customer loyalty and customer experience translate across, firms (or brands), industries, and countries. This question has immense practical importance for firms seeking to win and retain customers, to benchmark their performance against competitors, or to benchmark their performance in different markets. Without a rigorous basis for comparisons of customer loyalty and customer experience across competing brands, or comparisons of brand performance across countries – that is, without measurement invariance – one could truly be relying on an apples-to-oranges comparison to make critical business decisions. Indeed, a great deal of applied customer research is premised on the comparability of survey-based measures of customer sentiment, even though airline flights, hotel stays, smartphones, retail shopping, and everyday banking imply qualitatively different experiences.

The goal of this article, then, was to advance the literatures on customer research and consumer behaviour (as well as the literature on cross-national survey research more broadly) by presenting the first large-scale study of cross-firm, cross-industry, and cross-national measurement invariance of customer sentiment using MGCFA tools. The analyses presented here indicate that rigorous quantitative comparisons are in fact well-grounded. Whether examined across brands,

Table 3 MGCFA Standardised Factor Solutions (Scalar + Error Variance Invariance)

	By Brand (60)		By Industry (5)		By Country (3)	
	Customer Loyalty	Customer Experience	Customer Loyalty	Customer Experience	Customer Loyalty	Customer Experience
Factor Loadings						
Likelihood to recommend	0.829–0.921	–	0.846–0.898	–	0.861–0.881	–
Overall satisfaction	0.880–0.954	–	0.899–0.938	–	0.903–0.923	–
Likelihood to repurchase	0.828–0.920	–	0.843–0.892	–	0.855–0.876	–
Meet expectations	–	0.860–0.938	–	0.878–0.911	–	0.880–0.896
Value for money	–	0.770–0.921	–	0.830–0.881	–	0.827–0.853
Compare to other brands	–	0.846–0.931	–	0.866–0.901	–	0.862–0.884
Deliver what they promise	–	0.798–0.905	–	0.829–0.871	–	0.829–0.858
Care about you as a customer	–	0.757–0.903	–	0.814–0.866	–	0.805–0.836
Factor Means	-1.697–0.000	-1.396–0.110	0.000–0.554	0.000–0.549	-0.039–0.191	-0.043–0.197
Factor Correlation						
Customer Loyalty – Customer Experience	0.831–0.992		0.856–0.948		0.915–0.923	

Note: Factor loadings are standardised. Factor loadings for overall satisfaction (factor 1 – Customer Loyalty) and care about you as a customer (factor 2 – Customer Experience) are fixed to 1 in the unstandardised solution for model identification. Likelihood to recommend is recoded from 0–10 to 1–3.

industries, or countries, a simple two-factor model comprising customer loyalty and customer experience exhibits scalar plus error variance invariance (strict measurement invariance), providing a firm basis for cross-group comparisons. These results should be welcomed by customer research practitioners, since they imply that industry-standard measures of customer sentiment exhibit robust measurement properties.

These results should nevertheless be interpreted in light of the finite number of industries examined. Extending this research to other industries – for example, the automotive sector, insurance, computer software, restaurants, and consumer packaged goods, among others – would assist in reconfirming or qualifying the findings presented here. Similarly, the number of countries included in the analyses is finite. In particular, it is important to acknowledge that the US, UK, and Canada all comprise English-speaking majorities, meaning this study has largely set aside the question of cross-language measurement invariance (but see Yu & Yang, 2015). Testing the measurement invariance of customer loyalty and customer experience across a larger number of countries and languages, as others have done on other substantive topics (e.g., Davidov & De Beuckelaer, 2010; Gravelle et al., 2017, 2020), would be a valuable test of the model advanced here.

More broadly, this article argues that survey researchers engaged in applied customer research – perhaps employed as an in-house analyst charged with assessing their firm's position in the marketplace vis-à-vis competing brands, or as a marketing researcher consulting to a large firm with a global customer base – *should* be concerned with the question of cross-group measurement invariance, and ought to examine it explicitly instead of leaving it as an untested assumption. This article provides a demonstration of how to do so using current MGCFA techniques.

References

American Association for Public Opinion Research. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). AAPOR.

Ansolabehere, S., & Schaffner, B. F. (2014). Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis*, *22*(3), 285–303.

Ariely, G., & Davidov, E. (2011). Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research*, *104*(2), 271–286.

Asparouhov, T., & Muthén, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling*, *21*(4), 495–508.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., … Tourangeau, R. (2013). Summary Report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology*, *1*(2), 90–105.

Brakus, J. J., Schmitt, B. H., & Zarantonello, L. (2009). Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty? *Journal of Marketing*, *73*(3), 52–68.

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.

Callegaro, M., & Disogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods and Research*, *36*(4), 462–494.

Chen, J. K. T., Valliant, R. L., & Elliott, M. R. (2019). Calibrating Non-Probability Surveys to Estimated Control Totals Using LASSO, with an Application to Political Polling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(3), 657–681.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*(982), 1–10.

Cieciuch, J., Davidov, E., Vecchione, M., Beierlein, C., & Schwartz, S. H. (2014). The Cross-National Invariance Properties of a New Scale to Measure 19 Basic Human Values: A Test Across Eight Countries. *Journal of Cross-Cultural Psychology*, *45*(5), 764–776.

Clinton, J., Cohen, J., Lapinski, J. S., & Trussler, M. (2021). Partisan Pandemic: How Partisanship and Public Health Concerns Affect Individuals' Social Distancing During COVID-19. *Science Advances*, 7(2), eabd7204.

Davidov, E., & De Beuckelaer, A. (2010). How Harmful are Survey Translations? A Test with Schwartz's Human Values Instrument. *International Journal of Public Opinion Research*, *22*(4), 485–510.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, *40*(1), 55–75.

Dick, A. S., & Basu, K. (1994). Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of the Academy of Marketing Science*, *22*(2), 99–113.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.). Hoboken, NJ: Wiley,

Finney, S. J., & DiStefano, C. (2013). Non-Normal and Categorical Data in Structural Equation Modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 439–492). Charlotte, NC: Information Age Publishing.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Everitt Bryant, B. (1996). The American Customer Satisfaction Index: Nature, Purpose, and Findings. *Journal of Marketing*, *60*(4), 7–18.

Gravelle, T. B., Reifler, J., & Scotto, T. J. (2017). The Structure of Foreign Policy Attitudes in Transatlantic Perspective: Comparing the United States, United Kingdom, France and Germany. *European Journal of Political Research*, *56*(4), 757–776.

Gravelle, T. B., Reifler, J., & Scotto, T. J. (2020). The Structure of Foreign Policy Attitudes among Middle Power Publics: A Transpacific Replication. *Australian Journal of International Affairs*.

Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. P., … Smith, T. W. (Eds.). (2010). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. New York, NY: Wiley.

Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. P. (Eds.). (2003). *Cross-Cultural Survey Methods*. New York, NY: Wiley.

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

Johnson, T. P., Pennell, B.-E., Stoop, Ineke, A. L., & Dorer, B. (Eds.). (2019). *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*. Hoboken, NJ: Wiley.

Klaus, P., & Maklan, S. (2013). Towards a better measure of customer experience. *International Journal of Market Research*, *55*(2), 227–246.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York, NY: Guilford.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69–96.

Malhotra, N., & Krosnick, J. A. (2007). The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. *Political Analysis*, *15*(3), 286–323.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194–1218.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Oliver, R. L. (1999). Whence Consumer Loyalty? *Journal of Marketing*, *63*(S1), 33–44.

Reichheld, F. F. (2003). The One Number You Need to Grow. *Harvard Business Review*, (December), 1–10.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373.

Rutkowski, L., & Svetina, D. (2017). Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance. *Applied Measurement in Education*, *30*(1), 39–51.

Sanders, D., Clarke, H. D., Stewart, M. C., & Whiteley, P. (2007). Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study. *Political Analysis*, *15*(3), 257–285.

Sartorra, A., & Bentler, P. M. (1994). Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research* (pp. 399–419). Thousand Oaks, CA: Sage.

Schwager, A., & Meyer, C. (2007). Understanding Customer Experience. *Harvard Business Review*, (February), 116–128.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, *25*(1), 78–90.

Stegmueller, D. (2011). Apples and Oranges? The Problem of Equivalence in Comparative Research. *Political Analysis*, *19*(4), 471–487.

Ueltschy, L. C., Laroche, M., Tamilia, R. D., & Yannopoulos, P. (2004). Cross-cultural invariance of measures of satisfaction and service quality. *Journal of Business Research*, *57*(8), 901–912.

Watson, G. F., Beck, J. T., Henderson, C. M., & Palmatier, R. W. (2015). Building, measuring, and profiting from customer loyalty. *Journal of the Academy of Marketing Science*, *43*(6), 790–825.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, *75*(4), 709–747.

Yu, D., & Yang, Y. (2015). Measurement Equivalence of a Concise Customer Engagement Metric across Country, Language, and Customer Types. *Public Opinion Quarterly*, *79*(S1), 325–358.

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: exact vs. approximate measurement invariance. *Frontiers in Psychology*, *6*(733), 1–11.

Appendix A:

Survey Questionnaires

[AIRLINES:] In the past 12 months, which of the following airlines have you traveled on? (Please select all that apply.)

[HOTELS:] In the past 12 months, which of the following hotel chains have you stayed at? (Please select all that apply.)

[CONSUMER ELECTRONICS:] In the past 12 months, have you purchased any consumer electronics (for example, a television, desktop computer, laptop, tablet, smartphone, wearable device) from any of the following brands? (Please select all that apply.)

[RETAIL:] In the past 12 months, which of the following stores have you shopped at? (Please select all that apply.)

[BANKING:] In the past 12 months, have you done any banking (through a checking account, savings account, mortgage, or personal line of credit) with any of the following banks or financial institutions? (Please select all that apply.)

How likely is it that you would recommend [BRAND] to a friend or colleague?

0 – Not at all likely
1
2
3
4
5
6
7
8
9
10 – Extremely likely

Overall, how satisfied or dissatisfied are you with [BRAND]?

Very satisfied
Somewhat satisfied
Neither satisfied nor dissatisfied
Somewhat dissatisfied
Very dissatisfied

How likely are you to [AIRLINES: travel with BRAND again] [HOTELS: stay at BRAND again] [ELECTRONICS: purchase BRAND products again] [RETAIL: shop at BRAND again] [BANKING: continue to bank with BRAND]?

- Extremely likely
- Very likely
- Moderately likely
- Slightly likely
- Not at all likely

How well have your experiences with [BRAND] met your expectations?

- Much better than expected
- Better than expected
- About what I expected
- Worse than expected
- Much worse than expected

How would you rate the value for money provided by [BRAND]?

- Excellent
- Above average
- Average
- Below average
- Poor

How does [BRAND] compare to other [AIRLINES: airlines] [HOTELS: hotel companies] [ELECTRONICS: consumer electronics companies] [RETAIL: stores] [BANKING: banks]? Are they...?

- Much better
- Better
- About the same
- Worse
- Much worse

How often does [BRAND] deliver what they promise?

- Always
- Usually
- Sometimes
- Rarely
- Never

How much does [BRAND] care about you as a customer?

- A great deal
- A lot
- A moderate amount
- A little
- Not at all

Appendix B:

Brands Rated by Industry and Country

	United States (n = 12,392)	United Kingdom (n = 14,974)	Canada (n = 17,311)
Airlines (n = 5,756)	United States (n = 1,272)	United Kingdom (n = 1,915)	Canada (n = 2.569)
	American Airlines Delta Airlines JetBlue Airways Southwest Airlines United Airlines	British Airways EasyJet Jet2 RyanAir TUI Airways	Air Canada Air Transat Porter Airlines WestJet
Hotels (n = 6,796)	United States (n = 1,489)	United Kingdom (n = 2,126)	Canada (n = 3,181)
	Best Western Comfort Inn Hilton Holiday Inn Marriott Travelodge	Best Western Hilton Holiday Inn Marriott Premier Inn Travelodge	Best Western Comfort Inn Hilton Holiday Inn Marriott Travelodge
Consumer electronics (n = 8,347)	United States (n = 2,168)	United Kingdom (n = 3,785)	Canada (n = 2,394)
	Apple Fitbit LG Microsoft Samsung Sony	Apple Fitbit LG Microsoft Samsung Sony	Apple Fitbit LG Microsoft Samsung Sony

	United States (n = 12,392)	United Kingdom (n = 14,974)	Canada (n = 17,311)
Retail (n = 16,638)	United States (n = 5,017)	United Kingdom (n = 5,192)	Canada (n = 6,429)
	Best Buy Costco Macy's Target The Home Depot Wal-Mart	Asda Currys PC World Homebase Marks & Spencer Tesco	Best Buy Canadian Tire Costco Hudson's Bay The Home Depot Wal-Mart
Banking (n = 7,140)	United States (n = 2,446)	United Kingdom (n = 1,956)	Canada (n = 2,738)
	Bank of America BB&T Capital One Bank Chase Bank PNC Bank U.S. Bank Wells Fargo	Barclays Halifax HSBC Lloyds TSB Nationwide NatWest Royal Bank of Scotland Santander	BMO Bank of Montreal CIBC HSBC Royal Bank of Canada Scotiabank Tangerine TD Canada Trust

Appendix C:

Ratings by Brand

Brand	Countries	n =	Brand	Countries	n =
Apple	US, UK, Canada	3,675	American Airlines	US	466
Wal-Mart	US, Canada	3,195	Delta Airlines	US	466
Samsung	US, UK, Canada	2,287	Wells Fargo	US	448
The Home Depot	US, Canada	2,083	Scotiabank	Canada	440
Costco	US, Canada	1,783	CIBC	Canada	416
Tesco	UK	1,776	Barclays	UK	415
Marriott	US, UK, Canada	1,531	Macy's	US	403
Asda	UK	1,401	EasyJet	UK	393
Air Canada	Canada	1,339	Sony	US, UK, Canada	370
Hilton	US, UK, Canada	1,296	Santander	UK	370
Best Buy	US, Canada	1,291	United Airlines	US	362
Holiday Inn	US, UK, Canada	1,167	Lloyds TSB	UK	347
Canadian Tire	Canada	1,136	BMO Bank of Montreal	Canada	346
Target	US	1,078	NatWest	UK	327
Best Western	US, UK, Canada	1,047	Nationwide	UK	325
Marks & Spencer	UK	1,030	Homebase	UK	316
WestJet	Canada	884	HSBC	UK, Canada	316

Brand	Countries	n =	Brand	Countries	n =
TD Canada Trust	Canada	733	British Airways	UK	300
Microsoft	US, UK, Canada	711	Capital One Bank	US	286
Comfort Inn	US, Canada	681	RyanAir	UK	274
LG	US, UK, Canada	663	Halifax	UK	268
Hudson's Bay	Canada	652	Air Transat	Canada	222
Fitbit	US, UK, Canada	641	Jet2	UK	170
Royal Bank of Canada	Canada	630	Tangerine	Canada	142
Premier Inn	UK	548	U.S. Bank	US	136
Travelodge	US, UK, Canada	526	TUI Airways	UK	135
Southwest Airlines	US	506	PNC Bank	US	129
Currys PC World	UK	494	Porter Airlines	Canada	124
Bank of America	US	480	JetBlue Airways	US	115
Chase Bank	US	477	Royal Bank of Scotland	UK	109