The predictive ability of different customer feedback metrics for retention

de Haan, Evert; Verhoef, Peter C.; Wiesel, Thorsten

Published in:
International Journal of Research in Marketing

DOI:
10.1016/j.ijresmar.2015.02.004

Publication date:
2015

Intern. J. of Research in Marketing xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full length article

The predictive ability of different customer feedback metrics for retention ☆

Evert de Haan [a,*], Peter C. Verhoef [a,1], Thorsten Wiesel [b,c,2]

[a] Department of Marketing, Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
[b] Marketing Center Münster, Westfälische Wilhelms-Universität Münster, Am Stadtgraben 13-15, 48143 Münster, Germany
[c] Affiliate Researcher of Marketing, University of Groningen, Faculty of Business and Economics, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Article history:
First received on February 14, 2014 and was under review for 5 months
Available online xxxx

Area Editor: Koen H. Pauwels

Keywords:
Customer feedback metrics
Customer satisfaction
Net Promoter Score
Customer Effort Score
Customer retention
Firm performance

ABSTRACT

This study systematically compares different customer feedback metrics (CFMs) – namely customer satisfaction, the Net Promoter Score, and the Customer Effort Score – to test their ability to predict retention across a wide range of industries. We classify the CFMs according to a time focus (past, present, or future) and whether the full scale of the CFM is used or whether the focus is only on the extremes (e.g., top-2-box customer satisfaction). The data for this study represent customers of 93 firms across 18 industries. Multi-level probit regression models, which control for self-selection bias of respondents, investigate firm-, customer-, and industry-level effects simultaneously. Overall, we find that the top-2-box customer satisfaction performs best for predicting customer retention and that focusing on the extremes is preferable to using the full scale. However the best CFM does differ depending on industry and the unit of analysis (i.e., comparing customers or firms with one another). Furthermore, combining CFMs, along with simultaneously investigating multiple dimensions of the customer relationship, improves predictions even further.

1. Introduction

New customer feedback metrics (CFMs), as Morgan and Rego (2006) label them, including Reichheld's (2003) Net Promoter Score (NPS) and Dixon, Freeman, and Toman's (2010) Customer Effort Score (CES), are introduced frequently. These CFMs promise to be "the best" indicator of (future) firm performance, prompting leading companies in a wide range of industries to start using them (Bain and Co., 2013). Academic research challenges these promises (e.g., Keiningham, Cooil, Aksoy, Andreassen, & Weiner, 2007; Keiningham, Cooil, Andreassen, & Aksoy, 2007; Morgan & Rego, 2006); however, most studies investigate only a limited range of firms, industries, and settings, and they lack comparability, because they use different dependent variables, research settings, methodologies, units of analysis, and so on. Marketing managers thus lack guidance on which CFM to monitor and how to interpret changes in these CFMs, which can lead to uncertainty, frustration, and even abandonment of the CFMs in question. Such outcomes might lessen marketing departments' accountability and influence (Verhoef & Leeflang, 2009), hinder firms from becoming more customer centric (Shah, Rust, Parasuraman, Staelin, & Day, 2006), and negatively affect marketing-mix performance (Mintz & Currim, 2013).

This study aims to provide, for a wide range of industries, insights into the impact of different CFMs, including which (combinations of) CFM(s) a firm should monitor, how to interpret changes in CFMs, and how this differs across industries. We use actual customer retention data to compare the predictive power of various CFMs across a large number of firms and industries. Our focus on customer retention reflects three key considerations: (1) customers are among the most important marketing assets of the firm, (2) an almost one-to-one relationship exists between the value of the customer base and firm value, and (3) CFMs are frequently used as indicators of future loyalty (Gupta, Lehmann, & Stuart, 2004; Rust, Zeithaml, & Lemon, 2000). We simultaneously analyze customer-, firm-, and industry-level effects of CFMs on retention, using multi-level models to support comparisons of within-firm, between-firm, and between-industry effects. With this approach, we can provide generalizations and recommendations on which CFM(s) to monitor and how this differs across industries. We use surveys to collect CFM scores and customer background information from 6649 respondents, who in total filled out 8924 firm evaluations for 93 firms across 18 industries. In a follow-up survey two years later, filled out by 1308 respondents who provided 1375 firm evaluations (i.e., a 15.4% response rate), we measure our dependent variable, customer retention. We measure the usefulness of the CFMs in predicting

☆ Acknowledgement: This project is based on data collected through an annual survey on customer performance. The authors thank Metrixlab, MICompany, and the Customer Insights Center of the University of Groningen for data support. They also thank the participants of the 2013 GSBE Marketing-Finance Symposium in Maastricht, the participants of the 2013 INFORMS Marketing Science Conference in Istanbul, the two anonymous reviewers, the area editor, and the editor, Eitan Muller, for their valuable comments.
* Corresponding author. Tel.: +31 50 363 9621.
E-mail addresses: evert.de.haan@rug.nl (E. de Haan), p.c.verhoef@rug.nl (P.C. Verhoef), thorsten.wiesel@wiwi.uni-muenster.de (T. Wiesel).
[1] Tel.: +31 50 363 7320.
[2] Tel.: +49 251 83 28207.

http://dx.doi.org/10.1016/j.ijresmar.2015.02.004
0167-8116/

retention by analyzing both in-sample fit (with the Akaike information criterion [AIC] and Bayesian information criterion [BIC]) and out-of-sample fit (with the Gini coefficient, top-decile lift, and hit rate) for which we use a one-third holdout sample. In terms of the out-of-sample fit, the top-decile lift is an important criterion when the main goal is to identify customers most likely to churn, while the Gini coefficient and hit rate are important when the goal is to judge accuracy for all customers (i.e., make good predictions about both retainers and churners) (Blattberg, Kim, & Neslin, 2008).

Our results show that the top-2-box customer satisfaction score offers the single-best predictor of retention across industries. In general, transforming scales of the CFMs to capture the proportion of most satisfied customers (as is done with the top-2-box customer satisfaction) or splitting customers up into groups (as is done with the promoters and detractors of the NPS) is preferable to using the full scale of the CFMs. In addition, our results show that the CES in itself has little to no predictive power and performs the worst of all CFMs studied. Which CFM performs best in predicting retention is however industry dependent, and it also depends on whether the CFM is meant to be used for customer management (i.e., compare customers of the same firm with one another) or to analyze the competitive position of a firm (i.e., compare different firms in the same industry with one another). Combining metrics, especially the CES with the customer satisfaction-related CFMs, results in improved out-of-sample retention predictions. A dashboard of CFMs that measure different dimensions, as indicated in our conceptualization, is preferable to monitoring a single CFM.

Table 1 illustrates the study's contribution with a selective literature overview. This study is the first to investigate the predictive power of CFMs over three levels (i.e., customer, firm, and industry) simultaneously. In doing so, we can distinguish between the heterogeneity of customers (i.e., which CFM is most appropriate for customer management) and the heterogeneity of firms (i.e., which CFM is most appropriate for competitive positioning). Furthermore, this study is one of the first to use the official NPS, as Reichheld (2003) intended it, and to investigate the CES in line with Dixon et al.'s (2010) approach. As such, we test the ability of two recently introduced metrics that have become famous as key CFMs. We also combine CFMs to determine whether using multiple CFMs improves the predictive power, as often done in firms' dashboards. Furthermore, we predict actual future performance, in contrast with other studies that investigate only same-period correlations (e.g., Anderson, Fornell, & Mazvancheryl, 2004; Keiningham, Cooil, Andreassen, & Aksoy, 2007). In doing so, we can test the usefulness of CFMs for predictive purposes. Finally, we judge the (combination of) CFMs on their out-of-sample predictions to determine whether they have real incremental predictive power and to overcome over-fitting problems. In doing so, we can show the validity and robustness of our results.

In summary, we investigate how valuable various CFMs are in predicting retention in different situations, both to increase the academic understanding of these CFMs and to help managers select the best CFM(s) according to their situation and demands. To do so, we examine (1) the overall usefulness of different CFMs in predicting retention, (2) the differences of this usefulness between industries, (3) the differences between different units of analysis (e.g., customer- or firm-level retention for customer management and competitive analysis purposes), and (4) the incremental power of monitoring multiple CFMs over using a single CFM. These insights can help practitioners decide which (combination of) CFM(s) to use in different situations and help academics understand how valuable CFMs are and what the determining factors (e.g., industry differences, unit of analysis) for this are.

2. Conceptual background

As Gupta and Zeithaml (2006) note, it is critical to understand the relationships among CFMs, customer behavior, and firm performance. Although the positive relationship between customer satisfaction and firm performance is well established (Gupta & Zeithaml, 2006; Hanssens, 2009), a similar state does not exist for other CFMs. In this section, we classify the CFMs under study and highlight the importance of the unit (or level) of analysis.

2.1. Conceptual classification of metrics

Research in marketing has discussed a large number of metrics. Farris, Bendle, Pfeifer, and Reibstein (2006) classify these metrics as share-of-mind metrics and consider customer satisfaction and willingness to recommend a specific sub-group within these metrics. In marketing research practice, these metrics are known as customer feedback metrics (CFMs) (Morgan & Rego, 2006). These CFMs have specifically gained attention in the service and relationship marketing and customer (relationship) management literature. Given the broad number of CFMs, we distinguish between these metrics on two dimensions. The first dimension is introduced by Bolton, Lemon, and Verhoef (2004) and more recently by Zeithaml et al. (2006), who focus on the time span of measures and distinguish between more backward-looking (including the present) and more forward-looking metrics. Forward-looking CFMs focus on what customers plan to do in the future and may signal something about the future performance of the relationship. Reichheld's (2003) NPS is an example of a forward-looking CFM because it considers the willingness to recommend a firm in the future, which may also signal one's future relationship with the firm (e.g., Zeithaml et al., 2006). Backward-looking metrics focus on the past and current performance of a company toward customers. The CES is a typical backward-looking CFM because it measures the

Table 1
Literature overview on (predictive) performance of CFMs (selection).

Study	Level of analysis			CFM			Combine multiple CFMs	Predictive power	Out-of-sample prediction
	Customer	Firm	Industry	Satisfaction	NPS	CES			
Hallowell (1996)		√		√					
Ittner and Larcker (1998) (ch. 3)	√			√				√	
Ittner and Larcker (1998) (ch. 5)		√	√	√				√	
Mittal and Kamakura (2001)	√			√				√	
Anderson et al. (2004)		√	√	√					
Gruca and Rego (2005)		√	√	√				√	
Morgan and Rego (2006)		√		√	[1]			√	
Keiningham, Cooil, Aksoy, Andreassen, and Weiner (2007)	√		√	√	[1]		√	√	
Keiningham, Cooil, Andreassen, and Aksoy (2007)		√	√	√	[1]				
Rego et al. (2013)		√		√				√	
Van Doorn et al. (2013)		√		√	√			[2]	
Current paper	√	√	√	√	√	√	√	√	√

[1] A proxy question used, instead of the official NPS question developed by Reichheld (2003).
[2] Van Doorn et al. (2013) test predictive power but find no significant effects of the CFMs.

Please cite this article as: de Haan, E., et al., The predictive ability of different customer feedback metrics for retention, *Intern. J. of Research in Marketing* (2015), http://dx.doi.org/10.1016/j.ijresmar.2015.02.004

Table 2
Conceptualization of studied CFMs.

		Time dimension		
		Past focus	Present focus	Future focus
Part of the scale used	Full scale	CES	Customer satisfaction	NPS value
	Focus on extremes		Top-2-box customer satisfaction	Official NPS

perceived service performance from a specific past experience (Dixon et al., 2010). The CES is based on a single question ("How much effort did you personally have to put forth to handle your request?"), measured on a five-point scale. Dixon et al. (2010) suggest that the CES is a better predictor of repurchase (intentions) and increased spending than the NPS or customer satisfaction. Finally, customer satisfaction centers more on the overall evaluation of the interactions between the customer and the firm over time and tends to have a more present focus (Verhoef, 2003), though it is also based on past experiences.

The second dimension pertains to how the measurement scale of the CFM is used. Advocates suggest looking not at the value of the scale but at the proportion of people responding very positive and/or very negative. An example of this is the top-2-box customer satisfaction, which measures the proportion of customers filling in the two highest-scoring points of the overall customer satisfaction scale (Morgan & Rego, 2006). Morgan and Rego (2006) show that this transformation serves as a good predictor of future business performance. The transformation to come to the official NPS also distinguishes between very positive, moderate, and very negative responses (Reichheld, 2003). Transformations can theoretically be defended because research has shown that customers mainly focus on extreme experiences and therefore the effects of CFMs can be rather non-linear (e.g., Streukens & De Ruyter, 2004; Van Doorn & Verhoef, 2008). Moreover, service marketing experts have pledged to delight customers, implying that customers will evaluate firms with extreme scores on the CFM scales (Oliver, Rust, & Varki, 1997). Firms can however also choose not to use a transformation and instead make use of the full scale (e.g., the 0–10 scale of the NPS). If we combine the two dimensions, we end up with a three-by-two classification matrix, as provided in Table 2.

Combining CFMs could create a more powerful predictor, because the different CFMs all have their unique focus. Therefore, we also test whether using multiple CFMs simultaneously improves predictive power over monitoring only a single CFM.

2.2. Studying effects across customers, firms, and industries

In this study, we investigate the CFMs' impact on retention at three different levels: customer, firm, and industry. These levels provide different insights into how the CFMs influence customer retention—specifically, (1) which part of retention comes from the greater satisfaction of a customer than an average customer at the same firm, (2) which part comes from the outperformance of the firm over competitors, and (3) which part comes from industry differences. Failing to separate the effects across these levels can lead to misinterpretation of the outcomes (Hox, 2010).

At the customer level, in which we investigate how the CFMs provide information about one customer compared with another customer at the same firm, CFMs can be used for customer management purposes. The customer level provides insights into how loyal customers differ from disloyal customers and the extent to which CFMs help discriminate between these customers. For example, Keiningham, Cooil, Aksoy, Andreassen, and Weiner (2007) show that CFMs are significant but not strong predictors of individual customer behavior, given the relatively low R-squares. This low level of prediction can be expected because many other factors influence customer behavior (e.g., competition, customer heterogeneity), but Keiningham, Cooil, Aksoy, Andreassen, and Weiner (2007) also show that CFMs can be used to distinguish between loyal and disloyal segments of customers within a firm.

At the firm level, in which we investigate how the CFMs provide information about the firm compared with its competitors, CFMs can be used for analyzing the firm's competitive position. Most studies on the effectiveness of CFMs (e.g., Keiningham, Cooil, Andreassen, & Aksoy, 2007; Morgan & Rego, 2006; Rego, Morgan, & Fornell, 2013; Van Doorn, Leeflang, & Tijs, 2013) focus on this level of analysis. Between-firm differences in CFMs are more likely to affect retention than within-firm differences. For example, if a customer is relatively unsatisfied with the firm, but the firm outperforms its competitors on average customer satisfaction, the customer is left with few alternatives to choose from and will not likely switch to a competitor (Rego et al., 2013). Furthermore, a firm that performs better than its competitors on customer satisfaction is also likely to receive more positive word of mouth, which is likely to increase the retention rate of other customers at the same firm. This positive word of mouth will also likely lower the retention rates at competing firms because customers will switch to the firm with more satisfied customers.

Finally, at the industry level, in which we investigate how the CFMs provide information about the industry's relative performance compared with other industries, CFMs can be used as a benchmarking tool, for example, for investors and other stakeholders. For explaining these between-industry differences, the usefulness of CFMs in predicting retention is less clear however. In industries in which customers are less satisfied or industries that have fewer promoters, switching rates could be lower, and thus retention rates could be higher. However, many other factors play a role, such as industry competitiveness, switching costs, and product visibility (Ou, Verhoef, & Wiesel, 2014). At first glance, it is not clear how and if between-industry differences influence customer retention.

3. Data

The data were collected for an annual study on customer performance of Dutch service firms in September 2010 (Bouma et al., 2010). The data came from an online survey sent to a sample of customers who are representative of the Dutch consumer market. Each respondent could fill out the survey for multiple firms in his or her industry. In total, 6649 respondents filled out the survey, providing a total of 8924 firm evaluations (i.e., 1.34 firm evaluations per respondent) for 93 firms across 18 industries. The survey measures the CFMs and asks other questions as well (e.g., Bouma et al., 2010). Table 3 presents an overview of the CFMs we study and their measurement.

In November 2012—two years after they answered the initial survey—respondents received a request to participate in a follow-up survey. The subsequent survey asked whether the respondents remained customers at the firm for which they completed the first questionnaire. Among the 8924 firm responses from 6649 unique responders of the first survey, 1308 unique responders also filled out the second survey, providing us with 1375 firm responses (i.e., a 15.4% response rate).[3] Descriptive statistics of this sample of firm responses appear in Table 4. We also have CES data for 20.5% of the sample; 79.5% of the customers had not contacted the firm with a request and thus could not provide a CES, though all respondents answered all other questions. The first survey provides the independent variables, while the retention variable (dependent variable) comes from the second survey. From the 1375 firm responses, 370 (26.9%) customers churned within the two years between the surveys. Appendix A provides more details about the differences in CFM scores and retention rates among industries.

[3] We thank Onrust, Verhoef, Van Doorn, and Bügel (2014) for sharing these data.

Table 3
Questions underlying the CFMs.

CFM	Measurement
1. Customer satisfaction	"All in all, how satisfied or unsatisfied are you with [company X]?" (1 = *very unsatisfied*, 7 = *very satisfied*). Research has found that measuring customer satisfaction with one item is sufficient (e.g., Ittner & Larcker, 1998; Van Doorn et al., 2013). At the firm (industry) level, this translates into the average customer satisfaction score given within the firm (industry).
2. Top-2-box customer satisfaction	A dummy at the customer level indicates whether the customer has given a score of 6 or 7 on the customer satisfaction question. At the firm (industry) level, this is the proportion of customers of that firm (industry) that gave a score of 6 or 7.
3. NPS	"How likely is it that you would recommend [company X] to a friend or colleague?" (0 = *very unlikely*, 10 = *very likely*). Respondents who gave a score of 0–6 are "detractors," those who gave a 7 or 8 are "passives," and those who gave a 9 or 10 are "promoters." Subtracting the proportion of promoters by the proportion of detractors provides the NPS at the firm level (Reichheld, 2003). At the customer level, the NPS reduces to a value of −1 for detractors, 0 for passives, and +1 for promoters. At the firm (industry) level, this translates to a score ranging from −1 (only detractors) to +1 (only promoters).
4. NPS value	This is the untransformed NPS score (0–10 range) provided by the customer. At the firm (industry) level, this translates to the average NPS value given within the firm (industry).
5. CES	"Did you try to contact [company X] with any kind of request?" (yes/no) If yes, the following question is asked: "How much effort did you personally have to put forth to handle your request?" (1 = *very low effort*, 5 = *very high effort*). At the individual customer level, we only have a dummy variable for the first question and a score in the 1–5 range for the second question. At the firm and industry level, we have the proportion of people who answered yes to the first question and the average score of the second question.

For our analyses, we re-code some of the control variables, to increase the ease of interpretation and decrease the number of required parameters. For age, we re-code 18–29 years to 23.5 years, 30–39 years to 35 years, 40–49 to 45 years, 50–64 to 57 years, and ≥65 years to 70 years. In doing so, we only need to estimate one parameter that shows the change in retention when age increases by one year, rather than a parameter per age group, which must be compared with the baseline. For the same reason, we re-code the income groups: <€30,000 to 15, ~€30,000 to 30, €30,000–€60,000 to 45, >€60,000 to 75, and n/a (people who did not want to state their income) to 30. In addition, we create a dummy variable set to 1 if the respondent did not provide his or her income. Again, in this way we only need to estimate one parameter that shows the change in retention when income increases by approximately €1000 and one parameter indicating whether people who did not state their income differ significantly from people with an income of ~ €30.000 (i.e., Dutch modal income). Finally, we re-code relationship length as follows: <1 year to .5 year, 1–2 years to 1.5 years, 2–3 years to 2.5 years, 3–5 years to 4 years, 5–10 years to 7.5 years, >10 years to 12.5 years, and n/a to the average relationship length at the same firm.

Note that other variables may influence retention but are omitted from our study because of limited data availability and the scope of our study (e.g., firm marketing-mix actions). We have only limited insights into the marketing actions conducted by the different firms in the time between the two surveys. From previous studies, we know that marketing actions influence CFMs (Gupta & Zeithaml, 2006; Hanssens, Pauwels, Srinivasan, Vanhuele, & Yildirim, 2014; Srinivasan, Vanhuele, & Pauwels, 2010), which in turn influence behavior and firm performance. This last part, the impact of CFMs on behavior and performance, is the focus of our study. Therefore, the impact of marketing-mix actions on CFMs and on customer behavior and firm performance is outside the scope of this study. Including such variables in practice can however increase overall model fit and predictions (Hanssens et al., 2014; Srinivasan et al., 2010), so marketing practitioners should also take these variables into account.

The correlation matrix in Table 5 provides initial insights.[4] First, the length of the relationship is a good indicator of retention; not surprisingly, a person who already has a long relationship with a firm is more likely to stay with it than someone who has been a customer for only a short time. Among the customer satisfaction CFMs, the top-2-box customer satisfaction offers the highest correlation with retention, in line with the non-linear relationship of customer satisfaction with customer behavior found in previous studies (e.g., Dong, Ding, Grewal, & Zhao, 2011; Ittner & Larcker, 1998; Mittal & Kamakura, 2001; Van Doorn & Verhoef, 2008). Of the two NPS metrics, the official NPS correlates best; the size of this correlation is comparable to that of the top-2-box customer satisfaction. In addition, the CES significantly correlates with retention, but it is considerably lower than the correlation of the other CFMs. The CESyes dummy, which indicates whether the customer has contacted the firm with a request and provided a CES, is significant and positive, indicating that people who make requests are more likely to remain customers.

Table 4
Descriptive statistics (n = 1375).

Variable	Summary statistics			
Age	18–29 years (12.7%), 30–39 years (18.6%), 40–49 years (22.2%), 50–64 years (36.8%), ≥65 years (9.7%)			
Gender	Male (51.7%), female (48.3%)			
Income	<€30,000 (19.9%), approximately €30.000 (11.6%), €30.000–€60.000 (29.1%), >€60.000 (10.3%), decline to answer (29.0%)			
CESyes/no	Contacted company with a request (20.5%), did not contact company with a request (79.5%)			
Relationship length	<1 year (6.5%), 1–2 years (6.9%), 2–3 years (8.8%), 3–5 years (11.6%), 5–10 years (19.3%), >10 years (37.2%), don't know (9.5%)			
Retention	Still a customer (73.1%), not a customer anymore (26.9%)			
	Mean	SD	Minimum	Maximum
Satisfaction	5.14	1.56	1	7
Top-2-box	0.48	0.50	0	1
NPS official	−0.24.	0.69	−1.	1
NPS value	6.59	2.13	0	10
CES (n = 282)	2.41	1.26	1	5

4. Model

Given both the study's research objective to find which individual and combinations of CFMs performs best in predicting retention and given the data, we need a model that can handle the self-selection of customers to participate in the second survey, the hierarchical structure of the data, and, in some cases, multiple responses per participant. For this reason, we estimate two multi-level probit regression models that are allowed to correlate. The first equation predicts which customers

[4] For the correlation matrix, we use Pearson correlations. Note that we have three dichotomous variables in this correlation matrix: retention, top-2-box customer satisfaction, and CESyes. The Pearson correlation between a dichotomous variable (e.g., retention) and an interval variable (e.g., customer satisfaction) is mathematically equivalent to the point-biserial correlation (Warner, 2013). The Pearson correlation between two dichotomous variables is mathematically equivalent to the phi correlation (Warner, 2013). Using a Pearson correlation table in which both dichotomous and interval variables are included is thus appropriate. Furthermore, because the Pearson correlation is scale independent, it is appropriate to compare CFMs that are measured using different scales (e.g., NPS, which is measured on an 11-point scale, and customer satisfaction, which is measured on a seven-point scale).

Table 5
Correlation matrix (n = 1375).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Retention	1							
(2) Satisfaction	.151**	1						
(3) Top-2-box	.184**	.805**	1					
(4) NPS official	.170**	.343**	.454**	1				
(5) NPS value	.159**	.390**	.463**	.818**	1			
(6) CESyes	.085**	.008$^{\text{ns}}$	.041$^{\text{ns}}$	.107**	.079**	1		
(7) CES[a]	−.073**	−.165**	−.145**	−.090**	−.144**	.051$^{\text{ns}}$	1	
(8) Relationship length	.199**	.043$^{\text{ns}}$	.043$^{\text{ns}}$	.064*	.061*	.108**	−.081$^{\text{ns}}$	1

** $p < .01$.
* $p < .05$.
$^{\text{ns}}$ $p > .05$.
[a] For customers who did not provide a CES, we used the mean within the firm as a substitute.

participate in the second survey, and the second equation predicts retention. The multi-level structure of the models enables us to distinguish between the effectiveness of the CFMs at the three levels of analyses. We use clustered errors to account for the cases where there are multiple responses one respondent. The models are validated on the basis of in-sample fit (using the AIC and BIC) and out-of-sample fit (using the Gini coefficient, top-decile lift, and hit rate) for which we use a holdout sample containing one-third of the observations.

4.1. Self-selection

The self-selection of respondents in the second survey could potentially affect the findings; for example, customers who are more satisfied might be more likely to fill out the survey again (something we can observe). Similarly, customers who have churned might be less likely to fill out the survey (something we cannot observe). To investigate this potential bias, we evaluate whether the scores on the independent variables provided in the first survey differ between respondents and non-respondents of the second survey.

Table 6 shows that respondents in the follow-up survey do not differ from non-respondents in terms of the average customer satisfaction score, the top-2-box customer satisfaction, the CES, and income. People who provided a higher NPS are however significantly more likely also to take part in the follow-up survey. This can be an important insight for firms that use panel data to track their customers; customers who are more likely to promote the firm are also more likely to stay in the panel. This likelihood may provide an upward bias in the panel over time for this CFM, especially when the panel is not updated regularly. For customer satisfaction-related metrics, this does not seem to be the case. Older people, men, and customers who have a long relationship with the firm are also more likely to take part in the follow-up survey.

Table 6
Differences between respondents and non-respondents.

	Respondents		Non-respondents		
	n	Mean	n	Mean	p-Value
Satisfaction	1375	5.14	7549	5.11	0.563
Top-2-box	1375	0.48	7549	0.47	0.313
NPS official	1375	−0.24	7549	−0.30	0.005
NPS value	1375	6.59	7549	6.43	0.010
CESyes		21.0%		18.0%	0.030
CES	282	2.41	1362	2.52	0.109
Age	1375	47.2	7549	41.5	0.000
Gender (1 = female)	1375	0.483	7549	0.545	0.000
Income provided		29.0%		30.9%	0.160
Income	976	€ 38,483	5215	€ 39,057	0.408
Relationship length	1375	7.65	7549	6.96	0.000

The findings from Table 6 indicate that there is a self-selection bias, which affects some of the independent variables in our data set. Thus, we need to control for self-selection in our model. We do this by estimating a bivariate probit model with endogenous sample selection (Greene, 2012) with the following probabilities of the three possible outcomes:

$$\Pr\left(\text{Participate}_{\text{ijk}} = 0 | \text{X}_{1\text{ijk}}\right) = \Phi\left(-\text{X'}_{1\text{ijk}}\beta\right); \tag{1}$$

$$\Pr\left(\text{Retention}_{\text{ijk}} = 1 | \text{Participate}_{\text{ijk}} = 1,\ -\text{X}_{2\text{ijk}},\ \text{X}_{1\text{ijk}}\right) \times \Pr\left(\text{Participate}_{\text{ijk}} = 1 | \text{X}_{1\text{ijk}}\right) = \Phi_2\left(-\text{X'}_{2\text{ijk}}\delta, \text{X'}_{1\text{ijk}}\beta, -\rho\right); \tag{2}$$

$$\Pr\left(\text{Retention}_{\text{ijk}} = 0 | \text{Participate}_{\text{ijk}} = 1,\ \text{X}_{2\text{ijk}},\ \text{X}_{1\text{ijk}}\right) \times \Pr\left(\text{Participate}_{\text{ijk}} = 1 | \text{X}_{1\text{ijk}}\right) = \Phi_2\left(\text{X'}_{2\text{ijk}}\delta, \text{X'}_{1\text{ijk}}\beta, \rho\right); \tag{3}$$

and

$$\left(\begin{matrix}\varepsilon_{1ijk} \\ \varepsilon_{2ijk}\end{matrix} \middle| X_{1ijk}, X_{2ijk}\right) \sim N\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}\right], \tag{4}$$

where Φ_1 is the univariate standard normal cumulative distribution function (c.d.f.), Φ_1 is the bivariate standard normal c.d.f., ρ is the correlation between the error terms of the participation and retention equation, and X_{1ijk} (β) and X_{2ijk} (δ) are the covariates in (parameters of) the participation and retention equation, respectively. If ρ is not significantly different from zero, there is no self-selection bias in the retention equation. We estimate the model using iterative maximum likelihood estimation, in which the Newton–Raphson algorithm is combined with the Davidon–Fletcher–Powell algorithm, using the cmp package (version 6.6.4) from Stata. To estimate the random effects, we use simulation-based estimation (Greene, 2012). Finally, we use clustered errors to account for the multiple responses from some respondents (Greene, 2012).[5]

The model is only well identified if the participation equation has at least one variable that is not in the retention equation. Ideally, such a variable should be correlated with the self-selection procedure, but not with the retention probability, because otherwise it could be considered an omitted variable in the retention equation, which leads to omitted variable bias. The variable we chose for this is a dummy variable indicating whether the respondent also took part in the survey of the annual customer performance measurement among Dutch service firms in 2011 (i.e., a survey conducted between the first survey, in which we measured the independent variables, and the follow-up

[5] Given the low amount of firm evaluations per respondent, this only has a limited effect on the estimates.

survey, in which we measured the dependent variable). Because these respondents participated in two subsequent surveys, it is likely that they will take part in other future surveys as well, meaning that this variable is likely strongly correlated with the self-selection procedure. However, because respondents were not required to answer questions for the same firm in the 2010 and 2011 surveys, in principle this new dummy variable should not be correlated with firm retention.

Of the 8924 respondents in the 2010 survey, 742 also took part in the 2011 survey. In addition, 14.5% of the respondents who only filled out the 2010 survey also participated in the 2012 follow-up survey, while 25.5% of the respondents who filled out both the 2010 and 2011 surveys participated in the 2012 follow-up survey. This is a significant difference ($\chi^2 = 62.886$, $p < .001$). The retention rate from respondents who only filled out the 2010 survey is 73.3%, while that for respondents who filled out both the 2010 and 2011 surveys is 72.0%, which is not a significant difference ($\chi^2 = .143$, $p = .705$). Thus, this dummy variable is appropriate as an additional variable in the selection equation to better identify the model because it is correlated with the self-selection procedure, but not with the outcome (i.e., retention) equation.

4.2. Model estimation

Because our data set has a hierarchical structure (i.e., customers are part of a firm, and firms are part of an industry), we must accommodate all three levels in our analyses. Therefore, we consider both within- and between-firm effects of different CFMs. We chose to use multi-level[6] probit regression models with random intercepts[7] at the firm and industry levels to account for heterogeneity of retention and participation at these levels, as specified in Eqs. (5) and (6). Table 7 explains the variables in these two equations.

$$
\begin{aligned}
retention_{ijk} &\sim Binomial\left(\eta_{retention,ijk}, \pi_{retention,ijk}\right)\\
probit\left(\pi_{retention,ijk}\right) &= \beta_{x,0jk} + \beta_{x,1}\cdot\left(CFM_{x,ijk} - \overline{CFM}_{x,jk}\right)\\
&\quad + \beta_{x,2}\cdot\left(\overline{CFM}_{x,jk} - \overline{CFM}_{x,k}\right) + \beta_{x,3}\cdot\overline{CFM}_{x,k}\\
&\quad + \beta_{x,4}\cdot lenght_{ijk} + \beta_{x,5}\cdot age_{ijk} + \beta_{x,6}\cdot income_{ijk}\\
&\quad + \beta_{x,7}\cdot income_{naijk} + \beta_{x,8}\cdot female_{ijk} + \varepsilon_{x,1ijk}.
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
participation_{ijk} &\sim Binomial\left(\eta_{participation,ijk}, \pi_{participation,ijk}\right)\\
probit\left(\pi_{participation,ijk}\right) &= \delta_{x,0jk} + \delta_{x,1}\cdot\left(CFM_{x,ijk} - \overline{CFM}_{x,jk}\right)\\
&\quad + \delta_{x,2}\cdot\left(\overline{CFM}_{x,jk} - \overline{CFM}_{x,k}\right) + \delta_{x,300}\cdot\overline{CFM}_{x,k}\\
&\quad + \delta_{x,4}\cdot lenght_{ijk} + \delta_{x,5}\cdot age_{ijk} + \delta_{x,6}\cdot income_{ijk}\\
&\quad + \delta_{x,7}\cdot income_{naijk} + \delta_{x,8}\cdot female_{ijk}\\
&\quad + \delta_{x,9}\cdot panel_{ijk} + \varepsilon_{x,2ijk}.
\end{aligned}
\tag{6}
$$

To investigate the incremental power of each CFM, we estimate a baseline model in which only the control variables are included (i.e., no CFMs). We compare the models including the CFMs with each other and with the baseline model. As Eqs. (5) and (6) indicate, we group-centered the CFMs following Raudenbush's (1989) procedure, so that we can distinguish the CFMs' impact on retention at different levels of analysis. In other words, with this group-centering we disentangle the customer-, firm- and industry-level effects.[8] For example, the parameter $\beta_{x,1}$ is the effect on retention when a customer has a one-point-higher score on CFM variable x (e.g., customer satisfaction) than the average customer within the same firm. Parameter $\beta_{x,2}$ is the effect when a firm has a one-point-higher score on the CFM variable x than the average firm within the same industry. Finally, $\beta_{x,3}$ is the effect when an industry has a one-point-higher score on the CFM variable x than another industry. With this industry parameter, we can investigate whether customers are more loyal in industries with better CFM scores.

Table 7
Variable definitions.

Variable	Definition
$CFM_{x,ijk}$	Score on CFM x for customer i of firm j in industry k
$\overline{CFM}_{x,jk}$	Average score on CFM x for firm j in industry k
$\overline{CFM}_{x,k}$	Average score on CFM x in industry k
$lenght_{ijk}$	Length of the relationship (in years) for customer i with firm j in industry k
age_{ijk}	Age (in years) of customer i at firm j in industry k
$income_{ijk}$	Yearly income (in thousands of euros) of customer i at firm j in industry k
$income_{naijk}$	Dummy indicating whether income for customer i at firm j in industry k is unknown
$female_{ijk}$	Dummy indicating whether customer i at firm j in industry k is female
$panel_{ijk}$	Dummy indicating whether customer i at firm j in industry k participated in the 2011 survey

If we do not use group-centering, we cannot disentangle the different effects; in this case, $\beta_{x,1}$ would be a combination of the customer-, firm-, and industry-level effect, and $\beta_{x,2}$ would be a combination of the (remaining) firm- and industry-level effect. Thus, group-centering aids interpretation of the effects. With group-centering, the CFMs at the three levels are completely uncorrelated and do not violate any of the assumptions of the multi-level model. Indeed, the variance in the original CFM is due to group-centering allocated over the three levels.

Finally, we investigate industry heterogeneity by estimating a simplified version of the model in Eq. (5)—namely, a logistic regression model with industry-specific parameters for each CFM, as shown in Eq. (7).[9] We model this version separately from Eqs. (5) and (6) because including industry-specific parameters in these two equations would make the model extremely complex, given that the random effects in the selection equation may be correlated with the retention equation, with each other's (firm and customer level) intercepts, and so on. This would result in an extremely complex variance–covariance matrix and over-complicated models, which would likely result in over-fitting and provide lower out-of-sample fit. Thus, Eqs. (5) and (6) give a general picture about the average effect of each CFM, while Eq. (7) provides additional insights into industry heterogeneity.

[6] For all the different versions of this model, we performed a likelihood ratio test to investigate whether a multi-level probit is preferable to a normal probit. We find that this is indeed the case (all p-value are smaller than .001, indicating preference for the multi-level probit over the simpler but less complete normal probit).

[7] For all the different versions of this model, we performed a Hausman (1978) test to investigate whether random-effects models are preferable to fixed-effects models. All Hausman tests were highly non-significant, indicating that the parameter estimates of the random-effects models are consistent. Therefore, random-effects models are preferable because of their efficiency (Hausman, 1978). Note that even a significant Hausman test in a multi-level model would not automatically imply that a random-effects framework should be abandoned, but that the conceptual level should also be considered when making such a decision (Fielding, 2004; Snijders & Berkhof, 2007).

[8] Thus, the CFM variables are all measured at the customer level, and through the group-centering approach, we distinguish among customer-, firm-, and industry-level effects. This deviates from other multi-level modeling approaches (e.g., Steenkamp & Geyskens, 2014), in which higher levels in the hierarchy (e.g., firm level) include the parameters from the lower levels as dependent variables (e.g., customer-specific parameters) as well as the associated higher-level predictors (firm level in this example).

[9] We choose to estimate a simplified version of Eq. (5) because the results of the full model indicate no significant self-selection bias and no significant impact of the control variables on the CFM parameter estimates. Because the amount of observations per industry is relatively low (51–135 observation) and the retention rates are high (i.e., low number of churners), three parameters per industry is already quite high for a logistic regression model (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996).

$$retention_{ijk} \sim Binomial\left(\eta_{retention,ijk}, \pi_{retention,ijk}\right)$$
$$logit\left(\pi_{retention,ijk}\right) = \alpha_{x,0k} + \alpha_{x,1k} \cdot \left(CFM_{x,ijk} - \overline{CFM}_{x,jk}\right) + \alpha_{x,2k} \cdot \left(\overline{CFM}_{x,jk} - \overline{CFM}_{x,k}\right) + \varepsilon_{x,3ijk} \quad (7)$$

where $\alpha_{x,0k}$ captures the industry-level heterogeneity, $\alpha_{x,1k}$ captures the effect of differences between customers within the same firm, and $\alpha_{x,2k}$ captures the effect of differences between firms within the same industry. We investigate per industry which CFMs are useful for customer management within the firm (i.e., have a significant $\alpha_{x,1k}$) and which CFMs are useful to compare the focal firm's competitive position with its competitors within the same industry (i.e., have a significant $\alpha_{x,2k}$). Furthermore, we indicate per industry which CFM is the most useful (i.e., have the highest significance level) for these two levels of analyses.

4.3. Model validation and comparison

We estimate the models for all the CFMs, using all 8924 observations for Eq. (6) and 1375 observations for Eq. (5), as well as only customers who made a request – that is, 1644 observations for Eq. (6) and 282 observations for Eq. (5). In all models, we control for the length of the relationship, age, income, and gender. First, we estimate separate models for each CFM to compare how each performs in predicting retention (i.e., Eqs. (5) and (6) are estimated for each CFM). In doing so, we can compare the individual CFMs and provide insights into which single CFM is the best predictor of retention, in line with how some practitioners use CFMs in daily practice and the quest for the best predictor of future performance (e.g., Dixon et al., 2010; Reichheld, 2003). Because we are less interested in the size of the effects and more in the model (in- and out-of-sample) fit, omitting the other CFMs is not a concern (Blattberg et al., 2008). This procedure is also in line with studies comparing CFMs on their ability to explain firm performance (e.g., Morgan & Rego, 2006; Van Doorn et al., 2013). Second, after investigating the single CFMs, we investigate whether using one CFM is preferable to monitoring multiple CFMs by also estimating models that include multiple CFMs simultaneously.

If the models that include the CFMs perform better than the baseline model, which only includes the control variables, the CFMs have at least some incremental power. We test this by performing for each model a likelihood ratio test for nested models. We compare the in-sample fit of the different CFMs models, which are not nested, using the AIC and BIC, both of which penalize for over-fitting. To formally test which of the CFM is "the best CFM," we calculate the Akaike weights in line with Wagenmakers and Farrell's (2004) procedure. The Akaike weight indicates how likely it is that each CFM is the best out of all the CFMs tested.

In addition to the in-sample fit, we use three out-of-sample fit criteria: the Gini coefficient, the top-decile lift, and the hit rate (e.g., Lemmens & Croux, 2006; Neslin, Gupta, Kamakura, Lu, & Mason, 2006; Risselada, Verhoef, & Bijmolt, 2010). Accordingly, we used two-thirds of the observations per firm to estimate the models and one-third to validate them (i.e., we calculate the three fit criteria on the models' out-of-sample predictions). The models containing the CFMs should perform better on all fit criteria than the baseline model, and the CFM that dominates the criteria is deemed the best single CFM. Because only a sub-sample (282 of 1375) of respondents answered the CES question, we also calculate the AIC and BIC for this sub-sample, to determine whether the CES outperforms the other CFMs for customers for whom this metric was especially designed.

Table 8
Parameter estimates multi-level logistic regression models for retention.

Fixed part		Baseline	Satisfaction	Top-2-box	NPS official	NPS value	CES
	Intercept	1.059 (.527)*	.002 (2.314)^{ns}	.446 (.765)^{ns}	1.151 (.508)*	1.387 (1.107)^{ns}	1.651 (.778)*
	CFM.cust.		.117^{a} (.050)**	.482^{a} (.159)**	.333^{a} (.129)**	.096^{a} (.040)**	−.198 (.136)^{ns}
	CFM.firm		.746 (.300)**	2.446 (.798)**	1.665 (.546)**	.457 (.173)**	.042 (.160)^{ns}
	CFM.indu.		.194 (.438)^{ns}	1.270 (1.118)^{ns}	−.227 (.492)^{ns}	−.037 (.146)^{ns}	−.311 (.240)^{ns}
	CESyes.cust.						.242 (.187)^{ns}
	CESyes.firm						.955 (1.467)^{ns}
	CESyes.indu.						1.285 (.750)*
	Relation length	.064 (.018)**	.061 (.019)**	.061 (.019)**	.056 (.019)**	.057 (.020)**	.064 (.017)**
	Age	−.007 (.006)^{ns}	−.006 (.006)^{ns}	−.007 (.005)^{ns}	−.008 (.006)^{ns}	−.007 (.007)^{ns}	−.008 (.006)^{ns}
	Income	−.001 (.005)^{ns}	.000 (.005)^{ns}	.001 (.005)^{ns}	.000 (.005)^{ns}	.000 (.005)^{ns}	−.001 (.004)^{ns}
	Income.na.	−.065 (.171)^{ns}	−.039 (.175)^{ns}	−.025 (.173)^{ns}	−.025 (.176)^{ns}	−.036 (.180)^{ns}	−.078 (.157)^{ns}
	Female	.013 (.158)^{ns}	.005 (.161)^{ns}	−.010 (.160)^{ns}	−.012 (.163)^{ns}	−.021 (.167)^{ns}	.006 (.146)^{ns}
Random intercept		Baseline	Satisfaction	Top-2-box	NPS official	NPS value	CES
	Industry level	.191 (.086)**	.226 (.077)**	.224 (.075)**	.242 (.074)**	.243 (.079)**	.156 (.082)*
	Firm level	.421 (.063)**	.359 (.063)**	.344 (.063)**	.321 (.063)**	.352 (.063)**	.399 (.061)**
Cross-equation correlation							
	Industry level	−.246 (.589)^{ns}	−.225 (.505)^{ns}	−.242 (.497)^{ns}	−.239 (.480)^{ns}	−.236 (.491)^{ns}	−.414 (.606)^{ns}
	Firm level	−.674 (.212)*	−.651 (.244)*	−.650 (.255)*	−.714 (.251)*	−.698 (.241)*	−.620 (.230)*
	Residuals	−.304 (.201)^{ns}	−.295 (.198)^{ns}	−.289 (.202)^{ns}	−.367 (.176)^{ns}	−.336 (.180)^{ns}	−.318 (.204)^{ns}
In-sample	Log-likelihood	−4438.34	−4420.76	−4410.02	−4410.04	−4413.69	−4425.80
	LR test (baseline)		35.16**	56.64**	56.60**	49.30**	25.08**
	AIC	8916.67	8893.52	**8872.05**	8872.07	8879.38	8915.60
	BIC	9058.60	9078.03	**9056.56**	9056.58	9063.88	9142.69
	Akaike weights	.000	.000	**.496**	.491	.013	.000
Out-of-sample	Gini coefficient	.130	.141	**.167**	.159	.153	.141
	Top-decile lift	1.394	1.646	1.801	**2.241**	1.980	1.646
	Hit rate	.536	.588	**.651**	.557	.577	.550

** $p < .01$.
* $p < .05$.
ns $p > .05$ (one-tailed for the CFMs, two-tailed for the rest).
a Significantly ($p < .05$) smaller than the firm-level effect.

Please cite this article as: de Haan, E., et al., The predictive ability of different customer feedback metrics for retention, Intern. J. of Research in Marketing (2015), http://dx.doi.org/10.1016/j.ijresmar.2015.02.004

5. Results

5.1. Main results

Table 8 presents the parameter estimates and summary statistics of Eq. (5). We observe that customer satisfaction, as well as most of the other CFMs, have a positive and significant impact at both the customer and firm level. Across all CFMs, except the CES, a one-point increase at the firm level has a significantly larger impact on customer retention than a one-point increase at the customer level (indicated by the superscript "a" in Table 8). This finding corroborates our argumentation that a customer at a high-scoring firm has few alternative companies to do business with and therefore is less likely to churn, even if he or she is relatively unsatisfied. At the industry level, customer satisfaction ratings and most of the other CFMs are not statistically significant, so industries with higher customer satisfaction rates do not appear to differ significantly in terms of retention rates from industries with lower customer satisfaction rates.

Table 8 also shows that the intercept is significantly random at both the firm and industry level for all models, indicating that the CFMs and the control variables do not completely capture firm and industry heterogeneity with respect to retention. Furthermore, the participation and retention equations are negatively correlated with each other at the firm level, indicating that, in general, firms with a higher participation rate have a lower retention rate. The residuals of the participation and retention equations are not correlated however (ρ in Eq. (4)), indicating no significant self-selection bias (i.e., the parameters do not differ substantially from those had we not controlled for self-selection).

In terms of in-sample fit, the top-2-box customer satisfaction (which has a present focus and centers on the extremes) and the official NPS (which has a future focus and also centers on the extremes) perform best on both the AIC and BIC. This indicates that the transformations for these two CFMs make sense and that the focus on customers with more extreme opinions is important. All models perform statistically better than the baseline model in terms of the likelihood ratio test. Also in terms of the AIC, all models outperform the baseline model, while only the top-2-box customer satisfaction and official NPS outperform the baseline model in terms of the BIC. This finding indicates that the CFMs have some incremental power, but (in terms of in-sample fit) the extra complexity does not always outweigh the better fit.

To determine which CFM performs best, we can evaluate the Akaike weights. The weights reveal a 49.6% certainty that the top-2-box customer satisfaction is the best CFM and a 49.1% certainty that the official NPS is the best CFM; thus, these two CFMs are almost equally likely to be the best CFM, and it is very unlikely that one of the other CFMs is the best.

In terms of out-of-sample fit, all models perform better than the baseline model for all three out-of-sample fit criteria, with the top-2-box customer satisfaction performing best on both the Gini coefficient and the hit rate and the official NPS having the highest top-decile lift. The Gini coefficients and top-decile lift are comparable in value to previous studies on churn prediction. The statistics are somewhat higher than those in Risselada et al.'s (2010) study but somewhat lower than those in Lemmens and Croux's (2006) study. Differences in out-of-sample fit can be explained by the type of model, the type of input variables, the time horizon of predictions, and the study including one firm rather than customers from a wide range of firms.

We measure the CFMs on different scales, so the parameters from a probit model are not straightforward to interpret. Because of this, we depict the impact of a one standard deviation difference in the CFM at the customer level on retention probability for the average customer in Fig. 1. The impact is highest for the top-2-box customer satisfaction, followed by the official NPS. Changes in these CFMs have the highest impact on the retention probability of a customer. For the CES, and if a customer had a request (CESyes), the line in Fig. 1 would be flatter, indicating that a change in the CES has a far smaller impact on retention probability. This is in line with the other findings; the top-2-box customer satisfaction and the official NPS perform rather well, but the CES does not seem to have a strong impact in general.

A possible reason that other CFMs outperform CES is that not all customers had a service request, so not all customers provided a CES. Accordingly, we estimated the same models for only the sub-sample of customers who made a request (i.e., have a value of 1 for the CESyes variable). Table 9 shows that even for this sub-sample of customers, for whom the CES was especially designed, the CES is the only CFM that performs worse than the baseline model in terms of the AIC. All models perform worse than the baseline model in terms of the BIC. This result may be due to the limited number of observations in this sub-sample ($n = 1644$ for the participation equation, and $n = 282$ for the retention equation). Yet, by quite a large margin, CES is still the worst-performing CFM. Thus, even for the group for which the CES was designed, it is still outperformed by the other CFMs.

In terms of our conceptualization from Table 2, we can conclude that the "part of the scale used" dimension is the most important. We clearly find that the top-2-box customer satisfaction and the official NPS, which focus on the extremes, outperform the CFMs that use the full scale. We also find that in terms of the time dimension, a present and future focus both work well, but a strong past focus does not aid in predicting retention.

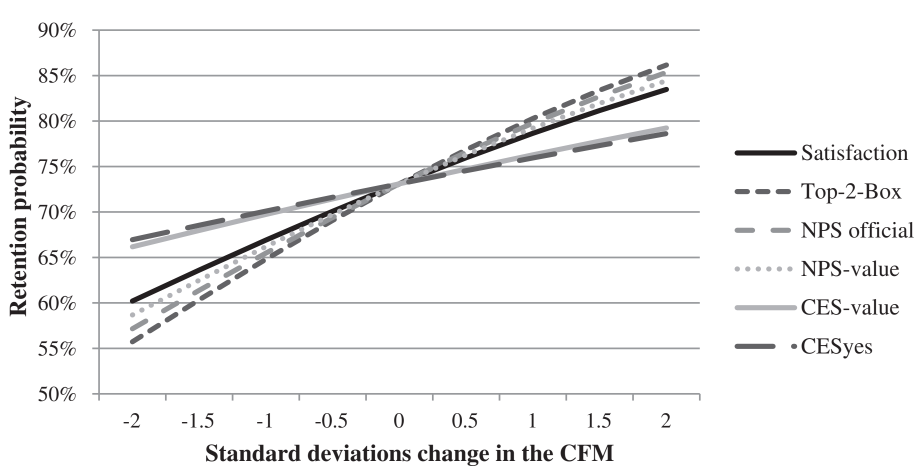

Fig. 1. Impact of differences in CFM scores on retention.

Table 9
CFM comparison for customers who made requests.

	AIC	BIC
Baseline	1742.55	1850.64
Satisfaction	1733.93	1874.46
Top-2-box	1728.56	1869.08
NPS official	1726.45	1866.98
NPS value	1719.49	1860.02
CES	1754.55	1927.50

5.2. Industry heterogeneity

The predictive power of CFMs might vary between industries. Table 10 shows the performance of the CFMs at the customer level per industry. The table indicates whether and to what extent the CFMs are useful to compare customers within a firm with one another (i.e., whether they are useful as a key metric for customer management purposes; $\alpha_{x,1k}$ parameters in Eq. (7)). The scores in Table 10 indicate the relative impact of the CFMs compared with the best CFM in that specific industry (i.e., the t-statistic of the indicated CFM in that specific industry divided by the t-statistic of the best-performing CFM in that industry). No score means that the CFM is not significant in explaining retention in the industry. In three industries (health insurance, energy, and supermarkets), none of the CFMs are significant; in the other 15 industries, at least one CFM is significant. The top-2-box customer satisfaction is significant in ten of the 18 industries and is the best-performing CFM in four industries (online booking, online shopping, drugstores and banks), though the relative differences from the raw customer satisfaction scores are not large. In the online world, in which customers can easily compare offers and switch firms, having highly satisfied customers (i.e., having a high top-2-box customer satisfaction) is of utmost importance rather than having "on average" quite satisfied customers. In terms of the classification of our conceptual framework, this indicates that in the online world, and also for drugstores, customers have a strong focus on the present and are driven by highly positive service experiences. The finding that customer satisfaction is important for loyalty in the online world is also in line with that of Shankar, Smith, and Rangaswamy (2003).

For the offline world, and especially for industries in which customers do not purchase frequently (i.e., mobile telecom providers, travel agencies, airline companies, and electronics stores), the customer satisfaction score is the best predictor. This indicates that (moderately) positive experiences in the present are the main driver of retention in these industries. The official NPS performs best in more traditional utility companies (general insurance companies, and fixed telecom) and experience-type companies (holiday parks, and furnishing stores such as IKEA). Conversely, the untransformed NPS value is the best-performing CFM for gasoline stations, department stores, and do-it-yourself stores. These CFMs are part of the future focus dimension of our conceptual framework. The CES value is only significant in one industry and performs best in none. Thus, the past focus is less relevant across industries.

Table 10
CFMs performance per industry at the customer level.

	Satisfaction	Top-2-box	NPS off.	NPS value	CES value
General insurance			**1.000**		
Health insurance					
Banks		**1.000**		.948	.985
Mobile telecom	**1.000**	.973			
Fixed telecom			**1.000**	.768	
Energy					
Gasoline		.751	.722	**1.000**	
Travel agencies	**1.000**	.791			
Holiday parks		.750	**1.000**	.781	
Airlines	**1.000**			.940	
Supermarkets					
Drugstores	.970	**1.000**		.827	
Department stores			.982	**1.000**	
Electronics	**1.000**	.950			
Do-it-yourself	.888	.766	.898	**1.000**	
Furnishing			**1.000**		
Online booking	.969	**1.000**		.840	
Online shopping	.884	**1.000**	.734		
Significant	**8/18**	**10/18**	**8/18**	**9/18**	**1/18**
Best performing	**4/18**	**4/18**	**4/18**	**3/18**	**0/18**

Table 11
CFMs performance per industry at the firm level.

	Satisfaction	Top-2-box	NPS off.	NPS value	CES value
General insurance					
Health insurance					
Banks					
Mobile telecom					
Fixed telecom					
Energy	.934	**1.000**		.915	
Gasoline		.847	.829	**1.000**	
Travel agencies		.974			**1.000**
Holiday parks					
Airlines	.544	.559	.996	**1.000**	
Supermarkets	**1.000**	.997	.955	.961	.387
Drugstores		.661	**1.000**	.908	
Department stores					
Electronics	.903	**1.000**	.821	.919	
Do-it-yourself					
Furnishing	.949	.909	**1.000**	.948	
Online booking		**1.000**			
Online shopping			.774	.755	**1.000**
Significant	**5/18**	**9/18**	**7/18**	**8/18**	**3/18**
Best performing	**1/18**	**3/18**	**2/18**	**2/18**	**2/18**

Table 11 shows the performance of the CFMs at the firm level per industry. The table indicates whether the CFMs are useful to compare firms within an industry with one another (i.e., whether they serve as a key metric to analyze the competitive position of firms; $\alpha_{x,2k}$ parameters in Eq. (7)). In 10 industries, at least one CFM is statistically significant at this level, the best performing CFM is more evenly spread over the industries. The top-2-box customer satisfaction is the best-performing CFM to compare firms in the energy, electronics, and online booking industries, while untransformed customer satisfaction is best for supermarkets. Conversely, the official NPS is the best CFM to compare firms in the drugstore, and furnishing industries; the untransformed NPS is the best CFM for gasoline stations, and airline. The CES, which handles requests, is the best CFM to compare travel agencies and online shopping websites.

Table 12
Predictive performance multi-CFM models.

		Satisfaction	Top-2-box	NPS off.	NPS value	CES
Satisfaction	Gini coefficient	*.141*				
	Top-decile lift	*1.646*				
	Hit rate	*.588*				
Top-2-box	Gini coefficient	.159	*.167*			
	Top-decile lift	**1.915**	*1.801*			
	Hit rate	.630	*.651*			
NPS off.	Gini coefficient	**.162**	**.186**	*.159*		
	Top-decile lift	2.241	1.947	*2.241*		
	Hit rate	**.591**	.595	*.557*		
NPS-value	Gini coefficient	**.156**	**.175**	.156	*.153*	
	Top-decile lift	1.947	1.947	2.241	*1.980*	
	Hit rate	**.589**	.599	.564	*.577*	
CES	Gini coefficient	**.151**	**.177**	**.163**	**.165**	*.141*
	Top-decile lift	**1.947**	**2.274**	1.947	1.646	*1.646*
	Hit rate	**.596**	.648	.553	.549	*.550*

Note: Bold figures indicate an improvement over using one of the two separate CFMs.

5.3. Combining metrics

Because the different CFMs may measure different dimensions and all CFMs have their unique focus (see Table 2), combining CFMs could create even more incremental power in predicting customer retention. Table 12 depicts the results of combining the CFMs using Eqs. (5) and (6). Combining customer satisfaction with top-2-box customer satisfaction and combining the official NPS with the NPS value do not much improve the out-of-sample predictions, probably because of multicollinearity (i.e., there is not much incremental information to make better predictions). For all other combinations, we can observe some out-of-sample improvements.

Most important, the Gini coefficient improves in all other cases; we find the highest Gini coefficient (.186) when we combine the top-2-box customer satisfaction and the official NPS. This means that by combining CFMs (i.e., having a dashboard of metrics that measure multiple dimensions, as indicated in Table 2), firms can obtain better predictions about their customer base as a whole. By contrast, the hit rate only improves when we combine the customer satisfaction score with another (non-customer satisfaction–related) CFM. However, in all cases the hit rate is lower than that for the top-2-box customer satisfaction as a single CFM, so in terms of hit rate, use of multiple CFMs is not necessary. In terms of the top-decile lift, the best improvement occurs when we combine the top-2-box customer satisfaction with the CES. Thus, the CES has incremental power, given that it is used in combination with a customer satisfaction–related CFM (i.e., combining the past and present focus of Table 2).

6. Discussion and conclusion

In the scientific marketing literature, multiple studies have evaluated the performance of CFMs (Gupta & Zeithaml, 2006). In addition, practitioner-oriented literature has proposed new CFMs that promise to be “the best” indicator of (future) firm performance on a regular basis (e.g., Dixon et al., 2010; Reichheld, 2003). We contribute to this literature by investigating the power of different CFMs to predict customer retention for a large number of customers, at a large number of firms, in a large number of industries. In doing so, we provide guidance to managers on which CFMs to use in which industries. Understanding this enables managers to (1) evaluate their existing system of CFMs or (2) decide which CFMs to use if introducing a new measurement and management system. Our study offers five key results. First, all CFMs, except the CES, have a significant impact on retention at both the customer and firm levels, but not at the industry level. The differences in performance of customer satisfaction and NPS are not substantial. Second, changes in top-2-box customer satisfaction, followed by the official NPS, have the highest impact on customer retention. This suggests that it is useful to transform CFMs and focus on very positive (or very negative) groups. Third, the predictive power of CFMs differs across industries, which is not reported in prior studies on the performance of different CFMs. Therefore, there is no single best metric to predict customer retention across industries. Fourth, the usefulness of CFMs differs depending on the purpose (key metric for customer management vs. key metric for analyzing the firm’s competitive position; see Table 13) because of the different performance of the different analyses. Fifth, combining CFMs tends to improve predictions. Therefore, firms might be better off using a dashboard of CFMs that includes different dimensions (Ambler & Roberts, 2008; Farris et al., 2006; Johnson & Schultz, 2004).

With our findings, we can state that, in contrast with previous research (e.g., Keiningham, Cooil, Andreassen, & Aksoy, 2007), monitoring NPS does not seem to be wrong in most industries. Our findings indicate that the NPS is an effective predictor of customer retention, though the top-2-box customer satisfaction is slightly better overall. In addition, with our findings, we can more specifically show the usefulness of the different CFMs for different purposes. In doing so, we distinguish among (1) the usefulness of these CFMs for discriminating between churners and retainers (or loyal customers) based on the Gini coefficient, (2) the identification of top churners based on the top-decile lift, (3) the usefulness for customer management based on the effects at the customer level, and (4) the usefulness for analyzing competitive positioning based on firm-level effects (see Table 13).

For differences across industries, we can state that customer satisfaction is useful as a key metric for customer management purposes in four industries (mobile telecom, travel agencies, airlines, and electronics), top-2-box customer satisfaction also in four industries (banks, drugstores, online booking, and online shopping), the official NPS score again in four industries (general insurance companies, fixed telecom, holiday parks, and furnishing), NPS value in three industries (gasoline, department stores, and do-it-yourself), and the CES in none of the industries. As a key metric for analyzing a firm’s competitive position, customer satisfaction is most useful in one industry (supermarkets), top-2-box customer satisfaction in three industries (energy, electronics, and online booking), the official NPS in two industries (drugstores, and furnishing), the NPS value in two industries (gasoline, and airlines), and the CES also in two industries (travel agencies, and online shopping). In terms of the limited overall incremental value of the CES in itself, managers should be reluctant to adopt any metrics that have a past focus and are limited in focus on one specific attribute and/or incident as an overall key performance metric. Furthermore, newly introduced CFMs should be classified in accordance with the existing CFMs, as we have done in our conceptualization in Table 2. From this classification, managers can judge how valuable the new CFM will be and what other CFMs the new CFM can be combined with to form a sufficient dashboard.

7. Research limitations and further research

One limitation of our study is that we use customers’ self-stated retention as a dependent variable. In contractual settings, this measure likely provides an accurate picture, but in non-contractual settings, the question can be difficult for customers to answer, especially if they do not often do business in the particular industry (Batislam, Denizel, & Filiztekin, 2007). Another limitation is that we only investigate business-to-consumer firms, mainly in the retail and service industries. Furthermore, the data include only Dutch participants, who may use scales differently than, for example, U.S. respondents (Van Doorn et al., 2013). The impact of the CFMs may also differ depending on the economic, political, and cultural factors of a country (Ou, De Vries, Wiesel, & Verhoef, 2014; Yeung, Ramasamy, Chen, & Paliwoda, 2013). Furthermore, the impact might change over time, which could be investigated in a more longitudinal study. Finally, because of limited data availability and the scope of our study, we omitted some important variables, such as marketing-mix variables. However, prior research has recommended that especially practitioners include these variables in their dashboards to better predict future performance (Hanssens et al., 2014; Srinivasan et al., 2010).

This study investigates one of the firm’s most valuable assets, the customer base (Rust et al., 2000), and, from that base, one of the most important drivers of customer and firm value, customer retention (Gupta et al., 2004). Further research could investigate other outcome

Table 13
Summary of the usefulness of the CFMs.

	Discriminating between churners and retainers	Identifying most risky customers (“top churners”)	Useful for customer management	Useful for competitive positioning
Satisfaction	+	+	++	+
Top-2-box	++	+	++	++
NPS official	+	++	++	+
NPS-value	+	+	++	+
CES	--	+	+/−	+/−

Appendix A. Statistics by industry

Satisfaction metrics							NPS metrics		CES metrics	
	n	Firms	Response	Retention	Satisfaction	Top-2-box	NPS	NPS value	CESyes	CES
General insurance	135	10	17.3%	74.8%	5.14	0.49	−0.39	6.05	17.8%	2.17
Health insurance	74	6	16.6%	83.8%	5.27	0.50	−0.31	6.35	31.1%	2.17
Banks	74	5	15.5%	89.2%	4.82	0.47	−0.47	5.80	25.7%	2.26
Mobile telecom	52	4	10.3%	78.8%	5.15	0.60	−0.29	6.40	36.5%	3.00
Fixed telecom	68	5	15.5%	77.9%	4.88	0.40	−0.26	6.29	39.7%	3.30
Energy	63	5	9.9%	77.8%	5.10	0.48	−0.43	6.00	17.5%	1.82
Gasoline	72	5	15.0%	47.2%	5.28	0.44	−0.50	6.08	2.8%	2.00
Travel agencies	111	8	17.4%	57.7%	5.39	0.59	0.06	7.43	27.0%	2.13
Holiday parks	74	4	21.6%	70.3%	5.23	0.61	0.23	7.93	29.7%	2.36
Airlines	81	5	19.2%	60.5%	4.64	0.36	−0.26	6.75	24.7%	3.00
Supermarkets	110	8	14.5%	79.1%	4.88	0.39	−0.25	6.67	14.5%	2.56
Drugstores	90	4	20.1%	80.0%	5.18	0.43	−0.44	6.24	1.1%	1.00
Department stores	73	3	17.0%	83.6%	5.36	0.51	−0.18	6.78	5.5%	2.00
Electronics	65	5	12.7%	70.8%	5.17	0.45	−0.32	6.72	21.5%	2.07
Do-it-yourself	54	4	12.6%	77.8%	5.35	0.57	−0.28	6.59	20.4%	1.91
Furnishing	76	4	15.7%	71.1%	5.08	0.39	−0.13	6.82	21.1%	2.31
Online booking	52	4	17.2%	63.5%	5.13	0.56	−0.13	6.67	13.5%	3.14
Online shopping	51	4	13.0%	76.5%	5.63	0.59	0.00	7.20	31.4%	1.94
Total sample	1375	93	15.4%	73.1%	5.14	0.48	−0.24	6.59	20.5%	2.41

variables, such as customer profitability, cross-buying, or customer lifetime value (e.g., Shah, Kumar, Qu, & Chen, 2012; Venkatesan & Kumar, 2004). An advantage of these outcome variables is that companies can more clearly understand the impact of the CFMs on the value of their customer base and how to manage it more effectively. Another direction might be to conduct a longitudinal study to provide more fine-grained insights into when companies should target which customers on the basis of (changes in) the CFMs. Furthermore, in this study we investigated how valuable the different CFMs are across industries. This can be a starting point for future studies to explain industry differences, using variables such as industry competiveness and product visibility (see Ou, Verhoef, & Wiesel, 2014). Finally, although we accounted for firm- and industry-level heterogeneity, customer-level heterogeneity might exist as well. Understanding how the different CFMs affect different (groups of) customers could help managers become more customer centric.

References

Ambler, T., & Roberts, J. H. (2008). Assessing marketing performance: Don't settle for a silver metric. *Journal of Marketing Management*, *24*(7–8), 733–750.

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, *68*(4), 172–185.

Bain and Co. (2013). *Companies that use Net Promoter®*. (http://www.netpromotersystem.com/about/companies-using-nps.aspx (accessed March 27, 2013)).

Batislam, E. P., Denizel, M., & Filiztekin, A. (2007). Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing*, *24*(3), 201–209.

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). *Database marketing: Analyzing and managing customers*. Philadelphia: Springer.

Bolton, R. N., Lemon, K. N., & Verhoef, P. C. (2004). The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science*, *32*(3), 271–292.

Bouma, J. T., Bügel, M. S., Verhoef, P. C., Alleman, T., Wiesel, T., & Wesselius, T. (2010, April). Dutch Customer Performance Index: Het nieuwe meten van klantprestaties. *Tijdschrift voor Marketing*, 58–63.

Dixon, M., Freeman, K., & Toman, N. (2010). Stop trying to delight your customers. *Harvard Business Review*, *88*(7–8), 116–122.

Dong, S., Ding, M., Grewal, R., & Zhao, P. (2011). Functional forms of the satisfaction-loyalty relationship. *International Journal of Research in Marketing*, *28*(1), 38–50.

Farris, P., Bendle, N. T., Pfeifer, P. E., & Reibstein, D. J. (2006). *Marketing metrics: Fifty + metrics every marketer should know* (1st ed.). Philadelphia: Wharton School Publishing.

Fielding, A. (2004). Role of the Hausman test and whether higher level effects should be treated as random or fixed. *Multilevel Modelling Newsletter*, *16*(2), 3–9.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Essex, UK: Pearson Education Limited.

Gruca, T. S., & Rego, L. L. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing*, *69*(3), 115–130.

Gupta, S., Lehmann, D. R., & Stuart, J. S. (2004). Valuing customers. *Journal of Marketing Research*, *41*(1), 7–18.

Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science*, *25*(6), 718–739.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *International Journal of Service Industry Management*, *7*(4), 27–42.

Hanssens, D. M. (2009). *Empirical generalizations about marketing impact*. Cambridge, MA: Marketing Science Institute.

Hanssens, D. M., Pauwels, K. H., Srinivasan, S., Vanhuele, M., & Yildirim, G. (2014). Consumer attitude metrics for guiding marketing mix decisions. *Marketing Science*, *33*(4), 534–550.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Ittner, C. D., & Larcker, D. F. (1998). Are non-financial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, *36*(3), 1–35.

Johnson, C. R., & Schultz, D. E. (2004). A focus on customers. *Marketing Management*, *13*(5), 20–26.

Keiningham, T. L., Cooil, B., Aksoy, L., Andreassen, T. W., & Weiner, J. (2007). The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet. *Managing Service Quality*, *17*(4), 361–384.

Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, *71*(3), 39–51.

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, *43*(2), 276–286.

Mintz, O., & Currim, I. S. (2013). What drives managerial use of marketing and financial metrics and does metric use affect performance of marketing-mix activities? *Journal of Marketing*, *77*(2), 17–40.

Mittal, V., & Kamakura, W. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, *38*(1), 131–142.

Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Science*, *25*(5), 426–439.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, *43*(2), 204–211.

Oliver, R. L., Rust, R. T., & Varki, S. (1997). Customer delight: Foundations, findings, and managerial insight. *Journal of Retailing*, *73*(3), 311–336.

Onrust, M., Verhoef, P. C., Van Doorn, J., & Bügel, M. S. (2014). *When doing good leads to increased customer loyalty: Why weak firms can benefit from CSR*. Working paper. University of Groningen.

Ou, Y. C., De Vries, L., Wiesel, T., & Verhoef, P. C. (2014). The role of consumer confidence in creating customer loyalty. *Journal of Service Research*, *17*(3), 339–354.

Ou, Y. C., Verhoef, P. C., & Wiesel, T. (2014). *The effects of customer equity drivers on loyalty intentions across industries and firms*. Working paper. University of Groningen.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379.

Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis. *Multilevel Modelling Newsletter*, *1*(2), 10–12.

Rego, L. L., Morgan, N. A., & Fornell, C. (2013). Reexamining the market share-customer satisfaction relationship. *Journal of Marketing*, *77*(5), 1–20.

Please cite this article as: de Haan, E., et al., The predictive ability of different customer feedback metrics for retention, *Intern. J. of Research in Marketing* (2015), http://dx.doi.org/10.1016/j.ijresmar.2015.02.004

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*(12), 46–57.

Risselada, H., Verhoef, P. C., & Bijmolt, T. H. A. (2010). Staying power of churn prediction models. *Journal of Interactive Marketing*, *24*(3), 198–208.

Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. New York: Free Press.

Shah, D., Kumar, V., Qu, Y., & Chen, S. (2012). Unprofitable cross-buying: Evidence from consumer and business markets. *Journal of Marketing*, *76*(3), 78–95.

Shah, D., Rust, R. T., Parasuraman, A., Staelin, R., & Day, G. S. (2006). The path to customer centricity. *Journal of Service Research*, *9*(2), 113–124.

Shankar, V., Smith, A. K., & Rangaswamy, A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, *20*(2), 153–175.

Snijders, T. A. B., & Berkhof, J. (2007). Diagnostic checks for multilevel models. In J. De Leeuw, & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–175). New York: Springer.

Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research*, *47*(4), 672–684.

Steenkamp, J. -B. E. M., & Geyskens, I. (2014). Manufacturer and retailer strategies to impact store brand share: Global integration, local adaptation, and worldwide learning. *Marketing Science*, *33*(1), 6–26.

Streukens, S., & De Ruyter, K. (2004). Reconsidering nonlinearity and asymmetry in customer satisfaction and loyalty models: An empirical study in three retail service settings. *Marketing Letters*, *15*(2–3), 99–111.

Van Doorn, J., Leeflang, P. S. H., & Tijs, M. (2013). Satisfaction as a predictor of future performance: A replication. *International Journal of Research in Marketing*, *30*(3), 314–318.

Van Doorn, J., & Verhoef, P. C. (2008). Critical incidents and the impact of satisfaction on customer share. *Journal of Marketing*, *72*(4), 123–142.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, *68*(4), 106–125.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, *67*(4), 30–45.

Verhoef, P. C., & Leeflang, P. S. H. (2009). Understanding the marketing department's influence within the firm. *Journal of Marketing*, *73*(2), 14–37.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196.

Warner, R. E. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Los Angeles: Sage.

Yeung, M. C. H., Ramasamy, B., Chen, J., & Paliwoda, S. (2013). Customer satisfaction and consumer expenditure in selected European countries. *International Journal of Research in Marketing*, *30*(4), 406–416.

Zeithaml, V. A., Bolton, R. N., Deighton, J., Keiningham, T. L., Lemon, K. N., & Petersen, J. A. (2006). Forward-looking focus: Can firms have adaptive foresight? *Journal of Service Research*, *9*(2), 168–183.

Please cite this article as: de Haan, E., et al., The predictive ability of different customer feedback metrics for retention, *Intern. J. of Research in Marketing* (2015), http://dx.doi.org/10.1016/j.ijresmar.2015.02.004