

Over-Time Variation in Individual's Customer Satisfaction scores

John Dawes*, Ehrenberg-Bass Institute

Lara Stocchi, Flinders University, South Australia

Francesca Dall'Olmo-Riley, Kingston University

* corresponding author John.Dawes@marketingscience.info

Keywords: Customer Satisfaction, Variability, Regression to the Mean, Psychometrics

Over-Time Variation in Individual's Customer Satisfaction scores

Abstract

Customer satisfaction is a commonly used business performance metric. Despite the widespread use of satisfaction surveys, little is known about how stable individual's satisfaction scores are. If individual's scores show instability, this has implications for market research design and managerial actions. To investigate the stability of satisfaction scores, this study uses data from a two-wave satisfaction survey in which the same respondents were interviewed six weeks apart. The respondents had no recorded purchase with the retailer between survey waves. The main finding is that only 49% of respondents give exactly the same satisfaction score on a 1-7 scale when re-surveyed. After aggregating the results into three simple categories of dissatisfied, neutral to somewhat satisfied, and satisfied, the proportions who stay in the same category from one survey to the next are 44%, 57% and 82% respectively, despite the overall average score for the sample staying the same. The changes in scores are a manifestation of *regression to the mean*, whereby those who give a low or a high score the first time tend to regress up or down toward the overall average score the next time. The main management implications are (1) interventions aimed at low or high-satisfaction customer groups need to take regression to the mean into account; (2) attempts to relate individual's satisfaction scores to future behavior (e.g., loyalty, brand-switching) should use scores averaged over two surveys (3) the oft-quoted belief that dissatisfied customers will tell more people compared to satisfied customers is less tenable, given that low satisfaction scores tend to regress upward more than high scores regress down.

Keywords: Customer Satisfaction, Variability, Regression to the Mean, Psychometrics

Over-Time Variability in Individual's Satisfaction scores

Introduction

Customer satisfaction is seen as an important business performance measure. It is one of the most frequently collected indicators of customer perceptions (Farris et al., 2016 p. 40).

Satisfaction scores are used to reinforce the idea that customers are important, and that all endeavours should be made to accommodate them.

Given that customer satisfaction monitoring is so widespread, the survey questions used should exhibit two key facets: validity and reliability. Validity is about whether something is measuring what is intended, while reliability is the extent to which a measure, such as a question and rating scale *accurately* reflects what it is measuring (Peter, 1979).

Much research has been conducted on the validity and reliability of satisfaction questions (e.g. Anderson & Fornell, 2000; Danaher & Haddrell, 1996; Sitzia, 1999). However, a gap exists in relation to how *consistent* or stable the responses are from the same people if they are asked again in a follow-up survey. Some academic research has been conducted on this issue, under the rubric of test-retest reliability. Studies show fairly high, but not perfect test-retest scores, for example Lam (1997) reported scores of around 0.7 to 0.8 (a perfect reliability score is 1.0) despite the overall sample score being the same from one survey wave to another. This suggests respondents are to some extent inconsistent in their satisfaction scores, but no work has examined exactly how many people give the same or different score when asked again in a follow-up survey.

The notion of instability in individuals' satisfaction scores reflects findings about the instability of brand attributes. Research has shown that when consumers are asked if they consider a brand has certain attributes (e.g., 'Good value for money') over multiple surveys, there will be a proportion of respondents who give different answers in a second survey to what they gave in the first (e.g. Castleberry et al., 1994). The effect has been called a 'repeat-rate' and it averages 50% (Dall'Olmo Riley et al., 1997).

The existence of this pattern in brand attributes research raises the question as to whether consumer responses to satisfaction questions also exhibit a similar repeat-rate. If they do, this might provide a new perspective on how to collect, interpret and use customer satisfaction data. A related question is whether there is more instability in low satisfaction scores than high ones. There are some grounds for thinking this might be the case. Research on brand attributes has found negative attributes are less stable among the same individuals than positive ones (Sharp & Winchester, 2002). Given the large investments made in customer satisfaction measurement and enhancement, the answer to these questions will be of interest to marketing and insight managers, as well as market research providers.

We next review relevant research on satisfaction measurement. We explain how related research implies there should be individual-level instability in responses. We then discuss how changes in responses by the same individuals (without intervening service experience) may be a manifestation of regression to the mean, and explain that concept. Next, we draw from the literature about brand attribute repeat-rates to form an expectation that low satisfaction scores on will be inherently less stable than high scores. We test the extent of a satisfaction repeat-rate and the comparative instability of lower and higher satisfaction scores using a large survey dataset of over 12,000 respondents in which the same individuals were surveyed twice, six weeks apart.

Customer Satisfaction measurement

Satisfaction is an unobservable concept, or ‘construct’, typically measured via the survey responses of customers. The extent to which a customer is satisfied is usually measured via either an agreement scale, i.e. ‘strongly disagree to ‘strongly agree’ (e.g. Nicholls & Gilbert, 1998) or other response scales such as ‘very dissatisfied ... very satisfied’ (e.g. Mihelis et al., 2001). Due to the fact that satisfaction is not directly observable and is inferred from survey responses, considerable attention has been paid to the measurement qualities of satisfaction questions. One psychometric measure of interest is *reliability*, defined as “the degree to which measures are free from error and therefore yield consistent results” (Peter, 1979 p. 6).

Reliability is often measured as the extent to which various scale items designed to measure the same construct correlate with each other (Churchill, 1979).

Reliability can also be measured as the correlation of individual’s scores over two surveys (Peter, 1979). This approach is called *test-retest reliability*. Several studies have examined

the test-retest reliability of customer satisfaction. They reported coefficients ranging between 0.65 and 0.80 (Westbrook, 1980); 0.6 to above 0.9 (Torkzadeh & Doll, 1991); and 0.7 to 0.8 (Lam & Woo, 1997). The overall sample average scores in these studies tend to be extremely stable. However, test-retest correlations smaller than 1.0 imply that some respondents gave different answers from one survey to another – in other words, some people who initially scored a business say, ‘5’ on a numeric scale or ticked ‘quite satisfied’ for satisfaction give it a different score the next time. The possibility that the same survey respondents could give a different satisfaction score has never been explicitly examined before, but poses several implications. First, and most elementary is the implicit assumption that customer satisfaction scores give an accurate indication of clients’ view on the provider. A business could reasonably assume dissatisfied customers represent a distinct segment, as per Baumann, Burton & Elliot (2005), or might target dissatisfied customers for intervention (Sulek et al., 1995), or very satisfied customers with rewards. However, the basis for intervention is less certain if scores by the same person on the same entity are different if they were asked again. Second, while customer satisfaction is seen as an extremely important business performance measure (Anderson & Fornell, 2000), past research has highlighted some counterintuitive facts. One is that many highly satisfied customers defect (Jones & Sasser, 1995), another is that customer satisfaction for firms and brands is not positively correlated with market share (Fornell, 1995). Exploring the degree to which individual’s scores are stable or unstable can shed light on these issues.

As mentioned, there is limited past research on the stability of customer satisfaction scores over time. However, we can draw on other literature to frame expectations and an approach. Past studies have examined brand attribute ‘repeat-rates’. For example, Dall’Olmo Riley et al. (1997) found that of all the respondents who agreed that a brand has some attribute (e.g., ‘Good value for money’ or ‘Removes stains’ for laundry detergents) the first time they were surveyed, only around 50% agreed again the second time. Dall’Olmo Riley (1997) confirmed that this result was robust across many different conditions, and the same empirical pattern is documented in other studies (Castleberry et al., 1994; Rungie et al., 2005; Sharp, 2002). The question arises, what proportion will give the same response for customer satisfaction questions? Stated formally, our first question is therefore:

RQ1. What is the repeat-rate for customer satisfaction scores from the same respondents over two surveys?

Note, the assumption is made that respondents have no intervening interaction with the service provider that would change their scores.

Low versus High scores

If there are differences in the stability of satisfaction scores depending on whether they are low or high, there is a clear managerial implication. For example, if low satisfaction scores (dissatisfaction) are very stable over time, then managers should worry even more about dissatisfied customers than they do currently, because this means the dissatisfaction will persist. We now consider the proposition that low satisfaction scores are likely to be *less* stable over time than high satisfaction scores. There are several grounds for thinking this could be the case.

Past research shows that respondents are less likely to give the same response for negative brand attributes when re-surveyed, compared to positive ones (Sharp & Winchester, 2002). Examples of negative brand attributes in a services context are statements such as ‘Doesn’t deliver on promises’ or ‘Doesn’t care about customers’ (Sharp & Winchester, 2002). Therefore, there are some grounds for assuming that a low satisfaction score, which is akin to a negative brand attribute, might be less stable than a high score.

In addition, research on brand attributes indicates the *repeat-rate* for an attribute is correlated with its initial *response level*. That is, if more people agree a brand has a particular attribute in an initial survey, a higher proportion of them will agree again in the second survey (Dall'Olmo Riley et al., 1997). The link between this and satisfaction is that satisfaction scores are usually negatively skewed - Peterson and Wilson (1992) and Fornell (1995). This means that there are usually more positive responses (e.g., 5, 6, 7 out of 7) than negative responses (e.g., not many scores of 1, 2, 3 out of 7). In other words, there is a higher initial *response level* for positive satisfaction scores. Therefore, based on the brand attribute repeat rate work, we should expect a correspondingly higher repeat rate for positive scores; and a lower repeat rate for negative satisfaction scores.

This line of logic leads to our second research question:

RQ2. Do low satisfaction scores by individuals have a lower repeat-rate than high scores?

We now report the results of an empirical study that addresses these hypotheses.

Method and results

To address the two research questions, we use two-wave survey data for a global furniture retailer. In one of the countries where the retailer operates, Germany, the firm ran a loyalty scheme to which all these respondents had joined. As part of the scheme, they provided contact details and were invited to participate in regular online surveys. Members of the loyalty program were issued cards to scan at the time of purchase, and this information allows us to identify respondents who did *not* record a purchase in the time period between surveys. Analysing only those who did not record a purchase means we can minimize the potential that an intervening service experience affects the satisfaction repeat-rate. This approach cannot *absolutely* eliminate intervening service experiences, however. For example, it is possible some members did not bring their loyalty card to make a purchase, however on such occasions retailer staff may offer to look up the customer's details to properly record and reward their purchase. It may also be that some people visited the retailer's premises and merely browsed without buying. That said, we feel these possibilities are unlikely to have significantly influenced the results.

The first survey wave was completed by 33,607 respondents. The second survey, six weeks later was completed by 20,642 respondents. A total of 12,876 responded to both surveys (all these figures relate to consumers who did not record a purchase in the intervening period between the two survey waves). Three customer satisfaction questions, derived from Homburg, Koschate and Hoyer (2006), were administered in the survey. The satisfaction questions were as follows.

In total, how satisfied are you with X

X is the ideal furniture store

Taking into account all aspects that are related to buying furniture, I am very satisfied with X.

Respondents answered on a 7-point, one to seven scale with anchors at 1 and 7 (1=fully disagree, 7=fully agree)¹. The reliability coefficient for these three items is 0.89. This indicates that the three items have appropriately high correlations with each other and that at face value, all three items measure the same underlying construct: customer satisfaction.

We checked that participation in the second survey was not related to satisfaction level in the first. The survey 1 satisfaction score averaged over the three questions was 5.61 for those who participated in both surveys, and 5.59 for those who participated only in survey 1 (t-test non-significant, $p=0.14$).

RQ 1: repeat-rates

To address RQ1, we calculated the proportion of respondents who gave the same or different score, for each of the three questions across the two surveys. Results are shown in Table 1.

¹ Note, the data had a very small proportion of responses (less than 0.1% for each question in either survey) recorded for a scale value of zero, which had no verbal anchor. We do not know if this was meant to represent 'don't know', so we therefore disregard these responses in the analysis.

Table 1. Repeat-Rates: the proportion of respondents who give same score in two surveys

	Score in first survey *	Total N first survey	Total % giving this response in first survey	Repeat-rate (% who give the same score twice)	Average score in second survey among those who gave * response in the first survey	Change in average score from survey 1 to 2	Change in score consistent with regression to the mean?
Q1							
	1	33	0.3	39	2.6	+1.6	yes
Mean Score	2	101	0.8	21	3.4	+1.4	yes
Survey 1= 5.8	3	293	2	22	4.0	+1.0	yes
Survey 2 = 5.8	4	842	7	29	4.7	+0.7	yes
	5	2679	21	39	5.3	+0.3	yes
	6	5478	43	54	6.0	0.0	no
	7	3450	27	68	6.6	-0.4	yes
Total		12,876	Avg.	52			
Q2							
	1	93	0.7	38	2.5	+1.5	yes
Mean Score	2	316	2	24	3.3	+1.3	yes
Survey 1= 5.4	3	658	5	24	3.8	+0.8	yes
Survey 2 = 5.4	4	1701	13	34	4.5	+0.5	yes
	5	3419	27	40	5.2	+0.2	yes
	6	4106	32	51	5.8	-0.2	yes
	7	2551	20	64	6.5	-0.5	yes
Total		12,844	Avg.	46			
Q3							
	1	41	0.3	34	2.6	+1.6	yes
	2	172	1	23	3.3	+1.3	yes
Mean Score	3	421	3	22	4.0	+1.0	yes
Survey 1= 5.6	4	1187	9	29	4.6	+0.6	yes
Survey 2= 5.6	5	2869	22	37	5.2	+0.2	yes
	6	4959	39	51	5.9	-0.1	yes
	7	3173	25	63	6.5	-0.5	yes
Total		12,822	Avg.	48			

The repeat-rates across the three questions are 52%, 46% and 48% for an average of 49%; therefore 51% gave a response that differed by at least one scale point. We also created an overall averaged score for the three questions, converted it to 1-to-7 score and calculated its repeat-rate to be 51%, the same as the single questions. This indicates the results are not due to an inherent level of random error associated with single measures (e.g. Diamantopoulos et al., 2012).

The figure of approximately 50% response instability is high, but we recognize that the precise repeat-rate depends on the number of scale point used – more scale points will mean a lower repeat rate. We therefore aggregated the responses into three classes – dissatisfied (1 to 3), satisfied (6 and 7), with the remainder, 4 and 5 representing neutral to somewhat satisfied.

This approach is commonly used to simplify satisfaction survey results, for example Anderson and Mittal (2000) discuss the ‘top-two box’ approach in which the top two scores in a scale are aggregated and reported. In a similar fashion, the calculation of Net Promoter scores involves aggregating responses on a 0-10 scale into only two categories (zero to six, then nine and ten) but discards responses between seven and eight out of ten (Reichheld, 2003). However, in our analysis we do not discard the middle responses.

After aggregating the responses in this way, we then calculated the repeat-rate for the resultant dissatisfied, neutral to somewhat satisfied, and satisfied responses. The results are shown below.

Table 2. Repeat Rates for Dissatisfied, Neutral to Somewhat Satisfied, and Satisfied respondents

	Satisfaction Question 1		Satisfaction Question 2		Satisfaction Question 3		Average repeat rate
	% response in Survey 1	% repeat rate in Survey 2	% response in Survey 1	% repeat rate in Survey 2	% response in Survey 1	% repeat rate in Survey 2	
Dissatisfied (1-3)*	3	42	8	47	5	44	44
Neutral to somewhat Satisfied (4-5)*	27	55	40	60	32	56	57
Satisfied (6-7)*	69	85	52	80	63	82	82
Weighted Average		76		69		72	

*Scores out of 7

We see that there is still considerable instability in the response rate for these aggregated categories. On average, 44% of dissatisfied customers are dissatisfied the next time they are surveyed, 57% of neutral to somewhat satisfied customers are the same the next time, and 82% of satisfied customers are still satisfied when re-surveyed. Given that a large proportion of initial responses are either 6 or 7 indicating satisfied, the weighted average repeat rate over

the three questions is 72%. These proportions are certainly large enough for the issue of response instability to be managerially relevant.

A phenomenon called *regression to the mean* helps explain how the average scores stay the same over the two surveys, but there is a high level of instability in individual's scores. Regression to the mean is when "unusually large or small measurements tend to be followed by measurements that are closer to the mean" (Barnett et al., 2005 p.34). The results for Q1 in Table 1 definitely show this effect. The overall mean score for Q1 was 5.8. Those who scored 1 in the first survey scored on average 2.6 in the second survey - a regression upward to the mean of 1.6 points. The same pattern is apparent for those who scored 2, 3, 4, and 5 - they all regressed upwards towards the mean score. Among those who scored 6, there was no change on average. The group who scored 7 regressed downward slightly to 6.6 points, a change towards the mean of -0.4 points². The same pattern is obvious in the responses to question 2 and 3. The changes in scores are consistent with a regression to the mean effect in 20 out of 21 cases in Table 1 ($p < 0.001$).

RQ2: Instability of low and high scores

We see strong evidence that low satisfaction scores are less stable than high ones. In Table 1, we see for Q1 the repeat-rates for the low scores of 1, 2 and 3 are 39, 21 and 22 percent respectively for a weighted average of 25%. The repeat-rates for the high scores 6 and 7 are 51 and 64 percent respectively for a weighted average of 56% (difference significant at $p < 0.01$). The same pattern is evident for Q2 (low 23%, high 56%) and Q3 (low 29%, high 59%). Likewise, we see the same pattern when the responses are aggregated: table 2 shows the repeat rate for dissatisfied responses is 44%, for satisfied responses it is 82% (difference statistically significant at $p < 0.01$).

This is an important finding: respondents who initially give a low satisfaction score are less likely to give the same low score in a subsequent survey, compared to respondents who initially give a high score and give a second high score in a subsequent survey.

² Those who scored 7 cannot go higher, but this type of ceiling effect is part of the regression to the mean phenomena.

Discussion and implications for theory & practice

The key results from this study are as follows. On average, 49% of respondents give the same customer satisfaction score on a 1-7 scale when re-surveyed. After aggregating the results into three simple categories of dissatisfied, neutral to somewhat satisfied, and satisfied, the proportions who stay in the same category from one survey to the next are 44%, 57% and 82% respectively.

This effect occurs for single questions, as well as for a score created from averaging the results of three satisfaction questions. Moreover, the changes in scores exhibit a regression to the mean effect, whereby respondents who initially give low scores increase them on average when re-surveyed; and respondents who initially give high scores tend to reduce their scores on average when re-surveyed. Lastly, the study finds that low customer satisfaction scores exhibit a lower repeat-rate, in other words more instability, than high scores.

The fact that satisfaction scores exhibit a repeat-rate poses several implications for research and practice. First, it suggests that when people are surveyed about their level of customer satisfaction they do not necessarily have a ready or thoughtfully considered idea in their mind of their perceived satisfaction³. Rather, they may browse through their *memory* to locate relevant salient information about the business running the survey. Since the information retrieved to give an answer to the survey question will be determined by whichever salient link in memory is thought of at the time of being asked, the respondent might access different information at a different time. This potential explanation resonates with what Wilson & Hodges (1992) state about attitudes: "... people often have a large, conflicting "data base" relevant to their attitudes on any given topic, and the attitude they have at any given time depends on the subsets of these data to which they attend" (p. 38). The upshot is that while the overall results of a survey of many individuals represents an accurate estimate of population level customer satisfaction, the individual scores are far less fixed or reflective of 'true' satisfaction levels than has been thought to date.

The focus of this study was individual-level variation. Businesses and market research agencies are usually more interested in overall average scores, or scores among groups and

³ Many service encounters are fairly innocuous or engender quite low levels of engagement between the customer and service provider, therefore impressions about satisfaction may be forgotten quite quickly for a large proportion of people.

how these change from survey to survey, rather than individual clients' variation over time. Nonetheless, the results of this study have implications for commercial market research. To illustrate, consider that businesses can now survey customers en masse and administer satisfaction questions, and use the results for intervention effort – i.e., to design rewards or special offers to assuage clients with low levels of satisfaction. Indeed, businesses might wish to see the results of their interventions later. However, our study suggests that if a business identifies low-satisfaction clients and targets them with an intervention and re-surveys them later, there is a good chance it could falsely conclude the intervention worked. This is because, as per the results returned here, low scores tend to naturally regress upwards. Similarly, a business might wish to reward highly satisfied customers with special offers to keep them satisfied. Yet, they might be disappointed to see that the scores among this group drop in follow-up surveys. To mitigate these risks, we suggest that in order to evaluate the targeted promotions, businesses and research providers need to calculate changes in customer satisfaction relative to an expected regression to the mean effect. Indeed, the idea of segmenting and targeting customers (Baumann et al., 2005; Sulek et al., 1995) based on their satisfaction level seems unadvisable; as many clients classified as satisfied or dissatisfied from a survey would not have been classified as such in a second survey.

In addition to these considerations, this study finds that low, or negative satisfaction scores exhibit lower repeat-rates, and so more individual-level instability. This finding is consistent with previous research on the instability of negative brand attributes (Sharp & Winchester, 2002). While there is a vast literature on attitudes and attitude measurement, there seems to be no previous work that has examined whether negatively held attitudes are more or less stable than positive ones. While the results of this study certainly do not imply that one should ignore unhappy clients, it appears that for a good proportion of such customers, dissatisfaction dissipates over time. Consider that a marketing organization might consider reparations or inducements to clients who report low satisfaction in a survey. These findings suggest it is worth considering, at least, how many clients express sustained low satisfaction a second time, and focusing on them.

Next, the finding pertaining to less stability for low satisfaction scores has an implication for the oft-quoted research that claims dissatisfied customers will tell more people how they feel than satisfied ones (e.g. Eccles & Durand, 1998). That is, if dissatisfied customers tend to notably regress upward towards the mean satisfaction level over time, then they should not

necessarily be any more inclined to spread (negative) word of mouth than satisfied clients – a conclusion in line with previous research by East, Hammond and Wright (2007) who reported that positive word of mouth is around three times as prevalent as negative.

In a similar vein, the present study has implications for customer loyalty research. Past work has suggested many satisfied customers defect (e.g. Jones & Sasser, 1995; Reichheld, 1996). From the results of this study, we now know that a significant proportion of customers who would have been classified as satisfied / dissatisfied in one survey would be rather less satisfied / more satisfied respectively if asked again in another. This variation may at least partially explain why there is a poor link between customer satisfaction and subsequent behavioral outcomes. Practitioners seeking to model the links between satisfaction and later behavior (stated or actual measures of brand switching or loyalty) should arguably employ the average satisfaction score of individuals over two surveys for such work.

This study was undertaken using one specific category, namely furniture. Future work should examine satisfaction score stability in services contexts such as banking, insurance, or telecommunications. Satisfaction surveys are widely used in these sectors and evidence about response stability would be very informative to researchers and clients in these markets.

Lastly, the approach used in this study could be applied to other prominent survey-based metrics. One such metric is the Net Promoter Score (NPS). Many businesses employ the Net Promoter Score to monitor the likely extent of positive word of mouth from satisfied clients (see East, Romaniuk and Lomax (2011) and Fisher & Kordulepski (2019) for critiques of this method). However, no work has ever been conducted on the extent to which ‘promoters’ still agree they would recommend the brand, if re-interviewed. Given the results of the present study, and the large investments in NPS, an investigation into its response stability would be timely.

Appendix: Raw scores for Surveys 1 and 2

Table 3. In total, how satisfied are you with X

		Response in second survey							
		1	2	3	4	5	6	7	Total
Response in first survey	1	13	6	4	5	3	1	1	33
	2	9	21	29	20	10	10	2	101
	3	7	25	64	93	69	29	6	293
	4	6	20	78	248	298	155	37	842
	5	3	30	83	353	1036	1001	173	2679
	6	5	27	41	175	861	2967	1402	5478
	7	5	9	7	41	147	894	2347	3450
Total		48	138	306	935	2424	5057	3968	12,876

Table 4. X is the ideal furniture store

		Response in second survey							
		1	2	3	4	5	6	7	Total
Response in first survey	1	35	18	15	16	5	3	1	93
	2	24	75	81	78	46	9	3	316
	3	22	69	159	230	135	34	9	658
	4	6	53	193	585	592	224	48	1701
	5	6	39	137	542	1371	1108	216	3419
	6	5	24	42	220	841	2091	883	4106
	7	1	4	4	45	137	724	1636	2551
Total		99	282	631	1716	3127	4193	2796	12,844

Table 5. Taking into account all aspects that are related to buying furniture, I am very satisfied with X.

		Response in second survey							
		1	2	3	4	5	6	7	Total
Response in first survey	1	14	10	6	6	2	2	1	41
	2	18	39	46	34	24	8	3	172
	3	9	43	94	140	86	39	10	421
	4	6	35	139	345	425	200	37	1187
	5	2	33	105	451	1058	1004	216	2869
	6	8	33	47	204	957	2545	1165	4959
	7	3	5	11	58	168	921	2007	3173
Total		60	198	448	1238	2720	4719	3439	12,822

References

- Anderson, E. W., & Fornell, C. (2000). Foundations of the American Customer Satisfaction Index. *Total Quality Management*, 11(7), S869-S882.
- Anderson, E. W., & Mittal, V. (2000). Strengthening the Satisfaction-Profit Chain. *Journal of Service research*, 3(2), 107-120. doi:10.1177/109467050032001
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215-220. doi:10.1093/ije/dyh299
- Baumann, C., Burton, S., & Elliott, G. (2005). Determinants of customer loyalty and share of wallet in retail banking. *Journal of Financial Services Marketing*, 9(3), 231-248.
- Castleberry, S. B., Barnard, N. R., Barwise, T. P., Ehrenberg, A., & Dall'Olmo Riley, F. (1994). Individual Attitude Variations Over Time. *Journal of Marketing Management*, 10(1-3), 153-162.
- Churchill, G. A., Jr. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of marketing research*, 16(February), 64-73.
- Dall'Olmo Riley, F., Ehrenberg, A., Castleberry, S. B., Barwise, T. P., & Barnard, N. R. (1997). The variability of attitudinal repeat-rates. *International Journal of Research in Marketing*, 14(5), 437-450.
- Danaher, P. J., & Haddrell, V. (1996). A comparison of question scales used for measuring customer satisfaction. *International Journal of Service Industry Management*, 7(4), 4-26.
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3), 434-449. doi:10.1007/s11747-011-0300-3
- East, R., Hammond, K., & Wright, M. (2007). The relative incidence of positive and negative word of mouth: a multi-category study. *International Journal of Research in Marketing*, 24(2), 175-184.
- East, R., Romaniuk, J., & Lomax, W. (2011). The NPS and the ACSI: A critique and an alternative metric. *International Journal of Market Research*, 53(3), 15.
- Eccles, G., & Durand, P. (1998). Complaining customers, service recovery and continuous improvement. *Managing Service Quality: An International Journal*, 8(1), 68-71. doi:<https://doi.org/10.1108/09604529810199421>
- Farris, P., Bendle, N., Pfeifer, P. E., & Reibstein, D. J. (2016). *Marketing metrics: The manager's guide to measuring marketing performance* (3 ed.). New Jersey, United States: Pearson.
- Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, 35(1), 138-151. doi:10.1002/asmb.2417
- Fornell, C. (1995). The Quality of Economic Output: Empirical Generalizations About Its Distribution and Relationship to Market Share. *Marketing science*, 14(No. 3, Part 2), G203-G211.

- Homburg, C., Koschate, N., & Hoyer, W. D. (2006). The role of cognition and affect in the formation of customer satisfaction: a dynamic perspective. *Journal of Marketing*, 70(3), 21-31. doi:<https://doi.org/10.1509/jmkg.70.3.21>
- Jones, T. O., & Sasser, W. E., Jr. (1995). Why Satisfied Customers Defect. *Harvard Business Review*(November-December), 88-99.
- Lam, S. S. K., & Woo, K. S. (1997). Measuring service quality: a test-retest reliability investigation of SERVQUAL. *International Journal of Market Research*, 39(2), 381.
- Mihelis, G., Grigoroudis, E., Siskos, Y., Politis, Y., & Malandrakis, Y. (2001). Customer satisfaction measurement in the private bank sector. *European Journal of Operational Research*(130), 347-360.
- Nicholls, J. A. F., & Gilbert, G. R. (1998). Parsimonious Measurement of Customer Satisfaction with Personal Service and the Service Setting. *Journal of consumer marketing*, 15(No. 3), 239-253.
- Peter, J. P. (1979). Reliability: A Review of Psychometric Basics and Recent Marketing Practices. *Journal of marketing research*, 16(1), 6-17.
- Peterson, R. A., & Wilson, W. R. (1992). Measuring Customer Satisfaction: Fact and Artifact. *Journal of the Academy of Marketing Science*, 20(No. 1), 61-71.
- Reichheld, F. F. (1996). Learning from customer defections. *Harvard Business Review*, 74(2), 56-69.
- Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*(December), 46-54.
- Rungie, C., Laurent, G., Dall'Olmo Riley, F., Morrison, D. G., & Roy, T. (2005). Measuring and modeling the (limited) reliability of free choice attitude questions. *International Journal of Research in Marketing*, 22(3), 309-318.
- Sharp, A. (2002). *Searching for boundary conditions for an empirical generalisation concerning the temporal stability of individual's perceptual responses*. (PhD Doctor of Philosophy). University of South Australia, Adelaide.
- Sharp, A., & Winchester, M. (2002). *The Temporal Stability of Negative Brand Image Attributes*. Paper presented at the ANZMAC, Melbourne.
- Sitzia, J. (1999). How valid and reliable are patient satisfaction data? An analysis of 195 studies. *International Journal for Quality in Health Care*, 11(4), 319-328.
- Sulek, J. M., Lind, M. R., & Maruchek, A. S. (1995). The impact of a customer service intervention and facility design on firm performance. *Management Science*, 41(11), 1763-1773. doi:10.1287/mnsc.41.11.1763
- Torkzadeh, G., & Doll, W. J. (1991). Test - Retest Reliability of the End - User Computing Satisfaction Instrument. *Decision Sciences*, 22(1), 26-37.
- Westbrook, R. A. (1980). A rating scale for measuring product/service satisfaction. *Journal of Marketing*, 44(Fall), 68-72.
- Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37-52). New Jersey: Lawrence Erlbaum Associates.