

How To Improve Customer Satisfaction Surveys:
It's about Ranking and Power Laws

By Jan Hofmeyr and Ged Parton



synovate
Customer Experience

How to Improve Customer Satisfaction Surveys: It's about Ranking and Power Laws

by Jan Hofmeyr and Ged Parton

Summary Although increased customer satisfaction leads to improved business results, research shows that the relationship between a customer's overall satisfaction and what they actually do is weak. Two reasons for this are: first, the failure to measure how customers feel about competitors; and second, the failure to recognize that what matters is a business's rank, not its rating. A customer's best rated option gets the lion's share of their loyalty; and the share each competitor gets, is governed by a power law.

Managers the world over believe that improving customer satisfaction leads to improved business results – and not without reason: there's a lot of research that proves the relationship. Yet the same research shows that the relationship is weak. As one researcher noted in an article published in the *Journal of Retailing* in 2003, overall customer satisfaction seems to have only a 'modest' impact on the share of wallet retailers get from their customers. And in an award-winning article published in the *Journal of Service Research* in the same year, Keiningham and his colleagues found a significant but weak relationship between customer satisfaction and share of wallet in financial services. In fact, research appears to show that up to 80% of what individual customers do with their money is unrelated to how satisfied they say they are in surveys.

Typically, two reasons are suggested for the weak relationship between customer satisfaction and individual behaviour. The first is heterogeneity: different kinds of customer appear to differ in their sensitivity to variations in satisfaction; or at least, in their use of scales. The second is non-linearity: research suggests that customer loyalty only increases when customers are 'extremely' satisfied.

The idea that loyalty improves when customers are 'extremely' satisfied led to the emergence in the mid 1990's of the doctrine of 'delight'. And the inclusion of heterogeneity and non-linearity has led to improved models. But some 70% of what people did with their money still remained unexplained.

The Dangers of the Doctrine of 'Delight'

We should pause for a moment to consider the doctrine of 'delight'. The challenge that 'delighting' customers poses for managers can be summed up in a word: habituation. As humans we become used to the things customer experience managers do to delight us. It's a tendency with which we're born and it's partly responsible for our evolutionary success – if it weren't for habituation we might never get restless or bored; and we might never learn anything. Habituation impels us forward and motivates us to improve things.

But for business managers, the fact that customers become 'habituated' is a problem because it creates a constantly shifting target. We (your authors) still remember the first time we walked into a hotel room in the 1990s and found a flower and a chocolate with a 'welcome' note on our pillows. It was a surprise. We were delighted. But quite quickly the novelty wore off as all good hotels did the same. Soon we came to expect flowers and chocolates. When it didn't happen, we became annoyed. And that's the problem: a customer's state of mind becomes "Oh, that's okay, but what are you going to do next." Finding ways to delight customers creates new forms of spending that are quickly taken for granted, and just as quickly matched by competitors if they pose a competitive threat. Badly managed, the doctrine of 'delight' can create a spending treadmill with a limited payback.

As we'll show however, the commitment by a business to be the best with respect to dimensions of product performance or service that are important to people does lead to greater customer loyalty. Businesses may use 'regular delight' as a means to get managers to think about how to create and sustain competitive advantage. But it's not the only mechanism. The key words are "being best at what's important to people."

As an example, consider the airline industry. When British Airways introduced flat beds in the 1990s it created a stir and lots of good publicity, and it considerably enhanced the experience of flying long distances. Soon the world's other airlines had to match the offer – and over time, they did. But the spending that the constant focus on 'enhanced customer experiences' created in the airline industry opened up a gap for 'no frills' airlines like Ryan Air. These airlines offer just one thing: they're cheap. Flying with them isn't delightful. But they have a competitive advantage when it comes to something that's important to people, namely, price.

*It's not about Delight,
it's about beating the
Competition*

The main reason that a lot of customer satisfaction research fails to relate strongly to what individual customers do is that it devotes too little time to asking respondents about the competition. Often when we're on a plane and we're coming in to land, a cabin attendant will say something like, "Thank you for flying with us today – we know you have a choice." And yet, many if not most of the customer satisfaction surveys ask mainly about 'this trip' or 'this airline' or this 'service provider'. Some in our industry might reply "But we do take the competition into account – we benchmark."

So let's think about benchmarking.

i. Benchmarking – or how not to take competition into account

Benchmarking is the practice of comparing a business's satisfaction scores with the scores collected for similar businesses in similar surveys. Scores above the benchmark are good. Scores below are obviously bad. Benchmarking requires a database and an investment in quality control when building the database, and to ensure that the correct benchmarks are chosen. And it creates expenses because of the logistics involved.

There are two problems with benchmarks. First, they're backward looking – by definition, you cannot have benchmarks for anything new a business might do to create competitive advantage; and second, they don't get us closer to understanding individual customer behaviour. Even when a business beats its benchmark and does better business, the weaknesses of the basic approach remain, and the correlation of satisfaction with individual behaviour stays poor. Customers are constantly comparing what they get from one product, service, or brand with what they get from others. And it's that comparison that needs to be taken into account if we're to get better at predicting how individuals will behave.

Like the doctrine of 'delight', benchmarks create a trap for customer experience managers – they lock managers into a 'backward looking' approach to customer management and into the use of dominant suppliers because good databases are needed if they're to work.

ii. Taking competition into account by asking the customer

There is a less clumsy and more informative way to get at competitor information and that's to 'just ask'. I'll illustrate with a concrete example: our company recently did a survey among grocery customers in Italy. The respondents were members of a well-known retail consumer panel. This means that we were able to combine

each panelist's attitudinal satisfaction as expressed in the survey with their actual share of wallet as recorded on the panel. We were therefore able to analyze the relationship between a customer's satisfaction with each retailer and the share of wallet they really allocated to each retailer.

We used the following question to measure customers' ratings of retailers: "Based on what you have seen or heard about each of the following retailers, even if you've never been there or never bought directly from them, I would like to know how you rate each retailer in terms of their ability to meet your needs when you do your shopping." We used a 10-point scale anchored by the words 'terrible' (1) and 'outstanding' (10). Panelists were asked to rate the retailers they used or would consider using – in other words, their evoked set. There were sixteen retailers in total. The average panelist used 3.2 retailers and considered a further 5.8. We therefore got ratings for an average of 9.0 retailers from each panelist. The inclusion of competitor ratings in this way may seem to lengthen the questionnaire considerably. But it adds less than 30 seconds because most of a respondent's time is spent reading the question rather than providing the ratings. And by focusing on the retailers that were relevant to each panelist, we didn't waste time collecting ratings for retailers that a panelist was unlikely to use.

Like researchers before us, we found a significant relationship between individual customer satisfaction scores and shares of wallet. But the relationship was poor with a correlation of just $R = 0.21$. Because of the possibility of scale response bias, we took the analysis further by mean-centering the data at respondent level. To do this we calculated the mean rating a respondent gave to the retailers they rated; and then subtracted the mean from each retailer's rating. Mean-centering in this way significantly improved the correlation between individual customer satisfaction and behaviour – from $R = 0.21$ to $R = 0.40$.

Obviously, it would not have been possible to mean-center the data and improve the relationship without collecting competitor information. The implication for businesses is that this route to improved measurement requires the collection of competitor ratings. Many managers may feel that this adds too much time to the questionnaire. In fact, however, it turns out to be quicker than conventional customer satisfaction approaches. Most commercial approaches measure customer satisfaction with three questions, typically, 'satisfy', 'recommend', 'buy again'. These three questions

are then used to derive an overall score or index for the client. In our experience it takes time for respondents to read each question and then rate the client. By contrast, our approach requires only one question. Less time is spent reading, and although more time is spent rating, people are so much faster at rating than reading, that they take less time overall.

The main weakness of the mean-centered approach is that it's not psychologically 'truthful'. Mean-centering is equivalent to having a person compare each option with a mental average. In this way it is both comparative and it takes competitor scores into account. But it's not what real customers do. Real customers go directly to a comparison of their options. They don't compare via an 'average'. A process that gets closer to direct comparison is ranking through rating.

Ranking through Rating

Ranking through rating involves the following: Ask each customer to rate the products, services, or brands in their evoked set. Use the resulting scores to create a ranking at respondent level. Then calculate the relationship between individual behaviour and the ranks.

To illustrate, the product, service or brand with the highest score gets a rank of '1'. The one that comes second is ranked '2'; and so on. This continues until all the rated options, i.e., the options in the evoked set, are ranked. Non-rated options are automatically assigned a share of wallet of zero based on the assumption that any product, service, or brand that doesn't get into the evoked set is unlikely to be bought.

An important question is: why go to the trouble of asking a rating question first and then only ranking? Why not ask respondents to rank their options directly? The answer is that, in real life, share of wallet ties are ubiquitous. In consumer panel data for packaged goods, for instance, we've found that up to 35% of panelists use two or more brands in their repertoires equally. In the retail data we're using in this example, some 31% of panelists had brands in their repertoires that tied for share of wallet. We therefore have to allow for the ties. Respondents find it easier to capture this indirectly through tied ratings; than directly through saying which products, services or brands are tied. The question then becomes, how do you turn tied ratings into ranks?

There are two main methods for dealing with ties, namely, the 'classroom' and the 'prize money' methods. By the 'classroom' method, if two options are tied for second place (for example), then they both get a rank of '2'; and the next best option gets a rank of '4'. This is the 'classroom' method because it's the way students are ranked in exams. By the 'prize money' method, the two tied options would have occupied second and third spots had they not tied – so they share the ranks and each get '2.5'. This is the 'prize money' method because it's the way prize money is usually shared.

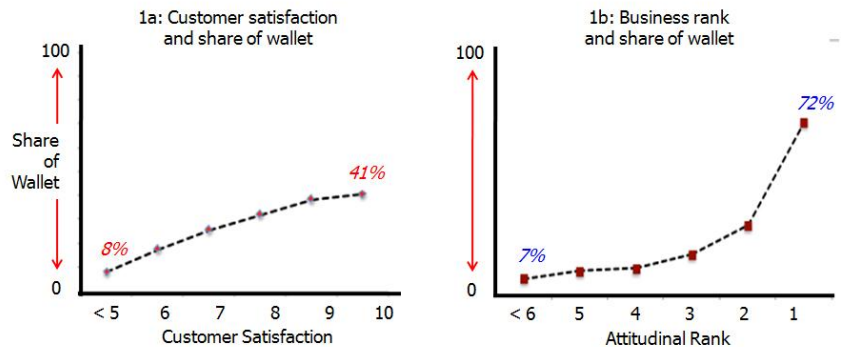
Our research shows that the 'prize money' method leads to better results.

i. Results: how ranking through rating improves prediction

In ranking, higher satisfaction scores lead to lower (meaning 'better') ranks. In our data, for instance, the best score a business can get from a customer is '10', but the best rank it can be is first. As a result, the correlation between the rank and share of wallet a business gets from its customers is negative. We can flip the negative into positive for comparative purposes. When we tested the 'prize money' ranking method on our data, we got an improved correlation with share of wallet of $R = 0.46$. By taking competition into account, we improved our ability to predict the share of wallet each individual customer would give each retailer by more than 115%, from $R = 0.21$ (old-fashioned satisfaction method), through $R = 0.40$ (mean-centering method), to $R = 0.46$ ('prize-money' ranking method).

Exhibits 1a/b show the relationship between share of wallet and both old-fashioned satisfaction (1a) and ranks (1b). At aggregate level, the relationship between satisfaction and share of wallet is very strong: retailers got an average share of just 8% from customers who rated them '5' or less, and 41% for retailers that were rated '10' (an increase of six notches). This is an improvement of more than 500%. But the relationship between rank and share of wallet is even stronger. Taking a retailer's rank from sixth to first (also six notches) leads to an increase in share of wallet of more than 1000% - from about 7% to 72%.

Exhibit 1: A comparison of the share of wallet gains that come from increased satisfaction scores versus improved ranks



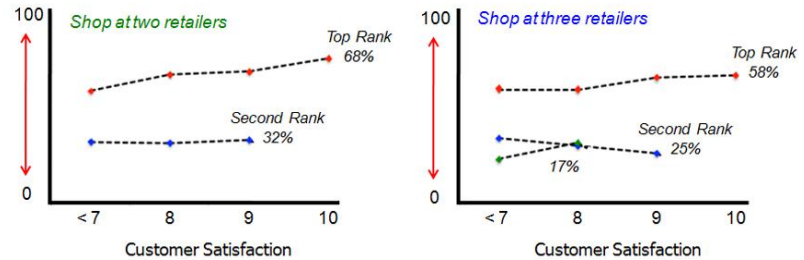
Ranking therefore leads to better results in at least two respects. First, the correlation with individual behaviour is some 115% better, and second, the payback curve is steeper.

ii. When scores are ranked, the relationship with scores disappears

One of the questions one might ask when seeing these data is “Why, when the relationship between old-fashioned satisfaction and average share of wallet is so strong, is the relationship between individual satisfaction scores and share of wallet so weak (R – 0.21)?”

The problem, as we show in exhibit 2, is that the increase in share of wallet isn’t explained by the increase in rating per se. It’s explained by the fact that the better a retailer’s rating, the higher its rank is likely to be. In other words, the improvements in share of wallet came mainly from improved ranks, not improved scores. No matter what a retailer’s score, if it was a panelist’s top ranked retailer and that panelist only had two retailers in their evoked set, it got between 59% and 68% share of wallet. By contrast, the second ranked retailer only got about 32% share of wallet no matter what its score.

Exhibit 2: An illustration of the fact that it is 'rank' that determines share of wallet rather than 'satisfaction score'



Put another way: if the top ranked retailer was rated '8', it got 68% share of wallet as long as it was top ranked. By contrast, even if the second ranked retailer got '9', it only got 32% share of wallet as long as another retailer – the top-ranked one – got '10'. Among panelists using three retailers, the share that went to the top ranked retailer fell – from about 68% to about 58%. But the pattern remained the same. The second and third ranked retailers scored about 26% and 16% respectively, irrespective of their scores. It is the fact that a high score like '8' can sometimes be associated with being the best, but at other times with being second or third that leads to the poor individual correlations for the old-fashioned 'rating' approach.

These results show that a business will do little to change the share of wallet it gets from individual customers unless that improved score leads to an improved rank. It's about being first, not scoring high – though scoring high helps a business to be first. This is also why 'extreme satisfaction' appears to drive high loyalty: the higher the satisfaction score, the more likely the business is to be a person's top ranked option. Some marketers erroneously concluded that it was only about 'delighting' their customers. What it was actually about was being seen to be 'number one'. 'Delight' might play a role in achieving this but it's not the main reason 'extreme' scores lead to more business.

iii. The relationship between rank and share of wallet is governed by a power law

Exhibit 1b shows a pattern of escalating ROI for higher ranks at individual customer level. This suggests that the relationship might be governed by a power law. That leads to the final refinement in this

approach, namely, the transformation of the ranked customer satisfaction score into an implied share of wallet by the use of a power law. We call the resulting number the business's attitudinal equity. The equation is:

$$AE_i = \frac{100}{Rank_i^{1.35} \left[\sum_{i=1}^n \left(\frac{1}{Rank_{i=1}^{1.35}} \right) \right]}$$

where 'n' is the number of options in a customer's evoked set; and 'i' is the 'ith' option in that set. The output from this equation is a number between '1' and '100' for every option – both client and competitors – in a customer's evoked set. When all the scores are summed, they add to 100. In other words, the equation turns ranks into a 'share of mind'. The correlation of this number – an attitudinal share of mind for each option – with share of wallet is R = 0.54. Overall, by taking competition into account, creating rankings through ratings, and using a power law, we improve the predictive power of satisfaction measurement by more than 150%, from R = 0.21 to R = 0.54.

Moreover, we do this with just one question, not three. And we get scores for both the client and the competitors.

*Summary: How to
Improve the Predictive
Power of Satisfaction
Research*

To improve the predictive power of customer satisfaction research:

- Use only one question (overall satisfaction), not three, but ask respondents to rate all of the products, services, or brands in their evoked sets.
- Although it may seem that this would take longer, in fact it tends to be quicker because it involves less reading (though it involves more rating).
- Having collected competitor ratings, a simple way to improve the correlation between satisfaction and behaviour is to mean-center the data at respondent level.
- However, the best way to improve the relationship is to turn the ratings into ranks and then to transform the ranks into 'shares of mind' using a power law.
- In this way the relationship can be improved while at the same time generating measures of strength for competitors.

- This allows for direct comparisons of your business with your competitors without having to use benchmarks (though they can still be used if you want).

Our research suggests that improved business results flow mainly from improved ranks rather than improved scores. Improved results only appear to flow from improved scores because higher scores tend to be associated with higher ranks. This being so, the idea that 'delight' alone drives success is wrong. What drives success is being 'number one'. Businesses are on a treadmill – the treadmill of maintaining a perceived competitive advantage in a world where innovation coupled to habituation creates a constantly rising performance 'bar'.

Jan Hofmeyr Jan Hofmeyr (Ph.D.) has developed numerous measurement systems that are widely used in marketing research including TNS's *The Conversion Modeltm* and Synovate's *Brand Value Creator*. He is currently the International Director of Synovate Laboratories and is responsible for research and development in the customer experience and branding spaces. He can be contacted at jan.hofmeyr@synovate.com.

Ged Parton A marketing services practitioner for more than 20 years, Ged Parton has worked with multinational clients across every industry throughout the world to help them understand data and utilise it to drive marketing strategy strategy. His personal obsession is data that drives smart business decision making. Ged can be reached at ged.parton@synovate.com.