D&M
RESEARCH
Since 1994

5 New Ways To Revitalise Your NPS Program

Research conducted by D&M Research

September 2018

Contents

Introduction

When Fred Reichheld first introduced the Net Promoter Score (NPS), he claimed it to be the "one number you need to grow", naming his paper as such (2003). Reichheld claims that customer loyalty is the best predictor of business success, as opposed to other standard metrics such as satisfaction (Reichheld, 1993).

The proposed index has linked loyalty with positive word-of-mouth or referrals, and it has gained widespread traction and still remains as one of the most commonly utilised measurements across a number of industries 14 years later.

Traditionally, NPS is calculated by taking the proportion of top two scorers on a 0–10 "Likelihood to Recommend" (LTR) scale (known as Promoters) and subtracting the proportion scoring in the bottom 6 (Detractors). While variations on this do exist, this paper will refer to the basic NPS as proposed by Reichheld.

NPS's simplicity, in that it aims to easily distil complex loyalty data into one easily digestible number, is one of the main appeals of the metric (Reichheld, 2006). Although the idea of only needing one number in order to grow a business is attractive, criticisms of the measure are abundant and diverse (see Chandon, Morwitz & Reinartz, 2005; East, Romaniuk & Lomax, 2011; Grisaffe, 2007; Keiningham, Aksoy & Cooil, 2008; Kristensen & Eskildsen, 2011; Morgan & Rego, 2006; Pingitore, Morgan, Rego, Gigliotti & Meyers, 2007; Pollack & Alexandrov, 2013; Poon, Koehler & Buehler, 2014; Schneider et al, 2008 for discussions).

The issues discussed throughout the literature presented three main areas of interest, resulting in three main questions:

1. Do NPS scores predict actual recommending behaviour?
2. Does the motivation behind scores match the NPS segments of Promoter, Passive and Detractor?
3. Can an empirically derived, actionable segmentation be built?

Not only are the answers to these questions inherently interesting, they also have propelled the development of solutions to many of the problems typically associated with NPS. In summary, these are:

1. Get under the hood
2. Get sentimental
3. Get in the driver's seat
4. Go to the top of the class
5. Get real

Research

Participants and Procedure

We used:

- Representative sample of 983
- Weighted to Census data by age, gender, location
- Online survey software optimised for mobile to allow for completion on all devices
- Introductions and honesty agreements were displayed, and demographic data was collected

Questionnaire

A series of carefully curated questions were asked of all participants. Identifying main bank was defined by everyday transaction account, and was assessed in terms of Satisfaction as well as the traditional LTR with reason for score. Behavioural elements were also assessed, as participants were asked if they actually had recommended their main bank in the last 12 months. If yes, the number of times this recommending behaviour occurred was collected. Finally, participants were asked to define how many of these real recommendations were reactive (i.e. someone asked for their opinion) or proactive (i.e. they mentioned it unprompted).

Statistical Analysis and Coding

All statistical analyses were performed in Q software. Chi squared analysis was utilised in almost every comparative aspect of the present study. There were two exceptions to this: the first being when comparing average NPS ratings, where an independent samples t-test was used, and the second being when assessing mean number of recommendations given, which utilised an F-test ANOVA. A correlational analysis between Satisfaction and LTR was also performed. The p-value throughout was set to 0.05, and corrected for multiple comparisons.

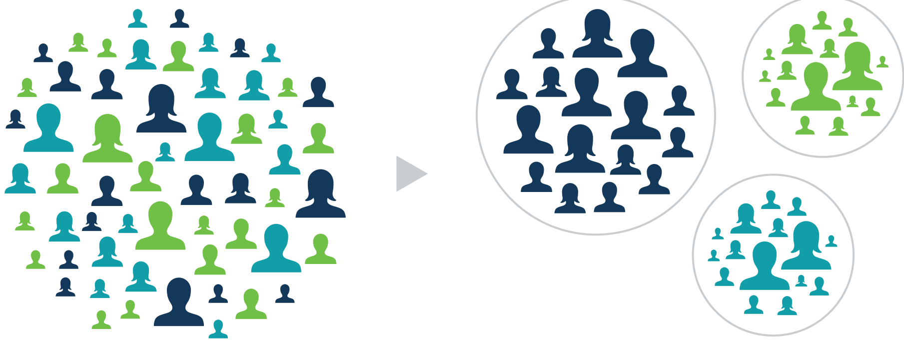
Each verbatim response was individually coded through the use of a code frame that was built based on the first 100 responses. Common themes for positive sentiment included good customer service, range of products, and reasonable fees/interest rates, with negative seeing the opposite of these. Examples of neutral sentiment include being unwilling to make recommendations, and all banks are the same.

The exploratory segmentation was completed using Latent Class Analysis. This allowed us to find the most mathematically sound "clusters" of scorers, known as segments (see Figure 1). Each of the solutions to these analyses were statistically tested to determine the goodness-of-fit for each model.

The analysis regarding recommending behaviour required some further sub-division, as recommending behaviour was assessed using three separate measures: P12M recommendation, number of recommendations, and type of recommendations.

1. First, the binary measure querying whether any recommendations had occurred in the past 12 months (P12M) was compared across the LTR scale. Although it was expected that those giving a higher LTR score would have a higher proportion of actual recommenders, it was also hypothesised that endorsing behaviour would be present throughout all segments. That is, it was anticipated that not only Promoters would contain substantial proportion of recommenders, but that Passives and Detractors would also (although at decreased rate).
2. Next, the numeric value assessing the number of total recommendations made was compared against LTR in a similar fashion. Again, higher scores were expected to see more recommendations on average. However, it was hypothesised that those scoring an 8 would not have a significantly lower average number of recommendations than those scoring a 9 for LTR.
3. While there was little expected differentiation with who might be making recommendations, it was expected that LTR score (and traditional NPS segments) may be distinguishable by what type of recommendations were given. It was hypothesised that Detractors and Passive recommenders would mainly be providing reactive suggestions, while higher scores would be making more proactive endorsements.

Fig. 1: An example of a segmentation that can be performed using Latent Class Analysis. This allows us to tease out the different kinds of people there may be based in the variables we are assessing.



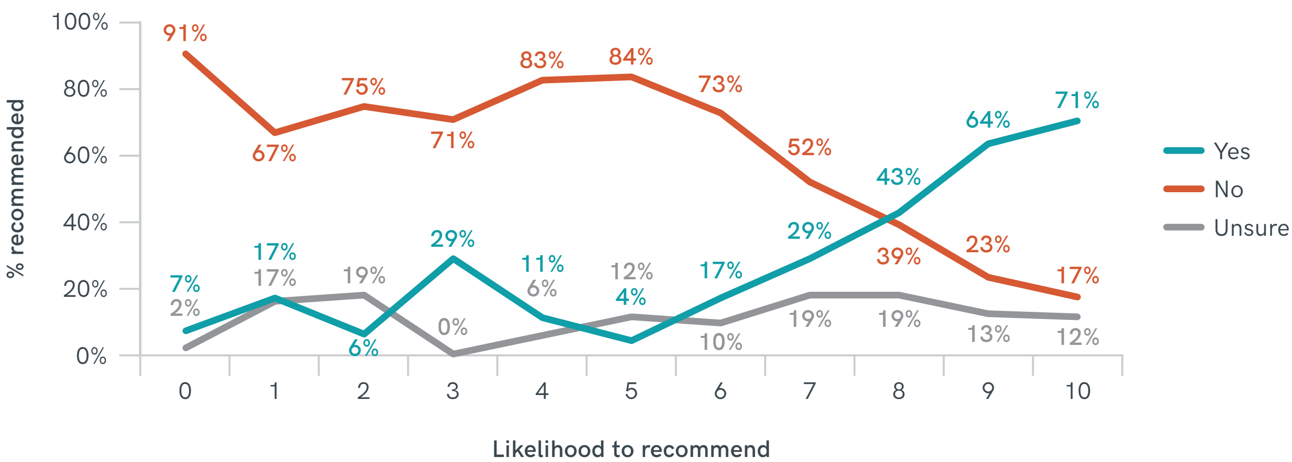
Results

The current study aimed to determine three key things: (1) how traditional NPS segments related to behaviour; (2) whether Promoters, Passives and Detractors behaved as expected in terms of motivation; (3) whether an empirically derived exploratory analysis could deliver a systematic segmentation.

NPS and Behaviour – Passives also recommend!

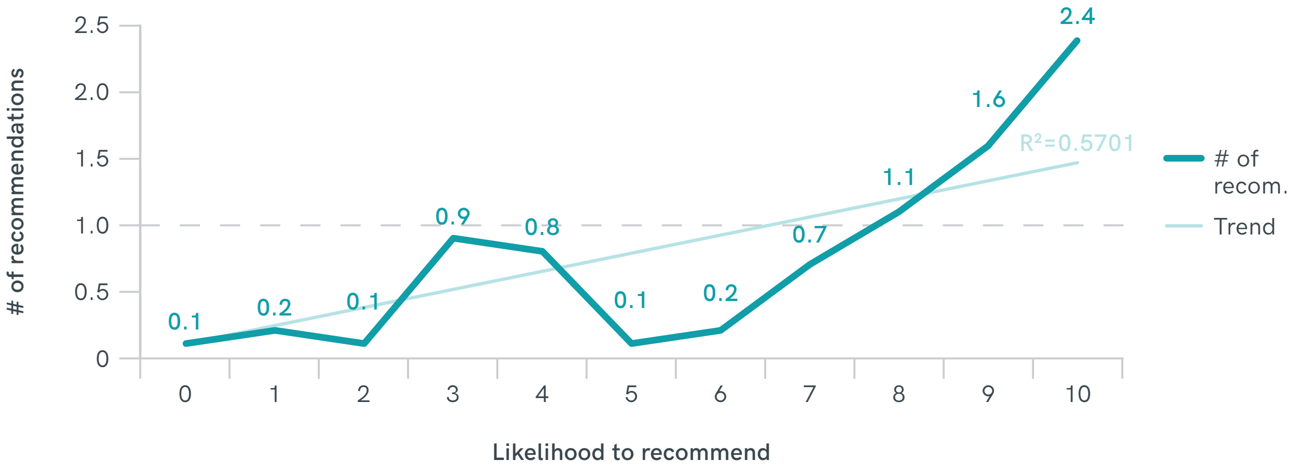
Traditionally, NPS claims that Promoters make recommendations, while anticipating that there would be few (if any) endorsements from Passives and Detractors (Reichheld, 2003). As expected, higher scores on the LTR scale displayed larger proportions of recommenders, as well as increasing in average number of recommendations. However, as seen in Figure 2, those rating an 8 on the LTR scale – a traditional Passive – were more likely (43%) than not (39%) to have actually made a recommendation in the past 12 months. Even 1 in 10 Detractors had contradicted their title. This aligns with numerous prior findings, in that participants typically struggle to link intent with actual behaviour (Chandon et al., 2005; Pingitore et al., 2007; Poon, Koehler & Buehler, 2014).

Fig. 2: Proportion of recommenders by LTR score



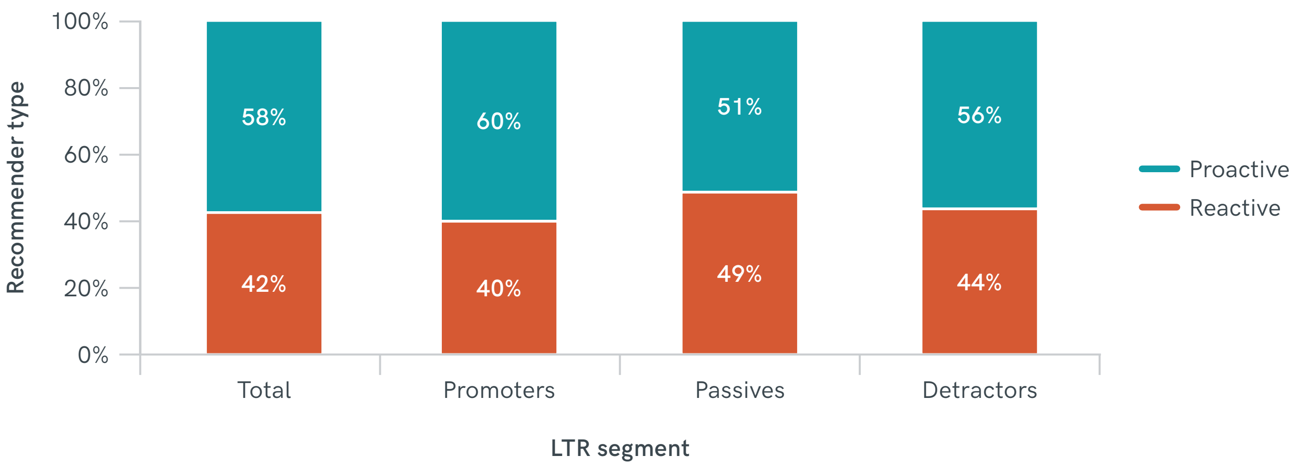
Similar patterns are displayed in average number of recommendations, where crossover to 1 or more average recommendations occurs at a score of 8, not 9 as expected from traditional NPS segments (Figure 3). Considering these rates, Passives cannot really be considered submissive at all, as many appear to be making an effort to speak highly of the brand. Furthermore, the linear trend data shortens this crossover closer to 6, which is typically deemed a Detractor.

Fig. 3: Average number of recommendations by LTR score, including a linear trend



Considering both Promoters and Passives comprise largely of recommenders, one must ask what truly differentiates these two segments. **While Passives do see fewer endorsers than Promoters, type of recommender is key distinguisher that is not commonly assessed in customer experience metrics.** Though Passives and Detractors display a relatively even spread of reactive compared to proactive recommendations, Promoters maintain significantly more proactive endorsements (Figure 4). **This may suggest that one of the major differentiators between Promoters and Passives is not *whether* recommendations are being made, but rather what *kind* of acclamations are being given.** Of course, those who are exceptionally satisfied with their brand are more likely to want to express such proclamations, although traditional NPS does not account for the context of recommendations.

Fig. 4: Proportions of proactive and reactive recommendations per NPS segment, amongst those who had made at least one recommendation in the past 12 months

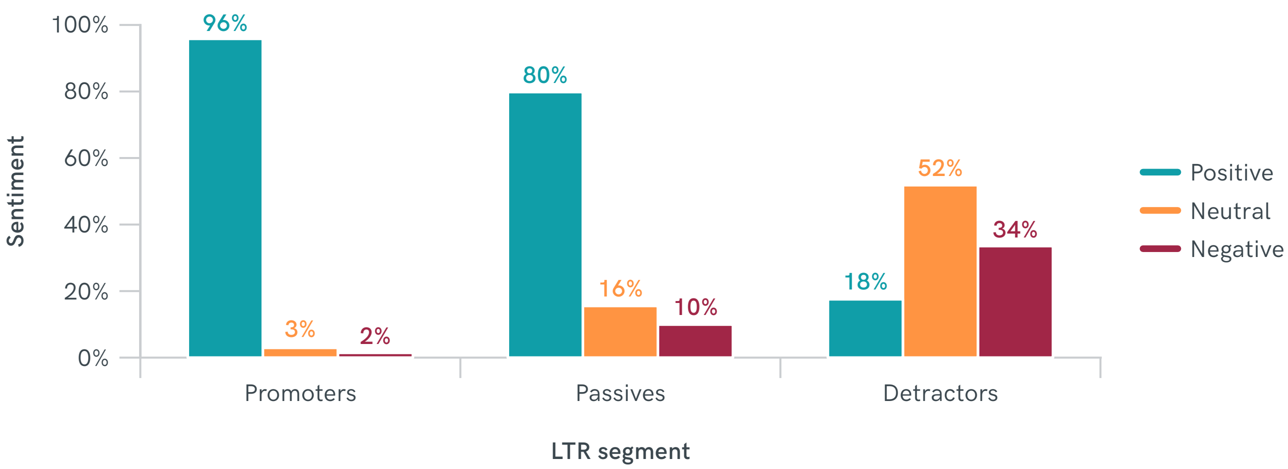


NPS and Sentiment – Passives actually much more likely to be positive than neutral!

The results from the sentiment analysis mirror those discovered in the addition of the behavioural metric, as it demonstrates that Passives are not truly passive, and are actually largely favourable of their chosen financial institution. Having 80% of Passives citing a positive reason for their score (Figure 5) is somewhat unsurprising considering 29% of 7s and 43% of 8s had made at least one recommendation in the last year alone (Figure 2) So again, considering their overwhelming positivity only adds to the point that Passives are not as disinterested as Reichheld (2003) proposed.

Even more surprising than Passives' positivity is the impartiality demonstrated by Detractors, where over half (52%) of all motivations were neutral in nature. While neutral mentions appear to peak at 5 along the LTR scale, it is also most prominent at the 0 mark, with negative sentiment overtaking in scores between 1 and 4. However, there were so few respondents in the 'Middle Detractor' range that they only represented a very small sector of Detractors. In total, only just over a third (34%) of those scoring between 0 and 6 for LTR attributed their score to something negative. Although this proportion is larger than what is seen for Promoters (2%) and Passives (10%), one cannot claim that the entire segment is detracting considering only a third said something negative. With few truly unhappy, Morgan and Rego (2006) recommend that complainers should be the core focus for improvement, as appeasing their doubts and retaining them as customers is crucial. This is particularly interesting in terms of the present findings, with the majority across the scale positively appraising the brand, and many Detractors simply neutral.

Fig. 5: Proportions of positive/neutral/negative sentiment by LTR segment



NPS and Segmentation – there are better ways to segment!

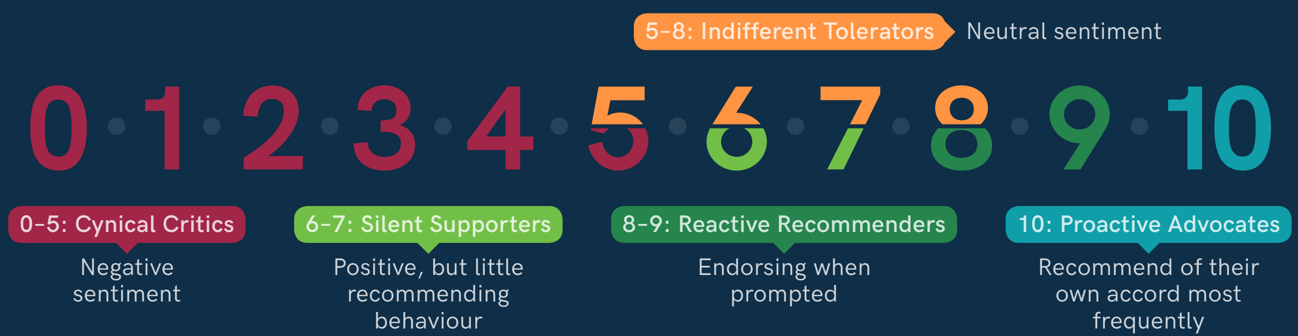
The purpose of market research as an industry is to base decisions and develop findings and conclusions based entirely on empirically derived data, free from subjective opinion, attitudes and bias. This begs the question as to why the current standard of the industry so wholly relies on a "one size fits all" metric, with segments that were not developed through any scientific or exploratory justification. In order to better the commonly utilised arbitrary segmentation that fails to appropriately categorise respondents as shown above, the current research harnessed the assessed variables to derive multiple new solutions.

The present research utilised a number of iterations for this segmentation, with each comprising of different variables. The first was the simplest, with just two variables: LTR and Satisfaction. The binary behavioural metric assessing whether or not respondents had made a recommendation was next, followed by an iteration that included sentiment. The binary behavioural metric was replaced by the number of recommendations given in the last 12 months, with the final analysis substituting this for the number per type of recommendation.

Overall, each of these analyses resulted in a fairly similar segmentation, with some key factors remaining the same throughout. First, these clusters typically resulted in five segments, though the second solution (which included only Satisfaction, LTR and the binary P12M scores) shortened this to only four groups. It should be noted that this then went back up to five segments once sentiment was included. The other defining feature throughout all of these segmentations was that those who scored a 10 on the LTR scale typically remained in their own cluster.

Culminating the information gathered from the above exploratory analyses results in the segmentation illustrated in Figure 6. The first segment, termed 'Proactive Advocates', generally consist of the highest score in both LTR and Satisfaction. Almost entirely positive sentiment is also a feature, though possibly the most interesting and definitive aspect of Proactive Advocates is the considerable share of proactive recommenders (59%). Though the second segment, the Reactive Recommenders also contain almost entirely positive sentiment and matches the Proactive Advocates on reactive recommendations (39%), the former's proactive proportion is substantially lower than the latter (59% vs. 45%). Silent Supporters, mostly consisting of LTR scores of 6-8, are mostly favourable in sentiment though offer fewer actual recommendations in comparison to the first two segments. Next, the Indifferent Tolerators are distinct in their neutrality, as LTR and Satisfaction scores mostly overlap with the Silent Supporters. The final segment consists of those providing negative sentiment, the Cynical Critics, with general LTR scores of 5 and below.

Fig. 5: Outcome of a Latent Class Analysis using all variables of interest and optimal population distribution. Although segments overlap in terms of LTR ratings, each cluster maintains their own defining features.



5 New Ways To Revitalise Your NPS Program

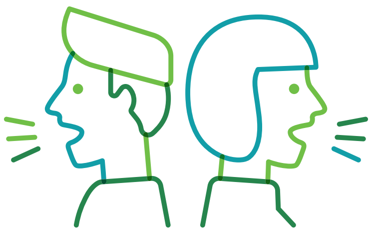
So, what are the next steps? In light of the research, and the resulting D&M Research NPS Revitalisation Program, we are helping clients get more from the NPS data they may already have.

1. Get under the hood

Use the whole scale: why do we reduce an 11-point scale to just 3 points when there is so much to gain from using the whole scale? This approach results in a lot of lost information, and relies on assumptions that bias already available data. Looking at the same variables in a slightly different way could therefore provide a much better understanding of how the metric is working, and has the added benefit of being specific to the industry or brand being assessed. Not only can this provide more actionable information, these kinds of analyses don't rely on the "one size fits all" approach harnessed by NPS.

There are multiple approaches that can be used to really "get under the hood" of the NPS, many of which can be applied on already existing data (e.g. adding a mean, median and mode). However, including a behavioural measure, as shown in the above research, provides an invaluable insight into where the true recommenders lie along the metric.

For example, many brands spend a lot of time and energy (not to mention cost) attempting to shift 8-scorers into the Promoter range. However, this may not be the most effective means of improving actual recommending behaviour if those scoring an 8 are already speaking positively and persuading others to try the brand. In this way, the extended behavioural knowledge acquired through a the addition of one simple metric could reduce time, cost and energy spent trying to shift customer dynamic in the brand's favour.



2. Get sentimental

As with behaviour, adding sentiment tags to the reasons for given LTR ratings reveals whether the Promoters, Passives and Detractors really are promoting, passive or detracting. The majority of NPS programs already collect this verbatim data (otherwise known as "reason for score"), although it is rarely utilised to its full capacity. Even coding a subset of the verbatim data could prove instrumental for tracking, as it could provide some insight into why a brand's NPS may be shifting.

Furthermore, comparing the common themes and sentiment that are present throughout each score across the LTR measure, as demonstrated by the behaviour test above, may uncover the true motivation behind each given score. For example, many of those who scored an 8 in much of our client research, across a number of industries, display positive motivation for doing so. This indicates that when respondents are giving a score of 8 out of 10, they likely think of this as a pretty good score. Therefore, relying on what customers are saying, rather than assumptions, could prevent the misinterpretation of consumer data that is so problematic in the traditional NPS.

3. Get in the driver's seat

In our extensive experience with tracking data, there is one question that clients ask us more than any other when it comes to NPS – why has the number moved? While sentiment analysis may tell us what respondents are saying, we need to uncover what factors are most important to customers in relation to their NPS. Being able to understand the importance of particular variables or 'drivers' in predicting an outcome (in this case, NPS) is critical in terms of continual monitoring as well as strategy development. It identifies key areas of impact, and allows a brand to focus on the elements that really matter, and as a result can be instrumental in determining which levers to pull to improve a brand's NPS.

1. Reasons for NPS

Review open enders, code and tabulate

Can use historical data

2. Run regression model

Test reasons to see which impact NPS using regression

Clues to the right levers will be revealed

3. "Factor" analysis

Review themes and draft statements to measure underlying constructs

Re-test using scales for each

4. Re-run regression

Using scaled data, re-run regression to understand which are the most important levers

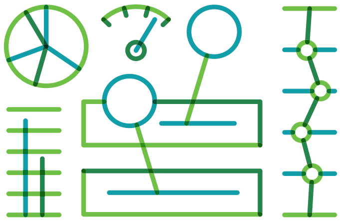
5. Monitor levers

Incorporate levers in NPS instrument

Measure performance and develop strategy to address areas for impact

4. Go to the top of the class

In some cases, particularly for ratings given in relation to a specific transaction, a full text or driver analysis may not be required. Coding all scores in detail may require an exceptional investment in transactional NPS, as these are often collected immediately after customer contact, thus resulting in a large database. Rather than coding all responses, analysing just those who gave a LTR score of 10 could still provide a wealth of information into what customers enjoyed about the transaction. Focusing on these high scores could therefore help develop a set of service standards to ensure an exceptional experience is achieved on every occasion. Not only will this improve transactional NPS ratings, but it will also heighten overall customer satisfaction in a broader sense.



5. Get real

Many of our clients already collect a wealth of information regarding average customer spend, rate of return or renewal, and frequency of contact. These factors paint a broader picture of the customer experience, and are real outcomes that don't rely on self-report or estimation. Furthermore, many of our projects also collect a suite of Key Performance Indicators (KPIs) such as Value for Money, Satisfaction, and Expectation. Combining all of these multiple sources of information can give us a better idea of what is most impactful on NPS in terms of measurable, observable and objective outcomes.

1. Test other metrics

Include a range of other metrics in your program like Satisfaction, Expectation, Future Propensity, Value for Price etc.

2. Include a behavioural metric

Include a metric which reflects real behaviour such as Spend or Frequency to Buy etc.

3. Regression analysis

Run a regression including NPS to see which metrics impact actual behaviour

4. Refine metrics

Incorporate only those metrics which impact desired outcomes into your Program

5. Monitor metrics

Monitor these metrics as part of your Program implementing programs to improve these

The Takeaway Message

Working in the Market Research industry over the years has enabled the team at D&M Research to experience a broad range of insights and learnings. However, the one that epitomises them all is expressed most poignantly through the phrase “knowledge is power”.

Means for managing our clients’ data to optimise their outcomes is forever evolving alongside customer change, making the need to reassess and improve upon traditional methods of analysis imperative.

Looking at the numbers from a slightly different perspective can maximise insights, even when considering already existing data. Slight tweaks to current practices in how the information is collection can further these benefits exponentially, particularly when focusing on measures of actual behaviour rather than intended behaviour.

Finding ways to better utilise the knowledge that has been gained through market research will inevitably result in optimally favourable business decisions, extending far beyond just improving NPS.

References

Chandon, P., Morwitz, V. G., & Reinartz, W. J. (2005). Do intentions really predict behavior? Self-generated validity effects in survey research. *Journal of Marketing*, 69(2), 1-14.

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), 327–346.

Grisaffe, D. B. (2007). Questions about the ultimate question: conceptual considerations in evaluating Reichheld's net promoter score (NPS). *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 20*, 36.

Keiningham, T. L., Aksoy, L., Cooil, B., Andreassen, T. W., & Williams, L. (2008). A holistic examination of Net Promoter. *Journal of Database Marketing & Customer Strategy Management*, 15(2), 79–90.

Kristensen, K., & Eskildsen, J. (2011). Is the net promoter score a reliable performance measure?. In Quality and Reliability (ICQR), 2011 IEEE International Conference on (pp. 249–253).

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.

Pingitore, G., Morgan, N. A., Rego, L. L., Gigliotti, A., & Meyers, J. (2007). The Single-Question Trap. *Marketing Research*, 19(2).

Pollack, B. L., & Alexandrov, A. (2013). Nomological validity of the Net Promoter Index question. *Journal of Services Marketing*, 27(2), 118–129.

Poon, C. S., Koehler, D. J., & Buehler, R. (2014). On the psychology of self-prediction: Consideration of situational barriers to intended actions. *Judgment and Decision Making*.

Reichheld, F. (2003). The One Number You Need. In: *Harvard Business Review*, 81(12), 46–54.

Reichheld, F. (2006). The ultimate question. *Harvard Business School Press, Boston, MA*.

Schneider, D., Berent, M., Thomas, R., & Krosnick, J. (2008). Measuring customer satisfaction and loyalty: improving the 'Net-Promoter' Score. In Poster presented at the Annual Meeting of the American Association for Public Opinion Research, New Orleans, Louisiana.