



Keeping Score

**Is the net promoter score a reliable
tool in a global recessionary
environment?**

**By Michael Feehan, Cristina Ilangakoon
and Penny Mesure**

The annual brand metric figures are coming out, and the brand team is desperate to see how its product stacks up from last year in terms of loyalty and customer retention. In particular, interest is high in the net promoter score (NPS) values year

over year—the firm having identified this as its key performance indicator. In the past heyday of strong company and brand growth, NPS improvement was used to benchmark annual bonuses. Now with the company share price tumbling due to the global recession, team leaders need the reassurance of solid NPS numbers to simply keep their jobs.

Everyone waits with bated breath for the “money slide” in the presentation ... and collectively gasp as they see a sudden and dramatic downward NPS drop from last year! Should the brand team be nervous? Well, only if the apparent NPS change is real. What if the change is simply an artifact of the measurement technique and any inherent unreliability in its use over time? In an up market when everyone is benefiting, perhaps positive NPS change is felt to be solid and viewed with a less critical eye.

Executive Summary

The net promoter score (NPS) uses a single indicator of customer loyalty to measure a brand's performance. To understand the stability and potential volatility of NPS in these recessionary times, we included an NPS measure in a recent survey evaluating the drivers of loyalty of patients to their primary care physicians. Our study demonstrates that NPS is "substantially reliable" over time, but there are inherent biases in repeat measurement of the same individuals.

However, in tough times, under increased pressure from senior management the first (defensive) action may be to suspect a problem with the score measurement itself. A second hypothesis may well be that the downward shift is not due to a meaningful change in perception of the company and its brands, but rather that the dip is an artifact of a general disinclination to recommend products and services when the economy is retracting. In this environment, consumers' moods may be adversely affected.

An entire consulting industry has grown around promoting the NPS and certifying companies in its use, all around the use of a single measure of customer loyalty to measure a brand's performance. Since the publication of Reichheld's "The Ultimate Question" (Harvard Business School Press, 2003), which promotes the measure as a "... foolproof test," many companies have adopted NPS as their core performance indicator. In marketing its adoption, Reichheld quotes Jeff Immelt, CEO of General Electric, after seeing the 2005 healthcare division scores: "This is the best customer metric I've seen. I can't understand why any of you wouldn't want to try it!" Reichheld goes on to say that Joe Hogan, CEO of the healthcare division, "made NPS a major component in the bonus formula for each of his direct reports." Simply put, NPS highlights the proportion of promoters of the brand relative to the detractors of the brand, in terms of their response to a single question on a 0-10 point scale: "How likely are you to recommend this company/product to a friend or colleague?"

Clearly the interpretation (and utility) of this NPS number and how it changes over time is predicated on a single premise: that the measure itself is fundamentally reliable and any change over time is due to change in the company or brand performance, not any unreliability of the measure. In any field where something is measured, one expects the "ruler" itself to not change over time. Where would we be if a tire gauge, carpenter's rule or blood pressure cuff gave different measures from day to day or week to week? If we expect precision and stability over time from these functional measures, we should expect the same of a measure of a brand performance that will be used to measure change in that brand over time.

To understand the stability and potential volatility of NPS in these recessionary times, we included an NPS measure in a recent survey evaluating the drivers of loyalty of patients to their primary care physicians. To measure the reliability of the measures in the survey, including NPS, the survey was re-administered to a subsample over a two-week period and standard indices of reliability were calculated.

The executives reviewing their downward-tending NPS may also suspect the change in the NPS score is a result of the depressed economy rather than a change in their product's actual performance. After all, when everyone is tightening their belts and feeling nervous and stressed about their own and the country's financial future, they may be less likely to recommend products (particularly expensive ones) to their friends.

To address the broader question as to whether consumers' likelihood to recommend and resultant NPS may be volatile and conditional on their personal mood (which might be expected to be more negatively tuned in tough times), additional measures of personal mood and the mood of the economy were included. The hypothesis here is simply that those individuals who are more depressed or feeling more pessimism about the economy will be less likely to recommend products and services and will thus artificially inflate the proportion of detractors and neutrals in any NPS calculation.

Approach

In May 2009, Observant LLC conducted an online survey of 394 respondents representative of the general population to determine the drivers of loyalty to their primary care physicians, leveraging three different arms to capture alternate methods of measuring confidence in their ratings of physician performance on a breadth of product attributes. Substantive findings on those methods will be reported in a subsequent analysis.

Here we focus on the NPS measure included. Each respondent was asked "How likely would you be to recommend your primary care doctor to a family member or friend who was looking for a new doctor?" and gave their response on a 11-point scale from 0 ("not at all likely") to 10 ("extremely likely"). The question was asked very early in the interview to avoid question-order bias. Respondents were then classified as being "promoters" (P: 9 and 10), "neutrals" (N: 7 and 8) and "detractors" (D: 0 through 6). NPS was calculated as per the standard approach advocated by Reichheld, $NPS = P - D$. Open-ended responses were gathered to "explain why you gave this rating." Respondents also indicated the degree to which they felt they would continue seeing their same PCP in the future.

Further measures were included later in the survey to gauge respondents' current mental health (on a 0-10 rating scale) and a composite score of symptoms of psychological distress experienced in their worst month within the past year. Their level of consumer confidence was also gauged—how optimistic they felt that their own "personal financial situation will get better over the next 12 months" and that the "outlook for the U.S. economy and business is going to get better over the



While many individuals give the same rating over this short interval, there are some who change their rating meaningfully.

next 12 months” (11-point agreement scales from “strongly disagree” to “strongly agree”).

To gauge the reliability of key survey questions, the entire survey was re-administered to a subset of 202 respondents (51.3 percent) two weeks later. All completed the second survey between 13 and 15 days later.

Key Findings

Test-retest reliability. In the subsample of 202 respondents who completed the retest survey, the correlation between the measure of likelihood to recommend over time was $r = 0.82$ ($p < 0.001$). The r^2 indicates that only 67 percent of the variance in the retest score can be accounted for by the original score. Clearly, while many individuals give the same rating over this short interval, there are some who change their rating meaningfully.

So how does this translate into the NPS? At test, 65.8 percent were promoters, and 12.4 percent detractors, giving an NPS of 53.4. At retest, 64.4 percent were promoters, and 12.4 percent detractors, giving an NPS of 52.0. In this comparison, 11.3 percent of promoters dropped to neutral, and 1.5 percent dropped all the way to detractor (12.8 percent ceased being promoters). On the other hand, this loss was over-compensated by change in the initial detractors, of whom 16 percent became neutral, and 4 percent became promoters (20.0 percent ceased being detractors).

The test-retest reliability of the derived NPS score was determined using a Spearman rank-order correlation, which indicates the degree of association between each respondent’s categorization (1, detractor; 2, neutral, and 3, promoter) at initial test and at retest ($r_s = .72$; $p < 0.001$). Using the categorical approach, half the variance in the retest NPS is accounted for by the NPS at time one ($r_s^2 = 0.52$). Despite the observation that there is substantial movement across categories within a two-week retest period, the resultant weighted Cohen kappa statistic of .70 indicates the measure is “substantially reliable” according to standard criteria. (See Landis, J.R. and G.G. Koch (1977), “The Measurement of Observer

Agreement for Categorical Data,” *Biometrics*, 33, 159-174, and Munoz, S.R. and S.I. Bangdiwala (1997), “Interpretation of Kappa and B Statistics Measures of Agreement,” *Journal of Applied Statistics*, 24(1) 105-111.)

This level of reliability is commensurate with other often used consumer attitude scales such as the “attitudes to the social role of corporations” with retest reliability for its subscales ranging from .40 to .79 and the “consumer alienation from the marketplace” scale with 3-week retest reliability of .75.

In this case, the two NPS scores appear relatively close (53.4 and 52.0); however, initial detractors tend to become neutral or promoters on retest. Moreover, when the test-retest sample ($n=202$) is compared with those who declined the retest ($n= 192$), those who declined the retest are less likely to recommend their doctor ($t = 2.45$, $p < 0.05$). Hence the

Exhibit 1

Tests of association with NPS

	NPS promoters	NPS detractors	Mean difference	t value	p value
Physical health rating	6.37	5.88	.49	1.65	n.s.
Mental health rating	7.33	6.8	.53	1.71	n.s.
Mental distress score	15.83	15.41	.42	.58	n.s.
Personal economic outlook	5.81	5.53	.28	.72	n.s.
U.S. economic outlook	5.26	4.38	.88	2.72	0.01

slight downward shift in total NPS in the retest is likely to be an underestimate of the negative true change if all potential respondents were retested. The net bias effect then of repeated measure of NPS is positive. Thus, when NPS is tracked by a company across the same individuals over time, any positive change seen in NPS is likely an overestimate, and, critically, any downward change in NPS is likely to be an underestimate of the true negative change.

Individual and Market Pessimism

Across all measures in Exhibit 1, detractors tended to have slightly lower scores than promoters. However, our initial hypothesis that those respondents with poorer mental health may be less likely to recommend was refuted. There were no significant differences found between promoters and detractors in ratings of their own mental health and mental distress.

In terms of pessimism about individual fortunes, again there was no association with the NPS groups. Promoters and detractors were equally likely to be optimistic about their immediate financial future. In terms of the U.S. economic outlook, there was a trend for significance here ($p=.10$) with detractors scoring almost a full point lower on the 11-point scale. In this study of the likelihood to recommend a physician, it may be that macro-economic pessimism may not play a role in influencing NPS as much as it might for high-cost consumer goods. Companies might wish to consider including measures of market optimism in their NPS surveys and analyses in order to tease out whether change in brand perception is truly due to the brand and not an artifact of market mood.

The Final Score

Any consideration of the strengths and weaknesses of NPS can stimulate intense argument and rhetoric, particularly when considering the validity of the measure in predicting company growth and profits. Reichheld takes an ideological stance in defending the use of NPS, casting as an “army of enemies” those that would critique the use of “a system of accountability that is based on a single number.” In his book, he cited a market research group’s opposition to the company’s plan to adopt NPS because it was “... based on flawed research, flagrantly illogical, statistically invalid, irresponsible and fundamentally flawed.” However, no exposition of the group’s rationale was given nor a rebuttal given.

Recent research has severely critiqued NPS in terms of its predictive validity in relation to company growth. Keiningham and colleagues conducted a longitudinal study of company performance and NPS measures across industries and were unable to replicate Reichheld’s 2003 findings. (See Keiningham, T.L., B. Cooil, T.W. Andreassen and L. Aksoy (2007), “A Longitudinal Examination of Net Promoter and Firm Revenue Growth,” *Journal of Marketing*, 71, 39-51.)

They concluded that “we find no support for the claim that net promoter is the single most reliable indicator of a company’s ability to grow.” They quote Reichheld as reporting on the net promoter Web site in 2006 that “Frankly we see

little value in continued debate about cause versus correlation, timeframes or statistical methods.” They go on to note that “unfortunately the statistics matter and they reinforce the point that executives adopting NPS do so with the belief that there is sound science underpinning the measure and its utility.” In the Winter 2008 issue of *Marketing Research*, Sharp goes so far as to describe NPS as “... snake oil, fake science” due to Reichheld’s selective reporting of cases that illustrate an association between NPS and company growth.

In this article we have taken the approach of looking at the NPS measure’s reliability as opposed to validity. In order for any measure of anything to be valid, it first must be reliable. Our study demonstrates that NPS is “substantially reliable” over time, but there are inherent biases in repeat measurement of the same individuals. In tracking NPS where there is a possibility of retaining the same consumers in the sample design, companies need to be cognizant of the positive bias in NPS generated by repeat measurement.

Some companies choose to establish longitudinal panels where consumers are expressly re-interviewed (to examine switching behavior and change in attitudes or beliefs etc. over time), whereas others expressly exclude any prior participants from repeat participation. Others may not address this issue in design—under the assumption that large samples would mitigate the likelihood and impact of any repeat participation. This may in fact fall down when company’s are tracking performance among highly select audiences where the population of interest may be relatively small (e.g., senior physicians, C-suite executives). This can be exacerbated when researchers are relying on the use of online panels. We would recommend that, in instances where prior participants are not expressly excluded, analyses of NPS statistically adjust for repeated measurement among any subset of repeat consumers.

Our data suggest that companies may be reasonably confident that NPS is robust in not being adversely affected by respondents’ individual moods. However, there is a suggestion that some variation in NPS in a down economy may be an artifact of a depressed market mood and that this should be parsed out when evaluating any individual brand’s performance. In our present study of the likelihood to recommend a primary care physician, cost-to-see physician differentials across respondents may not be large. Thus consumers may be relatively insensitive to market conditions. It would be interesting to see if this is true in other categories, where the likelihood to recommend higher cost items (e.g., automobiles, plasma TVs) may be more subject to variability in consumer confidence. ●

Michael Feehan is the CEO, Christina Ilangakoon is a senior associate and Penny Mesure a director at Observant LLC (www.observant.biz), a strategic market research partner to the pharmaceutical and healthcare technology industries. Correspondence regarding this article may be directed to mfeehan@observant.biz.