Choosing a Customer Experience Metric

Medallia
INSTITUTE

TABLE OF CONTENTS

AT A GLANCE

WHAT?

Choose the primary customer experience metric that your company will use to measure performance.

WHY?

Your primary metric anchors feedback design, internal reporting, goal setting, and employees' understanding of performance.

Your primary metric helps build a common focus

The primary customer experience metric you choose is one of the most visible elements of your program, focusing everyone from executives to the frontline on a common measure. A primary metric is defined by three elements:

1. The question asked. For example, do you ask your customers to rate their overall experience or report their expected future behaviors such as likelihood to recommend?
2. The answer options. You will need to select a rating scale (e.g., 0 to 10) and endpoint labels (e.g., "not at all satisfied" and "extremely satisfied") as part of your survey design.
3. The calculation. Will you take an average of all responses, calculate the percentage of high scores (e.g., Top-2 Box), or compare the percentage of high scores to the percentage of low scores as with Net Promoter®?

Common metrics used by Medallia customers include NPS® (the percentage of 9 and 10 ratings minus the percentage of 0-6 ratings on the likelihood to recommend question), Top-2 Box Satisfaction (the percentage of 9 and 10 ratings on the question, "How would you rate your overall satisfaction?"), mean scores for overall satisfaction, and mean Likelihood to Recommend (the average of the likelihood to recommend question).

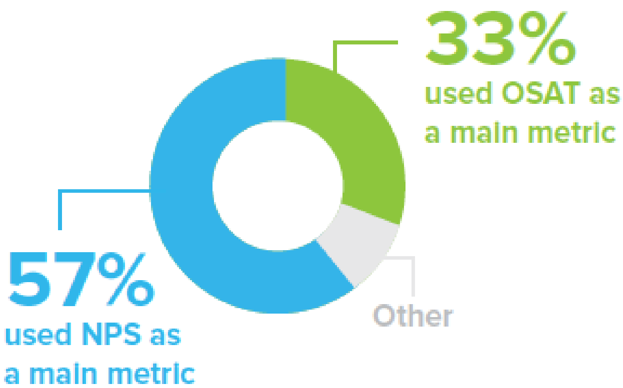
KEY QUESTIONS

How will we know if our scores are good or not?
Many companies prefer primary metrics that are widely adopted, such as NPS or Top-2 Box Satisfaction, because benchmark comparisons for these metrics are more readily available.

How do we assess differences in scores across units or over time?
Statistical tests can be used to determine whether differences in performance across units or time are significant and meaningful, or simply the result of random noise. It's important to note, however, that these tests are less reliable when sample sizes are small or response rates are low. Why? Because the distribution of responses from customers who choose to take the survey may be markedly different from the distribution of responses from those who choose not to take it. Also, with small numbers, scores tend to be more volatile.

How do cross-cultural differences affect scores?
Some companies report seeing systematic differences in scores across countries or regions. Many factors may contribute to these differences: Customers may have truly different experiences; attitudes and/or behavior may differ systematically across cultures; or response rates may vary. Because it's difficult to tease apart the reasons for these differences, many Medallia customers choose to track changes over time within a given country or region, rather than draw conclusions from comparisons across units in different countries.

In 2014, 57% of Medallia programs used NPS as a main metric (using the likelihood to recommend question) and 33% used overall satisfaction as a main metric. Of the programs that used overall satisfaction as their main loyalty question, 59% used a Top-x-box score and 33% used an average score.



What makes a primary metric "good"?

A good primary metric is reliable, strongly related to other outcomes you care about (e.g., financial performance, customer behavior), and relevant and meaningful for employees, motivating them to take action and improve.

Links to financial performance.

The metric you choose should have a demonstrated connection to the customer behaviors you want to drive (e.g., repeat purchase, retention, additional spend, word of mouth). If using a common metric such as NPS, you may be able to rely on published sources to make this link. For example, in a study looking across 8 different hospitality brands, Medallia found that for mid-tier brands, converting a passive into a strong promoter increased customer spending by 46% over the following year; and converting a strong detractor into a strong promoter produced spending increases of 175%[1].

Although considerable research has shown a link between customer experience and financial performance, it can be valuable to conduct company-specific analyses that demonstrate how your primary metric aligns with customer outcomes using your own data. Doing so will help your key stakeholders buy in to the customer experience program.

Survey questions about customers' future behavior such as "How likely is it that you would recommend us?" or "How likely are you to buy again?" are often the strongest predictors of future spending patterns. Asking about overall satisfaction is another common alternative that works well for both overall relationship measures and measures of transaction quality.

Differentiates excellence.

If your metric generates too many maximum ratings, distinguishing good performance from great is difficult. This is particularly problematic with short scales (e.g., 1-5). Longer scales such as 0-10 tend to result in fewer top-box (i.e., maximum) ratings, giving you greater ability to differentiate excellent performance.

The calculation you choose also matters. Calculating an average (e.g., 9.6 out of 10) often results in scores too close to the maximum. In this case, the best performers can have scores that appear very similar to those of average performers. A Top-2 Box or net score calculation will magnify differences in performance and will be more sensitive to changes in customer ratings. In addition, when calculating NPS, a 0-10 point rating scale is recoded to a scale that ranges from -100 (all detractors) to +100 (all promoters). This introduces more variance in scores. All else being equal, greater variation usually leads to greater explanatory power in predicting or explaining outcomes.

Easy to understand.

The primary metric you use should be easy for employees to understand and take action on. Net Promoter Score has become popular because

[1] The Medallia Institute (2014). The Revenue Impact of a Great Customer Experience.

its language of Promoters (9 and 10), Passives (7 and 8), and Detractors (0-6) describes tangible customer outcomes that your employees can relate to. Also, percentages are often more tangible than averages; for example, “12% of my customers are extremely satisfied” is more concrete than “my customers are 7.43 satisfied.”

Another common question is whether to use a primary metric based on a single question or an index based on multiple survey questions and ratings. In most cases, Medallia recommends using a single question as the basis for your primary metric while providing clear visibility to underlying drivers that influence the primary metric. However, as we explain below, including multiple loyalty items can also provide benefits when looking for customer insight.

The table below outlines the pros and cons of using a single metric versus an index.

Which metrics do a better job revealing insights?

The short answer is: It depends. Specifically, it depends on the type of changes you’re interested in detecting and where your scores fall on the overall continuum of possible scores. Different metrics are more sensitive at different points in the distribution and are better equipped to detect different types of change.

The Medallia Institute conducted an analysis comparing the statistical properties of mean likelihood to recommend scores (LTR), NPS, and Top2Box scores on a 0-10 or 1-10 scale. Here are the most important factors to consider when choosing or comparing metrics:

Type of insight

The most important consideration is the type of insight most relevant to your business objectives and improvement efforts. Are you trying to reduce the number of bad customer experiences? Improve every customer experience? Or, increase the percentage of outstanding customer experiences relative to poor experiences? Each metric detects a different insight and each can be useful depending on your goals.

CHOICE	PROS	CONS
Single Metric	Simple and easy to rally the organization around Conducting key driver analysis on a single metric enables you to observe changes in key drivers over time	The overall metric may not feel actionable on its own.
Indexed Metric	Minimizes fluctuations: reduces noise and only showcases large changes	Less sensitive to change; can downplay meaningful differences More difficult for the frontline to interpret and take action on
Weighted Index	Attempts to embed both drivers and their priorities into one measure	Even more difficult to interpret and act on; changes in score can be based on changes in underlying attributes

NPS and Top2Box are good metrics for detecting the ratio of outstanding experiences to poor experiences. These metrics will be the best measures for estimating how well you've increased your percentage of highly satisfied customers relative to others. This can be important because customers who are promoters are usually more loyal and more likely to generate positive word of mouth about your products and services.

Yet, while NPS and Top2Box are good at detecting an increase in outstanding experiences relative to others, they are less sensitive to incremental improvements across the entire distribution of customer experience. For example, Top2Box will not detect a shift from detractors to passives, which may be important to assessing the progress of specific improvement initiatives and their effectiveness over time.

Starting points matter

In general, when there is large room for improvement, a Top2Box calculation or net percentage such as NPS is often the most statistically sensitive to change and best highlight performance differences that yield actionable insights. While averages are more stable, they may downplay meaningful differences and make it more difficult to identify strengths or opportunities for improvement.

However, the difference between metrics depends on where you start. If your current performance is relatively low (e.g. mean LTR < 8.0, Top2Box < 55%, NPS < 30), then Top-2Box and NPS will likely still be better at detecting significant diferences, especially when those diferences are small (see Figure 1A).

For very high scores (e.g., NPS >70, Top2Box >80, mean LTR>9.0), then all three metrics perform similarly well in terms of statistical power.

Best practices for implementing metrics

Any metric can be challenging to implement and use well. Regardless of the metric you choose, Medallia recommends the following best practices.[2]

Identify performance drivers specific to your business.

Measuring a single overall metric to understand the customer experience is rarely enough to enable learning and appropriate action. It's important to also measure the service components or product attributes that drive your primary metric. These

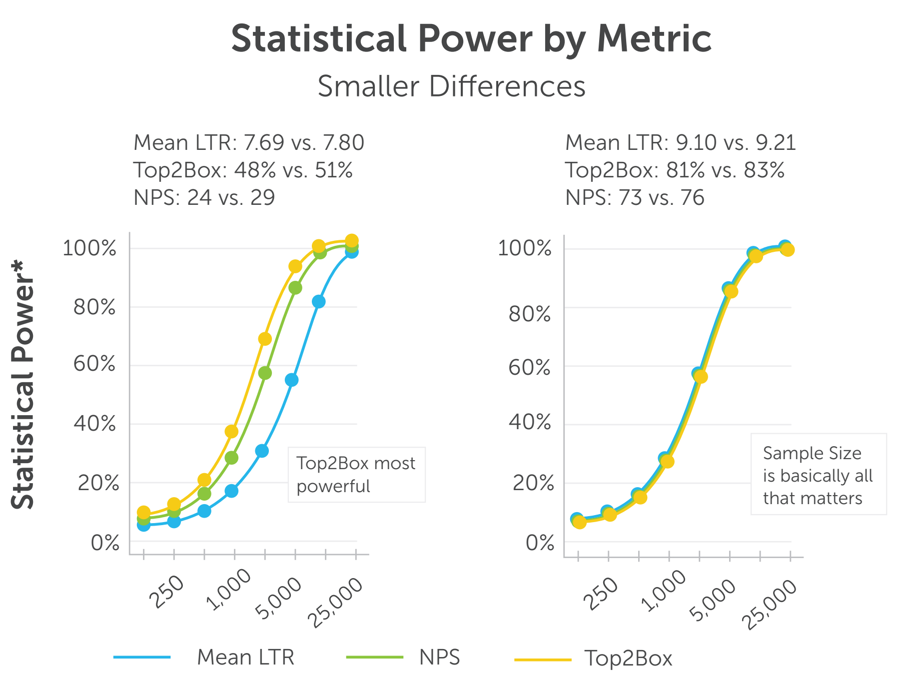


Figure 1: Each data point based on 1000 random samples from the given dataset. *p<.05

[2] Some adapted from David Whitlark and Gary K Rhoads (2011). "Scoring Success." Marketing Research, Spring: 8-13.

drivers are usually unique to your business and should be crafted as such. They are essential to getting to the core of your customers' needs and identifying improvement opportunities that can enhance business results.

Include open-ended questions.

To keep survey lengths manageable while enabling actionable insights, it is important to include open-ended questions to follow up when customers rate their experience poorly. With current text analytics, responses to open-ended questions can surface important trends and sentiments that might otherwise go unnoticed with a short survey and limited set of questions. Verbatim responses will often give managers the will to make changes and the actionable detail needed to make the right ones.

Make sample size king.

Regardless of the metric you choose, it's important to remember that a large representative sample is the key to drawing accurate conclusions. When sample sizes are small, sensitive metrics like NPS may appear volatile. However, with enough responses, any metric will be able to detect meaningful differences (if they exist).

There are several best practices that can help you maximize your sample size. First, you can optimize the design of your survey to make it easier for customers to respond. Minimizing survey length and streamlining the subject line are two common approaches. Second, maintaining a "healthy" email address book can make it possible to reach out to a large number of customers. And, third, using smart quarantine rules will ensure that you interact with customers at the right frequency, thereby preventing survey fatigue. Each of these practices will help you to increase the size and representativeness of your sample.

Strive for Excellence Across All Service Dimensions.

Assuming you're measuring the correct performance drivers specific to your business, it's important to scrutinize any service dimensions or product attributes that are not achieving top scores. For example, in a typical data set, NPS may be 80 among customers who rate customer support a 9 or a 10. However, for the same data set, NPS can drop to 10 among customers who rate the same dimension 8 or less and -15 for those that rate the dimension 7 or less.[3] Even when ratings for all performance drivers are consistently high, surprises may occur. Although customers may score all service dimensions quite high, they may inexplicably score their likelihood to recommend relatively low. New predictive modeling techniques and text analytics make it possible to turn these apparent aberrations into important insights, identifying hidden problems that are important to address.

> "...a large representative sample is the key to drawing accurate conclusions."

Include multiple CX items for additional insight.

In many cases, companies find it valuable to look at multiple loyalty or satisfaction items. Measuring multiple CX indicators can provide additional insight even for companies that focus on one primary metric like NPS. For example, research comparing the predictive performance of NPS and other commonly used loyalty metrics found that a single item measuring agreement with the statement "The offerings are the best in the industry" performed on par with NPS in predicting business unit revenue. The point here is not that this single-item consistently predicts performance better than others; rather, it simply tells us that when this item predicts performance better than other metrics it provides unique insight that may have implications

[3] Ibid

for messaging and marketing promotions. In this case, the fact that the “best in the industry” item had the largest predictive impact on revenue highlights the competitive nature of the industry and may suggest more aggressive promotions are in order. Similarly, if the item “I have very little interest in the products of competitors” was the best predictor, it might indicate pricing opportunities.

Consider distribution patterns.

Individual metrics typically tell only part of the story. In many cases, it can be valuable to look at how scores are distributed. For example, you may want to see if customer experience is normally distributed or highly bifurcated across locations or products (see figure 2). Are customers reporting very positive and very negative experiences and nothing in between? Or are they reporting experiences that vary along the entire continuum of satisfaction? These patterns may have very different implications for action, so it can often be useful to look at standard deviations and the extent to which there are different response patterns at different score ranges across the overall continuum. These can provide important insights beyond that which can be learned from NPS, Top2Box, or mean scores alone.

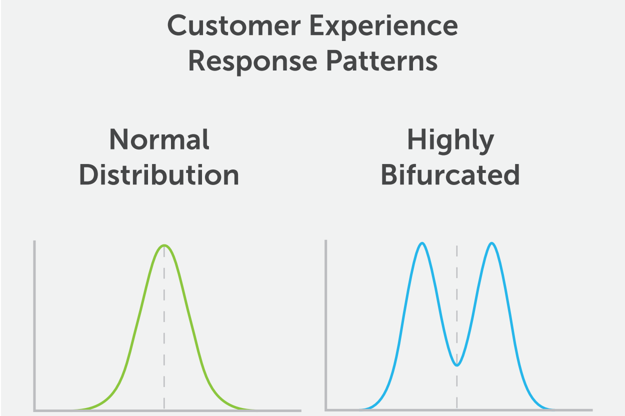


Figure 2: Two customer experience response patterns with the same mean but very different distributions.

About Medallia

Medallia is the pioneer and market leader in Experience Management. Medallia’s award-winning SaaS platform, the Medallia Experience Cloud, leads the market in the understanding and management of experience for customers, employees and citizens. Medallia captures experience signals created on daily journeys in person, digital and IoT interactions and applies proprietary AI technology to reveal personalized and predictive insights that can drive action with tremendous business results. Using Medallia Experience Cloud, customers can reduce churn, turn detractors into promoters and buyers and create in-the-moment cross-sell and up-sell opportunities, providing clear and potent returns on investment. www.medallia.com

Follow us: medallia blog.medallia.com @Medallia

Medallia
INSTITUTE