

Quantitative UX Glossary

Benchmark

(Not to be confused with benchmarking study)

A single number that you want to compare your site to. A benchmark could be a value that you are aiming for or a value obtained from a third-party study.

Benchmarking Study

A study intended to measure the user experience of a product over time. It usually involves looking at the same metrics over different iterations of the product. A benchmarking study might compare the most recent set of metrics (which represent your product's current performance) against the metric values collected in past studies. This comparison can help reveal whether and how much your product has improved over time.

Between-Subjects Design

A study design in which different participants are assigned to different conditions corresponding to a variable. For example, in a between-subjects study, you might recruit 80 participants for a quantitative usability study, and randomly assign each participant to either site A or site B. This type of study design is usually contrasted with within-subjects design.

Binary Metric

A metric that can have only two possible values. Examples include task success or conversion (that is, whether a certain desired event has happened — for example, whether a user has registered for an account).

Center of a Dataset

One number that summarizes the dataset. The average (i.e., arithmetic mean) is usually used for the center of a dataset. However, with skewed distributions like task times, the median or the geometric mean may be more appropriate.

Categorical Metric

A metric that can take only a limited, fixed number of values. For example, t-shirt size; (e.g., S, M, L) is a categorical variable.

Confidence Interval

Likely range for the true score of your entire population. In other words, the confidence interval is a range of values that is likely to contain the value that you're aiming to find. It is closely related to the concept of margin of error.

Confidence Level

How confident you can be in your confidence-interval calculation. This value can be chosen by the researcher. In UX, we usually use 95% or 90% confidence levels. Essentially, choosing a lower confidence level indicates that you're more tolerant of the risk that the confidence interval may not actually cover the true score of a metric, obtained across the whole user population.

Confounding Variable

A hidden variable that influences both the independent and dependent variable(s), causing a false relationship between the two.

Continuous Metric

A metric that can take any value between two possible values. In general, metrics that can be expressed with any number of decimals are continuous. Time on task is an example of a continuous metric.

Dependent Variable

A variable that is measured in a study and whose value is expected to vary based on the manipulation of the independent variable. For example, we may change our design (the independent variable) and look to see if that impacts how satisfied users are with the product (the dependent variable).

Discrete Metric

A metric that can take only a set of countable, prescribed values. A categorical metric is always discrete. Rating scales are discrete; other examples of discrete metrics include the number of user visits to a site and the number of errors made in a task.

External Validity

A quality of a study that ensures that the study setup and participants are naturalistic and reflect the real-world situation

Independent Variable

A variable that is manipulated by the researcher in a study. Then researchers look at the dependent variable(s) to see if the independent variable has had any impact.

Internal Validity

A quality of a study that ensures that the study setup does not favor any condition or participant response and that all conditions are treated in the same way.

Likert Scale

A 5-point rating scale in which respondents must indicate agreement to a statement. The points on a Likert scale are Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree.

Margin of Error

Half the width of a confidence interval for a given metric. If you have the margin of error and an observed score, you can compute the confidence interval. In other words, the margin of error and the confidence interval convey pretty much the same information and you will usually see reported one or the other.

Metric

A quantitative variable or indicator that is collected from a study. Typical UX metrics include task time, success, satisfaction, conversion, or ease of use. UX metrics capture some aspects of the user experience that can be observed and quantified.

NASA-TLX *(NASA Task Load Index)*

A post-task 6-question questionnaire that is meant to measure the perceived workload for a given task. It is used for studying complex tasks in high-consequence situations.

Normal Distribution

(Also known as Gaussian distribution or bell curve)

A type of distribution for continuous metrics that assumes that most values are relatively close to the center of the distribution and symmetrically distributed around it. Its graph is bell-shaped. Many of the metrics used in UX have a normal distribution.

Net-Promoter Score (NPS)

A score based on answers to a one-question survey that asks “How likely is it that you would recommend this site to a friend or colleague?” on a scale from 0 (not at all likely) to 10 (extremely likely). The NPS ranges from -100 to 100.

Observed Score

Value of a metric obtained from a sample of the user population. It is used as an estimate of the true score.

P-Value

The probability that a difference between two (or more) observed scores is due to chance. Statistical significance calculations produce p-values. The p-value must be less than 0.05 for the difference to be statistically significant.

Population *(or User Population)*

All the possible users of a product.

Performance Metric

A metric that captures how users interact with a particular product. Task time, number of errors, and task success are examples of performance metrics. Performance metrics are usually contrasted with self-reported metrics.

Post-Task Survey

A survey administered after each task that the user has performed in a quantitative study. Usually, post-task surveys need to be very short. The Single-Ease Question is a common post-task question.

Post-Test Survey

A survey administered at the end of the entire quantitative study. The System Usability Scale is a popular post-test survey.

Practical Significance

Whether a statistically significant difference obtained in a study is likely to have any practical implications in real life. For example, a study may find that a difference of a few seconds is statistically significant, but that might not make any difference to the users, and thus would not be practically significant.

Preference Metric *(See self-reported metric)*

Rating Scale

A closed-ended survey question that asks the respondent to assign a value to a specific concept. In UX, the most common rating scales are semantic-differential scales and Likert scales.

Sample *(or Sample of a Population)*

A subset of a user population. For example, you might recruit 40 participants (the sample) from your user base (the population).

Self-Reported Metric

(Also preference metric or subjective metric)

A metric that measures users' perception of the system and how they feel about it. Self-reported metrics are collected by asking users questions about the system, usually as part of surveys. They are contrasted with performance metrics.

Semantic-Differential Scale

A rating scale in which the two ends of the scale host antonym adjectives. Points in between the ends are usually not labeled. For example, you may ask your users to choose the option that best describes a site on a 7-point scale, where 1 is labeled "Unprofessional" and 7 is "Professional."

Single-Ease Question (SEQ)

A one-question survey instrument asked after the user has attempted a task. The question is "Overall, how difficult or easy did you find this task;" users must select an answer on a scale from 1 (very difficult) to 7 (very easy).

Skewed Distribution

A distribution that is asymmetrical relative to its center. Time-on-task data is often skewed. Skewed distributions are not normal.

Standard Deviation

A measure of the variability of a distribution that captures the average departure from the center of the distribution. The standard deviation is used when calculating confidence intervals.

Statistical Significance

Whether a perceived difference between two (or more) observed scores is due to chance. Usually, it is established through a statistical significance test, which returns a p-value.

Subjective Metric *(See self-reported metric)*

System Usability Scale (SUS)

A 10-item post-test questionnaire that measures the perceived usability of the system. SUS scores can range from 0 to 100.

True Score

The center of the distribution obtained by including all the users in a user population. This value can be known only if the user population is fairly small. In most cases, it is estimated by the observed score.

Variability

How different the data points in the data set are from each other and from the center of the distribution. Variability is usually measured by the standard deviation.

Within-Subjects Design

A study design in which the same participant tests all conditions corresponding to a variable. It is usually contrasted with between-subject designs. For example, you might recruit 40 participants for a quantitative usability study and have all 40 participants test both site A and site B. In this type of study, it's important to randomize the order in which users will see the designs.

Quantitative UX Glossary

Benchmark

(Not to be confused with benchmarking study)

A single number that you want to compare your site to. A benchmark could be a value that you are aiming for or a value obtained from a third-party study.

Benchmarking Study

A study intended to measure the user experience of a product over time. It usually involves looking at the same metrics over different iterations of the product. A benchmarking study might compare the most recent set of metrics (which represent your product's current performance) against the metric values collected in past studies. This comparison can help reveal whether and how much your product has improved over time.

Between-Subjects Design

A study design in which different participants are assigned to different conditions corresponding to a variable. For example, in a between-subjects study, you might recruit 80 participants for a quantitative usability study, and randomly assign each participant to either site A or site B. This type of study design is usually contrasted with within-subjects design.

Binary Metric

A metric that can have only two possible values. Examples include task success or conversion (that is, whether a certain desired event has happened — for example, whether a user has registered for an account).

Quantitative UX Glossary

Benchmark

(Not to be confused with benchmarking study)

A single number that you want to compare your site to. A benchmark could be a value that you are aiming for or a value obtained from a third-party study.

Benchmarking Study

A study intended to measure the user experience of a product over time. It usually involves looking at the same metrics over different iterations of the product. A benchmarking study might compare the most recent set of metrics (which represent your product's current performance) against the metric values collected in past studies. This comparison can help reveal whether and how much your product has improved over time.

Between-Subjects Design

A study design in which different participants are assigned to different conditions corresponding to a variable. For example, in a between-subjects study, you might recruit 80 participants for a quantitative usability study, and randomly assign each participant to either site A or site B. This type of study design is usually contrasted with within-subjects design.

Binary Metric

A metric that can have only two possible values. Examples include task success or conversion (that is, whether a certain desired event has happened — for example, whether a user has registered for an account).

Quantitative UX Glossary

Benchmark

(Not to be confused with benchmarking study)

A single number that you want to compare your site to. A benchmark could be a value that you are aiming for or a value obtained from a third-party study.

Benchmarking Study

A study intended to measure the user experience of a product over time. It usually involves looking at the same metrics over different iterations of the product. A benchmarking study might compare the most recent set of metrics (which represent your product's current performance) against the metric values collected in past studies. This comparison can help reveal whether and how much your product has improved over time.

Between-Subjects Design

A study design in which different participants are assigned to different conditions corresponding to a variable. For example, in a between-subjects study, you might recruit 80 participants for a quantitative usability study, and randomly assign each participant to either site A or site B. This type of study design is usually contrasted with within-subjects design.

Binary Metric

A metric that can have only two possible values. Examples include task success or conversion (that is, whether a certain desired event has happened — for example, whether a user has registered for an account).

Quantitative UX Glossary

Benchmark

(Not to be confused with benchmarking study)

A single number that you want to compare your site to. A benchmark could be a value that you are aiming for or a value obtained from a third-party study.

Benchmarking Study

A study intended to measure the user experience of a product over time. It usually involves looking at the same metrics over different iterations of the product. A benchmarking study might compare the most recent set of metrics (which represent your product's current performance) against the metric values collected in past studies. This comparison can help reveal whether and how much your product has improved over time.

Between-Subjects Design

A study design in which different participants are assigned to different conditions corresponding to a variable. For example, in a between-subjects study, you might recruit 80 participants for a quantitative usability study, and randomly assign each participant to either site A or site B. This type of study design is usually contrasted with within-subjects design.

Binary Metric

A metric that can have only two possible values. Examples include task success or conversion (that is, whether a certain desired event has happened — for example, whether a user has registered for an account).