Building Your Benchmark

How to Measure UX for Product Impact Over Time

Jennifer Otto | Yina Li Turchetti | Paul Sisler
Welcome

Jennifer Otto
Principal UX Researcher
Fidelity Investments

Yina Li Turchetti
Senior UX Researcher
Autodesk

Paul Sisler
Senior UX Researcher
Fidelity Investments
What We’ll Cover

2. Selecting Your Method
3. Determining Your Cadence
4. Choosing Your Recruit
5. Designing Your Study
6. Analyzing Your Data
7. Reporting Your Results
What Is a Benchmark Test?

- Testing an entire end to end IT system, website or application
- Helps get a pulse of how a product is doing overall, not to be used for a deep understanding of one area
- Repeatable and run on a regular cadence
- Helps to measure how one change affects overall performance
- Compares against own performance as well as competitors’

![NPS Chart]

NPS

- 2012
- 2013
- 2014
- 2016
- 2017

iOS

- 9.0
- 25.0
- 29.0
- 22.9
- 49.7
A Brief History of Our Benchmarks

- 100-200 participants per platform
- 10 most common tasks per platform, including some pre-login and post-login
- Demographics of participants line up with those of site/app visitors
- Run at year-end
- Quantitative and Qualitative task-based studies
What Our Benchmarks Measure

Task success  Ease of use  Time on task  SUS system usability score  NPS net promoter score

 Metrics are compared YoY and platform to platform
Why We Perform Benchmarks

Product Strategy

- Competition
- Market Research
- Customer Feedback (attitudinal)
- Analytics (behavioral)
- Benchmarking attitudinal & behavioral

use case

- 2016 Fidelity.com task success = 92.4%
- 2017 global navigation changes
- Feature testing went well
- 2017 benchmark test
- 2017 Fidelity.com task success = 75.9%
Building Your Benchmark

GETTING STARTED
Selecting Your Method

Choose an approach that’s repeatable and provides measurable results suitable for comparison from round-to-round of evaluation. It should leverage the strengths of your research team. For us, that’s mostly task-based and quantitative.

Primary components of our benchmark

<table>
<thead>
<tr>
<th>Qualitative Pilot</th>
<th>Quantitative Core</th>
<th>Qualitative Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmoderated online video study</td>
<td>Task-based online study</td>
<td>Unmoderated online video study</td>
</tr>
<tr>
<td>Catches glitches in the tools, task wording, or study design before launch</td>
<td>Tracks success, perceived ease of use, and task time</td>
<td>Provides explanation for patterns observed in core, with insights, and stories</td>
</tr>
<tr>
<td></td>
<td>Captures broad subjective measures: SUS and NPS</td>
<td></td>
</tr>
</tbody>
</table>
Determining Your Cadence

Develop a recurring cycle that respects your product release schedule and rhythm of your industry. Leave enough time to measure growth and effects of change in your product, your industry, and the broader environment.

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Our approach and concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once or twice per year</td>
<td>Once per year</td>
</tr>
<tr>
<td>Avoid unpredictable periods in industry</td>
<td>Tax season</td>
</tr>
<tr>
<td>Same time of the year, every year</td>
<td>November to December</td>
</tr>
<tr>
<td>Sufficient gap between releases</td>
<td>Skipped a year due to app redesign</td>
</tr>
<tr>
<td>Prepare for the unexpected</td>
<td>Market changes</td>
</tr>
</tbody>
</table>
Planning for Time-in-field

You want to be in the field for as little time as possible: updates and other factors can affect results. It may take you longer than you think to recruit, fill, and analyze.

### Factors likely to impact time

<table>
<thead>
<tr>
<th>Scope and number of tasks</th>
<th>Sample size, complexity, and incident rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of platforms</td>
<td>Recruit process (direct, panel)</td>
</tr>
<tr>
<td>Changes from previous rounds of study</td>
<td>Study requirements (log in to account)</td>
</tr>
<tr>
<td>Size and experience of team</td>
<td>Technical issues (product, tools, vendors)</td>
</tr>
</tbody>
</table>

**2017:**
- 2 months
- 3 researchers
- 3 platforms
Choosing Your Recruits

Keep your recruit consistent round-after-round. Match your sample to your users by platform, using info from analytics or other sources. Include cheater and speeder traps in the study tool if you can.

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Our approach and history</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruit users (not prospects, not co-workers, not the public, not your mom)</td>
<td>Customer panel and a third-party recruiter</td>
</tr>
<tr>
<td>Recruit for significant results - about 150 completes on each task per platform and per segment</td>
<td>About 300 mobile app users on two platforms and 300 web (on computer)</td>
</tr>
<tr>
<td>Prepare for cheaters and speeders</td>
<td>Removed 30% of sample after closing</td>
</tr>
</tbody>
</table>
Deciding What to Test

Test common and critical tasks. Cover the breadth of your information architecture based on both user and business needs. We selected tasks based on analytics covering specific user segments.

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use the same tasks every time</td>
<td>Tasks evolve with team and site/app</td>
</tr>
<tr>
<td>Use the same tasks across platforms</td>
<td>Core tasks and platform specific tasks</td>
</tr>
<tr>
<td>If measuring against competitors, ensure tasks can be done across all platforms</td>
<td>Selected 2-3 competitors that focus on our main tasks</td>
</tr>
<tr>
<td>Keep the total test time short</td>
<td>We limit it to 10 minutes, adjust tasks/participant and increase our sample size</td>
</tr>
<tr>
<td>Randomly assign tasks if you have too many and present them in random order</td>
<td>Login and logout are first and last. Other tasks are randomized</td>
</tr>
</tbody>
</table>
Building Your Benchmark

STUDY DESIGN CONSIDERATIONS
Define Your Goals

Identify clear your goals before creating your tasks.

Goal

How well do respondents understand the general market performance?

Examples

According to [Platform], is the NASDAQ up or down today?

According to [Platform], what is the current value of Dow Jones?
Be Sensitive to Personal Data

Have a way for participants to opt-out of the study if they are not comfortable continuing, due to potential sensitivity of their personal data.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Opt-out Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your study <strong>requires people to log in</strong> to their own accounts to complete some tasks which may make some uncomfortable</td>
<td>Study: <strong>Are you willing</strong> to log in to your account in your web browser on a computer and go through a few tasks for this study?</td>
</tr>
<tr>
<td>Participants may not want to provide <strong>personal information</strong></td>
<td>Task: Find your account with the smallest balance. What is its current value?</td>
</tr>
<tr>
<td></td>
<td>Provide <strong>ranged answers</strong> and an option of <strong>Prefer Not to Say</strong></td>
</tr>
</tbody>
</table>
Use Measurable Tasks

Use verifiable answers if you can. Self-report scores tend to introduce noise and produce inflated results.

**Example:** According to [Platform], is the NASDAQ up or down today?

<table>
<thead>
<tr>
<th>Type</th>
<th>Validation Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report – Likert:</td>
<td><strong>How confident</strong> are you that you have completed the task successfully?</td>
<td>1 (Not confident at all) – 7 (Extremely confident)</td>
</tr>
<tr>
<td>Self-report – binary:</td>
<td><strong>Did you</strong> complete this task successfully?</td>
<td>Yes</td>
</tr>
<tr>
<td>Semi-verifiable:</td>
<td><strong>Is</strong> the NASDAQ up or down?</td>
<td>Up</td>
</tr>
<tr>
<td>Verifiable:</td>
<td><strong>What</strong> is NASDAQ at today?</td>
<td>[Text entry]</td>
</tr>
</tbody>
</table>
Consider Measurable vs Realistic

Balance realistic task design and verifiable answers. Sometimes, you may have to choose between a more realistic task which only allows for self-reported answers versus a less realistic task which allows for measurable results.

Task goal: Can participants locate news relevant to a stock?

More Realistic User Task
Find a news article about [company X]. Were you able to complete this task?
Self-report: Y/N

More Measurable Research Task
Find the most recent article on [platform] about [company X]. What is the first word of the article?
Verifiable: [write-in]
Consider Write-in vs Multi-Select

Use open-ended answer format over multiple choice options to curb cheating behavior.

Task: Find the most recent price for Apple stock (AAPL).

<table>
<thead>
<tr>
<th>Easier to Grade</th>
<th>Easier to Verify</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Below $150</td>
<td>• [Open ended text answer]</td>
</tr>
<tr>
<td>• Between $151 and $190</td>
<td>• I don’t know</td>
</tr>
<tr>
<td>• Between $191 and $230</td>
<td></td>
</tr>
<tr>
<td>• Between $231 and $270</td>
<td></td>
</tr>
<tr>
<td>• Above $271</td>
<td></td>
</tr>
<tr>
<td>• I don’t know</td>
<td></td>
</tr>
</tbody>
</table>
Building Your Benchmark

ANALYZING & REPORTING RESULTS
Analysis

- Use same methodology year after year
  - Clearly define success
  - Don’t change scales (5- to 7-point)
- Determine and record task-level success for write-in responses
- Monitor how time-on-task is calculated if you change tools or methods
- Review supplemental qualitative studies to add quotes and video content to support conclusions

**Year-over-Year Average Success**

- 2012: 93%
- 2013: 94%
- 2014: 96%
- 2016: 83%
- 2017: 83%
Reporting

Start with overall findings, then drill into tasks

Include charts and verbatims

Describe factors that may have affected results: demographic changes, industry events

Include screen shots of before and after designs
In Summary

Take time when developing your study design

Leverage data from both qualitative and quantitative task-based methods

Consider your research goals when developing tasks
Questions
Keep in Touch

Jennifer Otto:  jennifer.otto@fmr.com
Yina Li Turchetti:  yina.turchetti@autodesk.com
Paul Sisler:  paul.sisler@fmr.com