

Building leadership support for good design with better metrics



Emily Lam,
User Experience Researcher, BenchSci



Marta Mazurkiewicz,
Senior UX Researcher, BenchSci

Building leadership support for good design with better metrics

Emily Lam and Marta Mazurkiewicz

BenchSci

Pistoia Alliance UXLS 2023

Outline

- 1 The challenge with metrics for leadership
- 2 Our solution – A recipe to follow
- 3 The impact of our work

sanofi

Bristol Myers Squibb™

BIOMARIN

Voyager
THERAPEUTICS

UNIVERSITY OF
CAMBRIDGE

UNIVERSITY OF
TORONTO

GILEAD

moderna™

AMGEN

abbvie

ETH zürich

UCL

Takeda

NOVARTIS

THE UNIVERSITY OF
CHICAGO

東京大学
THE UNIVERSITY OF TOKYO

GSK

AstraZeneca

BenchSci

HARVARD

CAROLINSKA INSTITUTET
ANNO 1810

UNIVERSITY OF OXFORD

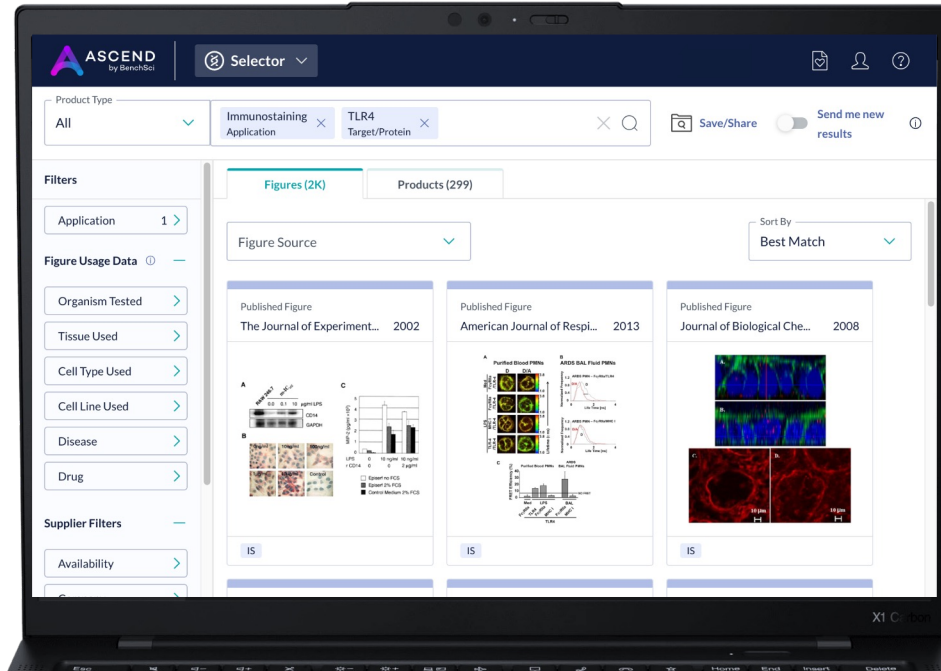
novo nordisk

Mit

UCSF

Daiichi-Sankyo

Stanford



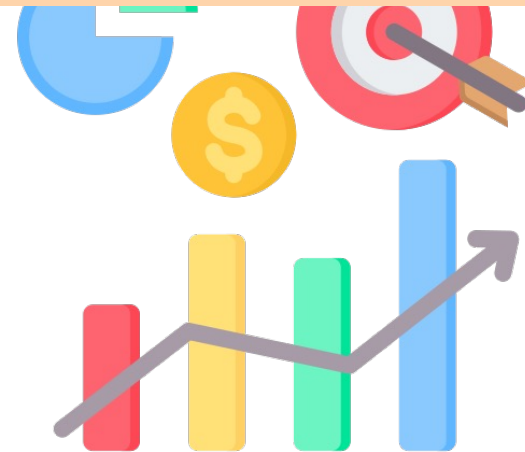
What does leadership respond to?

Building leadership support

for good design with better metrics

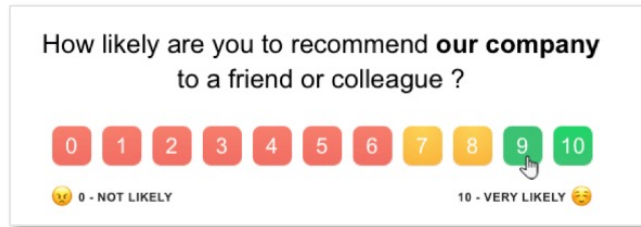
Numbers!

- Familiar metrics that are easy to understand
- Measurable over time
- Concise

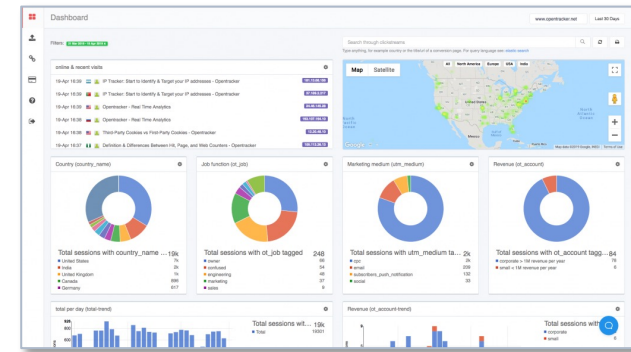


What tools already exist?

Single Survey-based Metrics



Engagement Metrics



Why is the number changing?

How can we solve the problem?

It's hard to say what all these numbers mean

We need more data...



For every 10 new users,

will find repeated success in their key tasks.

= usability score from 3 task evaluations \times completeness of our data

Recipe



- 1 Participants
- 2 Tasks
- 3 Scoring
- 4 Evidence
- 5 Reporting

1

Participants

- Pick a group that best reflects what you want to evaluate
- Decide how many participants
- Consider how much time and resources you have



x10

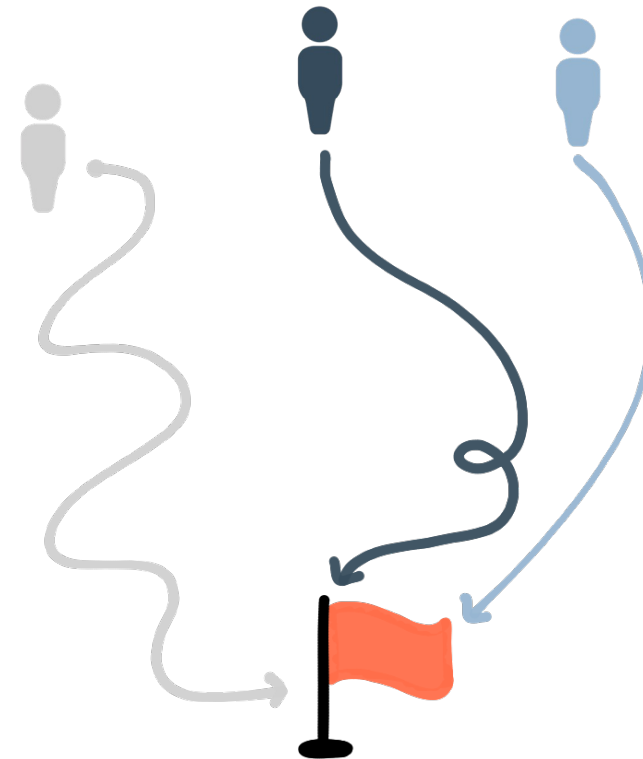
x 45 min/session
x 2 or 3 researchers per session
+ analysis
+ planning
+ alignment
+ reporting

= ~150 people hours over 1.5 months

2

Tasks

- Features can change, focus tasks on primary user goals
- Test in a live application
- There can be multiple paths to accomplish a user goal
- Prototypes may not have the data depth needed for a user to find value



3

Scoring

- Scores reflect how well the product supports your users
- Create a standardized scoring system



Would the user succeed if they returned to the app?

0	Unlikely to complete
0.5	May complete
1	Likely to complete

3

Scoring

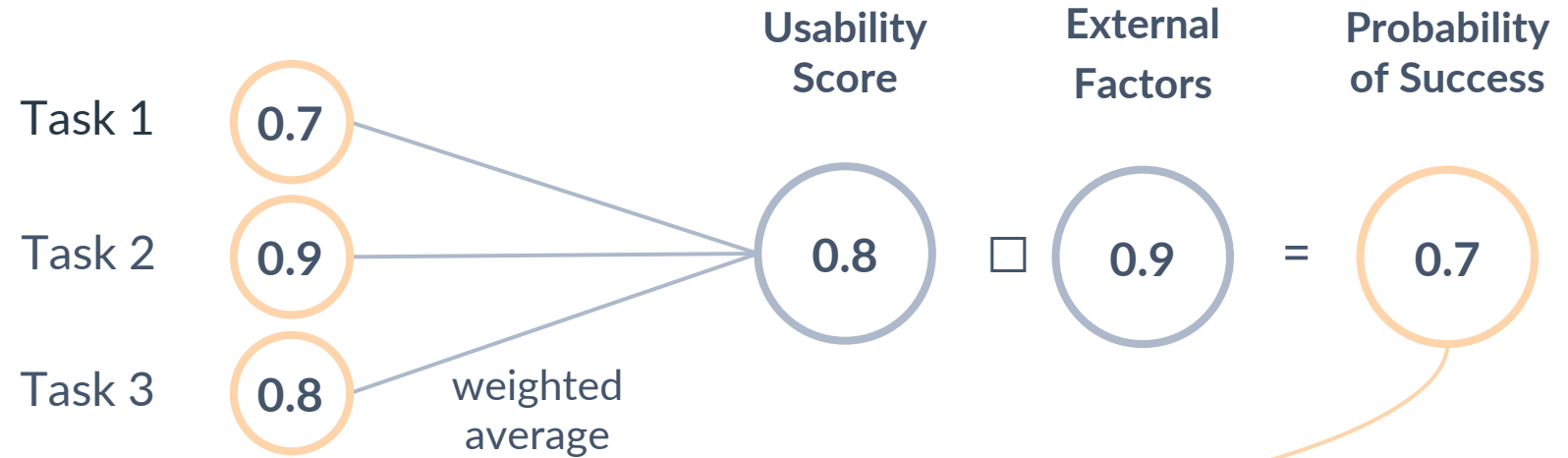
Task 1: Find a reagent for an experiment related to your research

Session	Participant	Score					Path
		Rating	Initials	Rating	Initials	Combined	
1	Alex						

Average **0.7**

3

Scoring



For every 10 new users,

7

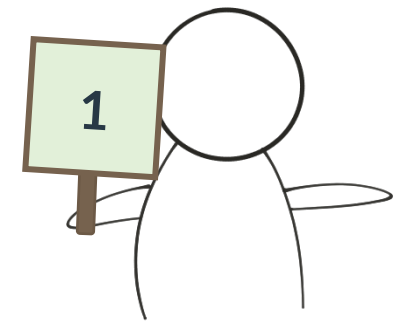
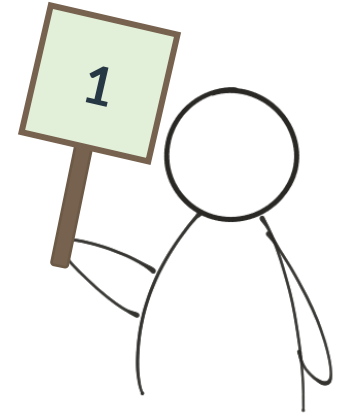
will find repeated success in their key tasks.

3

Scoring

To build trust and gain acceptance for the process, consider

- Scoring with 3 researchers
- Benchmarking/norming exercise
- Understanding and discussing variability and differences in scoring



4

Evidence

Identify and illustrate the reasons behind the scores

The screenshot shows a video conference in progress. The main window displays the 'ASCIEND Selector' software interface. The interface includes a search bar, a filter sidebar on the left, and a central grid of scientific figures. The figures are arranged in a 2x4 grid, each with a title and a thumbnail image. The titles include 'The Journal of Experimental...', 'American Journal of Resea...', 'Journal of Biological Che...', 'Journal of Biomedicine...', 'Molecular Medicine Report...', 'American Journal of Pharm...', 'Journal of Cellular Phys...', and 'Biochemistry & Biophysics...'. The date '2023-07-03 10:07:40' is visible in the bottom right corner of the software window. In the top right corner of the video conference, there are three video thumbnails of participants. Below the software window, there are five time-stamped markers: 11:22, 17:27, 22:42, 22:42, and 29:31, each with a small colored dot below it. To the right of the software window, there is a vertical list of reasons for scores, each in a colored box: Positive reaction (red), Negative reaction (yellow), Conceptual misunderstanding (green), Unmet expectation (teal), Learnability (blue), Discoverability (light blue), Questions (purple), Feature requests (dark blue), and Noteworthy (pink).

Positive reaction

Negative reaction

Conceptual misunderstanding

Unmet expectation

Learnability

Discoverability

Questions

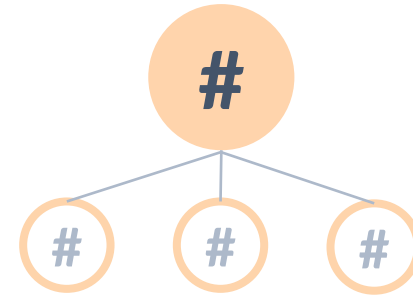
Feature requests

Noteworthy

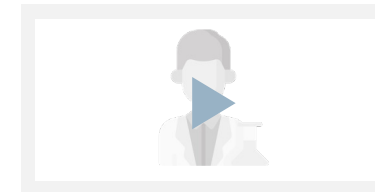
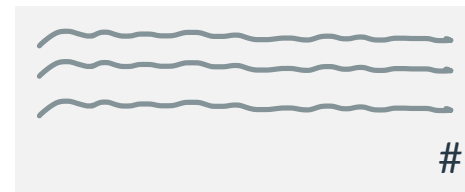
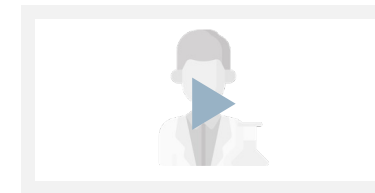
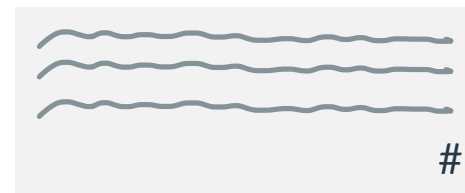
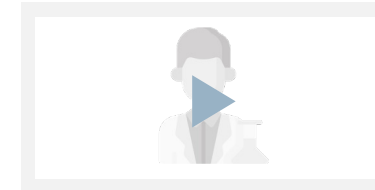
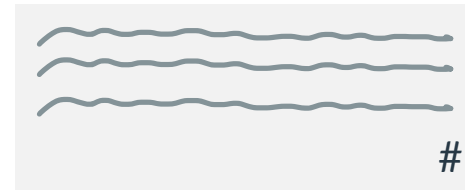
5 Reporting

Give the **team** the details to understand the scores and develop a mitigation plan

- Overall and task scores
- Main reasons
- Frequency counts
- Evidence



Main reasons:



5

Reporting

For **leadership**, be concise and focus on the big picture

- Overall score
- 3 main contributors for the score
- Steps to success
- Timeline



5

Reporting



Impact

- Leadership allocated resources to address issues
- Teams prioritized projects based on findings

BUT this takes a lot of time

I can see this scoring as a way for us to frame our work and say, “**Ok, which of these ideas in the backlog do we feel actually makes a tangible impact on the number?**”

The people that are directly involved with design and product research were **validated** [...] For the rest of the team, they knew what we were working on and what we believe would make a difference. I think they were **excited and felt at ease**.
What I was very excited about was that we were building **a tool to measure the state, maturity, health, and usability of our product.**

HUGE thanks to Jeremy Epstein, Mayte Gonzalez, Fariha Mosaddeque, and Keira Pereira

Try this out during our workshop!

We're hiring!

https://linktr.ee/benchsci_research



Image Attributions

- NPS Survey - <https://www.zonkafeedback.com/blog/why-measure-net-promoter-score>
- User Engagement Dashboard - <https://www.opentracker.net/article/user-engagement/>

From flaticon.com:

- Metrics by Uniconlabs
- Cooking by justicon
- Scientist by Freepik