Full case study available on request

Evaluating the Future of Financial Modeling for Microsoft's Enterprise Deals

Benchmarking a Next-Generation Financial Tool Against its Legacy Predecessor for Accelerated Deal Velocity at Microsoft

Microsoft Commerce + Ecosystems, 2024

Some content is obscured for NDA.

Microsoft's Commerce + Ecosystems (C+E) Studio is a core team of UX professionals providing user insights across the company's vast range of business tools and processes, serving strategic industry clients for licensing and Enterprise Agreements.

Photo by Sam Torres on Unsplash

Evaluating the Future of Financial Modeling for Microsoft's Enterprise Deals: Boosting Deal Velocity to Drive Microsoft’s Bottom Line (2024)

Project Summary

I used a mixed-methods approach to compare a legacy tool with its replacement, collecting quantitative performance metrics and qualitative usability feedback via post-task and post-product assessments.

Timeline & Team

Co-led with another UX Researcher over 6-months. Independently conducted all sessions.

Deliverables

- Benchmarking Framework + Metrics
- Usability Insights + Scorecard
- Report + Executive Summary

Impact

This study confirmed significant efficiency gains in the new tool, which will translate to increased deal velocity and volume for Microsoft.

Problem: Enterprise Agreement financial modeling at Microsoft is a critical yet notoriously slow and error-prone phase of deal-making. **How does the newly developed financial modeling tool perform against the established legacy tool?**

The overarching goal of this benchmarking study was to evaluate and compare the performance of the legacy and next-generation financial tools while gathering usability feedback to center user voice in product strategy. I accomplished this in 3 ways:

1

Define Core Performance Metrics

- Establish a comprehensive set of quantitative and qualitative metrics to evaluate performance of both the legacy and next-gen tools.

2

Benchmark Tool Performance

- Systematically measure and compare the two tools across critical dimensions such as speed, accuracy, overall experience.

3

User-centered Product Impact

- Gather usability feedback to ensure user-voice drives iterative product improvement, ensuring a smooth roll-out of the new tool.

I co-led this six-month study, collaboratively defining the framework and metrics, **then independently conducting** 12 remote benchmarking sessions (120 minutes each) across US and EMEA offices**.**

Collaborative Approach

1

Discovery & Definition

Identified core users and 6 critical tasks with Product Owners & Stakeholders.

2

Scenarios & Metrics

Crafted realistic test scenarios and defined key metrics like Time on Task & Error Rate.

3

Execution & Feedback

Executed tests, measured performance, and captured continuous user feedback for improvement.

4

Insight & Integration

Analyzed data, and fed insights back to the development team.

What we measured and why:

Task-Level Performance Metrics:

a. **Perceived Ease, Speed, and Confidence (Likert)** to measure subjective user experience in 3 critical dimensions for each task. (⚖️)

b. **Combined Experience Score (percentage)** to provide a single, normalized metric for overall experience. (%)

c. **Time on Task, Success & Failure Rates** to measure efficiency, speed, error incidence for each task. (00:00)

Product-Level Performance Metrics:

a. **Overall Experience Ratings** to gather broader user impressions on tool capabilities beyond specific tasks, identifying strengths and weaknesses. (⚖️)

b. **Net Promoter Score (NPS)** to measure overall user loyalty and willingness to recommend each tool. (-100 to 100)

c. **Tool Preference** to provide high-level, comparative insights into the overall value proposition and user acceptance of both tools. (💬)

A significant challenge emerged: Users were highly proficient with the legacy system but had no prior exposure to the new one, which risked skewing performance metrics based on familiarity rather than intrinsic tool capability. As a result, I recommended implementing an additional set of metrics on tool familiarity to ground all other metrics::

Familiarity Metrics: Quantifying user experience through a scale, years of use, and deal count

1. **Familiarity Scale**
   For each tool, rate your level of familiarity (1-5)

2. **Years of Experience**
   How many years of experience do you have with each tool?

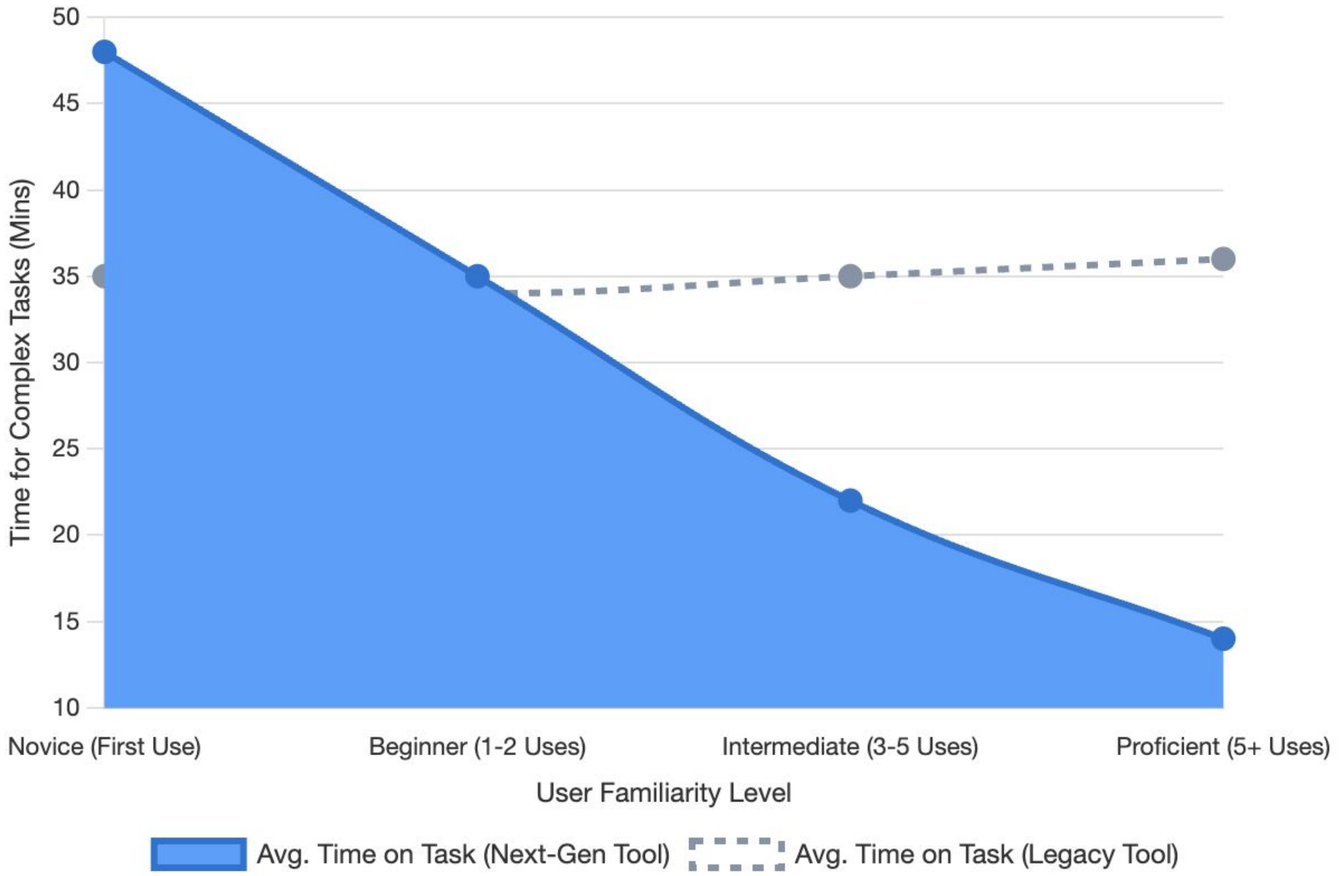
3. **Number of Deals in Each Tool**
   How many deals have you completed in each tool?

Contextualizing Performance: Side-by-side familiarity levels for each tool provide crucial context for interpreting performance metrics.

Attributes	Tool 1 (NDA)	Tool 1 (NDA)
Familiarity rating	4.6	2.9
Average number of deals created	480	8.8*
Median number of deals created	600	0
Number of years of experience	5.6 years	< 1 year**

N= 8
Scale used: *1-Not familiar at all, 2-Not familiar, 3-Neutral, 4- Somewhat familiar, 5-Very familiar*

By collecting tool familiarity data, **we could contextualize performance metrics with users' familiarity levels,** enabling us to project future gains in confidence and speed as new tool familiarity increased.



As user familiarity with the Next-Gen tool increases, we project a significant decrease in task completion time, surpassing the legacy tool's efficiency.

Having grounded users' responses within their level of experience with each tool, we then presented quantitative metrics and qualitative usability findings for each task.

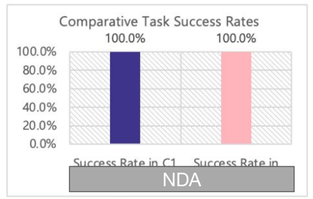
Task 2: Add Packages

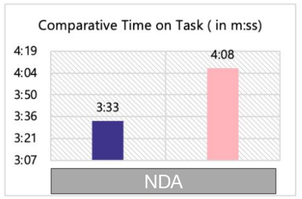
Instruction:

You want to add 1 custom base package and 2 add-on packages to this quote. Please rename the packages by adding "Year One" to each. For Azure package, change hours for Service Delivery Management Extended.

Key take aways:

- Users performed the significantly task faster in NDA
- Users appreciate new features: division of Support & Consulting packages and Add-to-Cart functionality.
- Users did not know about bulk add capability, but when shown, they were excited by the time-savings.
- Areas of improvement in UI: Search Box visibility, Packages hierarchy & navigation; ability to edit and organize packages on Package List view





Experience Score	Attributes	Experience score in C1
Task 2 Add packages	Ease of Use	83%
	Perceived speed on task	17%
	Confidence in task completion	100%

Experience score measures the percentage of participants who rated an attribute as 4

Scale used: *1-Not satisfied at all, 2-Not Satisfied, 3-Neutral, 4- Somewhat Satisfied, 5-Very Satisfied*

>= 80% rating *70 to 79% rating* *<70% rating*

1 We presented side-by-side comparison of quantitative performance metrics for each task

2 Qualitative usability findings were prioritized by severity to guide strategic product improvement

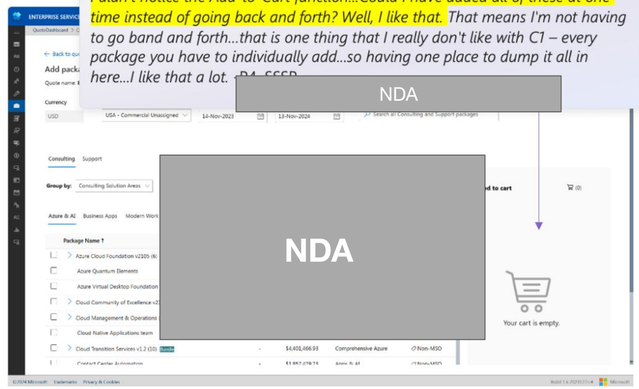
Task 2: Add Packages - FINDING #2

T2: Most users did not notice the Add-to-Cart functionality but liked it once knew they could bulk add packages

- Users said they liked being able to add packages in bulk and that this will save them time.
- NDA aid being able to check/edit what they are adding in the shopping cart before adding to deal is useful.

That's actually very useful [Add-to-Cart]. I like that a lot because right now in Co NDA ne, you have to go in and add the schedules individually." -P5, NDA

I didn't notice the Add-to-Cart function…Could I have added all of these at one time instead of going back and forth? Well, I like that. That means I'm not having to go band and forth…that is one thing that I really don't like with C1 – every package you have to individually add…so having one place to dump it all in here…I like that a lot. NDA



RECOMMENDATION

✓ Consider showing a prompt when users add one package that tells them that they can add multiple packages in one go in NDA

We also provided overall product-level comparisons, showing experience ratings across seven dimensions, Net Promoter Scores (NPS), and user preference ratings.

March 2024

CPQ vs. C1 Comparative Study

Post Product Metrics: Overall Experience & NPS

CPQ received higher post product ratings in all categories, with a notably higher NPS score, over C1.

Attributes	Tool 1 NDA	Tool 2 NDA
Overall Satisfaction	71%	100%
Ease of Use	71%	86%
Technical Reliability	43%	71%
Features and capabilities	29%	100%
Visual Appeal	14%	86%
Speed	14%	86%
Data Accuracy	29%	71%

Overall Experience score measures the percentage of participants who rated an attribute as 4 or 5.

N= 9
Scale used: *1-Not satisfied at all, 2-Not Satisfied, 3-Neutral, 4- Somewhat Satisfied, 5-Very Satisfied*

>= 80% rating *70 to 79% rating* *<70% rating*

	Tool 1	Tool 2
Net Promoter Score	-11	67

N= 9

Tool Preference		Synthesized reasons for preference
# of users that preferred Tool 1	1	High level of familiarity with the tool
# of users that preferred Tool 2	7	Speed; Visual Appeal, Time-saving bulk capabilities (add, copy, edit)

I mean, hands down, I would take Tool 2 the end of the day. You know anything that can save us time is super welcomed. Anything that can help minimize any mistakes we make. Of course is super welcome and anything that can make our jobs easier is super welcome, and I thi Tool 2 checks the box for all three.- P3, SSSP

1 Comparative scores on 7 dimensions of user experience

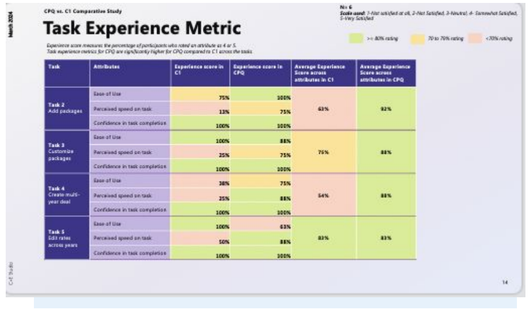
2 Comparative NPS scores

3 “Which tool do you prefer?”

Deliverables

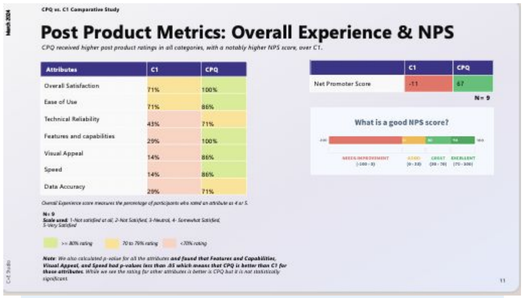
6 Task Level Benchmark Metrics

These metrics were contextualized by users' familiarity with each tool and measure how users performed and felt while interacting with each tool on specific, critical financial modeling tasks.



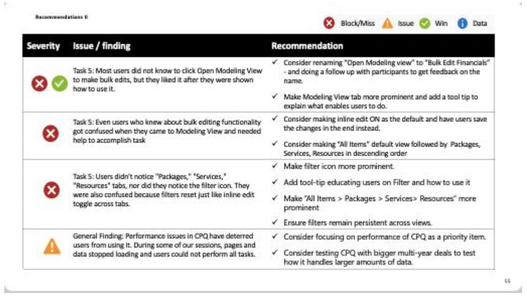
3 Product Level Benchmark Metrics

These metrics were used provide a holistic assessment of each tool, gathering broader user experience, including tool preference.



Set of Prioritized Usability Insights

Ranked findings by severity to directly guide product strategy and ensure user feedback drives iterative improvements.



Research Topline:

Key Finding: Users rated all metrics in post product survey better for Tool 2 *even after failing tasks* in it due to lack of familiarity. New features – specifically, bulk copy and bulk edit – will significantly reduce time and errors in financial modeling tasks.

Key Recommendation: Efficiency gains via all new features in Tool 2 might not be as visible until users become more familiar with it and are aware of optimized task flows. Invest in non-tool readiness trainings. Continue to implement usability feedback to ensure smooth transition.

Business Impact: While users may experience initial friction due to the learning curve, clear signals indicate that the new tool's features will significantly reduce financial modeling bottlenecks, leading to faster, more accurate, and higher-volume deals for Microsoft in the future.

IMPACT

Key Study Findings: Despite significant disparity in user familiarity, the next-generation tool demonstrated substantial overall time savings, largely attributed to new powerful bulk editing features.

Actionable Outcomes & Recommendations: The usability findings provided a detailed scorecard for design, highlighting specific areas for improvement across all tasks.

For Product & Business Stakeholders: We established a robust framework for long-term performance comparison, integrating continuous open-ended usability feedback to directly inform product improvements.

Impact on Research Capacity: We successfully developed the research studio's first-ever benchmarking study template, significantly enhancing our future research capabilities.

REFLECTION

This study taught me the critical balance required to meet diverse stakeholder priorities, from business metrics and product performance validation to design usability insights and research rigor, all while navigating significant constraints.

Despite limitations like small sample sizes, vast user familiarity disparity, and the logistical challenge of conducting lengthy, multi-faceted tests solo, we successfully delivered a robust set of metrics, definitive insights on the new tool's performance, actionable usability feedback that put users into the design and improvement process, and a valuable templatized benchmarking study, proving that meaningful findings can be achieved even without strict statistical significance when stakeholder alignment is present.