

The BLUE Framework: Designing User-Centered In-Product Feedback for Large Scale Applications

SERENA HILLMAN

Microsoft, Azure Data, serena.hillman@microsoft.com

SAMIRA JAIN

Microsoft, Azure Data, samira.jain@microsoft.com

VICHITA JIENJITLERT

Microsoft, Azure Data, vichita.jienjiltert@microsoft.com

PAULA BACH

Microsoft, Azure Data, paula.bach@microsoft.com

In-product feedback mechanisms allow for capturing user feedback while the user is engaging with the product or service. Traditionally in-product feedback has focused on metrics such as Net Promoter Score (NPS) [21] and Customer Satisfaction (CSAT) [5] which look to measure customer loyalty or overall product satisfaction. By introducing complementary user experience (UX) metrics that are focused on user outcomes, UX teams have greater insight into measuring the successes or challenges of their users in the context of use. This case study describes and discusses the process employed and the lessons learned while designing and implementing a user-centered in-product feedback system. We specifically call out challenges and opportunities around aligning with business outcomes, navigating current frameworks, unlocking self-serve data to stakeholders, informing strategy, and feeding additional research. In conclusion, we present these learnings as a framework, dubbed BLUE, to help other UX teams design and implement in-product feedback mechanisms.

CCS CONCEPTS • Human-centered computing • Human computer interaction (HCI) • HCI design and evaluation methods

Additional Keywords and Phrases: In-product feedback, UX outcomes, self-serve data, UX health metrics

ACM Reference Format:

Hillman, S., Jain, S., Jienjiltert, V., and Bach, P. 2022. The BLUE Framework: Designing User-Centered In-Product Feedback for Large Scale Applications.

1 INTRODUCTION

In-product feedback is a valuable tool for gaining insight into how users think and feel about a product or service while they are using it, rather than gaining insight from simulated environments (e.g., usability testing) or out of context surveys

asking about an experience that may not have happened or happened too far in the past to gain valid feedback. In a user-obsessed business world, connecting with users while they engage with a product or service provides access to large-scale data making this in context feedback a powerful and useful resource. In-product feedback mechanisms have a mix of possible prompts. Common prompts include user-initiated and system-initiated. A user-initiated prompt occurs when the user chooses to leave feedback when they want through a dedicated feedback collection interface designed into the system. System-initiated prompts are of two types: sample and behavior. Sample prompts occur when the system asks for feedback based on predetermined sample logic (e.g., people new to the system) whereas behavior prompts occur when a specific user behavior triggers a prompt for feedback in the moment (e.g., the user closes a window, and the system displays the feedback UI). The ability to control the sample allows the mechanisms to capture representative feedback on the product, or feature within the product, to understand constructs such as user sentiment, satisfaction, and loyalty. In contrast, the user-initiated prompt allows users to share their thoughts at any time, for example, bugs, annoyances, or feature requests. Our framework includes all three types of feedback mechanisms.

At Microsoft, our team of five researchers is responsible for a family of services related to managing, transforming, and storing data within the Azure Cloud. Some of the twenty key services we currently support include Azure Synapse Analytics, Azure Purview, Azure Data Studio, and Azure Cosmos DB. In March 2020, our team began designing an in-product feedback system that was consistent across six key services to capture feedback through a user-centered process. The need was driven by a mixture of ad hoc in-product feedback mechanisms and inconsistent metrics captured across several Azure Data services. In some cases, these inconsistencies resulted in unacceptable disruption to the user experience and inability to act on the feedback or analyze data across the services for trends.

Our user-centered process initially included seven high-level steps:

1. Internal Metrics and Mechanisms Review.
2. External Metrics and Mechanisms Review.
3. Identifying UX outcomes.
4. Mapping UX outcomes to UX metrics.
5. Designing initial feedback mechanisms and testing with users.
6. Socializing our designs and process to internal stakeholders.
7. Implementing the design. Sections 2-5 of this paper describe these seven steps in detail.

See Figure 1 for the process we used to design a user-centered, in-product feedback system.

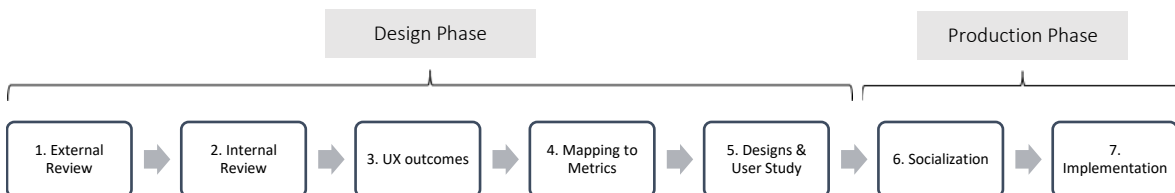


Figure 1 Our initial process for designing in-product feedback mechanisms and metrics

The final section of this paper presents our contributions to the space. Key lessons learned include the need to patch together multiple existing metrics frameworks, working with business objectives and metrics, and designing and planning for implementation to ensure the metrics would be useful and usable thereby contributing to our data driven culture. Given the lessons learned in executing the user-centered metrics design process, we propose a framework, named BLUE, for designing and implementing in-product feedback that supports the end-to-end design and implementation process. The pillars of the proposed BLUE framework include Build UX metrics, Leverage proven mechanisms, Unlock results, and Embed in practice.

2 REVIEW

2.1 Selecting user-centered metrics

Our external review of best practices around capturing metrics and feedback mechanisms yielded frameworks from both user centric and business centric perspectives. Over the last few decades, several frameworks relating to user-centered metrics have been introduced to the Human-Computer Interaction (HCI) and User Experience research domains such as the Software Development Productivity Framework [23], the See-Think-Care-Do-Framework [8], and Google’s HEART framework [21]. These frameworks focus on starting with an objective and then identifying signals and metrics that measure if that objective has been achieved. However, the HEART framework brings an additional level of focus on the user experience by identifying four distinct areas to measure: H – Happiness (the user’s overall emotion), Engagement, Adoption, Retention, Task (a catch-all for specific tasks). Using the framework requires defining objectives, signals, and metrics for each of the five areas within the context of a product or service.

From the business context, frameworks such as Grove’s Objectives and Key Results (OKRs) [11] and Seiden’s Outcomes Over Output [25] focus on stressing the importance of measuring business outcomes over business output. This point of view has recently regained popularity within the tech world with Doerr’s 2018 book: *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the Work with OKRs* [6]. This goal setting methodology determines the objectives for a set time, usually a quarter. Objectives (O) are high-level statements that are qualitative, while the Key Results (KR) are the measurable results of the objective. Doerr states that the four superpowers OKRs bring to an organization are as follows: Focus on Commitment to Priorities; Align and connect for teamwork; Track for accountability; and Strive for amazing. The book outlines each of these priorities with updated industry stories and examples.

Within the UX space, Jared Spool has recently defined an approach to UX metrics that highlights balancing the current over-indexing of capturing feedback from users to measure business outcomes. In recent articles [24][27] he echoes Seiden’s sentiment that metrics must focus on outcomes over outputs. He differentiates between business outcomes and UX outcomes. On one hand, business outcomes have meaningful outcomes for the organization – such as *How do we get subscriptions to go up?* or *How do we get customers to purchase from us again?* On the other hand, UX outcomes are meaningful outcomes for the user that can be both behavioral and attitudinal. To define UX outcomes, Spool’s strategy asks, *If we do a fantastic job on that solution, how would it improve someone’s life?* For example, if we do a great job designing this roadside assistance software, we are helping drivers feel safer, and from this we find good UX outcomes to determine our UX metrics. Spool’s UX outcomes fit into four buckets of UX health metrics:

- 1 UX Success Metrics: Tells us the moment when the user has achieved the outcomes.
- 2 UX Progress Metrics: Metrics that happen along the way to the success of the metric.
- 3 UX Problem Value Metrics: Metrics around the obstacles people run into.
- 4 UX Value Discovery Metrics: Metrics we surface to our users to improve UX.

Further, Spool stresses business outcomes and UX outcomes complement each other. The UX outcomes add a user-centered layer identified by user problems that need to be solved and achieving those outcomes should support the achievement of the business outcomes in a healthy way. In Figure 2, the bottom path represents what Doerr [1] and Grove [11] would describe in their pursuit of business outcomes. The top path illustrates the user’s problem as it relates to the solution between the problem, the solution, business outcomes, and UX outcomes.

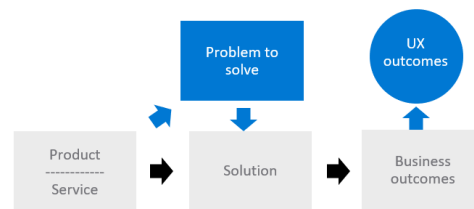


Figure 2 Spool’s Relationship between Business and UX outcomes. [26]

We saw strong alignment across all three types of frameworks and wanted a framework to include both business centric and user centric outcomes. As such we mapped the frameworks to see how they aligned or not. We mapped Spool’s UX

Success to Google’s H metric and the UX progress metrics as the subsequent EART metrics. With OKRs we see the Objectives as the UX outcomes, and the KR as UX success metrics, or, depending on the UX outcome, UX progress metrics with increased granularity. See Table 1 for the mapping.

Spool [26]		HEART [22]		OKRs [6]
UX Outcomes	=	Goals	=	Objectives
UX Success Metrics	=	Happiness metric	=	Key Results
UX Progress Metrics	=	EART Metrics	=	Key Results w/ more granularity
UX Problem Value Metrics	=	EART Metrics	=	Negative Key Results

Table 1 Our understanding around alignment of concepts across user-centered metric Frameworks

2.2 Mechanisms for capturing user-centric metrics

Google’s HaTS [18] method is perhaps the most well-known mechanism in the UX community for capturing in-product feedback. The Happiness Tracking Survey method lays out an approach using random sampling [15], a specific survey design, and some guidance on measuring satisfaction at both the product and task level. The HaTs method describes starting with an overall satisfaction question then provides an opt-in to answer additional questions around more finite tasks. The method also includes a recommendation for open-ended questions for users to vent frustrations, request new features and express appreciation. Prior to starting the design of our feedback mechanisms, we reviewed this method in detail to start from these best practices.

We also completed a review of the in-product feedback both at Microsoft and our competitors. From this, we were able to determine different types of in-product feedback mechanisms commonly employed in-house and in our product domain. Our review yielded common feedback mechanisms. The system-initiated feedback mechanism we discovered across our services relied on sample rules to gather feedback outside of product usage. The second type of in-product feedback mechanism was behavior based where the user provided feedback about the behavior they just completed. A simple example of this used the question prompt, *Was this documentation useful?* And the corresponding response item was a thumbs up or down or a Yes/No. The third type of in-product feedback mechanisms we identified was the user-initiated feedback mechanism. As the sampling is not controlled, the objective of this feedback mechanism was to provide sentiment around the satisfaction of a product but rather provide users with a mechanism to be heard. User-initiated feedback of this type is useful for bug reporting and capturing feature ideas that can be explored through additional efforts, but it is limited by those who are motivated to find the feedback UI and confident enough to have feedback, which are both a type of sample to be explored and understood but not the topic of this paper.

3 GETTING TO UX OUTCOMES

3.1 Objective

After reviewing the different literature and best practices around metrics and mechanisms, we set off to build our UX metrics by starting with UX outcomes. Our organization collects many business metrics such as CSAT, NPS, and Engagement, and was missing UX metrics. In this section we detail our approach for defining UX outcomes from an internal survey [4][8] as a first order understanding for how we can improve our users’ lives and what key moments are tied to this improvement. The results of the survey helped us identify UX outcomes and eventually derive UX metrics from the outcomes. This internal survey was completed in June and July of 2020.

3.2 Methodology for gathering UX outcomes

We surveyed our organization’s design and research team as they are experts in understanding our users’ journeys and objectives across our organization’s different products and services. We recognize that the ideal way to get UX outcomes is to derive them from primary research with the users of the products. However, we chose to use an expert approach due

to time and access constraints. We bootstrapped getting UX outcomes because it was more important to optimize getting the UX metrics system built and running rather than optimizing for the ecological validity or developing new constructs.

To gather the UX outcomes from our UX experts, we had them select core priorities defined by senior leadership that aligned most with the products they worked on so that we could optimize for depth of expertise. This meant that they chose at most two of the four core priorities. After this, the UX experts answered four questions for each of the two business goals they selected, all of which were open-ended responses. We derived some questions (Q3-Q5) directly from Jared Spool's UX Outcome Framework [27]:

Q1. Leadership has identified four core priorities. Please select 2 of these core priorities that you feel are most important to focus on for your product's UX for the semester. [*Multi-select two of the four core values*]

Q2. Please tell us briefly what <core value>, from a UX perspective, means to you for your product? [*Open-ended*]

Q3. If we do a fantastic job with <core value> in your product, how does this improve our user's life? [*Open-ended*]

Q4. What would be the key moment in our user's experience where they achieve this in your product? [*Open-ended*]

Q5. Are there problems our users regularly encounter that impact their ability to achieve this key moment in your product? [*Open-ended*]

3.3 Mapping outcomes to metrics

The survey collected a total of fifteen responses, representing around 75% of our research and design team. Answers spanned the six key products within our organization's family of services. From the data, themes emerged around our team's view of how we might improve our users' lives. We found that our UX outcomes focused on simplification of the complex, completing core tasks efficiently, and improvement of daily work routines. We mapped these outcomes to the following objectives usefulness, ease of use, efficiency, and user value and adapted existing metrics (e.g., [6]) to measure them.

3.4 Reflection

As discussed, we opted to use expert feedback to understand better the UX outcomes because of constraints such as aligning with release schedules, research budget, and access to highly specialized data professionals. Our derived UX outcomes reflected already known best-practice UX measurements. This bootstrapped method for deriving UX outcomes may have revealed a bias in the system showing where UX practitioners see user success from the lens of how the HCI community has defined them, i.e., useful, usable, efficient, satisfactory. An alternative approach to defining UX outcomes is to take an outside-in perspective [2] rather than an inside-out perspective as we did with BLUE. Taking an outside-in perspective to UX outcomes means that we generate the UX Outcomes from the users through in-depth journey mapping research where the users step us through their journey, and we learn what they are trying to achieve and why [28]. This approach ensures that we are evaluating our products and services from what matters most to users as they are working to accomplish their goals. As with most expert [19] methods, the first iteration of the BLUE framework generated UX outcomes via expert review from a principled approach. Further iterations of BLUE will include updating UX outcomes with success criteria derived from primary research rather than expert review. As such, the answers to the UX Outcome Framework questions will be derived from actual users rather than from within our UX team.

4 REFINING FEEDBACK MECHANISMS

4.1 Objective and methodology

Having our UX metrics in hand, we wanted to test the feedback system in which they were to be collected. To do this we conducted a remote user study with the target user group. Our goals for running the study were to collect initial impressions, test comprehension of the *give feedback* buckets proposed, and understand the impact that the system-initiated mechanism may have on users' workflows. We developed two interactive clickable prototypes, one for the system-initiated design (Figure 3) and the second for the user-initiated design (Figure 4). We recruited seven data engineers experienced in using a cloud-based data warehouse through a 3rd party recruitment agency. These data engineers were defined not by their job

titles but by tasks they completed at work (e.g., extraction, transformation, and loading of data). The protocol involved showing the participants the prototypes, interacting with them as they tried to complete a feedback submission, and encouraging them to think aloud. Remote sessions applied best practices for remote user research (e.g., [13][1]). The study sessions lasted around thirty minutes and were conducted in April 2020.

4.2 Findings & design implications

The user study helped us identify usability problems with the feedback system which were addressed in subsequent design iterations. One of the usability problems uncovered was user confusion around receiving a follow-up response from Microsoft after submitting feedback. Confusion also emerged around an option to send a compliment about something users liked in the product (see Figure 4). Users were not sure who would receive the compliment and did not see value in it, so we removed it and replaced it with a general feedback option. For the system-initiated prototype (Figure 3), we learned where placement of a pop-up would be least disruptive in their workflow.

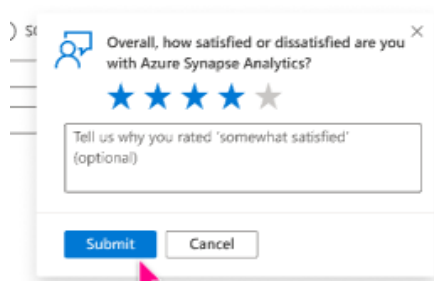


Figure 3 System-initiated Prototype Design, interaction included full label on roll-over.

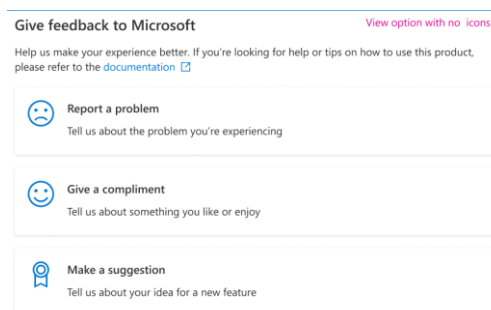


Figure 4 User-initiated Prototype Design

4.3 Reflection

Identifying the usability issues within the feedback system allowed us to prepare the product team with clarity for a design direction of the feedback mechanisms. Different products have different users, user goals, and workflows, therefore we don't presume that these findings yield design principles for feedback systems, per se, rather to highlight the importance of quick user studies to optimize the experience of giving feedback. In this round of designing a feedback mechanism, we opted not to test a prototype for the behavior-initiated feedback, because the implementation of it was too far away. When the development team is ready to add this kind of telemetry to the system we will create and test a prototype to help with the user experience of those designs.

5 PRODUCTION

5.1 Socialization

Although we communicated our plans and kept a handful of stakeholders in-the-loop throughout the design phase, we ramped up socialization efforts as we went into the implementation phase. We waited to bring in other parties until we had a proof of concept ready because creating, capturing, and reporting in-product feedback impacted many teams and partners, including product managers, designers, developers, researchers, and marketing across the six Azure Data services. To communicate and socialize the designs, research, and overall direction, we held over twenty-five briefing meetings, created two podcasts on the creation of the metrics and the feedback mechanisms, developed two videos reviewing different parts of the design, and gave two internal talks at Microsoft design and research events. While we planned and started the socialization process right away, we ramped up the engagement and storytelling with many stakeholders after the feedback system user study. Telling the story of our feedback system design process allowed us to gain buy-in with our initiative but also connect with other researchers. We further unpack socialization in the Lessons Learned section of this paper.

5.2 Implementation

We worked with our developers to implement the system-initiated feedback to capture both product and task level feedback with the final design of the feedback system completed in late May of 2020. This included small iterations to align with Microsoft-specific requirements (e.g. privacy statement placement). In November 2020, the designs went live. The in-product feedback data was stored in a SQL database, and we used Power BI to access the data and create interactive reports. These interactive reports visualized the data, including qualitative comments to allow our partners the ability to explore data based on their needs. We created a dashboard to visualize the results across multiple products and services, and an individual dashboard dedicated to each UX metric. As an ongoing practice we summarize the data and present it to stakeholders and leaders every three months to help them assess UX health.

While the data produced from the implementation of the designs helps our partners assess UX health, it also informs further research. Our team uses the data as part of our phased UX approach. As shown in Figure 5, our phased UX process includes Exploratory, Formative and Summative research. Summative initiatives are when focus research on evaluating a “complete” product to assess if we’re meeting our defined standards [3]. The in-product surveys fit into the summative phase. Our research team takes in-product data and investigates patterns and trends to generate new research questions and hypotheses for investigation in the Exploratory phase. Beyond showing where in-product surveys fit into our process, we also incorporate in-product findings into our research rollups. Research rollups synthesize multiple research efforts across our organization. Thus, the UX health data are also used to drive high-level product direction to extend the “shelf life” of research produced by our team [4].



Figure 5 Example of a research process and inclusion of In-product feedback

6 CONTRIBUTION

In this section, we describe the lessons we learned. We first reflect on the design process, specifically around where the current frameworks [18][22][27] fell short in guiding us and how we addressed those gaps. We then discuss lessons learned while implementing the design.

6.1 DESIGN LESSONS LEARNED

6.1.1 Patching together frameworks

The key learning from this UX health metrics design initiative was the lack of a framework that could support the full process of incorporating an in-product feedback system from start to finish. Instead, we found a series of frameworks, patched them together, and innovated to close the gaps. Based on this experience we saw the need for an end-to-end framework that guided practitioners through the entire process of metric creation that is both user-centered and customizable.

6.1.2 Marketing, business, & customer experience metrics

In Microsoft and other technology companies the need for business metrics has exploded. Every unit collects business metrics: marketing, customer experience, and business development units measure different and sometimes similar constructs at different altitudes. Understanding this metrics landscape allowed us to approach our UX Health metrics effort as complementary to others by telling the entire story of where the metrics came from, why they are different, what the plan is to act on the data and how they fit into current practices. Our careful approach yielded insights. First, capturing and

maintaining metrics can be expensive. Without having a persuasive and complete story around how getting experience metrics in-product to help us understand users better to ultimately improve our experiences on their behalf, can result in a lack of support from product teams, which can end in the design never getting built. Secondly, any UX metric initiative should not be approached with the objective of overhauling all department metrics but instead be approached with bringing UX metrics to the table. Changing culture around what data is captured and how it is measured is challenging. Often other department metrics have a process by which they design and consume their metrics, e.g., monthly business reviews, and attempting to disrupt established practices can be shortsighted. Therefore, we designed our strategy around filling a missing gap in the metrics landscape.

6.1.3 *Five stars*

While we closely followed HaTS [18] in terms of cadence for our system-initiated feedback, sampling both task and product level feedback, we deviated from using a 7-point scale recommended in HaTS for several reasons. First, other department metrics were captured on a 5-point scale, and we opted for consistency around number of scale points. Second, we decided to use five stars as an item response choice to align more with quick feedback systems [19] thereby both reducing the number of choices for users and making the experience familiar. Third, Lewis and Sauro's work around understanding star usage vs linear numeric rating scales supports the use of them for UX practitioners [16]. Overall, we echo Lewis and Sauro's message that while there is a growing list of scaling formats (e.g., stars, slider, numeric ratings, smileys) that can devolve into polarizing debates, rather than over-indexing on these discussions, which have trivial effect on measurement, we focused on being consistent with our scale choices [24].

6.2 PRODUCTION LESSONS LEARNED

6.2.1 *Democratizing data*

Providing a means for our data savvy colleagues to access, explore and visualize as they needed, was not only valuable but expected. Presenting the results in Power BI allowed all stakeholders the opportunity to “own” the data. They saw developments in real-time and drilled down into metrics and qualitative comments on their specific features within the product. Seeing task-level, usage and usability data brought additional context into their understanding of exactly where user experience problems were occurring in the wild and kicked off additional investigations to understand further who the people behind the data were and why were they having challenges, thus providing a user-centric point of view for our partners. This self-service dashboard allowed us to support a more user-centered, data-driven culture. The qualitative comments were also a key-win as they brought customer stories into feature and product level discussions. In retrospect we see how this aligns well to some of Lowdermilk and Hammontree's strategies for developing a data driven culture [17], as they call out bringing customer stories to everyday conversions and bringing transparency to data. Moreover, and perhaps the biggest lesson learned for us in this area, was the importance of telling the story of our in-product feedback. In conjunction with socialization, without the visualization piece to show the output of our design it would have been incomplete and theoretical. It allowed us to end the story in both a visual and realistic way. This way proved to be crucial in an engineering culture.

6.2.2 *Socializing the metrics*

As mentioned in the Socialization section, we spent time describing our process, communicating the designs, and getting buy-in and feedback from our stakeholders. Often these activities were to groups with different interests and objectives around needing to understand the design. For example, researchers were interested in our methodology as they were tackling similar issues with their organization, while engineers were focused on clarity around the pixel-level mockups and implementation, and partner organizations were looking to understand the relationship to their metrics. This socializing effort surprised us as we did not anticipate the time required. We learned the importance of having an upfront plan for socialization. This plan should include personalizing engagements based on stakeholder needs, which could impact delays around implementation. Further, the process of socialization does not stop after implementation. As new stakeholders join the product team, they bring different expectations and experiences around what metrics matter and what mechanisms

should be employed to capture these metrics. Because of this ongoing need to keep both existing and new people in-the-loop, we found revisiting the in-product feedback socialization plan to be ongoing and we are experimenting with ways to activate the ongoing stream of UX health metrics coming into our insight portfolio.

6.2.3 *Communicating insights & measuring impact*

Good UX research teams know the impact they are having, and great UX research teams know how to measure their impact by knowing how their research influences product user experiences and stakeholder decision making. When it came to assessing in-product feedback impact, we found that embedding insights from those data into our current process (Figure 5) worked well. It worked well because being transparent around how the in-product feedback fits into the entire UX process helped us communicate the story which in turn helped others understand the importance of the initiative, and how it fits into the larger UX Research picture. This process understanding and action taken both from our stakeholders and our research team on the output from the Power BI dashboards meant that we all had a new view on how our product experiences were performing. We know that the north star of impact for a UX research team is to show that our work results in better user experiences and having the in-product feedback, based on UX outcomes, gets us one step closer to reaching the north star. Finally, we also found that keeping track of our wins was important. Every time the findings or data from the dashboard was used by either us or other teams, we made notes to learn what was working so we can continue to improve the uptake of UX health data ultimately so that our stakeholders make better decisions to build products that make our users' lives better.

6.3 The BLUE FRAMEWORK

Throughout this case study, we discussed navigating real-world design and implementation challenges around building a user-centered metrics system. We also identified how existing metrics frameworks offer a piece of the puzzle, but lacked direction around the following:

- Describing both what UX metrics to capture and how to capture them.
- Clear actions a practitioner can take in helping create UX health metrics customized for their content.
- Guiding principles around the relationship between Business and UX metrics.
- Embedding into the context of real-world collaboration, socialization and the UX research process.

Based on these gaps, we're proposing the BLUE framework. This framework supports the design process for creating user-centered in-product feedback mechanisms by combining best-in-class frameworks (e.g. HaTS [18], UX outcomes [25], and The Customer-Driven Culture [16]). Created from our lessons learned the framework guides UX practitioners through the UX health metrics process giving them a roadmap for success. The pillars for the BLUE framework are as follows: Build UX Metrics, Leverage Mechanisms, Unpack Output, EMBED in practice. This framework provides enough scaffolding and flexibility needed for customizable application in industry.

6.3.1 ***Build***

Build UX metrics. As we learned through our case study, building metrics is more than thinking about outputs—it's about getting to UX outcomes, translating the outcomes into metrics and mapping those to metrics that matter within your product and organizational context. The aim of the Build pillar is to capture metrics that are tied to organizational and user goals that can be later translated into actionable insights for stakeholders to build products that make users' lives better. It is important to steer away from solely business focused metrics and ensure our suite of metrics includes user goals. The process we employed aligned to Spool's [26][27] direction which entails accessing business or organizations goals, viewing these goals through your UX lens to determine how they translate to their needs/problems/outcomes, and defining clear UX metrics to measure these. In addition, we see an important part of this pillar reflecting on not getting bogged down with selecting a debatable "right" metric or scale but focus on consistency of capture.

- Align with business goals.
- Investigate product UX outcomes, goals, and metrics.
- Focus on consistency of scales and ratings.

6.3.2 **Leverage**

Leverage proven mechanisms. We recommend starting with the HaTS framework [18] as a guide for the mechanism behaviors, specifically around utilizing random sampling [15] based on individual users instead of page product views. Design so that users can provide both feedback on high-level satisfaction questions, and feedback on more granular task-based questions. However, cadence and sampling choices can be impacted by external sources so adapting these might be necessary. In addition to the system-initiated feedback, it's also important to give users a voice by providing a user-initiated feedback option.

- Utilize HaTS and adapt as needed.
- Customize needs around sampling/cadence.
- Include a user-initiated feedback mechanism.

6.3.3 **Unlock**

Unlock the results. This pillar speaks to supporting stakeholders for how they currently or ideally would like to access the UX health data. After all, they are the users of the metrics and being user-centered here helps drive success. Customizing for the existing data culture might mean piping the data into an existing dashboard or creating a new one because the company's rules around data retention and privacy require it. The goals for unlocking should make the UX health metrics accessible so stakeholders can empathize with to the users' stories. Planning for the best way to serve up the results ahead of time is an important part to ongoing success and engagement with the data, contributing to a data-driven culture.

- Make the data accessible.
- Keep the user story central.
- Contribute to a data-driven culture.

6.3.4 **Embed**

Embed in practice. The final pillar requires embedding the UX health metrics into your UX research practice. As a UX researcher this might seem like an automatic step as UXR processes and tracking impact are top of mind. However, this pillar is also about including a focus on socialization because the story behind these data is an important piece to using the data and investigating the *why* behind some of the metrics. Furthermore, socializing is about transparency and inclusion. Showing how we can prioritize the user with these metrics and tracking their impact across the entire UXR process is a compelling story to share. Finally, with unique stakeholders, user goals and flows, there is no one-size fits all approach. Getting feedback on the design and presentation of UX health feedback systems helps drive the user perspective into places where business data are driving product decisions. Embedding business data with UX health data helps everyone make better decisions with the user in mind.

- Socialize for understanding so that everyone advocates for users.
- Combine with UXR process and track impact.
- Get feedback on the feedback.

7 CONCLUSION

In this paper we described our case study around the design of a user-centered in-product feedback initiative for large scale applications. We presented our initial process, the lessons learned along the way, and where gaps in current frameworks existed. From this process, learning, and gaps, we presented a framework outlining four pillars. These pillars include Building UX metrics, Leveraging proven mechanisms, Unlocking results, and Embedding in practice. We believe this framework provides the guidance necessary for research teams to lead the implementation of UX in-product feedback giving their users a stronger voice. We plan on iterating on this framework as appropriate to evolve our UX health metrics system and continue testing and evolving the BLUE framework along the way.

8 REFERENCES

- [1] Andreasen, S., Nielsen, V., Schröder, O., & Stage, J. (2007, April). What happened to remote usability testing? An empirical study of three methods. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1405-1414).
- [2] Bach, P., Create an Outside-in Organization, 2018, Microsoft XC Research, <https://www.microsoft.com/en-us/research/group/customer-insights-research/articles/create-an-outside-in-organization/>
- [3] Baxter, K., Courage, C., and Caine, K., 2015, *A Practical Guide to User Research Methods, Understanding Your Users*, Morgan Kaufmann.
- [4] Burghardt, J., Extending insight 'shelf life' to get more value from research in product planning, *Integrating Research Medium*, May 3rd, 2021. <https://medium.com/integrating-research/extending-insight-shelf-life-to-get-more-value-from-research-in-product-planning-c2fb7af5d42a>.
- [5] Coelho, P., and Esteves, S., 2007. "The Choice between a Fivepoint and a Ten-point Scale in the Framework of Customer Satisfaction Measurement". *International Journal of Market Research*. 49 (3): 313–339. doi:10.1177/147078530704900305. ISSN 1470-7853. S2CID 166325179.
- [6] Davis, F. D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, 13 (3): 319–340, doi:10.2307/249008, JSTOR 249008
- [7] Doerr, John, *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs*, Penguin Random House LLC, New York, NY. 2018.
- [8] Eriksson, J., *KPIs: Essential Framework, See-Think-Care-Do*, 2015, Google Framework, <https://www.thinkwithgoogle.com/intl/en-145/perspectives/global-articles/kpis-essential-framework/>
- [9] Farris, P., Bendle, N., Pfeifer, P., Reibstein, D., 2010. *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance*. Upper Saddle River, New Jersey: Pearson Education, Inc. ISBN 0-13-705829-2.
- [10] Fink, A., *The Survey Handbook, Second Edition*, Sage Publishing, October 2003.
- [11] Grove, A., *High Output Management*. Random House. Objectives, and key results goals framework, (1983)
- [12] Hick, W. E. (1952). "On the rate of gain of information". *Quarterly Journal of Experimental Psychology*. 4 (1):11–26.
- [13] Hillman, S., Forghani, A., Pang, C., Neustaeder, C., Judge, T., *Conducting interviews with remote participants, Studying and Designing Technology for Domestic Life*. Pp. 11-53.
- [14] Jansen, H., *The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods*, *Forum Qualitative Research*, Volume 11, No.2, Art. 11 – May 2010.
- [15] Larson, R., and Csikszentmihalyi, M., *The experience sampling method*. In H. T. Reis, editor, *Naturalistic Approaches to Studying Social Interaction*, volume 15 of *New Directions for Methodology of Social and Behavioral Science*, pages 41–56. Jossey-Bass, San Francisco, CA, USA, 1983.
- [16] Lewis J., and Sauro, J., *Are Star ratings better than numbered scales?*, *MeasuringU*, July 22, 2020, *Are Star Ratings Better Than Numbered Scales? – MeasuringU*
- [17] Lowdermik, T., and Hammontree, M., (2020) *The Customer-Driven Culture: A Microsoft Story, Six Proven Strategies to Hack Your Culture & Develop a Learning-Focused Organization*. O'Reilly.
- [18] Muller, H., and Sedley A., *HaTS (2014)*, Google, Case Study. <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43221.pdf>
- [19] Nielsen, J. 1992. Finding usability problems through heuristic evaluation. *Proceedings ACM CHI'92 Conference (Monterey, CA, May 3-7)*, 373-380.
- [20] Nielsen, J., 1994. *Usability Engineering*. Morgan Kaufmann, San Francisco, CA.
- [21] Reichheld, F., *The One Number You need to Grow*, *Harvard Business Review*, October, 2003.
- [22] Rodden, K., Hutchinson, H., and Fu, X., *HEART, Measuring the user experience on a large scale: user-centered metrics for web applications (2010)*, Google Framework, <https://dl.acm.org/doi/10.1145/1753326.1753687>
- [23] Sadowski C., Storey MA., and Feldt R. (2019) *A Software Development Productivity Framework*. In: Sadowski C., Zimmermann T. (eds) *Rethinking Productivity in Software Engineering*. Apress, Berkeley, CA.

https://doi.org/10.1007/978-1-4842-4221-6_5

- [24] Sauro, J., and Lewis, J, Comparing Fully vs Partially Labeled Five-and-Seven Point Scales. January 22, 2020, Comparing Fully vs. Partially Labeled Five- and Seven-Point Scales – MeasuringU
- [25] Seiden, J., Outcomes over output: Why customer behavior is the key metric for business success. (2019). Sense & Respond Press.
- [26] Spool, J., Shifting our team goals to be UX outcomes, January 15th, 2020, Shifting Our Team Goals to be UX Outcomes — UX Articles by UIE
- [27] Spool, J., Why UX Outcomes Make Better Goals Than Business Outcomes, May 22, 2020, Why UX Outcomes Make Better Goals Than Business Outcomes — UX Articles by UIE.
- [28] Talboy, A., McVey, K., and Bach, P., Pillars of Customer Success. Medium: UXR at Microsoft. August 19, 2021.