Humanizing Quant and Scaling Qual to Drive Decision-Making

LAUREN MORRIS
Amazon Prime Video

REBECCA GATI
Amazon Prime Video

The Amazon Prime Video User Experience (UX) Research team endeavored to balance qualitative and quantitative insights and translate them into the currency that drives the business, specifically customer engagement, to improve decision-making. Researchers conducted foundational qualitative research to uncover what matters most to Prime Video customers, translated resulting insights into a set of durable, measurable customer outcomes, and developed a global, longitudinal online survey program that validated the importance and perception of these outcomes at scale. Researchers then systematically linked customers' attitudinal survey results to their usage patterns and overall satisfaction with the service. The resulting data showed how investing in improving a customer outcome is likely to increase service engagement, thus closing the loop between insights and business metrics for the first time. Prime Video executive leadership has not only embraced this integrated qualitative-quantitative system, but now uses it to prioritize projects on behalf of the customer. UX Research has shifted Prime Video culture from relying on analytics data alone and toward seeking comprehensive evidence to drive strategic decision-making.

INTRODUCTION

In the 2016 letter to Amazon shareholders, Jeff Bezos stated, "Good inventors and designers *deeply* understand their customers. … They study and understand many anecdotes rather than only the averages you'll find on surveys. … A remarkable customer experience starts with heart, intuition, curiosity, play, guts, taste." The concept of "anecdotal" evidence, or qualitative data, is embedded in Amazon culture—it is mentioned in everything from the company's guiding principles to training documents. However, when the company's initial streaming service launched in 2008, the team primarily used quantitative behavioral analytics (i.e. usage data) in decision-making and often used A/B testing for risk management before launch. After features shipped, the team used customer service contacts and app store reviews as a barometer for customer satisfaction.

As Prime Video grew, the team recognized the importance of expanding their understanding of the customer and subsequently invested more deeply in qualitative usability research and quantitative attitudinal research, primarily surveys, within the organization. Ultimately, UX Design hoped to evaluate design work more often and provide a method of measuring customer satisfaction in a way that could inform prioritization. However, the business did not have the expectation that research should provide strategic recommendations at the front end of the product development process. The greater organization recognized qualitative data included customer service contacts, app store reviews, and survey responses, and were intrigued by incoming usability test findings; however, they did not fully leverage results from these sources and questioned the validity of

small sample size data, particularly when the findings did not match commonly held organizational beliefs. There was still a comfort and familiarity with extremely large sample sizes—e.g., A/B testing with millions of customers—that prevented teams from acting on the UX Research team's qualitative insights:

> "Amazon overall—not just [Prime] Video—has been really good at the intersection of behavioral and quantitative [methods]. This is something we actually pioneered in the industry—the use of extensive A/B tests to make determinations. … It has been established as a scientific method with rigor and we have relied on that approach. That has been pretty well established for a really long time." – A Director of Engineering, Amazon Prime Video

Still, UX leadership saw the potential in investing in research to provide evidence that UX design quality and customers satisfaction are equally as important to address as moving key business metrics. At the time, product decisions *were* made on behalf of the customer, but from the perspective of increasing the number of customers and broadening content selection.

> "If you would have asked the folks at the time, 'How are you applying customer obsession?' They likely would have answered 'Well, for a customer that does not have access to our service on the device that they prefer, the best thing we can do for that customer is give them access to our service. For a customer who is in a market which we do not service today, the best thing we can do for that customer is exist in that market and give them access to our service'… So, those were the levers that seemed to have the largest customer impact and largest business impact at the time." – Principal Product Manager, Amazon Prime Video

UX leadership was hoping to inform product recommendations in a more holistic manner; the team believed that they could grow the impact of research at Prime Video and influence product stakeholders to focus on customer experience (CX) improvements with a diversely-skilled research group. The team hired two additional senior researchers—experts in qualitative user research and survey science—and they began gathering focused insights in their respective disciplines.

As the team grew, UX Researchers continued to collect and add to the existing wealth of independent insights and would occasionally come together to employ more traditional triangulation methods. They would, for example, examine the findings of a qualitative study and cross-reference with a similar quantitative study to validate results with the hopes of increasing confidence in results found in both studies. Researchers included inputs from other teams such as business analytics and market research to tell a more complete story. Converging inputs helped provide some recommendations to the business, but these attempts ultimately fell short to impact strategic business decisions. Studies were not always designed to have analogous objectives and operated on different roadmap timelines with different stakeholders; thus, the piecemeal nature made it difficult to compare results to create a comprehensive recommendation. For example, when examining how often customers clicked on certain elements on the website, analytics data showed heavy usage, leading stakeholders to believe this element was effective. There were also app store reviews that described a difficult UX and pointed to certain negative experiences but lacked detailed explanations. Additionally, a later qualitative study provided evidence that there was customer confusion around these elements, contributing to this heavy usage. Because these

studies occurred on different timelines and were owned by different teams, the organization could not take clear action.

These triangulation attempts helped develop incrementally better solutions that statistically increased customers' usage of the service, but it was unknown if these solutions had a broader impact on the holistic experience or customers' overall perception of the service. The UX Research team was armed with qualitative insights about customer needs and pain points with the potential to spur exciting new ideas and concepts, but often asked themselves: "how could we quantify, more comprehensively measure, and prioritize what matters most to customers?" The research team recognized that data triangulation was not enough and sought out partnership with the analytics team. Together they aimed to more directly link all the inputs available, not only to gain a better understanding of Prime Video customers, but to help the business understand how to measure and act on insights and to influence strategic business investments that would drive innovation.

To develop a systematic approach for measuring and scaling insights, the team realized they needed to be more integrated in their research approach, enabling direct mixed-method research across qualitative and quantitative methods. The team sought to learn about each other's disciplines, design studies collaboratively, conduct fieldwork jointly, review data simultaneously, and leverage partnerships that each discipline unlocks. Direct participation in fieldwork provided crucial nuance that helped research scientists interpret quantitative data. Similarly, qualitative researchers leveraged quantitative insights to understand which patterns generalize to customers and scaled the impact of their findings, mitigating uncertainty of this method. By integrating qualitative and quantitative research disciplines and dedicating the roadmap to mixed-method research, the team was set up to build an integrated system that capitalized on each other's expertise. With a new team approach and new partners on board, they began the journey to answer what matters most to customers and which investments to prioritize. They quickly realized they needed a new, directly integrated method to do it.

METHODOLOGY

Determining What Matters Most to Customers

To determine what matters most to customers, the researchers aggregated existing knowledge from three years of generative and evaluative qualitative research, which included data from in-home interviews, longitudinal diary studies, usability studies, and customer service contacts. They mined for pain points, explicit and latent needs related to video services generally, and brand attitudes and perceptions. There were clear, consistent patterns across these various sources—enough for the team to begin confidently hypothesizing what matters most to current and potential Prime Video customers.

Initially, the team developed a framework that illustrated the hierarchy of customer needs for a video service, inspired by Maslow's Hierarchy of Needs theory (1954). Maslow's theory states that humans are governed by a hierarchy of needs; once their essential needs are met, they are motivated to pursue more advanced needs, but are unable to value those advanced needs until essential needs are met. The team found through their existing research that a similar paradigm existed for customers and had a strong hypothesis for which needs were essential and which were advanced. Maslow's theory is illustrated as a pyramid, which the team also leveraged for their framework.

In this initial framework, the base of the pyramid was named "Connection" which housed the customer need of "Awareness"; in other words, customers need to know the service exists. Once a customer is aware of the service, they can move through the next levels of the pyramid, as long as their needs at every level are adequately met. The top of the pyramid was named "Affinity," where the video service solves for needs that have not yet surfaced for customers—customers feel taken care of. At this phase, customers become loyal to the service and share their love for it with others. The goal is for the customer to reach "Affinity." For example, when customers become loyal and enthusiastically recommend Prime Video to others, the business can redirect resources toward further delighting customers.
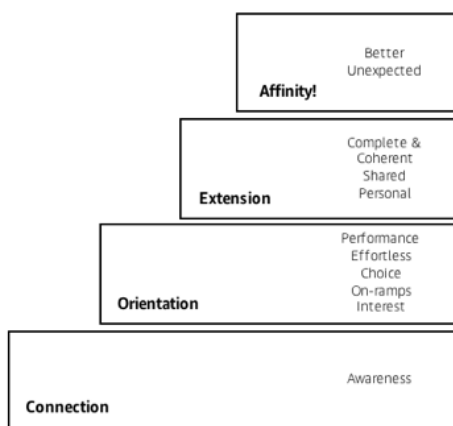

Figure 1. Prime Video's Hierarchy of Customer Needs Framework

The goal of this framework was to comprehensively and systematically capture and surface existing customer pain points, leveraging years of qualitative insights. The team translated these pain points into themes of customer needs, enabling teams to understand the breadth of customer needs that exist and consider this breadth when developing solutions in their domain. The team asserted the hierarchy of these themes to help the business prioritize investments, focusing on foundational customer needs first.

> "It was obvious that you would start at the bottom of the hierarchy. You wouldn't invest in solving some affinity problem if you couldn't even use the service, which was down in the basic level." — Director of UX Design, Amazon Prime Video

After collecting feedback internally, the research team uncovered three challenges with the framework: 1) utility—stakeholders, including members of the UX design team, did not understand how to leverage the framework because it did not provide specific recommendations, 2) theoretical basis—there was no quantitative validation of this hierarchy and Prime Video had not historically incorporated psychological theory-based frameworks into business decisions, and 3) ambiguity—the themes in the framework were scoped at a high level and thus open to interpretation, therefore stakeholders did not share a collective understanding of the themes. This limited the ability to measure whether or not customer needs were being met at each level.

"While the [pyramid framework] was interesting for creating a way of thinking about and conceptually prioritizing customer problems, it wasn't necessarily useful in terms of changing behavior – actually getting things done. We had the big band of things that were foundational, but which one of those was the most important thing to go focus on? We didn't have that yet." — Director of UX Design, Amazon Prime Video

Hoping to address these limitations, the team explored other effective, actionable frameworks that capture customer needs. A senior team member had applied and translated the "jobs-to-be-done" theory in multiple domains and recognized this could be a good fit for Prime Video culture. This theory was first articulated by Clayton Christensen (2003, 2016), and built upon by Anthony Ulwick (2005) and others. This theory asserts that people, in effect, "hire" products to do important jobs, and that they will "fire" a product if an alternative solution does these jobs better. The key to delivering successful products is identifying the jobs that are most important to customers and developing innovative solutions that do these jobs better than alternatives. The team set out to develop a more systematic approach that would allow them to not only build a framework of what matters most to customers with their insights, but to also quantify the magnitude and generalizability of these insights.

The research team began by engaging in multi-day working sessions over several weeks to develop a new approach. They built off of their existing Hierarchy of Customer Needs framework and applied the "jobs-to-be-done" theory lens, while also recognizing there were certain attributes of a framework that would be most effective (e.g., being able to appropriately scope customer needs, measure validity and generalizability of customer needs at scale, and ensure the future capability to provide recommendations). The result of these working sessions was the development of a set of "Customer Experience Outcomes" (CXOs) to articulate customer needs in an actionable, measurable way. A CXO captures a discrete customer need, derived from research insights, as 1) a "job" a product must do to address the need and 2) the criteria customers use to judge how well a product does the job. The anatomy of these CXOs, adapted from Ulwick (2005), included a *job statement* (e.g., "Protect my family from unwanted content") and a *success criterion* (e.g., "…with more confidence"), which enabled the team to measure how well the service does the job for the customer—both quantitatively and qualitatively.
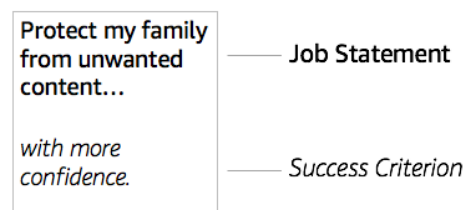

Figure 2. Anatomy of a Customer Experience Outcome (CXO)

By the end of this exercise, the research team had identified about 40 unique CXOs, loosely related to each section of the user journey that the organization was familiar with. The CXOs were all scoped at a granular level; the job statements contained only one verb that represented the smallest meaningful, but actionable, unit of customer value. Each one was measurable; they provided the basis for measuring the experience holistically. They

Humanizing Quant and Scaling Qual – Morris and Gati

intentionally had no mention of implementation or a solution, as they were meant to describe durable customer problems.
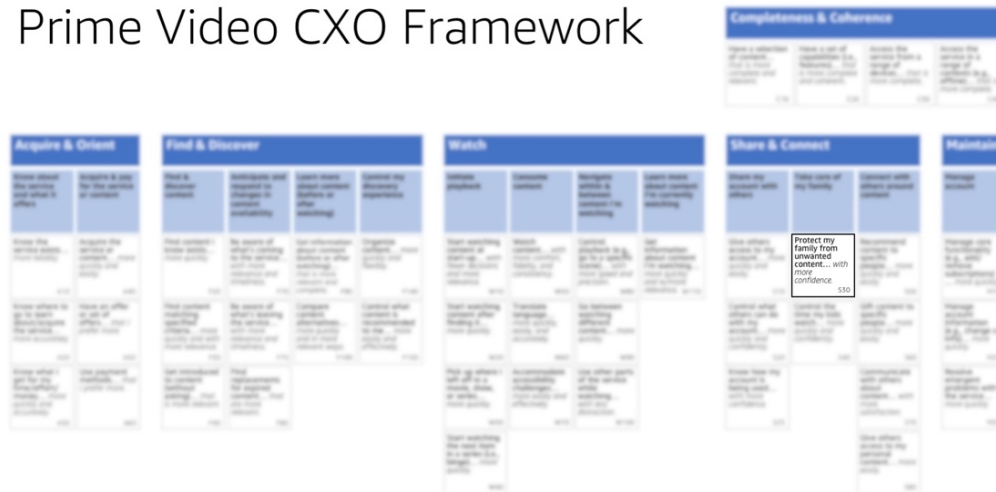
Prime Video CXO Framework

Figure 3. Prime Video's Customer Experience Outcome Framework

With the evolution of this framework, stakeholders were able to understand how to leverage it to make more customer-informed decisions and began to assess if their product features or designs did or did not map to these customer needs.

> "What started to change, in my mind, was when we started turning anecdotes, insights, survey results, whatever, into frameworks…it certainly resonated with me because I'm a strong believer in mechanisms. Mechanisms have an enduring power. They don't depend solely on one person's ability to evangelize them. If they make sense, they enable people to pick up those mechanisms and run with them, and essentially scale the impact of that mechanism." — Principal Product Manager, Amazon Prime Video

Validating What Matters Most to Customers

Although confident in the aggregated qualitative insights, the team believed it was necessary to validate these CXOs with customers and identify any missing pieces. The qualitative researchers designed a field study to validate the CXOs, using a semi-structured interview approach. Participants in two demographically-disparate cities were recruited for 2.5 hour-long in-home interviews. CXOs were written on cards and altered slightly to reflect customer-facing language rather than more technical job statements, although the team intentionally made sure the internal CXO framework language did not veer too far from an actual customer's voice. For example, the CXO "Protect my family from unwanted content…with more confidence" was changed to "Assure my kids have a safe and positive experience" in the interview. Much of the discussion focused on deeply understanding why certain CXOs were more important to participants than others.

Following a brief, unstructured interview consisting of a walk-through of the participant's home and their devices to ground and contextualize the discussion, participants

were presented with the CXO cards. They were encouraged to read the CXOs, confirm whether or not they were relevant, and even suggested ways to clarify the CXO phrasing. They were also provided with blank cards to add any CXOs they judged to be missing. This helped ensure that the CXO framework comprehensively represented what mattered most to customers.

Participants were then asked to sort the CXOs into three roughly-equal groups: "most important," "next-most important," and "least important." It was key for the groups to be roughly equal because it forced the participant to make trade-offs and think critically about the most important aspects of a video service. From their "most important" group, the participant was asked to choose her or his top five CXOs. Again, the researchers observed the trade-offs that the participant made. While the goal of this exercise was to understand *why* certain CXOs were more important than others, the team's objective was to identify patterns across all sixteen participants' top five CXOs during analysis.

After the sorting exercises, participants were asked to comment on their satisfaction with all of their video services for their top five CXOs. A hand-drawn spectrum was provided to the participants for this exercise. The ends of the spectrum were defined by a question that participants had answered at the beginning of the interview—"Which brand/service/company have you had an amazing experience with, and which brand/service/company have you had an awful experience with?" Examples of amazing experiences included Taylor Guitars, Coach purses, and Nike running products. A notably high number of participants mentioned companies with poor customer service as their most awful experiences. The team did not limit participants' thinking to just video streaming companies, as the goal was to encourage participants to reflect on their emotional associations with products and services.

Using this spectrum, the participant set their own bar for an exceptional experience and assessed both Prime Video and other video services according to this bar. For example, if one of their top five CXOs was "Protect my family from unwanted content," they were asked "How good of a job does [service] do at assuring you that your kids are having a safe and positive experience?" The participant would then place each service somewhere on the spectrum. This process often elicited a visceral reaction from participants and helped the research team better understand how Prime Video compares to competitors on the things that matter most to customers.

After the field study, the entire team engaged in data synthesis. Every member of the team—including research scientists—watched recordings of multiple sessions and coded the interviews with the CXOs in mind. During synthesis, one member of the team was responsible for creating video clips that aligned to early themes and patterns as the team debriefed and discussed. Ultimately, synthesis of this research led the team to formalize a set of key themes, each of those tying back to specific CXOs. Now the team could illustrate the linkage between a qualitative insight and its corresponding measurable, actionable CXO.
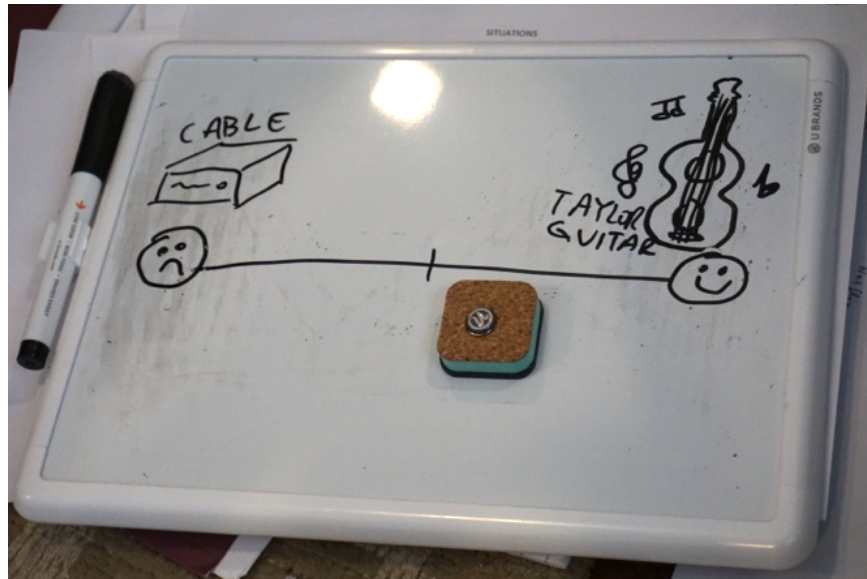
Figure 4. Example of a participant-defined experience spectrum. In this example, the participant drew their "worst experience," illustrated on the left, and their "best experience," illustrated on the right.

Influencing the Organization

The field study presented the first opportunity to fully integrate the qualitative and quantitative arms of the UX research team. The research scientists joined qualitative researchers in field, operating camera equipment and taking notes. Through their participation, they gained perspective that later directly influenced both quantitative study design and data interpretation. Additionally, the in-home interviews were a great opportunity to evangelize the value of applying ethnographic methods. The team invited senior leadership from a variety of disciplines into the field as notetakers, helping them build empathy with customers and gain a better understanding of field study as a valid research method. One senior leader later reflected on his appreciation of the experience, as he was reminded that customers live a completely different life than him. By participating, senior leaders were also reminded that customers' familiarity with technology and awareness of features are generally lower than the average Prime Video employee and are ultimately rooted in the context of their lives. At the conclusion of this study, leaders strongly encouraged their staff to become more involved in qualitative research opportunities.

Figure 5. A Director of Engineering taking notes in a participant's home

The team believed it was extremely important to curate and share high-production videos from this fieldwork to build empathy for customers and help stakeholders to challenge their assumptions about the customer. With Amazon being a heavy written-document culture, there was concern that sharing videos to stakeholders would not be effective or understandable. The result was a combination of the two approaches: 1) an immersive insights report, separated into sections for each of the key themes, with the integration of supplemental quantitative data to validate the insight from the research, and 2) five-minute customer videos correlating to the key themes. The team scheduled recurring customer insights immersion sessions where stakeholders from various teams were invited to watch these customer videos, with the UX Researchers in attendance to provide any necessary additional context. The researchers believed that these immersion sessions, specifically hearing directly from customers, would be a refreshing change for stakeholder attendees. It was also an opportunity to convey customer sentiment and emotion, which was a more effective tactic than having the researchers relay anecdotes from the field themselves. Session attendees were able to ask the research team questions about ethnographic methods, the study design, or the insights. The team followed up by sending the insights report to all session attendees, adhering to the Amazon written-document standard. The videos were embedded in the insights report, giving viewers a multi-faceted way to better understand their customers' needs and opportunities for innovation.

The outcome of this approach was the beginning of a culture shift. Leadership was intrigued and began requesting to attend future field studies. Key stakeholders left these workshops inspired by the opportunities to innovate. Teams that had never engaged with the research team in the past were now reaching out about how to work together. One of the researchers was even invited to speak at a Prime Video "all-hands" meeting, exposing the discipline of UX Research to thousands of employees. While some uncertainty of how to use qualitative data still existed, the team felt that ethnographic methods were gaining momentum.

Humanizing Quant and Scaling Qual – Morris and Gati

> "It had the additional benefit for people like me. To put me face-to-face with a customer has a certain impact value that you don't really figure out looking at any number of metrics in a conference room. You just don't get that same feel for things. … It's a cool area that we work on, in terms of how directly it impacts consumers, and this is a very in-your-face reminder of that in a helpful way." — A Director of Engineering, Amazon Prime Video

The team acknowledged that in order to influence business strategy and organizational priorities, qualitative data must be validated at scale and linked to the currency that drives the business—customer engagement.

Scaling Qualitative Insights

After completing the qualitative study and working sessions, the research team needed to determine if findings generalized to the larger customer base. Historically, this is where the quantitative and qualitative arms of the research team would diverge and later triangulate data. To keep the research tightly aligned, they used a common currency—the CXOs. Using field study insights and collaborating closely with the qualitative researchers, the research scientists designed a cross-sectional, longitudinal survey program and disseminated the survey to major markets around the world. Its aim was to measure how well Prime Video does on a subset of CXOs, selected based on the qualitative ranking exercises, aggregated insights, and the project priorities for the year.

Developing this survey program was challenging, but necessary to scale qualitative findings. Survey research was still considered qualitative, low-sample size research by some stakeholders, as they were most familiar with reviewing usage data and multivariate analyses using several million "respondents" versus several thousand used in survey research. In response, research scientists conducted multiple educational seminars providing details around the scientific process of survey design, sampling, and analysis, which helped quell apprehension and laid the foundation to use this method for decision-making. It also served as a mechanism to contextualize, and ultimately humanize, usage data the business relied heavily on by providing a window into the attitudes and perceptions of those using Prime Video.

Some stakeholders were also convinced there was "more art than science" in designing surveys, which prompted the team to qualitatively test the survey with customers to ensure questions were interpreted as intended. This went beyond standard quality assurance measures the quantitative team typically used. They dug into both how questions were interpreted but also why participants selected their responses. This allowed stakeholders to hear directly from customers about how they were interpreting the questions and showed stakeholders the survey instrument was reliable, which ultimately increased their confidence. This was another instance of the qualitative and quantitative research experts working in tandem to help quell concerns about methodological rigor from stakeholders as well as generate empathy by hearing directly from customers.
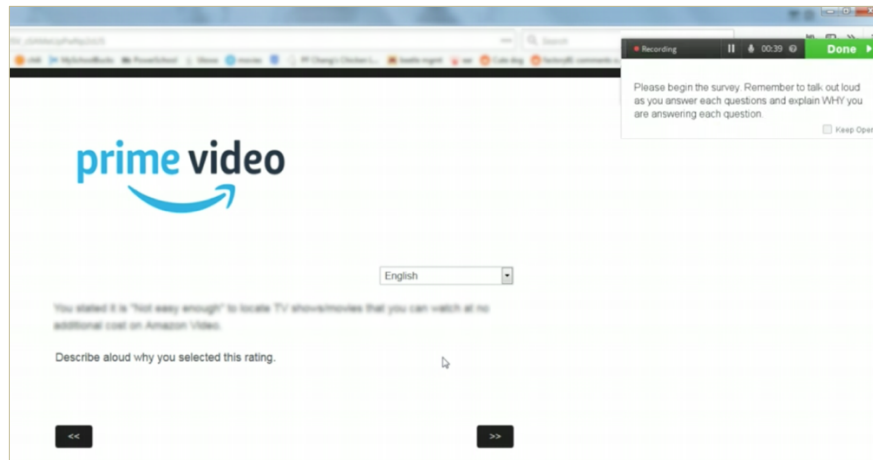
Figure 6. Screenshot of a participant using usertesting.com, the platform the team used to collect audio and video feedback from customers taking the survey. Participants were asked to talk aloud as they completed the survey. This helped the team understand how customers interpreted the survey questions and why they chose their responses.

There was also debate about which CXOs to prioritize in the survey instrument. There are always competing priorities on the product roadmap and product stakeholders wanted to ensure that their planned feature shipments would be covered. However, to maintain survey length best practices, informed by Dillman (2007), the research team had to assert that only around twenty-five CXOs would be covered. Although this left some stakeholders resistant to adoption, it forced teams to recognize that researchers would prioritize what to measure based on evidence of what mattered to customers instead of proposed product solutions.

The team also experimented with two waves of data before publishing to the organization, the first using a standard bipolar 5-point satisfaction scale (very satisfied to very dissatisfied) (Singleton & Straits, 2010) to assess CXO satisfaction and the second using a less traditional, categorical 3-point scale to assess performance of Prime Video on the same CXOs (very good, somewhat good, not good enough). Because the team piloted the program with both scales, research scientists were able to illustrate the similarity in results, ensure the robustness of the scale, and showcase the primary benefit of using the 3-point scale: persuasion. By requiring respondents to rate more extreme categories, it was much easier to interpret and determine whether the business should invest in an aspect of Prime Video when, for example, 20% customers stated it was "not good enough" vs. 4 on a satisfaction scale. The team wanted to reduce debate during results interpretation and encourage the business to act. Having polarizing categorical responses helped achieve this goal.

Through navigating these challenges, the research team showcased their expertise, helped stakeholders understand their processes, and ultimately elevated the opportunity to use survey data in decision-making. By bringing stakeholders in early and allowing them to challenge the teams' methods, adoption and evangelism was easier and not the sole responsibility of the research team.

After the data was available, the team was set up to integrate the survey dataset with behavioral usage data to help determine the relationship between satisfaction and

Humanizing Quant and Scaling Qual – Morris and Gati

engagement. This had been an ongoing topic of conversation amongst the research team and analytics partners. Both teams believed that this would be the most direct way to measure how customers were feeling about the service and if that had an impact on what they were doing on the service. Because the business focused so heavily on driving engagement, understanding the relationship of satisfaction to engagement would also be the most effective way to illustrate the importance of measuring and driving satisfaction, and that the two combined could determine which investments the business should make. However, complex new data infrastructure had to be built in Prime Video to enable this investigation and create a semi-automated process for ongoing measurement.

Creating New Systems to Support Integrated Data

In parallel to the survey program development, the research scientists knew they would need to work with partners outside the research team to develop a new system that could support both survey data and behavioral usage data. They would need to secure several months of technical resources which would overlap with the development and launch of major product releases. The research team had already garnered interest from the analytics team and engaged with them to discuss the details of leveraging each other's expertise to develop an integrated data system; they realized they would still needed engineering resources to develop the infrastructure. In conversation with engineering stakeholders, the initial response was that while the project was interesting, it would have to compete against other established priorities for the year. They continued to pitch the project, laying out their vision and the potential for the result. An engineering executive agreed to resource the project. With this support, the research scientists, analysts, and engineers were set to tackle this infrastructure challenge.

Since customer usage data was already being tracked for other reporting use cases, it made sense to try to ingest survey data into existing data warehouse structures. However, survey data is inherently flexible, semantic, and variable over time, while database structures are inherently inflexible, requiring rigid rules and consistency over time. The project team brainstormed for weeks and were ultimately able to create an adaptable system. The data engineers were able to extract and manipulate the new data, ultimately surfacing it in dashboards for the team to more easily work with. Connecting these data types was a huge victory, as this connection had never been done before in Prime Video. It enabled the team, for the first time, to begin investigating how the perception of Prime Video affected customers' usage of the service.

> "Before working with UX Research, our team would analyze data, hypothesize about root causes, and use our best judgment to describe results … but we never really knew if our judgment was correct. When we added data about what customers said, we no longer had to guess … we knew why there were differences in our datasets. This fundamentally changed the way we analyze and value data" — Senior Analytics Manager, Amazon Prime Video

Bringing Data Together

Once the data was connected, the team created a glanceable scorecard that illuminated areas that the business was doing well in and other areas that needed additional investment. There was so much new data that the team wanted to ensure stakeholders could quickly understand

what mattered most to customers. To start, they displayed "live" satisfaction scores (versus a traditional survey report) housed within the CXO framework. This reinforced the CXO framework as a mechanism to highlight comprehensive customer needs, now backed by quantitative data. This dashboard was also built to be interactive, including features like survey question and response distribution details on hover, clickable CXO boxes that revealed the underlying data, and calculations that automatically coded the satisfaction red to green based on the responses from the latest survey. This was a departure from the standard published dashboards that primarily included rows of data, filters, and occasional annotations.
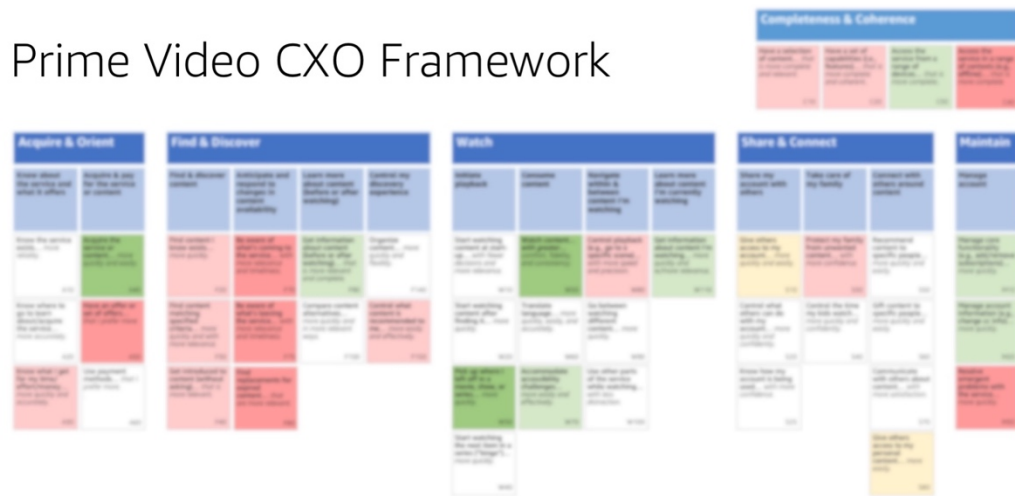

Figure 7. Prime Video Customer Experience Outcome Framework dashboard, including sample satisfaction ratings.

Because survey responses were linked to usage data, the team began investigating responses alongside corresponding engagement data, ultimately finding that customers who were unhappy with the service generally used the service less. They also pinpointed which aspect of the service, or which CXOs, correlated mostly strongly with engagement. It was the first time Prime Video had quantitatively-backed evidence that satisfaction was positively correlated with engagement, and the first time patterns of satisfaction and engagement could be examined together. The team could start to prioritize customer needs based on these correlations. It set the foundational understanding that satisfaction is equally important to measure and use in decision-making, and that integrating data could provide a framework for prioritization.

In order to make best use of the connected dataset, the team built several calculations to determine the "opportunity to influence overall service satisfaction" and the "opportunity to influence engagement." These calculations quantified the gap between satisfied customers and unsatisfied customers, giving each CXO a score. The result was an ordered list of CXOs based on 1) satisfaction with that CXO, 2) overall satisfaction with the service, and 3) overall engagement with the service, enabling ranking of CXOs across any of these dimensions.

To reduce complexity, the team decided to categorize CXOs and assign recommended actions based on uncovered patterns so that partners could more easily consume and act on

Humanizing Quant and Scaling Qual – Morris and Gati

this data. For example, an area that had low CXO satisfaction, a high correlation with overall service satisfaction, and a high correlation with service usage would be an area to *prioritize*. This pattern indicates that customers are generally unhappy with Prime Video performance on this CXO, have a relatively negative perception of the service, and those same customers use the service less. Alternatively, some CXOs had high satisfaction scores and high correlations with service satisfaction and usage. This pattern indicates that customers are happy with this aspect of the service, but if that CXO satisfaction score drops it will likely result in a decrease in service usage and overall satisfaction. This is an area to *fortify*.

Perhaps one of the more interesting patterns that could not have been uncovered examining satisfaction and engagement data independently was when a CXO had low satisfaction ratings and low correlations with service usage and overall satisfaction. This indicates that customers are unhappy with an area of the service, but it did not appear to influence their usage of the service. Puzzled, the research scientists consulted the qualitative researchers who had a bank of knowledge from the CXO validation fieldwork that included customers commenting on these particular aspects of Prime Video, and how no other service did a better job on this CXO. Customers had nowhere else to turn for a better alternative, so they were going to continue to use Prime Video (until something better came along). Keeping the two arms of the UX Research team in lockstep during development enabled the research scientists to interpret their data with fewer assumptions and allowed the qualitative researchers to use the aligned quantitative evidence to boost the impact of their findings. CXOs with this data pattern became known as *prospects*, or areas of opportunity for Prime Video to tackle to better compete and better serve unmet customer needs. These three categories became the primary recommended actions the team coined and supplied to the business as areas to invest in. With these recommendations automatically categorized and available in interactive dashboards, executives' prioritization questions were answered, truly, at a glance.

Challenges Integrating Data

Several challenges arose while developing this tool, many of which are still being iterated on. It was, and continues to be, difficult to distill multiple complex calculations into a glanceable scorecard that is useful to a range of stakeholders, including executive leadership, product managers, designers, marketers, and researchers. Most organizations request this type of scorecard to determine the health of their businesses; however, few are able to achieve a deliverable that is both comprehensive and usable. The quantitative team, guided by the qualitative team, usability tested this dashboard system with several stakeholders. They iterated on the tool's design multiple times, balancing methodological transparency with glanceability. Another challenge the team faced was training the organization on this tool. It required a dedicated communication strategy, with workshops and insights share-outs for more than 25 teams as well as the entire organization. After months of training, the organization now regularly uses the tool and its insights across a range of projects and job functions. It has also garnered interest from other Amazon businesses. Research scientists regularly consult with researchers, marketers, and product managers on how to build a similar system relevant to their respective businesses.

IMPACT

Since this integrated qualitative-quantitative program has come online, it has become a key input into the Prime Video decision-making process at all levels of the organization. The Prime Video roadmap is closely examined to understand which CXOs a given project influences. The quantity and priority of CXOs influenced determines how much support a project receives. Strategic goals have been set to improve satisfaction with particular CXOs, which are regularly reviewed with leadership. The CXO dashboard has the some of the highest views of all dashboards across the organization. Having CXO goals for each project helps all team members—from design, to product, to engineering—understand and stay aligned on the customer needs being addressed.

In addition to providing consistent updates to executive leadership, the UX Research team has concentrated on helping stakeholders think more holistically and more long-term about the customer experience—focusing on both satisfaction and engagement with the service. The team closely partners with product stakeholders to help them determine how their area of focus or specific feature compares to all other aspects of the service that customers care about. Engagement reporting meetings now include dedicated time to highlight satisfaction metrics because of the recognized relationship between business and attitudinal metrics. Additionally, they help product teams realize how introducing a new feature or changing one aspect of the service can come at the expense of negatively impacting a CXO.

> "[Skepticism from others was mitigated by] a) the apparent level of rigor that seems to have gone into the articulation of that framework and b) it's undeniable that linking quantitative data to qualitative data was a big bridge to lead people over … it's likely that no product manager in Prime Video has ever had the luxury of having the time or the resources in place to say 'I'm going to look at the entirety of the experience and consider all the needs our customers have AND I'm going to make an attempt at getting a sense at how good or bad we're doing at these needs AND I'm going to try to see what the business opportunities are associated with each of these needs. That then becomes an overwhelming amount of evidence, and something that doesn't seem like it could be quickly replicated by anyone on the team … linking the attitudinal to behavioral data becomes more of an exercise in prioritization." — Principal Product Manager, Amazon Prime Video

From a research methods perspective, one of the most valuable aspects of this integrated data system is the ability to "close the loop" between behavioral usage data, survey data, and data from subsequent qualitative evaluative research. In usability testing, the qualitative team measures project-related CXOs using the same 3-point scale that the ongoing survey program uses, which enables direct comparison between data sources. Interestingly, when participants are asked to rate their satisfaction for key CXOs on this 3-point scale, the distribution of responses from usability testing are often similar to responses from 5,000-person surveys. This demonstrates the power of qualitative, smaller sample-sized data and increases the confidence stakeholders have in using this method.
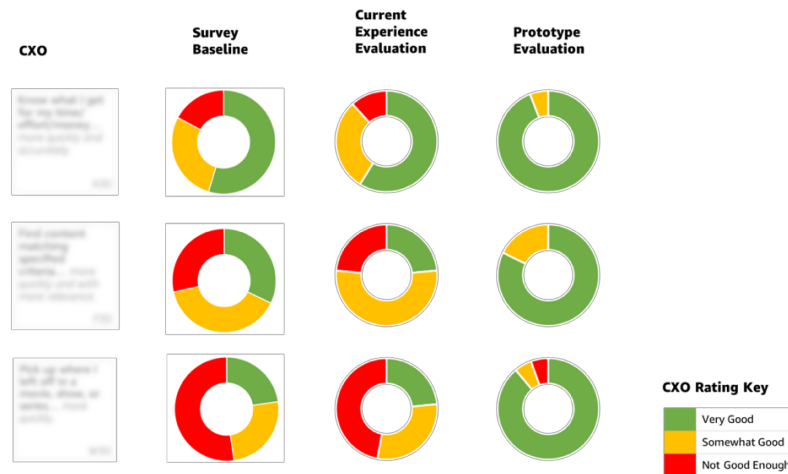
Figure 8. Sample comparison of survey, current experience evaluation, and prototype evaluation results. There are significant similarities between survey results and current experience evaluation, which increases confidence in qualitative findings.

As teams progress to using A/B and multivariate testing with greater sample sizes for their projects, results from usability testing have increased the confidence and decreased the cost of what the team includes in the tests. The team now provides recommendations for what should be included in multivariate or A/B testing based on confidence level and what is or is not well-suited for a usability lab setting. This has recently resulted in a multivariate test that had the highest increase in key business metrics in Prime Video history.

> "It's now clear that this has a level of rigor and thinking behind this that makes sense … this is a whole new way of approaching UX development and even product decisions … *That*, I feel, is a crucial inflection point. Now you have an approach that tells you what to prioritize in a data-driven way, which I think is going to be much less open to individual opinions than it had been before … Once you have an approach, doing user research to see if this actually helps that CXO improve … so you have high confidence before you launch that this is going to work because of the work that you've done up front, that is just completely new … I think that is a step function improvement in how we've gone about doing things." — A Director of Engineering, Amazon Prime Video

This is an ongoing program which directly informs what Prime Video prioritizes; the impact is continuously developing, and the team's knowledge is constantly evolving. However, there is clear evidence that with the integrated, closed-loop data system that the team has built, the organization is better equipped with the tools and insights they need to make confident, well-informed decisions on the customer's behalf. The customer's "voice" now has a seat at the table prioritizing roadmap investments. A senior leader summed up the research team's impact:

> "Seeing that we can build the case and hold our own using customer data, translated into business data … it's the first time I saw customer data at the business decision-making table and it breaking through. It basically bought its way onto the roadmap by being able to talk about it in the currency of the business." – Director of UX Design, Amazon Prime Video

Though this program has already been effective, in many ways it is still day one. The team is now embarking on extensions of this program and improving it as they learn more. Research scientists are working with other data scientists to explore how to capture customer satisfaction measures in multivariate tests. They are working with economists to determine the financial value of movement in satisfaction on key CXOs. Qualitative researchers are exploring ways to recruit participants who have responded to the survey in the past to test new concepts so they can compare their survey results to their evaluations of new concepts. The team is also continuing to qualitatively validate the CXO framework in new markets across the world—with recent studies completed in Germany and Japan—and spread the survey program into other key markets as well. These extensions help the UX Research team continue to expand their integrated, closed-loop system, and as more inputs are added to the system, the team will continue to expand how evidence is defined, balancing qualitative and quantitative inputs to drive decision-making for Prime Video.

NOTES

Note that neither Amazon.com, Inc. nor any of its affiliates is a sponsor of this conference and this case study does not represent the official position of Amazon.com, Inc. or its affiliates.

All customer data collection methods follow standard best practices and is anonymized.

REFERENCES CITED

Christensen, Clayton M. and Michael E. Raynor
2003   The Innovator's Solution: Creating and Sustaining Successful Growth. Boston: Harvard Business School Publishing Corporation.

Ulwick, Anthony
2005   What customers Want: Using Outcome-Driven Innovation to Create Breakthrough Products and Services. New York: The McGraw-Hill Companies, Inc.

Christensen, Clayton M. and Taddy Hall, Karen Dillon, and David S. Duncan
2016   Competing Against Luck. New York: HarperCollins Publishers.

Dillman, Don A.
2007   Mail and Internet Surveys: The Tailored Design Method, 2007 Update with New Internet, Visual, and Mixed-Mode Guide. Hoboken: John Wiley & Sons, Inc.

Singleton, Royce A., Jr. and Bruce C. Straits
2010   Approaches to Social Research. New York: Oxford University Press.

Maslow, Abraham H.
1954   Motivation and Personality. New York: Harper & Row.