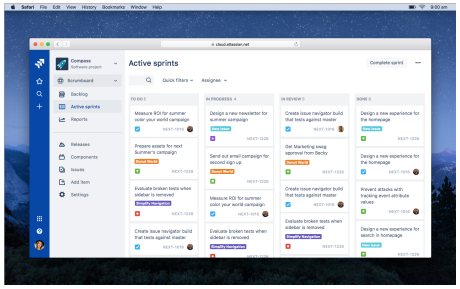

Figure 1: Case Study: Usability Improvements in Jira. Jira is a project management software created by Atlassian. Around mid-2016, the product leadership team has set a strategic goal to improve the Net Promoter Score of the product. The NPS was derived by conducting a periodic in-product survey asking users how likely they are to recommend the product to their peers, on a scale of 0 to 10, alongside optional free-text feedback. The overall NPS score is the percentage of users who provided a score of 9 or 10 minus the percentage of users who provided a score between 0 to 6. The NPS goal had to be translated into an actionable plan that UX designers and the engineering teams could execute. With the aid of the free-text feedback the teams identified Usability as the main focus area. To track progress, they used UMUX-LITE [6], which is based on a two questions survey: *"This product's capabilities meet my requirements"* and *"This product is easy to use"* that users rate using a seven point Likert scale. Multiple teams worked following the Agile paradigm to improve product usability. Statistical correlations linked improvements in usability to NPS scores.

*Currently at Google.

Bridging the Gap Between Business, Design and Product Metrics

Ilias Flaounas*
Atlassian
Sydney, Australia
iliasfl@gmail.com

Arik Friedman
Atlassian
Sydney, Australia
afriedman@atlassian.com

ABSTRACT

The integration of User-Centered Design with Agile practices studies the interactions between designers and developers and the alignment of the design and development processes. However, beyond the interactions with the development team, designers are often required to operate within a wider business context, driven by goals set on high-level metrics, like Monthly Active Users, and to show how design-led initiatives and improvements address those metrics. In this paper we generalize learnings from prior work on applying usability improvements to Jira, a project tracking software tool created by Atlassian, and we describe a structured approach to bridging the gap between feature work and business metrics.

INTRODUCTION

The creation of new products or improvement of existing ones under conditions of uncertainty is a common challenge for software companies. The lean startup approach [7] offers a methodology to progress in these conditions by testing assumptions and visions continuously, through rapid experimentation that aims to maximize learning. This approach advocates quick adjustments through

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313013

KEYWORDS
quantitative methods; metrics; UMUX-LITE; usability; KPI; NPS; agile programming; scorecards

build–measure–learn cycles, in which quick iterations maximize learning through incremental and iterative engineering.

Unfortunately, the scope of some problems is too wide to explore through quick iterations. Mature products will typically set and track goals around high-level business metrics or Key Performance Indicators (KPIs), such as Daily or Monthly Active Users (DAU/MAU), Net Promoter Score (NPS), Customer Satisfaction (CSAT), revenue, buying rate or some other measure that captures an intrinsic value proposition of the business [2]. These metrics are typically hard to move and they are rarely affected in an observable way by relatively small product changes. Consider, for example, customer feedback from surveys that points to usability issues. Improving product usability to address that feedback and drive up CSAT may require changes such as a visual redesign of the product or a new Information Architecture through considerable investments that cut across the product. While user tests can hint at whether a certain approach is a move in the right direction, affecting a change in CSAT in the final product will likely require ongoing efforts and considerable investment.

Advanced techniques like live-traffic experiments can track changes on user behaviors at the level of visual element impressions or user interactions, but they are not effective for assessing impact on lagging metrics like MAU or CSAT in a reasonable time frame. The gap between business metrics and feature-level metrics calls for the introduction of intermediary metrics and goals that would bridge the distance between the two ends. These intermediary goals enable iterations where two kinds of questions can be revisited on a regular basis: *"Did we choose the right metrics to capture our goals?"* and *"How did we affect through product changes the metrics that we chose?"*.

We propose a framework that encapsulates this process and aims to bridge the gap between business metrics and the efforts of product development teams. An example application of this framework was described in a detailed case study [3], outlined in Figure 1.

RELATED WORK

User-Centered Design (UCD) focuses on users and their needs within an iterative design process. The process and its effectiveness have been studied extensively, including the integration of UCD and Agile practices (AUCDI) [9]. The AUCDI line of work focuses mainly on the interactions between designers and developers, and how to align the design process with the development process. Gulliksen et al. [5] discussed the overlap of UCD principles with the Agile process, and offered "process customization" as one of the guiding principles for UCD. However, they did not address how the customization can be implemented in the organization context to address management needs and KPIs. Buur and Bødker [1] describe in "the Danfoss case" continuous engagement of stakeholders, but refer to interaction with business stakeholders with the goal of shaping the design itself, rather than measuring and reporting on its outcomes.

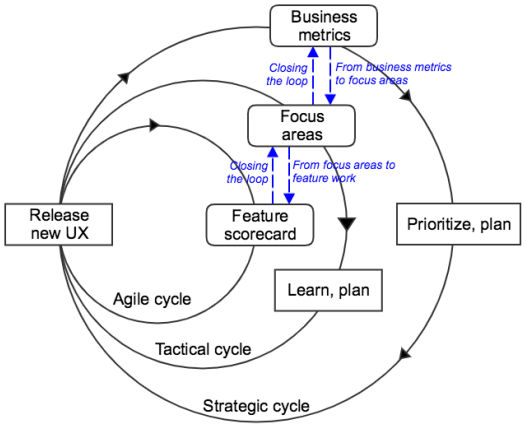

Figure 2: Three cycles in the software development process and their connections: the strategic cycle, the tactical cycle and the agile cycle. Our main focus is on the introduction of the tactical cycle and the transitions between the cycles (marked with dashed arrows).

While much of the prior work is relevant to the Agile cycle we describe in our framework, our main focus is on the interactions between the designers and the business, and tying gains in usability to the high-level metrics that the executives care about. This work aims to address the wide gap between the design process, which is the main subject of existing literature, and what management requires from designers. In that sense, this work is closer to works like Rosenbaum et al.'s [8], who mentioned this gap as one of the obstacles to strategic usability, or Vrendenburg et al. [10], who pointed out the gap between cited measures of UCD effectiveness (including business metrics) and those applied in practice. Despite the considerable number of designers that work in large corporations where this is a major concern, we are not aware of recent work in the HCI community that considers this aspect.

Practitioners experience a constant pressure to explain why the implemented improvements based on usability studies and related metrics are not reflected in improvements to their managers' KPIs. Failing to address such challenges could ultimately lead to loss of management's trust in design efforts. Our work proposes a process intended to bridge these gaps and help designers tie improvements they track in scorecards to the metrics that the business targets.

LINKING METRICS ACROSS CYCLES IN THE PRODUCT DEVELOPMENT PROCESS

Our framework identifies three nested cycles of the software development process as depicted in Figure 2.

The Strategic Cycle: Improving business metrics

The first cycle is focused on improving a high level business metric, i.e., some KPI, usually set by management. It is a metric that is typically hard to move, and parallel efforts across several business units may be needed to change its trajectory over multiple quarters or even years. Stakeholders here are executives that set the product or company long-term strategy. This cycle may go beyond the scope of work of design and product development teams. For example, the marketing department may launch a new targeting campaign, or the sales department may start promoting specific use cases with potential customers, alongside efforts from product teams. Even when teams, including UX teams, have the flexibility to set their own metrics and goals, there is an implicit expectation that their efforts contribute to improving the KPIs of the product. We refer to this cycle as the **strategic** cycle.

The Tactical Cycle: Breaking down business goals to focus areas

The second cycle is based on the effort of the product development teams and it is directly impacted by product design improvements. The introduction of this intermediate cycle is the core idea of this paper. The product teams need to align their road-maps ("what should we work on next?") with the business goal, and improve the product in the areas that are most likely to affect the business metric.

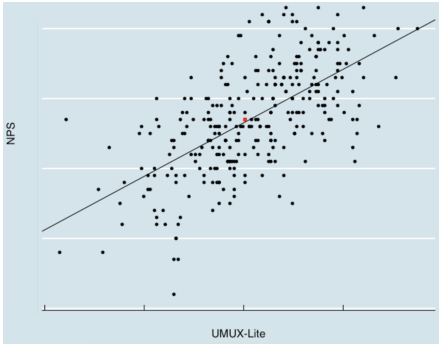

Figure 3: An example of a linear model that correlates UMUX-LITE to NPS for a segment of Jira users. In the Jira case study [3], NPS was the metric of choice at the strategic level, while UMUX-LITE was the metric of choice at the tactical level. This chart is based on answers from 2023 users and reflects the bootstrapping statistical technique (sampling with replacement) for simulating the repetition of the survey 300 times. We measured a correlation of 0.62 between NPS and UMUX-LITE, which was considered satisfactory for our problem and formed the "bridge" linking the strategic metric to the tactical metric in this case. The scales of the axes were removed to protect potentially sensitive product information. This type of modeling can help the teams form better understanding of the relation of the metrics at strategic and tactical cycles, set realistic goals and plan accordingly for the magnitude of the impact they should aim at in terms of UMUX-LITE, in order to achieve the goal which is set on NPS at the strategic level.

For example, the product team may aim to improve the usability or the reliability of the product. We refer to this cycle as the **tactical** cycle.

The first step to deriving actionable insights consists of breaking down the business metric into several focus areas. The specific breakdown may change depending on the nature of the high-level metric, the product, and the business needs. A possible approach is the use of FURPS [4], a framework for classifying software requirements. FURPS stands for Functionality, Usability, Reliability, Performance and Supportability. FURPS is adjustable to the requirements of a specific product, and can be extended to include additional types of requirements. Identifying focus areas that promise the highest impact to a high-level business metric may still be insufficient to drive concrete improvements in the product: in some cases the gaps may be related to a specific product feature, in others they may be broader in scope, pointing to fundamental deficiencies in the product. The exact process by which focus areas are tied to specific feature work may change depending on the goals that were set and the product in question. For example, the Jira case study [3] describes how customer feedback was labeled and broken down to FURPS-based categories. This process surfaced Usability as a focus area, and UMUX-LITE [6] was set as a metric to track product usability improvements directly. UX research demonstrated the correlation of NPS to UMUX-LITE for our case providing confidence to the team for the selection of the tactical level metric (see Figure 3).

In general, typical user research methodologies such as surveys or user interviews can be used to identify specific areas in the product, or recurring themes that should be addressed, in relation to the target metrics. User research, domain knowledge and customer feedback can therefore guide relevant breakdowns and potential focus areas. For example, if a product manager identified that the greatest opportunity for a high-level MAU goal is reducing user churn, then churn surveys and interviews with churned customers can reveal gaps in the product, such as missing functionality.

Once specific feature areas or aspects of the product use have been identified, they can be combined with quantitative data sources like log-based user behavioral analytics to highlight the importance of each. The combination of quantitative and qualitative inputs can point to specific product areas that represent the highest friction point, or conversely an opportunity to enact a positive change in the product that is most likely to impact the target metric. That focus area can then be targeted for iterative improvements via Agile cycles run by the product team.

Agile cycle: Tracking Feature Work with Scorecards

The third cycle is the work of each product team. This is the build–measure–learn cycle of the lean methodology [7], which we refer to as the **Agile** cycle. In this paper we focus on the measurement aspect of the cycle and specifically we focus on the use of scorecards. A scorecard is a set of metrics that cover the areas of interest in the product development life cycle. Ideally, the same key metrics should be monitored throughout all the product development phases, from early prototyping all the

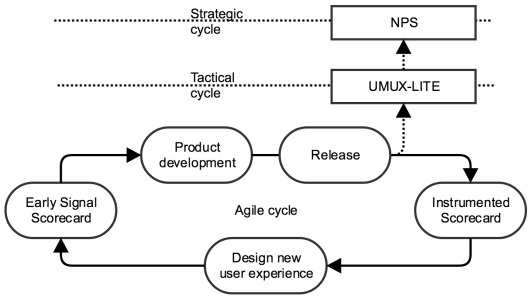

Figure 4: Within the Agile cycle, early signal testing scorecards help guide the design of feature work, and instrumented scorecards allow to track the impact of those changes. In an ideal situation, Agile iterations would improve the low-level metrics tracked in instrumented scorecards. Each iteration constitutes a concentrated software development effort of one to three weeks, with the goal of incrementally improving one or more low-level metrics. Over the span of a few months or a quarter, these changes will accumulate and affect the intermediary metrics tracked in the tactical cycle. Multiple tactical efforts would add up to change the trajectory of a high level business metric over several quarters, driving forward the strategic cycle.

way to shipping to real users. However, in practice this is not always possible: for example, a *time to task completion* metric is meaningful during usability tests of prototypes where users are asked to complete a specific task, but hard to measure in the final product, as we do not know at any given time what the users try to achieve as they interact with the product. The difficulty of attaining meaningful metrics that would apply across the whole product development cycle, end-to-end, leads to the distinction between two kinds of scorecards, one for the early signals testing phase, before the product is built, and one for the validation of results via behavioral analytics at scale when the product is deployed to real users. We refer to the former as the *early signal testing scorecard* and to the latter as the *instrumented scorecard*. Their impact in the agile lifecycle is visualized in Figure 4.

Closing the loop

Figure 4 illustrates the Agile cycle and how it fits within the tactical and strategic cycles. Tactical metrics are leading indicators for business metrics, which are lagging by nature. In a similar way, scoring cards can serve as leading indicators for the tactical metrics. There are two possible outcomes in this process. Either the business will achieve the goals for the metric at the strategic level, or it will become obvious that the efforts do not deliver the anticipated outcomes. The first case of success is easy to spot and the follow-up actions are straightforward: the metrics that the business cares about start moving in the right direction, and soon the question becomes whether the business should invest in additional iterations to further improve the metrics, or whether they have reached a satisfactory level in which case a new strategic cycle begins. In the case of unsatisfactory results a systematic analysis can reveal the underlying problem by asking a sequence of questions, starting from the outcomes of the Agile cycles, and expanding outwards up to the strategic cycle:

(1) Did the product changes affect the feature scorecard metrics as intended? A negative answer to this question draws the attention to the assumptions that have driven the product changes. Additional user research may reveal whether previously identified friction points have been solved, or whether additional improvements are required to achieve the stated goals.

(2) Did the tactical level metrics change following positive changes in the feature scorecard metrics? A negative answer to this question draws the attention to the relation between the feature level metrics and the tactical level metrics. As in the case of feature changes and scorecard metrics, it is possible that improvements in one area of the product will only uncover issues in other areas, in which case new feature goals should be set, with additional Agile iterations.

(3) Do the strategic level metrics correlate with changes in tactical level metrics? A negative answer to this question draws attention to the relation between the tactical level metrics and the strategic goal. In some cases, trends to high-level metrics may lag behind the product changes that drive them, so the decision makers should be clear about the time frame in which they expect to see these trends.

Step-by-step process and example

(1) At the strategic cycle, executives set the long term strategy, select a high level metric to track progress and set a goal, e.g., "improve product NPS by 10%."
(2) At the tactical cycle, product teams prioritize focus areas to achieve the business goal, identify a metric for tracking progress in those areas, and set a goal for that metric. For example, if usability issues are identified as the main driver for low NPS scores, the product team may set a goal to improve product usability by 20%, tracked by UMUX-LITE surveys.
(3) At the Agile cycle, engineering and design teams identify feature work and derive scorecards of user interactions to monitor their progress, e.g., "improve the on-boarding experience of new users and monitor their interactions with the new user experiences via behavioral analytics and satisfaction surveys."
(4) With improvements in the scorecard, the teams assess the impact on the tactical goal, e.g., a linear statistical model shows correlation of the use of an on-boarding wizard to high UMUX-LITE scores.
(5) As tactical goals are achieved, the teams can assess the impact on the strategic metric, e.g., correlation of UMUX-LITE scores to NPS scores.

CONCLUSIONS

We presented a framework that aligns metrics set by different stakeholders so that product teams can identify and track product work such as design and usability improvements towards accomplishing business goals. Our work is not intended as an automated process, but rather as a "playbook" that outlines a structured approach, which incorporates a strong human component to drive its success. In this approach, teams articulate their assumptions and goals explicitly, draw a line in the sand that defines what success looks like, and hold themselves accountable to either achieve the goals that they set, or to revisit their assumptions and apply their learning to improve the process on an ongoing basis. Decisions trickle down from the business level to the product team level, as the goals at each level are derived from the higher-level goals, and each of the inner cycles assumes that the right decisions were made in the outer cycles. However, the structure of the process forms a feedback loop in which these decisions are continuously put to the test, leaving room to pivot and self-correct if new learnings contradict any of the prior assumptions and decisions.

REFERENCES

[1] Jacob Buur and Susanne Bødker. 2000. From Usability Lab to "Design Collaboratorium": Reframing Usability Practice. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '00)*. ACM, New York, NY, USA, 297–307. https://doi.org/10.1145/347642.347768
[2] Robert G. Eccles, Joseph A. Ness, Thomas G. Cucuzza, and Robert Simons. 1998. *Harvard Business Review on Measuring Corporate Performance*. Harvard Business Press.
[3] Arik Friedman and Ilias Flaounas. 2018. The Right Metric for the Right Stakeholder: a Case Study of Improving Product Usability. In *Proceedings of the 30th Australian Computer-Human Interaction Conference (OzCHI 2018)*. ACM, New York, NY, USA, 602–606.
[4] Robert B. Grady. 1992. *Practical Software Metrics for Project Management and Process Improvement*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
[5] Jan Gulliksen, Bengt Göransson, Inger Boivie, Stefan Blomkvist, Jenny Persson, and Åsa Cajander. 2003. Key principles for user-centred systems design. *Behaviour & IT* 22, 6 (2003), 397–409. https://doi.org/10.1080/01449290310001624329
[6] James R Lewis, Brian S Utesch, and Deborah E Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2099–2102.
[7] Eric Ries. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business (USA).
[8] Stephanie Rosenbaum, Janice Anne Rohn, and Judee Humburg. 2000. A Toolkit for Strategic Usability: Results from Workshops, Panels, and Surveys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 337–344. https://doi.org/10.1145/332040.332454
[9] Dina Salah, Richard F. Paige, and Paul Cairns. 2014. A Systematic Literature Review for Agile Development Processes and User Centred Design Integration. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*. ACM, New York, NY, USA, Article 5, 10 pages. https://doi.org/10.1145/2601248.2601276
[10] Karel Vredenburg, Ji-Ye Mao, Paul W. Smith, and Tom Carey. 2002. A Survey of User-centered Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 471–478. https://doi.org/10.1145/503376.503460