Actionable Software Metrics: An Industrial Perspective

Prabhat Ram
M3S, Faculty of ITEE
University of Oulu
Oulu, Finland
prabhat.ram@oulu.fi

Pilar Rodríguez
M3S, Faculty of ITEE
University of Oulu
Oulu, Finland
pilar.rodriguez@oulu.fi

Markku Oivo
M3S, Faculty of ITEE
University of Oulu
Oulu, Finland
markku.oivo@oulu.fi

Silverio Martínez-Fernández
Dept. ESSI, GESSI Research Group
UPC-BarcelonaTech
Barcelona, Spain
smartinez@essi.upc.edu

Alessandra Bagnato
Softeam
Paris, France
alessandra.bagnato@softeam.fr

Michał Choraś
ITTI Sp. z o.o, Poznań and UTP
Bydgoszcz, Poland
chorasm@utp.edu.pl

Rafał Kozik
ITTI Sp. z o.o, Poznań and UTP
Bydgoszcz, Poland
rafal.kozik@itti.com.pl

Sanja Aaramaa
Nokia
Oulu, Finland
sanja.aaramaa@nokia.com

Milla Ahola
Bittium
Oulu, Finland
milla.ahola@bittium.com

ABSTRACT

Background: Practitioners would like to take action based on software metrics, as long as they find them reliable. Existing literature explores how metrics can be made reliable, but remains unclear if there are other conditions necessary for a metric to be actionable. *Context & Method:* In the context of a European H2020 Project, we conducted a multiple case study to study metrics' use in four companies, and identified instances where these metrics influenced actions. We used an online questionnaire to enquire about the project participants' views on actionable metrics. Next, we invited one participant from each company to elaborate on the identified metrics' use for taking actions and the questionnaire responses (*N=17*). *Result:* We learned that a metric that is practical, contextual, and exhibits high data quality characteristics is actionable. Even a non-actionable metric can be useful, but an actionable metric mostly requires interpretation. However, the more these metrics are simple and reflect the software development context accurately, the less interpretation required to infer actionable information from the metric. Company size and project characteristics can also influence the type of metric that can be actionable. *Conclusion:* This exploration of industry's views on actionable metrics help characterize actionable metrics in practical terms. This awareness of what characteristics constitute an actionable metric can facilitate their definition and development right from the start of a software metrics program.

https://doi.org/10.1145/3383219.3383244

CCS CONCEPTS

• **General and reference** → **Empirical studies;** Measurement; Metrics; • **Software and its engineering** → Empirical software validation.

KEYWORDS

actionable metrics, data quality, context, metrics program

ACM Reference format:

Prabhat Ram, Pilar Rodríguez, Markku Oivo, Silverio Martínez-Fernández, Alessandra Bagnato, Michał Choraś, Rafał Kozik, Sanja Aaramaa, Milla Ahola. 2020. Actionable Software Metrics: An Industrial Perspective. In *Evaluation and Assessment in Software Engineering (EASE '20), April 15-17, 2020, Trondheim, Norway. ACM, New York, NY, USA, 10 pages*. https://doi.org/10.1145/3383219.3383244

1 Introduction

One of the benchmarks of a software metrics program's effectiveness is how actionable it is [22], since software practitioners are not interested only in mining data for insights but also in using those insights to guide their actions [20, 24]. Software metrics can be used to empower software practitioners to gain and share insights from their data for better decision-making [13, 25]. In addition to being trustworthy [17] or reliable [20], such metrics need to manifest practitioners' practical requirements in order to contribute any actionable information [25]. Essentially, a metric that can influence action [2] to facilitate process improvement or remedy a situation [23] is an actionable metric. Existing literature posits data quality dimensions like accessibility, completeness, timeliness, relevancy, etc. to evaluate a metric's reliability, and enhance its suitability as a contributor to decision-making [20].

However, it is unclear if a metric's reliability is both a sufficient and necessary condition for it to be actionable, or if there are additional conditions that dictate a metric's actionability.

A common misconception in software development is that a metric is always actionable, regardless of whether that metric is merely an indicator of a condition, or if it can inform a practitioner on what to do next [12]. However, literature suggests that a metric is actionable only in the latter case [2]. There are studies that recommend success factors for a metrics program [17, 20, 21], and suggest measures to increase responsiveness to customers' needs [15], while broaching the topic of using metrics for taking actions. However, these studies do not discuss what makes a metrics truly actionable, beyond their requirement to be reliable [20]. The case study by Vacanti and Vallet [23] at Siemens Health Services (SHS) attempt to characterize actionable metrics using a real-world implementation as an example. The authors adopt *flow*[1] metrics, which they claim to be actionable, to improve overall predictability and process performance at the company [23]. The authors propose that an actionable metric should be able to suggest specific interventions if and when needed [23]. Beyond this all-encompassing property of an actionable metric, the authors do not discuss any other specific characteristics of an actionable metric. Based on the state of the art, we gather that an actionable metric should be contextual, reliable, and provide an impetus to act. In this view, we aim to investigate further to obtain an industrial perspective on actionable metrics, which can validate the above findings and even supplement them. We argue that this knowledge will help practitioners focus their efforts on defining and developing such metrics right from the start of a metrics program.

The goal of the EU H2020 Project, Q-Rapids, is to develop an agile-based, data-driven, and quality-aware rapid software development process [3]. Four software-intensive companies are involved as industry partners in the Project. These companies have been using the Project solution (Solution), including its metrics, in their daily work. They have also initiated improvements by acting on some of the metrics. These developments and the research aim to characterize actionable metrics from an industrial perspective motivated the following research questions (RQs):

1. **RQ1**: What role did the metrics play in the improvement actions taken by the participants of the study?
2. **RQ2**: What is an actionable metric according to the participants of the study?
    a. *RQ2.1*: What are the characteristics of an actionable metric?
    b. *RQ2.2*: What is the utilitarian perspective of an actionable metric?
    c. *RQ2.3*: What are the high-quality data requirements for an actionable metric?

The participants from the Project are in an ideal position to provide rationale for their decisions, especially to highlight metrics' role in the improvement actions. These rationales inform our following main contributions:

1. Providing empirical accounts of software metrics that are actionable, contributing to practitioners' decision-making
2. Analyzing practitioners' perspective on what an actionable metric is, and the characteristics that make a metric actionable in the context of software development.

In the remainder of the paper, we synthesize related work in Section 2, describe the research method in Section 3, followed by the results in Section 4. Discussion of the results is in Section 5, with Section 6 discussing limitations and threats to our research's validity, and conclusion and future research directions in Section 7.

2 Related Work

Research on metrics program (MP) has a long history [9], but what makes a metric truly actionable remains partially explored. We first look at the literature proposing characteristics of a successful metrics program, followed by the literature that explicitly address actionable metrics in some capacity. The objective is to aggregate characteristics of an actionable metrics reported in the state of the art, if any, and position our study accordingly.

Studies by Mendonça and Basili [11], Hall and Fenton [4], and Staron and Meding [21] propose success factors for an MP, but do not explore the characteristics of metrics that can make them actionable and contribute to an MP's success. The case study in [20] highlight the need for a metric to be reliable to make them actionable. Similarly, the case study in [17] highlight five success factors for effective operationalization of an MP, and the authors emphasize on metrics trustworthiness in order for practitioners to find them useful. Therefore, beyond a metric's reliability, a product of metric's trustworthiness, the above-mentioned studies do not tackle what makes a metric actionable. These studies address the bigger picture of eliciting factors that can help successfully deploy an MP and sustain it. An organization's default position on any MP is to have useful metrics, but even non-actionable metrics can prove useful [12]. Our study complements above studies, and investigates the specific characteristics of an actionable metric.

Port and Taber [16] present examples of metrics and analytics program used to support strategic maintenance of a critical system at NASA's Jet Propulsion Laboratory. The authors claim that these metrics are actionable, as they have been validated by their proven utility in maintaining the critical systems. In a brief account of different types of metrics, Croll and Yoskovitz [2] posit that an actionable metric should have the potential to induce change by suggesting a course of action. At the same time, the authors caution that actionable metrics cannot be "*magic*", as they can only provide guidance and not precise instructions. Buse and Zimmerman [1] surveyed 110 practitioners from Microsoft to understand their decision-making process, and found that managers rate data/metrics highly for taking actions. The authors posit that even

[1]Workflow metrics such as Work in Progress, Cycle Time, and Throughput are *flow* metrics

common metrics can be actionable, as long as they are context-driven. The above studies indicate the importance of a metric's actionability for the decision-makers, but discussion on what is needed to make a metric actionable has received scant attention.

The case study by Vacanti and Vallet [23] at SHS is among the recent studies that attempts to characterize actionable metrics in an industrial environment. In order to realize the predictability and transparency promised by the traditional agile metrics such as story points and velocity, SHS decided to use Kanban, along with much simpler and more actionable metrics of *flow* (Work in Progress, Cycle Time, Throughput). The *flow* metrics are tied together by Little's Law[2], and change in one metric influences the other two. If the said change is undesired, then the inherent relationship among these *flow* metrics will suggest what steps are needed to overcome that state. Beyond the inherent actionability due to Little's Law and use of simpler metrics, the authors do not discuss any further the characteristics that make the *flow* metrics, or any metric, actionable.

The state of the art highlights several factors that drive an MP's success, and a metric's usefulness is integral to that success. However, if a real metric has to be actionable [2], then practitioners should be aware of the steps needed to ensure that right from the start of an MP. Our research is an attempt to delineate characteristics that make a metric actionable, and contribute to the abovementioned steps.

3 Research Method

Following the guidelines recommended by Runeson and Höst [18] we conducted a multiple case study to answer the two RQs.

3.1 Research context

The following table characterizes the context of the four companies, to which the participants of this study belong:

Table 1 Case Company Characteristics

Parameters	Case Company 1	Case Company 2	Case Company 3	Case Company 4
ID	CC1	CC2	CC3	CC4
Size	Large	Medium	Large	Small and Medium sized Enterprise
Domain	Commercial services and solutions	Defense & Telecom.	Telecom	Multi-industry
Use case	Software modeling tool	Product information system	Hardware-oriented project	Warehouse Mgmt. System
Length of Solution use	~ 2 years	8 months	*deployment only*	~ 2 years
Use case team size	9	15	~120	10

CC1 used the Solution in the context of a modeling tool for model-driven development, which is part of a mature product line, with multiple releases already in the market. The company had used Solution in the course of development of past three releases of the above tool, with the fourth underway. Their views on actionable metrics is informed by this cumulative experience.

In case of CC2, we focused on the metrics, and their use, from the second use case (UC), as the first (pilot) UC was no longer in active development. Taking together the two UCs, duration of Solution use at CC2 is comparable to the other case companies.

A comparable Solution use at CC3 was hindered due to several unforeseeable technical and managerial circumstances. Therefore, in contrast to the other case companies, CC3's views on actionable metrics are based on the Solution deployment and preliminary use.

Following a positive and a formative experience from their pilot use case, CC4 continued to use the Solution in the next UC, albeit with customizations that are exclusive to this case company only. The experience accumulated from these two UCs inform CC4's views on actionable metrics.

3.2 Data collection

The case companies shared metrics data with the Project researchers on a monthly basis, along with a short report documenting their use of the Solution. In the course of several follow-up interactions with them, we learned that three of the case companies (CC1, CC2, and CC4) had utilized metrics to undertake process improvements. This knowledge triggered the need for a questionnaire to gather the rationale that underlie the aforesaid improvement actions.

We relied on the reviewed literature to design the questionnaire. We formulated a total of 15 statements, measured on a 1 – 5 Likert scale (strongly disagree – disagree – neutral – agree – strongly agree – don't know). These 15 statements and their sources of reference are shown in the table below:

Table 2 Questionnaire

ID	Questionnaire statements	Reference
	An actionable metric…	
Q1	...is practical	[2, 25]
Q2	…informs decision-making	
Q3	…is project/process /product specific	
Q4	…can be universal	[2]
Q5	...has to have high data quality	[20]
Q6	Every metric has to be actionable	[2]
Q7	Only certain metrics need to be actionable	
Q8	Without interpretation, a metric cannot be actionable	[1]
Q9	Aggregated metric is more actionable than a low-level metric (e.g. "resolved issues throughput" more actionable than "number of open issues")	[1, 19, 25]
Q10	A non-actionable metric can be useful	
	High-quality data are…	
Q11	…accurate	[6]
Q12	…consistent	
Q13	…current	
Q14	…credible	
Q15	…complete	

[2] Little's Law: Avg. Cycle Time = Avg. Work in Progress / Avg. Throughput.

With Q1-Q5 statements, we enquire about the general characteristics of an actionable metric. Statements Q6-Q10 enquire an actionable metric's *utilitarian perspective*, which characterizes the practicalities under which a metric can be actionable. Both these sections were followed by an option for additional comments. Q11-Q15 statements capture general characteristics of high-quality data, which can be attributed to actionable metrics data. The five data quality characteristics used in the questionnaire refer to the inherent data quality, as prescribed in ISO/IEC 25012:2008 [6]. This section was followed by the question "*Are there any other data quality characteristics we missed? Please mention them below, along with the rationale*".

The questionnaire was administered online to the users of the Solution from the four companies. The aim was to elicit responses from the users that were involved in the metrics-driven improvement actions at their companies. We received responses from 17 participants.

Following the questionnaire, we invited one participant from each case company to co-author this study. The aim was to gather first-hand insights about the actions influenced by the metrics, and to provide rationale to the questionnaire responses from the participants of their respective companies.

3.3 Data analysis

For analyzing the questionnaire, we used the data analysis approach adopted in [5]. Similar to our research, the questionnaire in [5] (*N=15*) is also characterized by small sample size. We calculated the following indicators to interpret the responses:

1. *% Agree*: Percentage of participants responding with either '*Agree*' or '*Strongly agree*'
2. *Top-Box*: Percentage of participants responding with '*Strongly agree*'
3. *Net-Top-2-Box* - Percentage of participants that chose bottom two responses (*Strongly disagree* and *Disagree*) subtracted from the participants that chose top two responses (*Strongly agree* and *Agree*).
4. *Coefficient of Variance (CV)*: Standard deviation divided by the mean. A higher CV indicates higher variability.

The first three indicators are a measure of central tendency, which indicates a single value that helps to identify the central value in a set of data.

4 Results

Answer to RQ1 is provided by the invited participants. Results from the questionnaire helps answer RQ2, which are also complemented by inputs from the invited participants.

4.1 RQ1. What role did the metrics play in the improvement actions taken by the participants of the study?

In this section, the participants elaborate on the instances where they used metrics for taking improvement actions. A complete list of metrics for each case company can be found in *Appendix A* (https://doi.org/10.5281/zenodo.3580893).

4.1.1 Metrics' role in CC1. The UC development team holds a meeting every day to discuss their next release and the next steps in the context of the forthcoming product release. In one of these meetings, the team used the '*non-blocking files*' metric to identify problems that were blocking certain development tasks, which were critical for developing features for their upcoming product release. To avoid delays, the team prioritized their development activities based on the tasks the above metric helped identify.

The Solution provides a quality-alert feature, where once a metric crosses a user-defined threshold, indicating violation of a quality goal defined by the UC Quality Engineers (QEs), it triggers an alert. The alert is accompanied by a recommendation to correct the situation, which is recorded as an abstract quality issue (quality requirement) in CC1's project management tool OpenProject[3]. The Project Manager (PM) decides if the quality issue should be integrated as a concrete development task in the project or not. For example, the PM accepted the alert generated by the metric '*critical issues ratio*', which tracks completed development tasks with critical severity issues, and converted it into a development task that called for review and update of the team's validation process. Owing to the above described Project feature, CC1 has formalized their quality requirements management, and moved away from the informal approach of mobilizing resources on an ad-hoc basis to address a quality issue.

In case of the '*non-blocking file*' metric, the UC team was familiar with the metric and knew it to be reliable. In case of the '*critical issues ratio*' metric, again the familiarity of the metric and PM's close scrutiny of the generated alert played an important role. In summation, the metrics' role in the improvement actions is undeniable, but they were among several other contributing factors.

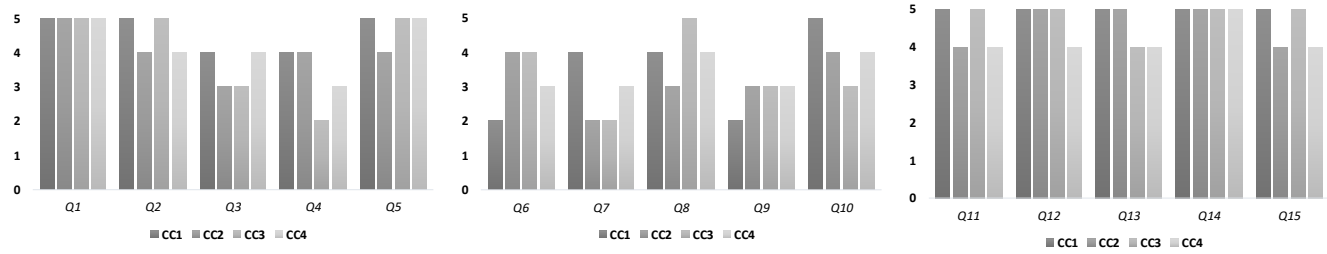
4.1.2 Metrics' roles in CC2. Using the '*well-defined issues jira*' metric, the UC Champion learned that the developers were not always following the practice of maintaining the '*Definition of Done*' (*DoD*)[4] field in Jira[5]. *DoD* is a concise list of requirements that a product should meet for a development team to call it complete, and the above metric helps track the Jira issues[6] where the field is defined. The metric enabled the UC Champion to take the decision of reinforcing the practice of maintaining *DoD* across the team. Soon thereafter, the metric's value started increasing, reflecting improvement in the practice of maintaining the *DoD* fields.

[3] https://www.openproject.org/
[4] https://www.atlassian.com/blog/jira-software/8-steps-to-a-definition-of-done-in-jira
[5] https://www.atlassian.com/software/jira
[6] https://confluence.atlassian.com/adminjiracloud/issue-types-844500742.html

Table 3 Overview of the questionnaire results

ID	Questionnaire statement	Scale (1 – 5)	N	% Agree	Top-Box	Net-Top-2- Box	Coefficient of Variance (CV)
	An actionable metric…						
Q1	...is practical		16	94%	63%	94%	14%
Q2	…informs decision-making		17	88%	47%	76%	23%
Q3	…is project/process /product specific		15	47%	20%	20%	33%
Q4	…can be universal		16	56%	6%	25%	35%
Q5	...has to have high data quality		17	82%	71%	76%	21%
Q6	Every metric has to be actionable		17	47%	18%	12%	35%
Q7	Only certain metrics need to be actionable		17	35%	0%	0%	37%
Q8	Without interpretation, a metric cannot be actionable		15	73%	27%	67%	22%
Q9	Aggregated metric is more actionable than a low-level metric (e.g. "resolved issues throughput" more actionable than "number of open issues")		17	18%	6%	-24%	35%
Q10	A non-actionable metric can be useful		16	75%	25%	69%	22%
	High-quality data are…						
Q11	…accurate		17	100%	59%	100%	11%
Q12	…consistent		17	100%	59%	100%	11%
Q13	…current		17	82%	29%	76%	20%
Q14	…credible		15	100%	80%	100%	9%
Q15	…complete		17	88%	41%	82%	20%

*The graph distribution scale from left to right: *Strongly disagree, Disagree, Neutral, Agree,* and *Strongly Agree*



General characteristics of an actionable metric *Utilitarian perspectives for an actionable metric* *High-quality data for an actionable metric*

Figure 1 Questionnaire responses (case company-wise)

Right after deploying the metrics, the UC team started customizing them to have them reflect their development practices. However, the '*well-defined issues jira*' metric did not warrant any change, as the team found it simple enough to rely on them. This perception and the trust it instilled among team members, made it the main driver of the undertaken improvement action.

4.1.3 Metrics' potential role in CC3. CC3 engages in very large and complex projects, involving globally distributed development teams. Combining all of CC3's business lines, terabytes of test log data are generated on a daily basis. It is humanly impossible and unfeasible to sort and parse all that data to infer insights, and act on them to improve processes, if needed. In absence of specific solutions to guide practitioners in their decision-making, they are

likely to rely on subjective criteria and intuition. In addition to the project complexity, the tool landscape at CC3 is equally diverse. Different teams could be using different tools, or use the same tool, but in a different fashion. This further exacerbates the subjectivity situation. In an organization like CC3, subjectivity is neither ideal nor an optimal working culture, and data-driven decisions are preferable to intuition. Metrics act as an objective benchmark for everyone, regardless of their extent of involvement in a project.

CC3 practitioners prefer high-level and simple metrics. Considering the development scale, insights generated by low-level metrics may be insufficient to address potential improvement scenarios at the requisite scale. Similarly, simple metrics negate unnecessary interpretation of what action a metric is suggesting, mitigating the risk of subjectivity. Essentially, the actionable information indicated by the metric should be explicit and uncomplicated enough to be implemented without much time-consuming deliberation. Equally important is that the actionable information is reliable, and can stand scrutiny when compared against the original data sources used to calculate the metric. Furthermore, an ideal actionable metric should be intuitive to the target users, inherently reflecting the development context for these users to infer appropriate actionable information immediately.

4.1.4 Metrics' role in CC4: The UC team used metrics like '*estimated ticket density*' and '*spent density*' to track original effort estimated to complete a task and actual effort spent on a task, respectively. As a result, the team identified gaps between their effort spent and effort estimated, thereby providing an impetus to improve their effort estimation process. Soon, the team started noticing the narrowing in the above gap, indicating that the Product Owners (POs) and Senior Mangers (SMs) now estimate better. In a bid to improve developers' efficiency, the POs and SMs used the above metrics to learn that four days should be the optimal reported time/effort spent on a task.

Furthermore, the POs and SMs identified a process bottleneck, where the development tickets would stay in the '*merge request*' phase of their development process longer than necessary. Upon investigation, the cause was attributed to experienced developers having inadequate time and resources to review code and perform merges. This was resolved by granting the medium-experienced developers the rights to perform merges.

CC4 is undergoing a gradual transformation towards becoming Agile. To facilitate this, POs and SMs want to ensure software product quality, and improve and sustain the effectiveness of the development team and the development process. *CC4* wants to enable even the POs with mid-level experience to make quick team-oriented decisions, without involving too many decision-makers like SMs or the company board. They find that process metrics (metrics measuring development process) allow them to do that. After realizing the actionability of these process metrics, especially low-level process metrics, *CC4* has started preferring them to the quality metrics from tools like SonarQube[7]. In addition to the metrics being low-level, the UC team's familiarity and the subsequent trust in them made these metrics a potent source of actionable information.

[7] https://www.sonarqube.org/

4.2 RQ2. What is an actionable metric according to the participants of the study?

We report the results for *RQ2* in the sequence of its research sub-questions *RQ2.1 – RQ2.3*. The overall outcome of the questionnaire responses is shown in Table 3, followed by the charts in Figure 1 showing median (*Md*) questionnaire responses.

4.2.1 RQ2.1: What are the characteristics of an actionable metric? Based on the *'Net-Top-2-Box*' (net agreement) criterion, most of the participants agree that an actionable metric should be practical (94%) (*Q1*), inform decision-making (76%) (*Q2*), and exhibit high data quality (76%) (*Q5*). There is a unanimous consensus among all the case companies, when it comes to these three characteristics. In additional comments, one of the CC4 participants commented that, "*Metrics should be simple to understand, conclusive, and as much as possible automatically gathered and processed*". Another participant added that, "*Not all metrics are straightforward and can mean different things according to circumstances. Yet it is better to have them than not*".

Support for an actionable metric to be project/process/product specific (contextual) (*Q3)* and universal (*Q4*) has a very low net agreement of 20% and 25%, respectively. CC1 participants prefer an actionable metric to be both contextual (*Md*=4) and universal (*Md*=4). The former is because no two projects are the same, and the latter is because of the amount of effort invested in defining, developing, and maintaining metrics. Therefore, metrics should be versatile enough to be applicable across multiple projects. For similar reasons, CC2 participants prefer that an actionable metric be universal (*Md*=4), the project, product, and process exclusiveness notwithstanding. After having implemented metrics for multiple use cases, CC2's UC team has the informed view that customizing all, or majority, of the metrics as per individual project and development processes can be unfeasible and counterproductive. At the same time, they are neutral (*Md*=3) about an actionable metric being contextual. In contrast, CC4 participants agree (*Md*=4) that an actionable metric has to be context-driven and not necessarily always universal (*Md*=3). Due to a relatively shorter development period and smaller project size, CC4's UC team have been able to contextualize their metrics, without putting a strain on the available resources. Participants' comments like, "*Some metrics are strongly correlated with the circumstances they refer to and can mean different things in different settings",* and, "*It all depends on the use case in project*", provides a basis for CC4's above stance. Responses from CC3 participants are exceptional, as they are neutral (*Md*=3) on an actionable metric's characteristic of being contextual, but disagree (*Md*=2) that an actionable metric should be universal. CC3 has global processes, tools, organization model, etc., which individual business units interpret in a fashion that can be applied to the local case only. However, this inherent diversity and project complexity necessitate context-driven metrics. At the same time, based on the experience of deploying the Solution and the entailed resource consumption, the participants are circumspect if the effort spent justifies a metric that can be used only in certain contexts.

4.2.2 RQ2.2: What is the utilitarian perspective of an actionable metric? Many participants believe that even a non-actionable metric can be useful (69%) (*Q10*). In support of this view, one of the CC2 participants commented that, "*Non-actionable metrics might still be informative*", which is also reflected in the question, "*Does a non-actionable metric provide any valuable information?*" from one of the CC3 participants. These justifications suggest that as long as a metric provides valuable information, then it may be immaterial if that metric is actionable or not. Overall, every case company, except CC3, share similar views on *Q10*. CC3 participants are neutral (*Md=3*) in this regard. They see a clear distinction between an actionable and a non-actionable metric. The former is meant to keep practitioners informed about the current state, while the latter influences improvement actions. Ideally, CC3 participants prefer the latter case.

There is a strong support for the perspective that a metric cannot be actionable without interpretation (67%) (*Q8*). The comment, "*Not all metrics are straightforward and can mean different things according to circumstances. Yet it is better to have them than not*", from a CC4 participant provides a good basis for the overall stance on *Q8*. In contrast, CC2 are neutral (*Md=3*), and the comment that an actionable metric should be, "*easy to understand what it means, quickly*", provides a rationale for the said perspective. CC2 participants prefer metrics that are simple and straightforward. The CC2 UC Champion stated that this was the driving factor behind undertaking improvements based on the "*well-defined issues jira*" metric. The metric was simple enough for them to not be skeptical towards the suggested actionable information.

There is moderate support for the utilitarian perspective that every metric has to be actionable (12%) (*Q6*), and the reason could be inferred from the responses for the perspective that only certain metrics need to be actionable (*Q7*). For instance, CC1 participants disagree (*Md=2*) with *Q6*, because they agree (*Md=4*) with *Q7*. In contrast, CC2 (*Md=4*) and CC3 (*Md=4*) are in favor of *Q6*, because they disagree (*Md=2*) with *Q7*. Owing to the resources consumed in developing metrics and the scale and complexity of most projects, CC2 and CC3's stance on *Q6* appears justified. However, CC1 is in favor (*Md=4*) of only certain metrics being actionable (*Q7*), and opposed (*Md=2*) to *Q6*. CC1 exercises caution in labeling a metric actionable, as most metrics can only be one of the several contributors and not the primary driver of action, an argument justified by their experience of metrics' use in the Project so far.

Relatively, there is no support for the utilitarian perspective that aggregated metric is more actionable than a low-level metric (-*24%*) (*Q9*). CC2-CC4 are neutral (*Md=3*), while CC1 disagrees (*Md=2*) with *Q9*. For CC2, the need to have simpler metrics takes precedence over the *Q9* perspective, as aggregated metrics typically involve complicated calculations, which makes their maintenance difficult and require interpretation to infer actions. Additionally, one of the participants opined that an aggregated metric's actionability is, "*very project/process/product specific dependent".* CC2's rationale can be extended to CC1 and CC3, too. This rationale may not apply in case of CC4, as the participants prefer low-level metrics, which can be easily computed from the data available from the tool itself. On the other hand, high-level metrics require interpretations, which may make it difficult to extract actionable information from them.

4.2.3 RQ2.3: What are the high-quality data requirements for actionable metric? There is unanimous agreement among the participants that an actionable metric must be accurate (100%) (*Q11*), consistent (100%) (*Q12*), and credible (100%) (*Q14*). There is also strong support for an actionable metric to be complete (81%) (*Q15*) and current (75%) (*Q13*). These results are further supported by the low variability in their respective CV values.

With respect to Q13, a CC2 participant commented that, "even incomplete and fuzzy (metric) data is better than no data at all. It just needs to be consistent and continuously monitored". Another CC2 participant added that an actionable metric needs to be, "valid (for its use)…time-stamped (on top of current)…available (+source known)". In reference to Q13, one of the CC1 participants equated 'current' with 'timeliness', and argues that it, "refers to the expectation of when data should be received in order for the information to be used effectively. The expectation and reality often do not align, leading to ineffective use of the data and a lack of data-driven decisions". A CC3 participant agrees with Q13 on the condition that, "…you can see also how the metric has evolved, i.e. if it is already improving (based on actions taken) why to interfere again?" Therefore, the participants support the 'current' high-quality data requirement for an actionable metric strongly, but qualify that strong support with other finer considerations.

5 Discussion

First, we discuss the results of *RQ1*, as presented above, informed by the views and rationale provided by the invited participants. In the second part of the discussion, we discuss results from *RQ2.*

5.1 RQ1. What role did the metrics play in the improvement actions taken by the participants of the study?

Judging by the varied contexts and stances on a metric's role for each case company, a multi-faceted view can be suggested for *RQ1*.

For CC1, a standalone metric cannot compel them to act on the suggested actionable information. Other supporting factors are needed to justify a metric influencing any improvement action. The participants subscribe to the idea that actionable metrics are not "*magic*" that can explicate what action should be taken [2], but it can only guide them in the right direction, albeit only after being corroborated by other means such as original data sources or an authority at the company. This watchfulness is also a product of the default requirement that these metrics should be reliable to support decision-making [20].

In case of CC2, the argument made in [23] about simple metrics being more actionable finds validation. CC2 participants could act on the '*well-defined issues jira*' metric, as they found them simple and even reliable. In addition, the straightforward calculation for the metric enabled them to become familiar with it more easily, which is one of the success factors for use and sustainability of an MP [17]. A simple metric and a reliable metric are not necessarily mutually inclusive, but the former heavily influences the latter.

Simpler metrics also involve less maintenance, which improves their adaptability. Meeting above conditions and in contrast to CC1, a metric acted as the primary driver of CC2's improvement action.

Choice of metrics and their utility can be dictated by company size and project characteristics [8]. Nowhere is this more evident than in CC3. According to the CC3 participants, any metric can be, and should be, actionable, validating the argument that any real metric has to be actionable [2]. The resources at stake are too precious to be wasted on a metric that could, potentially, be useful. Therefore, a standalone metric can influence actions, as long as it meets the requirements of being simple, contextual, and reliable. These metrics will, most certainly, require interpretation, but that should not come at the cost of the above mandatory requirements.

Just like CC3, company size and project characteristics significantly influence a metric's role in CC4. Unlike rest of the case companies, CC4's metrics use and actions derived need to be validated by their customers first. Furthermore, in order to support rapid software development, CC4's decision-making needs to be short and quick. These unique criteria impress upon the type of metrics that can dictate improvement actions at CC4. The participants believe that low-level metrics are better as actionable metrics, which is how they have used the metrics so far, and successfully.

5.2 RQ2. What is an actionable metric according to the participants of the study?

RQ2 can be answered by combining the answers to the following three research sub-questions:

4.2.1 RQ2.1: What are the characteristics of an actionable metric? In complete agreement with the state of the art, the participants believe an actionable metric should be practical, inform decision-making [2, 25], and exhibit high-quality data characteristics [6].

A potential source of confusion about actionable metrics is whether they should directly drive decisions, or if they can only be an indicator for making decisions [12]. Based on the metrics usage at the three case companies, a case can be made for both the possibilities. CC1 adopts the latter view, whereas both CC2 and CC4 have provided examples that fall in the former category. Essentially, it is the organization's prevalent view, informed by the context and the overall trust in the metrics, which decides what an actionable metric means to them. More importantly, any improvements driven by metrics are a result of several factors working in harmony [12], which is evident in CC1.

There is little difference in agreement over an actionable metric being project/process/product specific (contextual) and universal. Most participants do not agree as strongly to these two characteristics as they do to others. However, the empirical account of the case companies using metrics for taking actions contradict the low support for *Q3*. Even the literature supports that an actionable metric is typically characterized by the development process where it is intended to be used [12]. It is crucial that users understand what a metric truly entails to be able to take any actionable approach [23], and that cannot be accomplished without embedding contextual information in that metric, and users being aware of it. In contrast, the low support for *Q4* is justified, as none of the empirical account alludes to such a characteristic. As lessons learned from one project cannot always be applied to another [13], the same could be said for an actionable metric's utility and relevance across different projects.

4.2.2 RQ2.2: What is the utilitarian perspective of an actionable metric? The two topmost utilitarian perspectives for an actionable metric is that a metric need not be actionable for it to be useful, and that an actionable metric requires interpretation. The former perspective is true, as long as the metric is providing information of some value. For instance, Meneely [12] argues that non-actionable metrics are quite common, and can be useful in communicating symptoms of a problem (e.g. number of bugs). However, unlike actionable metrics, they do not provide any diagnosis. Next, empirical account of the case companies' metrics use support the *Q8* perspective. As long as one is aware of the context inherent in an actionable metric, the interpretation should be as less as possible. This is one of the reasons why CC2 participants could use the metric without any hesitation. The metric was simple, and reflected the process context accurately and sufficiently, with which the participants became familiar quickly.

There is some support for the perspectives that every metric has to be actionable, and conversely, only certain metrics need to be actionable. Judging by the metrics' role in the case companies, both perspectives are justified, albeit not as important as *Q8* and *Q10*. Ideally, practitioners would like every metric to be actionable. Reflecting this view, Croll and Yoskovitz [2] distinguish between a vanity metric and a real metric. The authors claim that an actionable metric is a real metric, whereas a vanity metric provides only cosmetic value, with no impetus to undertake any change. Some participants argue that a non-actionable (or vanity) metric can still be useful, supporting the proposition made in [12].

Relatively, the perspective that aggregated (or high-level) metrics can be more actionable than low-level metrics does not find much support in our study. CC4 is a proponent of low-level metrics, and that has proven beneficial for them. In contrast, most of the metrics used to generate quality alert [14] in CC1 are aggregated metrics. This is also the case with CC2's use of the metric to improve their practice of maintaining the *DoD* field in Jira. In general, higher level managers prefer aggregated metrics, while lower-level managers and developers tend to rely on detailed (low-level) metrics [1]. The size of the company and project characteristics can also dictate what type of metrics are preferable for inferring improvement or corrective actions. This claim is reflected in CC3 (large size, complex projects) and CC4's (small size, small projects) choice of metrics for generating actionable information. However, to the best of our knowledge, we have not come across any literature exploring the interesting influences of CC4's decision-making structure on a metric's actionability. CC4 is likely to be an exceptional case, and their circumstances and context may not be generalizable to other similar organizations.

4.2.3 RQ2.3: What are the high-quality data requirements for actionable metric? One of the success factors for a metrics program is management commitment [7], and that is easier to obtain if

metrics data can be used in the process of decision-making [20]. However, managers need assurances of the data quality in order to trust the metrics enough to act on them [20]. Evaluating a metric's data quality against an established data quality standard can facilitate all of the above. The Data Quality Model prescribed in ISO/IEC 25012:2008, SQuaRE [6] includes inherent data quality requirements that our questionnaire refers to, and to which the participants agree. There is unanimous agreement that actionable metrics data should be *accurate* (data is precise), *consistent* (data agrees with its context) and *credible* (data is correct), followed by a strong agreement for the *complete* (no missing data) and *current* (data is recent/timely) requirements.

The '*current'* high-quality data requirement is critical for ensuring effective use of data for data-driven decision-making. Current literature supports this stance [20], where data timeliness is considered as one of the criteria to determine if information is of sufficient quality for a manager to rely upon for decision-making. Ideally, a metric should provide actionable data faster than the rate of change of effects within a project to facilitate real-time decisions [13]. Next, one of the participants also pointed out that a metric should be valid for its intended use, which suggests that the metric should be contextual. This quality requirement is similar to *Q3* (project/process/product specific) characteristic, which both our study and the state of the art [2, 25] support.

An interesting view expressed by one of the participants was that an actionable metric data can be incomplete, as long as it is consistent and continuously monitored. The '*complete*' data quality requirement can be classified as a contextual quality [10], because the data quality requirement's relevance and importance is based on the context of the task at hand. Therefore, based on the context, it is possible that a metric's *completeness* can be traded for consistent and continuously monitored incomplete data. Such modulation of data quality requirements based on context is an interesting avenue warranting further and extensive exploration.

6 Threats to validity

We address threats to our study's validity based on the guidelines recommended by Runeson and Höst [18].

The questionnaire statements have been derived from the state of the art, but they are not validated constructs. Although this data collection instrument is not used for a survey, this shortcoming can still pose a threat to our study's construct validity. Combining the questionnaire responses with the empirical accounts of actionable metric use, and further validation by a practitioner from each case company help us mitigate this threat.

There is a possibility that the empirical accounts of a metric's use for taking actions are influenced by other unforeseeable factors, thereby posing a threat to the internal validity of our study. We mitigated this threat by asking the practitioners involved in using metrics for decision-making, to report on those experiences and help validate the findings.

Our study involves only four companies, and the participants for the questionnaire are based on convenience sampling. This affects the external validity of our study. However, owing to the diverse nature of the case companies, our study's findings could be applicable and beneficial to organizations that are similar in context to the four case companies. More importantly, our study's aim is to trigger further discussion and investigation on the research topic, rather than just generalizability.

Multiple researchers and practitioners helped elaborate the reported use of metric for decision-making, and validate the findings from the questionnaire. However, only one researcher was involved in data collection, which may affect the reliability of our study.

7 Conclusion

Ideally, a metrics program should facilitate data-driven decision-making. However, practitioners are wary of relying on metrics for decision-making, unless they are reliable. Besides reliability, the state of the art is not clear on what other factors govern a metric's property to be actionable. There is a long history of research on MPs [9], and success factors necessary for them [4, 11, 17, 21]. However, the literature on identifying the specific characteristics that make a metric actionable is scarce. In the context of the EU H2020 Project, Q-Rapids, we attempted to address this research gap by collaborating with the four industrial partners.

Building upon the empirical accounts of metrics' use for decision-making at the case companies, we administered an online questionnaire to document the involved practitioners' views on actionable metrics. Furthermore, we invited a practitioner from each case company to validate the role of metric's use for taking actions, and elaborate on the views recorded in the questionnaire. We found that some practitioners believe that a metric can mostly contribute to an action, while others acknowledge the potential of a standalone metric to drive improvement actions. According to the practitioners using the Solution, an actionable metric should be practical, inform decision-making, and exhibit high data quality. Despite moderate support, empirical evidence suggests that an actionable metric must be project/process/product specific (contextual). In utilitarian terms, even a non-actionable metric can be useful, but a metric cannot be actionable without interpretation, unless the metrics are simple. The more simple and contextual a metric is, the less interpretation is required to infer actionable information from that metric. There is some support for the perspectives that every metric has to be actionable and only certain metrics can be actionable. Despite the contradiction, the evidence suggests that both can be valid. Similarly, despite the relative lack of support for the perspective that an aggregated metric is more actionable than a low-level metric, we found that depending on the company size and project characteristics, both aggregated metrics and low-level metrics can provide actionable information. Intriguingly, a company's decision-making requirement can also influence the type of metric that can prove actionable. Among the high-quality data requirements for actionable metrics, the practitioners report that an actionable metric should be accurate, consistent, credible, current, and complete. However, the context can help trade the '*complete*' requirement in favor of another.

A potential future work could involve evaluating the reported viewpoints of actionable metrics across several other organizations to validate and improve upon them. Furthermore, a large-scale

survey could be employed to collect more responses for the questionnaire to improve current findings' generalizability.

Acknowledgments

This work is a result of the Q-Rapids Project, funded by the European Union's Horizon 2020 research and innovation program, under grant agreement No. 732253.

REFERENCES

[1] Raymond P.L. Buse and Thomas Zimmermann 2011. Information Needs for Software Development Analytics - Microsoft Research. *MSR Tech Report 2011-8*. (2011), 1–16.

[2] Alistair Croll and Benjamin Yoskovitz 2013. *Lean Analytics: Use Data to Build a Better Startup Faster*.

[3] Xavier Franch, Claudia Ayala, Lidia López, Silverio Martínez-Fernández, Pilar Rodríguez, Cristina Gómez, Andreas Jedlitschka, Markku Oivo, Jari Partanen, Timo Räty, and Veikko Rytivaara 2017. Data-driven requirements engineering in agile projects: the Q-rapids approach. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference Workshops, REW 2017*. (2017), 411–414. DOI:https://doi.org/10.1109/REW.2017.85.

[4] Tracy Hall and Norman Fenton 1997. Implementing effective software metrics programs. *IEEE Software*. 14, 2 (1997), 55–64. DOI:https://doi.org/10.1109/52.582975.

[5] Hennie Huijgens, Davide Spadini, Dick Stevens, Niels Visser, and Arie Van Deursen 2018. Software analytics in continuous delivery: A case study on success factors. *International Symposium on Empirical Software Engineering and Measurement*. (2018). DOI:https://doi.org/10.1145/3239235.3240505.

[6] ISO 2008. *ISO/IEC 25012:2008, Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model*.

[7] Tapani Kilpi 2001. Implementing a Software Metrics Program at Nokia. *IEEE Software*. 18, 6 (2001), 72–77. DOI:https://doi.org/10.1109/52.965808.

[8] Eetu Kupiainen, Mika V. Mäntylä, and Juha Itkonen 2015. Using metrics in Agile and Lean software development - A systematic literature review of industrial studies. *Information and Software Technology*. 62, 1 (2015), 143–163. DOI:https://doi.org/10.1016/j.infsof.2015.02.005.

[9] Frank Van Latum, Rini Van Solingen, Markku Oivo, Barbara Hoisi, Dieter Rombach, and Gunther Ruhe 1998. Adopting GQM-Based Measurement in an Industrial Environment. *IEEE Software*. 15, February 1998 (1998), 78–86. DOI:https://doi.org/10.1109/52.646887.

[10] Yang W. Lee, Diane M. Strong, Beverly K. Kahn, and Richard Y. Wang 2002. AIMQ: A methodology for information quality assessment. *Information and Management*. 40, 2 (2002), 133–146. DOI:https://doi.org/10.1016/S0378-7206(02)00043-5.

[11] Manoel G. Mendonça and Victor R. Basili 2000. Validation of an approach for improving existing measurement frameworks. *IEEE Transactions on Software Engineering*. 26, 6 (2000), 484–499. DOI:https://doi.org/10.1109/32.852739.

[12] Andrew Meneely 2016. Actionable metrics are better metrics. *Perspectives on Data Science for Software Engineering*. Elsevier. 283–287.

[13] Tim Menzies and Thomas Zimmermann 2013. Software analytics: So what? *IEEE Software*. 30, 4 (2013), 31–37. DOI:https://doi.org/10.1109/MS.2013.86.

[14] Marc Oriol, Pertti Seppänen, Woubshet Behutiye, Carles Farré, Rafal Kozik, Silverio Martínez-Fernández, Pilar Rodríguez, Xavier Franch, Sanja Aaramaa, Antonin Abhervé, Michal Choras, and Jari Partanen 2019. Data-Driven Elicitation of Quality Requirements in Agile Companies. *Communications in Computer and Information Science* (2019), 49–63.

[15] K Petersen and C. Wohlin 2011. Measuring the flow in lean software development. *Software - Practice and Experience*. 41, 9 (2011), 975–996. DOI:https://doi.org/10.1002/spe.

[16] Dan Port and Bill Taber 2017. Actionable Analytics for Strategic Maintenance of Critical Software: An Industry Experience Report. *IEEE Software*. 35, 1 (2017), 58–63. DOI:https://doi.org/10.1109/MS.2017.4541055.

[17] Prabhat Ram, Pilar Rodriguez, and Markku Oivo 2019. Success Factors for Effective Process Metrics Operationalization in Agile Software Development : A Multiple Case Study. *Proceedings of the 2019 International Conference on Software and System Process* (2019).

[18] Per Runeson and Martin Höst 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*. 14, 2 (2009), 131–164. DOI:https://doi.org/10.1007/s10664-008-9102-8.

[19] Vibhu Saujanya Sharma and Vikrant Kaulgud 2013. Adoption and use of new metrics in a large organization: A case study. *International Workshop on Emerging Trends in Software Metrics, WETSoM*. (2013), 21–27. DOI:https://doi.org/10.1109/WETSoM.2013.6619332.

[20] Miroslaw Staron and Wilhelm Meding 2009. Ensuring reliability of information provided by measurement systems. *Software Process and Product Measurement, International Conferences IWSM 2009 and Mensura 2009* (Berlin, Heidelberg, 2009), 1–16.

[21] Miroslaw Staron and Wilhelm Meding 2012. Factors determining long-term success of a measurement program: An industrial case study. *e-Informatica Software Engineering Journal*. 1, 1 (2012), 7–23. DOI:https://doi.org/10.2478/v10233-011-0027-z.

[22] Miroslaw Staron and Wilhelm Meding 2015. Transparent measures : cost-efficient measurement processes in SE. *Software Technology Transfer Workshop* (Kista, Sweden., 2015), 1–4.

[23] Daniel Vacanti and Bennet Vallet 2014. Actionable Metrics at Siemens Health Services. September 2011 (2014).

[24] Ye Yang, Davide Falessi, Tim Menzies, and Jairus Hihn 2018. Actionable analytics for you. *IEEE Software*. 35, 1 (2018), 51–53.

[25] Dongmei Zhang, Shi Han, Yingnong Dang, Jian Guang Lou, Haidong Zhang, and Tao Xie 2013. Software Analytics in Practice. *IEEE Software*. 30, 5 (2013), 30–37. DOI:https://doi.org/10.1109/MS.2013.94.