Socially-Constructed Metrics for Agile Quality: An Action-Research Study

Sharon Coyle
Business Information Systems
University of Sydney Business School
Australia, NSW 2006
Email: sharon.coyle@sydney.edu.au

João Barata
CISUC, ISMT, and ESTGOH
University of Coimbra
Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
Email: barata@dei.uc.pt

Abstract

We present a method to develop socially-constructed metrics for ascertaining agile software development quality. Canonical action research (CAR) is our mode of inquiry, conducted in a key European player of healthcare information systems. The result is a set of meaningful metrics that are built according to three interrelated dimensions: (1) evidence from practice; (2) stakeholders' expectations; and (3) stakeholders' evaluation. Our contribution suggests simple artifacts to create socially-constructed metrics and the main guidelines for applying them. Agile teams struggle with quality measurement, often supported by a plethora of metrics that do not adhere to rapidly changing project environments. We argue that socially-constructed metrics can address this problem, offering a contextualized perspective of quality that can improve tacit knowledge transfer; critical reflection about quality; and effective support in daily meetings, retrospectives, and audits. Moreover, it suggests a participative approach for continuous improvement in agile software development.

Keywords Socially-constructed metrics, agile development, quality management, action-research.

1 Introduction

The Agile Manifesto was introduced in 2001 with a goal of "*uncovering better ways to develop software*" (Beck et al. 2001) endorsing an iterative process involving intense stakeholder interaction throughout so as to develop a product of high quality that meets customer's expectations. Metrics have been established in order to determine the meaning of 'quality' and are a popular research topic in agile software development (Agarwal et al. 2014; Hayes et al. 2014; Kupiainen et al. 2015). There are studies focusing on product or software-related metrics (Kupiainen et al. 2015; Mishra et al. 2012), tests and quality control (Agarwal et al. 2014; Janus et al. 2012), software defects (di Bella et al. 2013), stakeholder expectations (Boerman et al. 2015) and the role of auditing (Scharff 2011). Literature provides guidance about metrics used in practice however, there are difficulties in adopting these quality metrics in dynamic project environments (such as agile), which are significantly different from their traditional ('waterfall') counterparts. Traditional assessments of quality focus primarily on outcome-related indicators such as product or overall project quality. Kupiainen et al. (2015) found that even in the application of identified metrics, almost 40% of these were customized. Their conclusions state that the majority of existing metrics are non-inclusive of people. This creates significant challenges given the nature of agile projects, which are inherently people-focused. Other metrics focus on the development process (Gruschwitz and Schlosser 2012) yet a lack of solutions exist that integrate different types of metrics in a single method that can be practically applied in agile projects. Moreover, existing methods do not promote stakeholder engagement for metric construction yet individuals and their interactions are a key principle of agile methodologies (Beck et al. 2001).

For the purpose of our research, a metric is socially-constructed (Berger and Luckmann 1991) when users have the capacity to adjust its dimensions and critically evaluate the results. In this context, the metric is not a mere observation of a fact because stakeholders' opinions are intrinsic to metric construction. Unlike traditional approaches, which compare against predefined goals, stakeholders are not just included at the end of measurement analysis; they become involved in constructing relevant metrics for their processes, project, and product(s). The difficulties of including quality assessment in agile development teams with the vision of socially-constructed metrics inspired our first research question (RQ1): How can agile teams develop socially-constructed metrics during their deployment of agile methodologies? Moreover, it is essential to test those metrics in practice, which we aim to address with RQ2: What are the advantages and disadvantages of using socially-constructed metrics in agile software development (ASD) teams?

The next section outlines the research background including quality management in agile and a review of different approaches for quality assessment and improvement including their importance and limitations. Next we present the selected research approach that is action research in its canonical form (Susman and Evered 1978). We subsequently detail a complete canonical action research cycle conducted in a leading IT supplier of healthcare information systems. The lessons learned and the results are presented afterwards, concluding with our study's limitations and future research.

2 Background

2.1 Quality Management in Agile

In information systems (IS), quality management is multidimensional including social and technical aspects. According to (Stylianou and Kumar 2000), holistic enterprise quality is a combination of IS quality and the quality of business processes. A strong quality culture encompasses customer orientation, continuous improvement, utilisation of data (and analysis) to support decisions and the involvement of people in quality problems (ISO 2015). This aligns closely with agile principles and practices. For example, the notion of continuous improvement is embedded in the practice of retrospective meetings in agile projects (Babb et al. 2014; McHugh et al. 2012). Therefore, the pattern of common quality principles determined by ISO 9001 that certified companies learn and internalize in their daily practices can be aligned with agile values (Stålhane and Hanssen 2008). In highly regulated development projects however there are reported difficulties of adopting different quality standards and improvement frameworks (including ISO 9001, ITIL, COBIT, and CMMI) with that of agile. Stålhane and Hanssen (2008) for example discuss difficulties in documentation requirements when combining ISO 9001 with agile approaches. However, a technical report provided by Hayes et al. (2014) identifies different moments in agile projects where it is possible to get customer feedback to assess their satisfaction. These are illustrated in Figure 1.

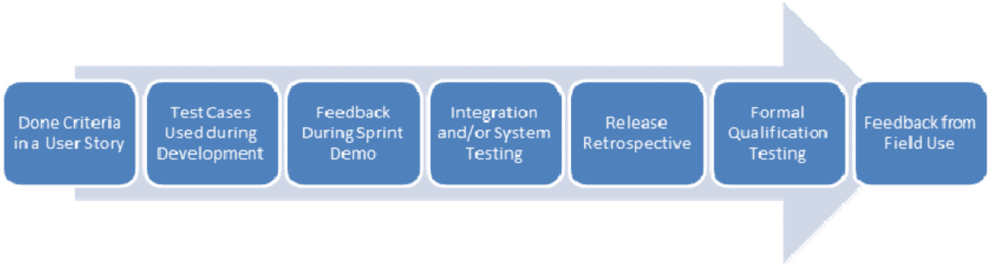

Figure 1: Quality touch-points in agile development (Hayes et al. 2014)

Figure 1 shows that quality requires a continuous effort during the entire project. The evaluation of results within specific meetings (e.g. retrospective) can be important in promoting discussion and conducting critical reflections about quality. Retrospectives allow reflection about previous iterations to identify subsequent actions needed and there are authors such as Péraire and Sedano (2014) who conclude that artifacts and guiding steps for retrospective meetings can provide distinct advantages. Nevertheless, as Baxter and Sommerville (2011) put it "*the agile approach of involving end-users as 'owners' of requirements is a good one but needs to be extended to take into account a broader set of system stakeholders*". There are threats to quality management in agile due to the constant pressures that can make reflection and analysis difficult in practice (Babb et al. 2014; McHugh et al. 2011). Moreover, ASD projects present different challenges when compared to traditional approaches, namely, "*the traditional approach of tracking progress against a pre-made plan and measurable goals conflicts with the Agile value of embracing the change* [... and its] *rather comprehensive set of metrics, which does not align well with the Agile principle of simplicity*" (Kupiainen et al. 2015). Agile methods go beyond the traditional views of quality such as measuring defects or functionality problems (Hayes et al. 2014). Quality concerns appear in the early stages of agile projects, proceeds in the complete documentation of user stories and "*can be supplemented with a more direct measure of customer-perceived value—using customer satisfaction feedback*" (Hayes et al. 2014).

2.2 Approaches for Quality Measurement and Improvement in Agile

Several approaches have been proposed for establishing quality in the context of agile practices. For example the 3C approach proposed by Janus et al. (2012) which combines software metrics and continuous integration, concluding that interpretation of results is necessary to promote continuous improvement actions. The model proposed by Hongying and Cheng (2011) include 20 key areas for agile software quality assurance. These authors suggest best practices for each area and a maturity model approach for evaluation and improvement. An earlier approach proposed by Sidky et al. (2007) to adopt agile quality principles, consists of two components. The first component includes an agile adoption index for the principles of "Embrace change to deliver customer value", "Plan and deliver software frequently", "Human centric", "Technical excellence", and "Customer collaboration". The second component is a four-stage process for agile adoption, guiding companies to (1) identify discontinuing factors that can prevent agile success, (2) conduct project-level assessment, (3) organizational readiness assessment, and (4) reconciliation to ensure that the organization implements practices required for the project. Sidky et al. (2007) present one of the few examples that includes guidance for assessment and improvement according to the goals established for agile practices. A distinct hierarchical model was developed by Bansiya and Davis (2002) to assess object-oriented design quality of software products and obtain a total quality index. The first level of the model includes product related attributes such as functionality, effectiveness, understandability, extendibility, reusability, and flexibility. The second level details properties that can affect the attributes (such as complexity), defining weights for each property and its positive or negative influence for each attribute (e.g., complexity has a negative influence on understandability). The model has two more specific levels, namely (3) design metrics (such as 'number of methods'), and (4) design components that are needed for metrics.

Social aspects, process, and outcome are deeply intertwined in iterative agile development projects. Recent studies to assess agility in enterprises (e.g. Tseng and Lin 2011) include social aspects such as personal skills, technology awareness, trust-based relations with customers, collaboration, empowerment, and motivation. Gren et al. (2015) identify different social approaches for assessing agility in teams, for example, using interviews or maturity models to guide the adoption of agile techniques but stress that "*more work is needed to reach the point where a maturity model with quantitative data can be said to validly measure agility, and even then, such a measurement still needs to include some deeper analysis with cultural and contextual items*". This research aims to help address this gap. Existing models do not incorporate people or their interactions into metric construction. According to Ghobadi and Mathiassen (2016) in order "*to bridge communication gaps and create shared understanding in software teams, it is critical to take the revealed concerns of different roles into account*". To date, a model that integrates different views of people, process, and outcome is absent in literature. In addition, the perspectives outlined above are applied singularly. They are also usually applied '*after the fact'* and therefore are very difficult to apply in iterative, dynamic ASD environments.

3 Research Approach

According to Baskerville (1999), from a socio-organizational viewpoint it is essential to study new techniques in practitioner environments. Action-research is well suited for this purpose in the field of IS (Baskerville 1999) as it is performed "*collaboratively in an immediate situation using data feedback in a cyclical process*" (Hult and Lennung 1980). Action-research encourages the interaction between the researcher and external clients, consequently contributing to some current challenges encountered in IS research (Gill and Bhattacherjee 2009). Amongst the multiple forms of action-research, we selected canonical action research (CAR) as one of the most popular and well documented (Davison et al. 2004). CAR cycles are conducted according to five phases (Lindgren et al. 2004; Susman and Evered 1978):

1. Diagnosing, identifying, or defining the problematic situation, as a shared task by the researcher and practitioner. The actors holistically interpret the phenomenon and formulate working hypothesis to be used in the subsequent phases of the cycle;
2. Action planning, specifying possible courses of action to improve the problematic situation;
3. Action taking, referring to the implementation of the course of action, causing change to occur and trying to create improvements to the situation;
4. Evaluating, assessing the consequences of the actions, involving a critical analysis of the results;
5. Specifying learning, identifying the findings, documenting and defining the outcomes that will add to the body of knowledge. Although appearing last, this phase is a permanent activity (Baskerville and Wood-Harper 1996; Cunha and Figueiredo 2002).

The perception of CAR as "context-bound" creates problems in generalizing the findings (Avison and Wood-harper 2003), however, there are different views regarding the degree to which generalization is required (Gregor 2006); the action researcher should look for transferable results. For example, Eden and Huxham (1996) assert that (1) there must be implications beyond those required for action in the specific project context, allowing it to inform other contexts; (2) there is a need to produce theory that is significant to others; (3) in the case of designing tools, techniques, models, and methods, its basis must be clear and linked to theory; (4) theory emerges from action and previous knowledge; and (5) theory building is incremental in action research, moving gradually from the particular to the universal. To ensure rigor and validity we evaluated our research according to the principles suggested by Davison et al. (2004), specifically for CAR: Principle of the Researcher–Client Agreement; Principle of the Cyclical Process Model; Principle of Theory; Principle of Change through Action; and Principle of Learning through Reflection. In the next section we depict the complete CAR cycle (Susman and Evered 1978).

4 Data Collection

4.1 Client-system Infrastructure

Our client is a European software provider of healthcare information systems for hospitals and clinics. Founded 25 years ago they are present in four continents, serving over 120,000 users and 25 million clinical processes. The company has migrated its quality management system to the recently revised 2015 version of ISO 9001:2015. Their regulatory space includes other specific standards for innovation management and healthcare standards that apply to their software product lines and operating context (such as data quality and record privacy). Quality management is essential to remain competitive and compete in different regions, as with its high-growth American market where the company achieved important contract agreements in recent years. Their global presence increases pressure for short development cycles and immediate feedback to their costumers and national partners conducive to an agile approach.

4.2 Diagnosing

The diagnosis included interviews with the quality manager and IT infrastructure manager. Simultaneously, we conducted a literature review to identify best practices for ascertaining quality and the role of metrics in quality assessment and improvement (Section 2). Metrics in the health sector are plentiful however, as stated by the quality manager, the company "*has numerous indicators but only a few are valid for agile quality*". The reasons vary because in some cases "*the numbers are highly dependent on the context and must be carefully interpreted*". In other cases "[she] *does not think it is fair to establish goals, for example regarding number of defects or features implemented; these type of metrics depend on multiple factors*". Agile quality is problematic to them because "*40% of our major customers* [representing 80% of income] *require quality indicators and evidence for each iteration, due to the critical nature of healthcare IT*". We confirmed the importance of retrospectives for quality in agile because in this case a lack of adequate implementation of retrospectives contributed to "*difficulties in creating improvement on our project and without appropriate communications we are not sharing knowledge which is a critical aspect of our business due to the complexity of product lines*". This research participant also talked about the importance of being able to change metrics for each project or team, in an "agile way" that coincides with agile principles. According to our interviewee, quality metrics provide interesting dashboards "*but what we need is to assess and improve quality; it cannot be done with ceremonial conformity or high level metrics that do not have correspondence with practice*". Even worse, *"template"* metrics "*and unrealistic goals can reduce the team commitment to quality during agile projects*". When we asked how user intervention might assist in constructing metrics she stated how "*this would be a very useful, inclusive approach [and that] it has the potential to address our main issues of (1) knowledge sharing, (2) obtaining quality evidence for our team and external audits, (3) re-invigorating our retrospectives, (4) providing support for weekly meetings and customer requests, (5) and 'provide meaning' to our agile numbers!"*

4.3 Action planning

Our action plan for this CAR cycle included four main activities:

- Establish a model to create metrics. The model should assist project participants in the identification of the types of metrics and how to calculate them;
- Define the indicators that should be included for each metric type (people, process, outcome);
- Establish the structure of each indicator (how it is calculated) according to three possible dimensions of (1) evidence from practice; (2) stakeholders expectations; and (3) stakeholders evaluation;
- Develop a tool to manage metrics that can be useful for daily meetings, retrospective, and quality audits.

The plan, agreed by researchers and practitioners, aimed at solving a practical problem while contributing to research in the form of a new method to use metrics that adhere to the principles of agile, in particular people and interaction. Moreover, we wanted to provide practical tools in the form of tables accessible to agile teams and not dependent of specific technologies. The CAR cycle started in March 2016 and ended on July 2016. The next section presents the results of action taking.

4.4 Action taking

First, we agreed on a reference model to guide the construction of metrics, presented in Figure 2.

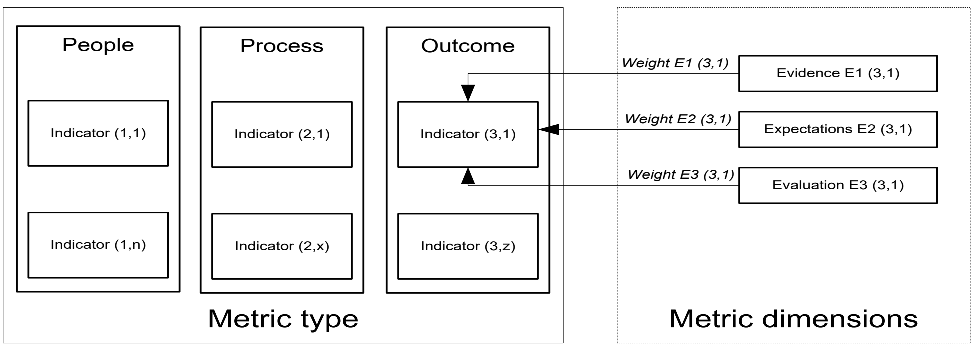

Figure 2. Model to create socially-constructed metrics for agile quality

According to our review, a comprehensive assessment of agile quality requires three main types of metrics, represented to the left of Figure 1: (1) people-related, (2) process-related, and (3) outcome-related pertaining to the specific project and the product. Moreover, our proposal of socially-constructed metrics allows users to create composite metrics, as presented to the right of Figure 2. The resulting indicator will include a comparison with past results to identify if improvement occurred (evidence); a comparison with the expected result according to the stakeholders' initial plan (expectations); and finally, critical analysis performed by the agile team (evaluation). The final result of each indicator is in fact a weighted average of each of its dimensions E1, E2, and E3. As a reference to weight the dimensions of each selected indicator, we used the suggestions included in Table 1 while Table 2 describes guidelines to evaluate each indicator according to the selected dimensions.

Dimension	**Definition**	**Potential ways to consider weightings**
Evidence	Quality is based on facts. Evidence represents the effective improvement of the indicator comparing it with the backlog.	If the indicator is not significantly affected by uncontrolled aspects, the weight can be higher.
Expectations	There are goals to achieve in agile development. There are technical goals (e.g., reduce defects), social goals (e.g. improve motivation), or other.	If the indicator is mostly influenced by stakeholders' decisions, the weight can be higher.
Evaluation	Agile quality requires reflection and debate (e.g. about the meaning of the data) and to identify lessons learnt.	If the indicator is not consensual or it is highly variable according to external factors, the weight can be higher.

Table 1. How to weight each dimension of (1) evidence, (2) expectations, and (3) evaluation

Dimension	**0 (regression)**	**50 (no improvement)**	**100 (clear improvement)**
Evidence	Worse comparing to last measurement	Similar to last result	Better than last measurement
Expectations	Below expectations	Within expectations	Better than expected
Evaluation	Negative opinion	Neutral opinion	Positive opinion

Table 2. How to calculate / value dimensions of each indicator

The second activity was to establish the indicators. We faced several difficulties at this stage because the company had dozens of indicators but did not have the practice of using them as an improvement tool. We decided to use specified indicators for each metric type and establish the rule that each type should have at least 1 indicator. Thirdly, for each indicator, the team decided the weights that needed to apply for the dimensions of evidence, expectation, and evaluation. Figure 3 presents the tool that was developed for using socially-constructed metrics in practice and as a result constitutes our fourth activity in CAR action phase.

	People					
	Costumer Satisfaction	*wP1*	*Team Satisfaction*	*wP2*	*Improvement Suggestions (Internal)*	*wP4*
Evidence	100	*0,5*	0	*0,2*	0	*0,2*
Expectation	100	*0,2*	50	*0,3*	50	*0,4*
Evaluation	100	*0,3*	50	*0,5*	100	*0,4*
Total (0-100)	***100***		***40***		***60***	
Comments					Failures in implementation	
Corrective Actions						
Improvement Actions					Contest ideas	

Figure 3. Socially-constructed metric in practice – Template for People

	Process							
	Automated Tests Success Rate	*wPr1*	*Open Incidents*	*wPr2*	*% Incidents - Expired Due date*	*wPr3*	*Schedule Efficacy*	*wPr4*
Evidence	50	0,5	100	0,3	100	0,3	100	0,3
Expectation	50	0,5	100	0,6	100	0,6	100	0,4
Evaluation	0	0	50	0,1	50	0,1	100	0,3
Total (0-100)	***50***		***95***		***95***		***100***	
Comments			Holydays decrease number of incidents		Holydays decrease number of incidents			

Figure 4. Socially-constructed metric in practice – Template for Process

	Outcome							
	Implemented Features	*wO1*	*Failed Features*	*wO2*	*Critical defects sent by customer*	*wO3*	*% Improvement Features*	*wO4*
Evidence	50	0,2	50	0,2	100	0,3	100	0,5
Expectation	50	0,6	50	0,4	100	0,5	50	0,3
Evaluation	50	0,2	100	0,4	50	0,2	50	0,2
Total (0-100)	***50***		***70***		***90***		***75***	
Comments			Lacking complete information		Few updates sent to customers		Holydays decrease requests (increase % of improvements)	

Figure 5. Socially-constructed metric in practice – Template for Outcome

Figures 3-5 include tables that we used to assess (1) people-related, (2) process-related, and (3) outcome-related metrics. We selected three for people (the columns were provided by the team: customer satisfaction, team satisfaction improvement, suggestions (internal)); four indicators for process (Figure 4) and another four concerning outcome (Figure 5). The aggregated result of each socially-constructed metric (line Total ranging from 0 to 100 that exists in each Figure is a weighted average. For example, for "customer satisfaction" in Figure 3 (column 1), evidence is weighted 0,5; expectation 0,2; and evaluation 0,3. In order to simplify grading in this instance each dimension can have a measure of 100 (clear improvement), 50 (no improvement), and 0 (regression) however deployment of a continuous scale is also an option. Below each Figure, the project stakeholders can provide comments about the interpretation of results and propose actions that remain in the table as long as they are active. Our model does not define or prescribe metrics; therefore, each team can select metrics according to their project or client priorities.

4.5 Evaluating

To ensure rigor and relevance we adopted the following principles (Davison et al. 2004):

- *Principle of the Researcher–Client Agreement*

Researchers and practitioner agreed that CAR was an appropriate approach to study socially constructed metrics in practice. The practitioner made an explicit commitment to the project and to adopt our proposed solutions within their teams. Their main objective is to improve quality assessment and improvement, making use of meaningful metrics that they can apply simply to their project. Data collection included interviews, observation, and document collection, safeguarding confidentiality.

- *Principle of the Cyclical Process Model*

Our research followed the five stages of CAR according to Susman and Evered (1978). We created our frame of reference for CAR with a literature review and semi-structured interviews (Barata and Coyle 2016). Then, we made a diagnosis of the situation in the selected company. During action taking, researchers and the quality manager developed an action plan and conducted a continuous evaluation according to the principles suggested by Davison et al. (2004). To minimize threats to validity two researchers proceeded in parallel, constantly contrasting data sources and challenging the results. Due to time constraints, we considered that one CAR cycle was appropriate, however, we identified opportunities for future research (presented later).

- *Principle of Theory*

Theory guided our research providing a theoretical frame of reference via the literature review. We were guided by existing theory in agile metrics and models to improve agile quality. We then proposed a new solution to share within the scientific community. Our proposal can support agility by (1) introducing flexibility in indicators' selection and weightings, (2) promoting continuous interaction and (3) critical evaluation and debate to accommodate variable factors of project environments.

- *Principle of Change through Action*

Change occurred in a number of situations. First, we created a new way of using and calculating metrics in the practitioner organization, including self-evaluation within a metric structure. We have created artifacts and promoted new routines (Pentland and Feldman 2008) to guide the development team and the quality department. The situation of this IT organization and its context was evaluated before, during, and after the intervention, ensuring that change was analysed and properly documented.

- *Principle of Learning through Reflection*

Progress reports were provided to the client. Learning trough reflection occurred as a joint activity by researchers and practitioners in different stages of CAR. There was a joint reflection to ensure that our results would be relevant for science and help to improve the client situation and ensure project results. We learned about the benefits of the method but also the challenges emerging from critical analysis and composite metrics that require explanation. The feedback was positive but new questions emerged, for example, *"should we create rules to enforce corrective actions bellow value X? Should we enforce explanations if the evaluation differs from the other two dimensions (e.g., evaluation 0 when the other dimensions receive 100)?"* These are questions we plan to tackle in our next research cycles.

5 Discussion

This research included self-evaluation by development team members. Future research cycles will include customer assessments to cross-check different perspectives and promote the quality debate. As this is one of the first studies aimed at unravelling socially-constructed metrics for agile quality we encountered some challenges. Firstly, questions emerged regarding the selection of indicators for people, process, and outcome. Our option was to look across literature and within the organisation for existing indicators. This minimized the overhead in deploying the tool in practice and we reduced the number of indicators to a maximum of four for each type of metric. We selected indicators that were directly relevant to the project and company's priorities at that time. However, we have constructed the model so that in future, the set of indicators can adapt. How to allocate weight to each dimension of the indicator and its grading was also cause for debate. The weights were selected by the managers in this cycle but we intend to provide a workshop in future cycles to define the indicators and weights, according to the guidelines presented in Tables 1 and 2. Allocating values to each indicator proved to be an incredibly insightful process. It opened up discussions as to what constitutes agile quality, the prescribed or extended practices, organizational goals and so on. In Figure 3, the weighted value of the indicator (60) is the least important part when compared to the [process] debate that included the search for solutions and opening communication between team members and management.

On analysing the "process" metric in Figure 4, indicators "Open Incidents" and "% Incidents - expired due date"; initially, there was a decrease in both indicators for the last week ('100' for evidence) and it was clearly below their established target ('100' for expectations), but the team highlighted that customer holidays are usually a period of less incidents so their number and percentage allocations are not justifiably comparable with other periods. They considered this as 'normal' but not excellent, the latter of which would be interpreted if we only looked at the value comparing to a pre-determined target. Outcome-related metrics (Figure 5) are also insightful for example, (1) On initial inspection, "failed features" present worrying results but the reason attributed to this was external to the team (problems in information completeness); (2) "critical defects sent by customers" clearly improved comparing to target (expectation) and past values (evidence) but the team attributed this to a reduction in system updates; and (3) "% of improvement features" increased compared to previous periods ('100' in evidence), while still not on target ('50' in expectations). The main reason attributed to this improvement was that by being a % of value, it increased because the total number of features decreased, making the number of improvements more significant in an artificial (rather than meaningful) way.

Corrective actions and improvement actions are important to this process (for readability purposes in this paper, we only include an example in Figure 3 for people to "contest ideas"). According to Oza and Korkala (2012) "*it is not sufficient to merely collect all possible metrics but driving the culture of continuous measurement is imperative*". The partitioners in this study consider this model an improvement for agile metrics that adhere to agile principles particularly that associated with interaction. There are also difficulties that are inherent to our use of composite metrics, namely (1) it is always necessary to see the values of the three dimensions to understand the result, (2) it is a contextualized evaluation and cannot be used to compare different companies – although it may be used to compare different in-house projects and (3) it includes a subjective part of evaluation that makes the value representative of the team's reality. The same difficulties can simultaneously provide potential improvements for quality in teams because they (1) require team's to specify their own metrics, (2) provide ongoing adherence to practice, and (3) promote debate and critical reflection that is an intrinsic part of our method which complements agile techniques.

6 Conclusion

This action research project was set up to develop socially-constructed metrics for agile quality. We conducted a diagnosis at our practitioner organization and a literature review to establish the theoretical frame of reference. Then, we designed and implemented our action plan to (1) propose a model to create socially-constructed metrics for agile quality, (2) define indicators (3), establish their structure, and (4) create simple tools to assist participants in using metrics. The findings suggest that socially-constructed metrics can provide a new way of assessing and improving agile quality, adhering to the most crucial values of agile: the focus on people and their involvement; simplify support processes for quality management; stakeholder collaboration; and accepting change as a part of the development process (Beck et al. 2001).

We concluded that it is necessary to consider three main types of socially-constructed metrics: (1) people-related, (2) process-related, and (3) outcome-related. We suggest that a small set of indicators should be used but in line with adaptive project management, companies should allow these to change over time according to the project requirements. Moreover, we suggest that socially-constructed metrics should include three interrelated dimensions: (1) evidence from practice; (2) stakeholder expectations; and (3) stakeholder evaluation. The metrics dashboard includes indicators that are easy to obtain and allocates rules to promote improvement and critical analysis. The limitations of this study act as a starting point for planning future research. First, this is the first CAR cycle of our research and, although developed in a highly demanding context (healthcare development) it is necessary to test our model in different settings. Secondly, the benefits of our method are only assessed by the researchers and the organizational team, omitting auditors, partners, other teams, or customers. The next cycle may include other stakeholders and explore the contrast of viewpoints within and among agile teams. Thirdly, we also identified difficulties during method execution (such as defining which indicators to use) that could benefit from a taxonomy of metrics for the three types. Fourth, there is opportunity to achieve a rich agile quality index that organizations can use to (self-) evaluate their improvement efforts and the efficacy of improvement actions by comparing how the indicators change over time. The index can be the result of weighted average of all the indicators in the company, opening opportunities for agile quality dashboards. Finally, due to our focus in developing metrics and tools, we could not fully explore the social changes (e.g., knowledge transfer, team motivation) involved in the systematic debate using metrics in daily meetings, retrospective, and audits. Future research can help in addressing these challenges and contribute to understanding the effect of using the artifact in organizations. We see potential for socially-constructed metrics to inspire other researchers to use, improve, change, and extend metrics in other fields, for example in other software development approaches, fostering participative assessment and improvement of quality or for business processes management as a participative form of evaluating and improving business processes.

7 References

Agarwal, A., Garg, N. K., and Jain, A. 2014. "Quality assurance for Product development using Agile," in *International Conference on Reliability Optimization and Information Technology (ICROIT)*, Faridabad, pp. 44–47.

Avison, D., and Wood-harper, T. 2003. "Bringing Social and Organisational Issues into Information Systems Development: The Story of Multiview," in *Socio-Technical and Human Cognition Elements of information Systems*, S. Clarke, E. Coakes, G. M. Hunter, and A. Wenn (eds.), , pp.

5–21.

Babb, J., Hoda, R., and Norbjerg, J. 2014. "Embedding Reflection and Learning into Agile Software Development," *IEEE Software* (31:4), pp. 51–57.

Bansiya, J., and Davis, C. G. 2002. "A hierarchical model for object-oriented design quality assessment," *IEEE Transactions on Software Engineering* (28:1), pp. 4–17.

Barata, J., and Coyle, S. 2016. "Developing Socially-Constructed Quality Metrics in Agile: A Multi-Faceted Perspective," in *Proceedings of Thirty Seventh International Conference on Information Systems (ICIS)*, Dublin.

Baskerville, R. 1999. "Investigating information systems with action research," *Communications of the Association for Information Systems* (2:3), pp. 1–32.

Baskerville, R., and Wood-Harper, A. T. 1996. "A critical perspective on action research as a method for information systems research," *Journal of Information Technology* (11:3), pp. 235–246.

Baxter, G., and Sommerville, I. 2011. "Socio-technical systems: From design methods to systems engineering," *Interacting with Computers* (23:1), pp. 4–17.

Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J., and Thomas, D. 2001. "Manifesto for Agile Software Development," *http://agilemanifesto.org/*, [date accessed 2016-09-07].

di Bella, E., Fronza, I., Phaphoom, N., Sillitti, A., Succi, G., and Vlasenko, J. 2013. "Pair Programming and Software Defects-A Large, Industrial Case Study," *IEEE Transactions on Software Engineering* (39:7), pp. 930–953.

Berger, P. L., and Luckmann, T. 1991. *The social construction of reality: A treatise in the sociology of knowledge*, Penguin Books.

Boerman, M. P., Lubsen, Z., Tamburri, D. A., and Visser, J. 2015. "Measuring and Monitoring Agile Development Status," in *6th International Workshop on Emerging Trends in Software Metrics (WETSoM)*, Florence, pp. 54–62.

Cunha, P. R., and Figueiredo, A. D. 2002. "Action Research and Critical Rationalisationism: A Virtuous Marriage," in *Proceedings of the 10th European Conference on Information Systems (ECIS)*, Gdańsk, Poland.

Davison, R., Martinsons, M. G., and Kock, N. 2004. "Principles of canonical action research," *Information Systems Journal* (14:1), pp. 65–86.

Eden, C., and Huxham, C. 1996. "Action research for management research," *British Journal of Management* (7:1), pp. 75–86.

Ghobadi, S., and Mathiassen, L. 2016. "Perceived barriers to effective knowledge sharing in agile software teams," *Information Systems Journal* (26:2), pp. 95–125.

Gill, G., and Bhattacherjee, A. 2009. "Whom Are We Informing? Issues and Recommendations for MIS Research from an Informing Science Perspective," *MIS Quarterly* (33:2), pp. 217–235.

Gregor, S. 2006. "The nature of theory in information systems," *MIS Quarterly* (30:3), pp. 611–642.

Gren, L., Torkar, R., and Feldt, R. 2015. "The prospects of a quantitative measurement of agility: A validation study on an agile maturity model," *Journal of Systems and Software* (107:2015), pp. 38–49.

Gruschwitz, S., and Schlosser, F. 2012. "Towards an integrated model for managing product and process quality in agile software projects," in *7th International Research Workshop on Information Technology Project Management (IRWITPM 2012)*, pp. 147–155.

Hayes, W., Miller, S., Lapham, M. A., Wrubel, E., and Chick, T. 2014. "Agile Metrics: Progress Monitoring of Agile Contractors. CMU/SEI-2013-TN-029," Pittsburgh, PA.

Hongying, G., and Cheng, Y. 2011. "A customizable agile software Quality Assurance model," in *The 5th International Conference on New Trends in Information Science and Service Science*, pp. 382–387.

Hult, M., and Lennung, S.-Å. 1980. "Towards a definition of action research: a note and bibliography," *Journal of Management Studies* (17:2), pp. 241–250.

ISO. 2015. *ISO 9001:2015 Quality management system – Requirements*, International Organization for Standardization, Geneva.

Janus, A., Dumke, R., Schmietendorf, A., and Jager, J. 2012. "The 3C approach for Agile Quality Assurance," in *3rd International Workshop on Emerging Trends in Software Metrics (WETSoM)*, pp. 9–13.

Kupiainen, E., Mäntylä, M. V., and Itkonen, J. 2015. "Using metrics in Agile and Lean software development - A systematic literature review of industrial studies," *Information and Software Technology* (62:1), pp. 143–163.

Lindgren, R., Henfridsson, O., and Schultze, U. 2004. "Design principles for competence management systems: a synthesis of an action research study," *MIS quarterly* (28:3), pp. 435–472.

McHugh, O., Conboy, K., and Lang, M. 2011. "Using agile practices to influence motivation within IT project teams," *Scandinavian Journal of Information Systems* (23:2), pp. 59–84.

McHugh, O., Conboy, K., and Lang, M. 2012. "Agile Practices: The Impact on Trust in Software Project Teams," *IEEE Software* (29:3), pp. 71–76.

Mishra, D., Balcioglu, E., and Mishra, A. 2012. "Measuring project and quality aspects in agile software development," *Technics Technologies Education Management* (7:1), pp. 122–127.

Oza, N., and Korkala, M. 2012. "Lessons Learned in Implementing Agile Software Development Metrics," in *UK Academy for Information Systems Conference (UKAIS)*, Oxford.

Pentland, B. T., and Feldman, M. S. 2008. "Designing routines: On the folly of designing artifacts, while hoping for patterns of action," *Information and Organization* (18:4), pp. 235–250.

Péraire, C., and Sedano, T. 2014. "Essence reflection meetings: Field study," in *18th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, London, pp. 1–4.

Scharff, C. 2011. "Guiding global software development projects using Scrum and Agile with quality assurance," in *Software Engineering Education and Training (CSEE&T)*, Honolulu, pp. 274–283.

Sidky, A., Arthur, J., and Bohner, S. 2007. "A disciplined approach to adopting agile practices: the agile adoption framework," *Innovations in Systems and Software Engineering* (3:3), pp. 203–216.

Stålhane, T., and Hanssen, G. 2008. "The Application of ISO 9001 to Agile Software Development," in *Product-Focused Software Process Improvement. Lecture Notes in Computer Science 4589*, Springer, pp. 371–385.

Stylianou, A. C., and Kumar, R. L. 2000. "An integrative framework for IS quality management," *Communications of the ACM* (43:9), pp. 99–104.

Susman, G. I., and Evered, R. D. 1978. "An Assessment of the Scientific Merits of Action Research," *Administrative Science Quarterly* (23:4), pp. 582–603.

Tseng, Y.-H., and Lin, C.-T. 2011. "Enhancing enterprise agility by deploying agile drivers, capabilities and providers," *Information Sciences* (181:17), pp. 3693–3708.