Developing Socially-Constructed Quality Metrics in Agile: A Multi-Faceted Perspective

Research-in-Progress

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

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

Abstract

This research proposes development of socially-constructed metrics for quality assessment and improvement in Agile Software Development (ASD) projects. The first phase of our research includes an extensive literature review, which indicates that traditional (outcome-focused) metrics that evaluate quality are not directly transferable to adaptive, ASD projects. We then conduct semi-structured interviews confirming the necessity of considering people and process aspects for quality considerations in agile. We propose three dimensions for composite metrics in ASD, namely, (1) evidence (2) expectation and (3) critical evaluation. This combines quantitative and qualitative information drawn from people, process, and outcome-related factors. The proposed model allows ASD teams to concurrently conduct quality assessment and improvement during their projects, producing innovative metrics, adhering to the core principles of the agile manifesto. In our next research stage, this reference model will be tested and validated in practice.

Keywords: Agile software development, quality improvement, socially-constructed metrics

Introduction

Quality is a multi-dimensional concept, combining information, technology, social, and organizational aspects (Stylianou and Kumar 2000). In software development projects, quality has been defined, analyzed and theoretically evaluated for many decades (Sarigiannidis and Chatzoglou 2014). There are influential studies about the complexity of quality metrics, defined as the degree in which a system or process holds specific quality attributes (Kaner and Bond 2004). Yet, very few have assessed quality metrics in the context of agile software development (ASD) projects (e.g. Jamieson and Fallah 2012; Kupiainen et al. 2015; Mishra et al. 2012).

According to Sarigiannidis and Chatzoglou (2014) and Kupiainen et al. (2015), a software project's total quality should be measured using three sub-factors; people, process, and product. In the context of software development the Software Engineering Institute's (SEI) Capability Maturity Model Integration (CMMI) embodied the premise that "*the quality of a system or product is highly influenced by the quality of the process used to develop and maintain it*" (CMMI 2010, p.5). Research by Jamieson and Fallah (2012) suggests that when compared against CMMI practices, the very use of agile techniques promotes quality in projects. To-date, quality in ASD has been studied from specific, singular perspectives ranging from product or software-related metrics (Kupiainen et al. 2015; Mishra et al. 2012) including; tests and quality control (Agarwal et al. 2014; Janus et al. 2012) and software defects (di Bella et al. 2013) to quality relating to stakeholder expectations (Boerman et al. 2015) and the role of auditing (Scharff

2011). Despite these important research investigations into quality metrics, more research is needed to improve our understanding as to how we can holistically assess quality in these rapidly changing project environments that differ vastly from their traditional ("waterfall") counterparts. Research purports that traditional techniques used to assess and improve quality are not directly transferable or entirely relevant in the context of ASD projects.

Agile methodologies may lead to quality improvements (Boerman et al. 2015; Jamieson and Fallah 2012) however, Reeves and Bednar (1994) explain that it is not always clear what "improving quality" means. They define multiple views of quality: (1) quality as excellence, suggesting that quality is assessed against some specific standard; (2) quality as value, extending the excellence view and introducing the cost-benefit of quality implementations; (3) quality as conformance with specifications, requiring one to consider whose needs are being satisfied and through which product; and (4) quality as meeting or exceeding stakeholder expectations. In ASD, "*quality as conformance*" addresses metrics of software quality and satisfaction of requirements where Boerman et al. (2015) discuss the alignment of quality principles and standards ("*quality as excellence*"), and the potential quality benefits provided by utilizing ASD techniques ("*quality as value*"). To date, a model that integrates these different views does not exist. The perspectives outlined above are applied singularly and we agree with Kaner and Bond (2004), that "*there are too many simplistic metrics that don't capture the essence of whatever it is that they are supposed to measure*". They are also usually used retrospectively (or "after the fact") and therefore are difficult to apply in iterative, adaptive ASD environments.

According to ISO 9001 (2015), a strong quality culture involves customer orientation, continuous improvement, using 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 ASD projects and collective, devolved decision making exists throughout (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).

Regardless of the methodology deployed, quality assessment in project environments usually involves an evaluation of outcomes that are often related to performance indicators (Agarwal and Rathod 2006; Heston and Phifer 2011). However, in the context of ASD, quality is an iterative, ongoing concern related to continuous effort (Dybå and Dingsøyr 2008), focusing primarily on interactions between individuals; their collaboration efforts, and their ability to be proactive and reactive in the face of change (Beck et al. 2001; Ghobadi and Mathiassen 2016; Kropp and Meier 2015). Therefore, traditional assessments of quality focusing primarily on outcome-related indicators are not conducive to ASD environments. Furthermore, existing studies that examine quality in ASD projects do not address interactions or critical reflections of project stakeholders for quality assessment or aspects associated with continuous improvement. Thus far, research has not integrated quantitative evidence with qualitative assessments, the latter of which are so crucial to the core principles set out in the agile manifesto (Beck et al. 2001).

This study is a starting point in proposing an approach to construct meaningful quality metrics that adhere to the core principles of ASD projects incorporating quantitative and qualitative approaches drawn from people, process, and outcome-related sub-factors pertaining to the project and the product (e.g. software quality, end-user satisfaction). Two underpinning research questions are formulated, namely, *RQ1: How do researchers address quality assessment and quality improvement in ASD literature?* and *RQ2: How do practitioners address quality assessment and quality improvement in ASD projects?* To answer *RQ1* we survey the state of the art in agile quality and to address *RQ2* we investigate the outlook of experienced ASD practitioners about quality assessment and improvement.

The remainder of this paper is presented as follows. The next section introduces our research approach to analyze the literature and conduct qualitative interviews in two companies. A review of 55 of the most relevant papers is presented, followed by the preliminary findings of two interviews with ASD experts. We conclude that existing quality assessment metrics do not adapt to ASD contexts. Afterwards, we propose the approach to design socially-constructed metrics. We close this paper stating the limitations of our research-in-progress and the directions for future research.

Research Approach

We started with a literature review (Kitchenham 2004; Okoli and Schabram 2010; Webster and Watson 2002) in February 2016, addressing our first research question. The research investigation included journals and conference proceedings, using different search engines of Google Scholar, EBSCOhost, Science Direct, Web of Science, IEEE Xplore, and Mendeley. We decided to start with broad search terms and progressively refine the results (Boell and Cecez-Kecmanovic 2014). First we screened the title and abstracts obtained with the search terms "agile quality" and "quality in agile". We found advantages in using different databases, for example, Google Scholar is less specific but presented the most extensive result lists: 278 results for "agile quality" excluding patents and citations, compared to EBSCOhost (4), ScienceDirect (18), IEEE Xplore (10) and Web of Science (1 in the title search but increased to 40 upon removal of quotation marks), and finally, Mendeley (81). Additionally, this provided opportunity to iterate the "*search and acquisition circle and the wider analysis and interpretation circle*" (Boell and Cecez-Kecmanovic 2014, p.263) by comparing databases. Second, we tested a combination of other related search terms in the same databases, for example, "agile development" + "quality metric"; "agile project" + "quality metric"; "agile development" + "quality monitoring"; and "quality improvement" + "agile practice". These terms emerged while we were reading initial studies and provided additional research focus. Complementarily, we performed citation analysis (Webster and Watson 2002) to identify related studies. We excluded papers that did not address quality assessment or quality improvement within ASD contexts and practices. Additionally, we did not include books, non-English papers, tools presentation, editorials, posters, patents, keynotes or panels conclusions. After eliminating duplicates, a total of 55 papers were reviewed and discussed by two researchers.

We followed a concept-centric review according to Webster and Watson (2002) and agreed that there are two main concepts in the literature addressing: (1) Quality Assessment in Agile and (2) Agile Practices for Quality Improvement. We subsequently identified units of analysis (Webster and Watson 2002) within each concept: (1) Quality Assessment in Agile included process, outcome, and metrics in the form of evidence, goals, and quality indices; and (2) Agile Practices for Quality Improvement included stakeholders interaction within agile projects, and the benefits and pitfalls of agile practices for quality improvement.

In the second phase we conducted two exploratory qualitative interviews (Myers and Newman 2007) in two geographically-dispersed organizations. The purpose was to indentify how quality metrics are used in practice, ascertaining difficulties of quality assessment in agile, and obtaining initial insights for improving metrics. We created a script for semi-structured interview sessions including questions for (1) company identification, (2) experience and practices of agile quality, (3) forms of measuring agile quality, (4) metrics, (5) improving agile quality, and (6) expectations for a new approach for agile quality. The authors conducted one in-depth interview each, averaging two hours. Subsequent analysis of interview data and project documentation was conducted. Afterwards, we discussed the findings with the interviewees to confirm our evaluation. We were able to compare findings in the literature with preliminary insights from the field to help generate rich data (Schultze and Avital 2011) and compare perspectives of both interviewees.

Literature Review

Quality Assessment in Agile

Quality metrics documented in agile literature have come from numerous sources. Some studies (e.g. Concas et al. 2012) select metrics from traditional software development (TSD) approaches. Mishra et al. (2012) argue that TSD metrics do not suit ASD projects and specific metrics are suggested for progress in an iteration, progress in a release and feature correction rate in an iteration or release. Padmini et al. (2015) propose a combination of TSD metrics (such as delivery on time or within budget) with agile-specific metrics (such as sprint-level burndown charts). Listing a total of 89 metrics, Kupiainen et al. (2015) present one of the most comprehensive reviews of agile quality metrics found in industrial studies. The majority of these metrics however are quantitative and relate to software defects, velocity, cycle times, faults, lead time, burndown statistics and so on. According to Kupiainen et al. (2015, p.150) ASD teams use metrics for "*Sprint and Project Planning, Sprint and Project Progress Tracking, Understanding and*

Improving Quality, Fixing Software Process Problems, and Motivating People". They also found that almost 40% of their identified metrics were customized. Their conclusions state that the majority of existing metrics relate to outcomes and process, but are non-inclusive of people. This is surprising considering the fundamental agile value associated with people over process.

This work is further supported in other literature that includes examples of quality indices for outcome and process. The quality control system presented by Rauch et al. (2008) includes ten indices to assess product and process quality in agile projects. Later, Gruschwitz and Schlosser (2012) suggest 40 indices for quality assessment of process and product in agile projects. However, the metrics selected by these authors are independent and do not allow for an integrated index as occurs in the work of Bansiya and Davis (2002). As before, these metrics are also quantitative in nature.

Moreover, existing studies do not address the interaction of stakeholders for quality assessment and improvement, namely, (1) how the evaluation of stakeholders can be included in the indices, (2) the utility for continuous improvement actions, or (3) the acquisition and sharing of tacit knowledge (Ryan and O'Connor 2013) during communication and the implementation of agile practices (Hummel et al. 2015). Finally, we did not find a social dimension which is required based on agile principles, namely, in the suggestion of "social metrics" that can be used to assess and improve agile quality, for example, "team capabilities" (Ghobadi and Mathiassen 2016), or "user involvement" (Begier 2010). There are inherent cultural and social aspects involved in determining quality that are core to the agile manifesto (Beck et al. 2001). Siakas and Siakas (2007, p.607) compare agile success factors with quality models such as ETHICS and Total Quality Management and conclude that: "*There is an obvious need for more scientific evidence and further research in order to understand the requirements of agile quality [operating] in different cultural and social contexts*". Social metrics are needed in ASD to address people's behavior and assess communications or aspects of interactions (Dorairaj et al. 2012; Gren et al. 2015; Ibrahim et al. 2010; Wiese et al. 2014). Recent studies to assess agility in enterprises (Tseng and Lin 2011) include several social aspects for example, personal skills, technology awareness, and trust-based relations with customers, collaboration, empowerment and motivation. Gren et al. (2015, p.38) also 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*". Wiese et al. (2014) identifies 51 social metrics in the field of software engineering and groups them into three main categories for communication, project, and cooperative work. It is surprising that similar perspectives have not been adopted in ascertaining quality in ASD.

Agile quality should assess interrelated dimensions (Meso and Jain 2006): people (Begier 2010; Fernandez-Sanz and Misra 2011; Ghobadi and Mathiassen 2016); process (Hayes et al. 2014; Kupiainen et al. 2015); outcome related dimensions of a project (Agarwal and Rathod 2006) and product (Kupiainen et al. 2015). Quality is a concern throughout the entire iterative, ASD lifecycle: In early phases (such as planning or risk analysis), during the development iterations and after product delivery (such as final tests, procedures for review, lessons learned and KPI evaluation). Outcome-related metrics are important but are not sufficient in assuring ongoing quality improvement or assessing the complex, iterative socio-technical nature of ASD-related quality (Maruping and Venkatesh 2009).

Agile Practices for Quality Improvement

The technical report provided by Hayes et al. (2014) identifies different moments in agile projects when it is possible to get end-user feedback to assess their satisfaction. Baxter and Sommerville (2011, p.13) explain that the involvement of the product owner "*needs to be extended to take into account a broader set of system stakeholders*". In addition, ASD practices differ significantly when compared to traditional approaches: "*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, p.144). Research has indicated that when comparing a 'waterfall' process to that of 'agile'; development performance and product quality are far superior in agile (Tarhan and Yilmaz 2014), highlighting attempts by agile methodologies to embed quality efforts into processes that are iteratively evaluated.

Retrospective meetings are an essential ASD practice promoting continuous improvement, however, development pressures can make this task difficult in practice (Babb et al. 2014; McHugh et al. 2011). Retrospectives allow reflection about previous iterations to identify subsequent actions needed. There are different forms of retrospective meetings presented by Peraire and Sedano (2014) and their work concludes that the use of artefacts and specific steps to guide retrospectives has advantages. According to Hayes et al. (2014) however, there are other quality touch-points in ASD besides retrospectives. Their research found that there is an emphasis on quality in the early project phases but there is potential and opportunity for adopting a quality-focus in the later stages to include documenting and user stories. This coincides with our earlier observations that quality needs to be considered throughout the entire iterative, ASD process. According to the authors: "*traditional views of quality (i.e., measured defects) can be supplemented with a more direct measure of customer-perceived value—using customer satisfaction feedback*" (Hayes et al. 2014, p.29). There is therefore a need to incorporate adaptive quality approaches with traditional techniques that are relevant and suitable for ASD projects (Salo and Abrahamsson 2007).

Backlogs and quantitative metrics are important for assessing past projects, but are limited in helping to prevent future problems or support continuous improvements. Interactions between team members and other project stakeholders to address quality concerns can inform preventive actions (Drury-Grogan 2014). In ASD projects there are daily opportunities for this due to ongoing intense stakeholder interactions among developers, facilitators, testers, and end-user representatives. Ghobadi and Mathiassen (2016, p.95) suggest this approach when addressing knowledge sharing barriers in agile projects. According to these authors 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*". Despite this, we could not find any literature guidelines for fostering quality assessment and evaluation techniques in ASD projects that incorporate such qualitative criteria with that of quantitative measures such as defects, velocity or cycle time.

It is evident that people, process, and outcome are deeply intertwined in ASD projects, as we confirm in our semi-structured interviews that are presented in the next section.

Preliminary Results from Qualitative Inquiry

We conducted qualitative semi-structured interviews (Myers and Newman 2007) in two geographically-dispersed organizations adopting ASD practices. The companies were selected from our list of contacts that recently revealed interest in building new quality metrics (company A), had experience in delivering ASD projects (A and B), and practical knowledge of quality in multiple contexts of ASD (company B). The two settings are presented in Table 1.

Table 1. Qualitative Interviews

	Company A	**Company B**
Industry	Healthcare	Software Consultancy
Employees	60	230
Interviewee	Quality manager	Senior Developer
Agile practices	Scrum	Scrum; Kanban
Quality Standards	ISO 9001:2015; NP 4457 for innovation management; multiple healthcare standards	None specified

Company A is a European software provider of healthcare software for hospitals and clinics. Founded 25 years ago they are present in four continents, serving over 120.000 users and 25 million clinical processes. When we conducted the interview with the quality manager she was preparing the external audit for the recently revised 2015 version of ISO 9001:2015. According to the quality manager, Company A "*has numerous indicators, however, only a few are valid for agile quality*". The reasons vary because in some cases "*the numbers are highly dependent of the context and must be carefully interpreted*". In other cases "[she] *does not think it 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 the income] *require quality indicators and*

evidence for each iteration, due to the critical nature of healthcare IT". Moreover, their certifications also include healthcare standards compliance and "*the need to provide quality evidence to external assessors"*. 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"*. The research participant also talked about the importance of being able to change metrics for each project or even team, in an "agile way". 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*".

Interestingly, when we asked about social metrics the answer was that: "*in my previous work our motivation was evaluated by top managers in a quantitative scale ranging from 1 to 5… How could they know my motivation or the factors affecting it? I simply hated to be evaluated that way*". According to the quality manager, it is difficult to measure social metrics. 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 request, (5) and "provide meaning" for our agile numbers, according to the team's perspective*". This company agreed to participate in the next stages of our research and would like to include our findings to inform potential improvement actions for their ISO 9001:2015 migration.

Founded fifteen years ago, Company B is an Australian-based software consultancy firm that helps clients develop innovative software projects. Our interviewee has over eight year's experience in working on ASD projects across many sectors in the consultancy industry and stated that when it comes to measuring quality in agile *"the biggest and most important test is what the user thinks"* highlighting the importance of stakeholder involvement in quality evaluation. The research participant described three quality metrics. The first relates to setting benchmarks for quality at the start of the project, which is *"extremely difficult to do when you have nothing to compare it to"*, but in his opinion, *"mature teams (in performing stages) are much better at doing this"*. The second relates *to "comparison exercises that we do; so we compare the latest increment against the previous one"* and finally, *"our centralized continuous integration server runs automated tests which is a very good way of measuring software quality"*. However, the participant stressed that *"even when each of those metrics are showing us very positive results, if the user isn't happy, then we haven't reached our quality pinnacle"*. This shows that quantitative-driven assessments alone are not complete indicators of quality. When asked about social metrics the participant stated how they *"could see this making a lot of sense and a huge difference for our agile projects but then…how we apply those metrics in the context of teams during storming or norming stages is the challenge because these teams often have misaligned observations for quality assessment"*. This highlights the complexity of this research endeavor but also the importance and benefits of including composite metrics for assessing quality in ASD teams while simultaneously raising quality awareness.

Developing Socially-constructed Metrics: A Reference Model

For the purpose of our research, a metric is socially-constructed (Berger and Luckmann 1991) when stakeholders have the ability to alter its structure and critically evaluate its results. In this context, stakeholders' perceptions and participation are intrinsic to metric construction particularly in ASD projects. Unlike traditional approaches, which compare against predefined goals (e.g. number of defects below X% when compared to lines of code), stakeholders are not extracted from the process or just included at the end. They are inherently involved in constructing relevant metrics for their project.

We propose three dimensions in socially-constructed metrics for ASD quality, namely (1) *evidence* from practice, (2) stakeholders *expectations*, and (3) stakeholders *evaluation*. We are guided by Baxter and Sommerville (2011) to include different stakeholders in metric construction; addressing the social nature of ASD teams (Whitworth 2008). We agree with Gren et al. (2015) that quantitative data needs to be enriched with contextual and cultural issues. Moreover, we are inspired by Kupiainen et al. (2015) to include metrics for people, processes, and outcomes. Our model is non prescriptive about the types of metrics that can be socially-constructed, therefore, it may potentially apply to any metric selected by agile

teams. On one hand, we follow numerous authors (Bansiya and Davis 2002; Janus et al. 2012; Sidky et al. 2007) in the need to obtain metrics for quality assessment, as an important step of ASD quality improvement. On the other hand, we argue that assessment and improvement can occur simultaneously during an ASD project and there may be other socio-technical approaches that aggregate different metrics, quantitative and qualitative, objective and subjective. Metric construction should be a moment of critical reflection in ASD teams (Hodgson and Briand 2013) and has the potential to increase team commitment in ascertaining quality.

Figure 1 presents our reference model of socially-constructed metrics for agile quality. It includes an illustrative example pertaining to end-user complaints. Identifying specific indicators for use is currently outside the scope of this research; our aim is to propose an approach for metric construction and identify the dimensions that should be considered, independently of the agile indicator.

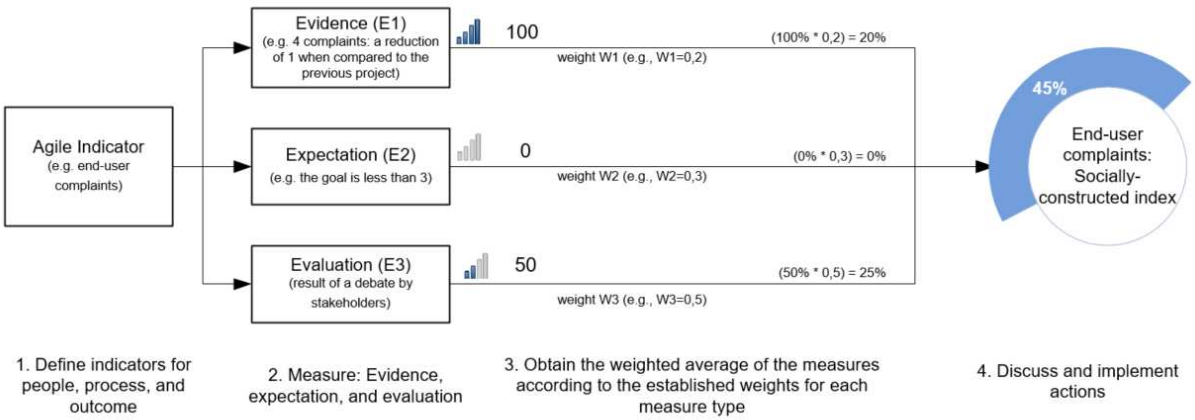

Figure 1. Socially-constructed Metrics for Agile Quality Assessment & Improvement

In Figure 1 we represent the indicators selected by the stakeholders (step 1). Each indicator is measured according to the three dimensions of (E1) *evidence*, (E2) *expectations*, and (E3) *evaluation*. On the right (step 4) we represent the resulting composite socially-constructed index. Steps 1-4 are a sketched pattern for action (Pentland and Feldman 2008) to encourage a reference framework for a metrics program (Oza and Korkala 2012). As stated by Oza and Korkala (2012, p.12): "*it is not sufficient to merely collect all possible metrics but driving the culture of continuous measurement is imperative*". Socially-constructed metrics are not the result of project evidences or a mere comparison with predetermined goals for ASD process or outcome. It includes both while adding context-bound, critical reflection and evaluation provided by stakeholders. Assessment and improvement are therefore intertwined. The weights of each dimension (W1 to W3) are established by project stakeholders. Each weight can vary from zero to one and the overall total of the three weights (W1+W2+W3) must be one. Table 2 presents these three dimensions for socially-constructed metrics (column 1); definitions (column 2); weighting comments (column 3) and rationale of measurement (column 4).

Table 2. Three Dimensions of Socially-constructed Metrics for Agile Quality			
Dimension	**Definition**	**Potential ways to consider weightings**	**How it was calculated in Figure 1**
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.	According to the history of the indicator. If current project has 4 complaints and the previous 5 then the indicator is 100% (improvement occurred).
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.	The team defines less than 3 complaints as a goal. In this case it was above (4) so the expectation indicator would have the value "0%" (did not improve). We are using 3 grades (success 100%, more or less 50%, unsuccessful 0%) but it could be a continuous scale.
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.	The team agrees that the end-user was a difficult one so 4 complaints in this case is a success (the reverse could also occur). However they agree that it is necessary to improve. They allocate 50% to evaluation.

In the case that we illustrate in Figure 1, the team considered that complaints are highly variable according to multiple factors including the type of project. Therefore, they decided that W1 = 0,2; W2 = 0,3; W3 = 0,5. The resulting metric would be: evidence (100%*0,2) + expectation (0%*0,3) + evaluation (50%*0,5)= 45% as represented in the figure. The team must justify the grade and propose actions to improve it. Combining the three forms of evaluation we reach a balanced index that includes (1) *evidence* comparing past results of similar projects, (2) conformance to *expectations* (goals), and (3) we ensure debate and critical *evaluation* of the indices in the quality model. Moreover, it is participative, aiming to build rich metrics that can be tailored (1) in early phases of ASD projects, (2) useful during project execution, and (3) provide valuable insights for retrospectives. Many differences exist when comparing ASD to traditional approaches: ascertaining and evaluating quality requires a specific, tailored approach to be effective in practice.

Conclusions and Outlook for Future Work

We propose an approach to create socially-constructed metrics, involving stakeholders (e.g. team members) in early phases to ascertain metrics and empower them to build metric-related results. In our proposal, agile quality is context-bound and not necessarily value-free, because in ASD teams the variety of stakeholders' perceptions, perspectives and priorities influence the meaning of quality, which can change from one release (or project) to the next. Without compromising proposed metrics (which can be decomposed to extract meaning), we can capture stakeholders' opinions and values (encouraging information and knowledge sharing) to construct relevant composite metrics.

There are a number of metrics available in the form of quality goals or quantitative indicators emerging from ASD practices, however, (1) few metrics are related to people, (2) do not comprehensively provide contextualized information within agile projects, and (3) do not result from a reflection made by project teams, which is essential in the context of both *being* agile and ascertaining quality. Our initial results emerging from analysis of the qualitative interviews suggests that traditional singular perspectives pertaining to quality are not sufficient nor practical for ASD. We argue that it is necessary to obtain synergies from (1) evidence that emerges from ASD practices, (2) expectations expressed in the form of goals and (3) evaluation resulting from a critical reflection. Ultimately, the combination of perspectives will promote sensemaking (Weick et al. 2005) with contextualized metrics and suggestions for quality improvement.

Our main conclusion for phase one is that socially-constructed metrics are essential for both endeavors: quality assessment and quality improvement in ASD projects. According to our interviews, socially-constructed metrics (Berger and Luckmann 1991) provide promise as they are composed of objective and

subjective parts. This combination however, needs to be cautiously evaluated for inaccuracies or bias particularly for teams during "storming" and "norming" stages as indicative of our early findings.

There are several opportunities for future research in socially-constructed metrics. First, it is necessary to test our multifaceted model within ASD contexts, for example, to understand (1) how agile teams define the relative importance of each dimension of the metric; (2) the potential advantages of our composite metrics when compared with "traditional" metrics; (3) the potential conflicts that can emerge in the debate and construction of metrics, and how to solve them; (4) the potential difficulties of metrics that require metadata – because there are a combination of evidences, expectations, and critical evaluation of stakeholders; and (5) the benefits of socially-constructed metrics for improving communication between different stakeholders, for example, in daily meetings, retrospectives or quality audits.

Our research-in-progress has limitations that must be considered. First, despite the care taken in literature screening, content analysis, and complementary searches, the selection of databases and keywords raises limitations for the literature selection. Second, we restricted the scope of our study to agile quality. It would be interesting to consider which metrics should be included in each dimension of our model and at which stage of the project they should be revised. Thirdly, the authors agreed with the concepts and units of analysis but other researchers may prefer distinct classification schemes. Forth, we gathered initial inspiring insights from two companies across two continents to obtain research breadth; however, our participants were limited to two persons and it is necessary to proceed with data gathering to include further interviews, and other sources such as documents and observation. To minimize threats to validity of our study we followed existing guidelines for literature reviews (Kitchenham 2004; Okoli and Schabram 2010; Webster and Watson 2002), elaborated an interview protocol, compared the interviews with other sources in literature, and two researchers proceeded in parallel during the entire research, contrasting different data sources and constantly challenging the results in distinct steps of this research-in-progress. Our current results provide a frame of reference for action research (Susman and Evered 1978) that is already planned in organizations adopting ASD practices, aiming at a pilot evaluation of the reference model in industry. Action research allows us to study complex problems in organizational settings, ensuring rigor and validity (Davison et al. 2004). We have selected this research approach for our next phase, with the dual aim to contribute to science while solving real organizational problems (McKay and Marshall 2001).

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.

Agarwal, N., and Rathod, U. 2006. "Defining 'success' for software projects: An exploratory revelation," *International Journal of Project Management* (24:4), pp. 358–370.

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.

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].

Begier, B. 2010. "Users' involvement may help respect social and ethical values and improve software quality," *Information Systems Frontiers* (12:4), pp. 389–397.

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.

Boell, S. K., and Cecez-Kecmanovic, D. 2014. "A hermeneutic approach for conducting literature reviews and literature searches," *Communications of the Association for Information Systems* (34:1), pp. 257–286.

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.

CMMI. 2010. "CMMI for Development, Version 1.3," *Carnegie Mellon University*, Pittsburgh.

Concas, G., Marchesi, M., Destefanis, G., and Tonelli, R. 2012. "An Empirical Study of Software Metrics for Assessing the Phases of an Agile Project," *International Journal of Software Engineering and Knowledge Engineering* (22:4), pp. 525–548.

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

Dorairaj, S., Noble, J., and Malik, P. 2012. "Understanding team dynamics in distributed agile software development," *Lecture Notes in Business Information Processing (LNBIP 111)*, pp. 47–61.

Drury-Grogan, M. L. 2014. "Performance on agile teams: Relating iteration objectives and critical decisions to project management success factors," *Information and Software Technology* (56:5), pp. 506–515.

Dybå, T., and Dingsøyr, T. 2008. "Empirical studies of agile software development: A systematic review," *Information and Software Technology* (50:9–10), pp. 833–859.

Fernandez-Sanz, L., and Misra, S. 2011. "Influence of Human Factors in Software Quality and Productivity," in *Computational Science and Its Applications (ICCSA)*, Santander, pp. 257–69.

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

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)*, Orlando, 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.

Heston, K. M., and Phifer, W. 2011. "The multiple quality models paradox: how much 'best practice' is just enough?," *Journal of Software Maintenance and Evolution: Research and Practice* (23:8), pp. 517–531.

Hodgson, D., and Briand, L. 2013. "Controlling the uncontrollable: 'Agile' teams and illusions of autonomy in creative work," *Work Employment and Society* (27:2), pp. 308–325.

Hummel, M., Rosenkranz, C., and Holten, R. 2015. "The role of social agile practices for direct and indirect communication in information systems development teams," *Communications of the Association for Information Systems* (36:2015), pp. 273–300.

Ibrahim, W. M., Bettenburg, N., Shihab, E., Adams, B., and Hassan, A. E. 2010. "Should I contribute to this discussion?," in *IEEE Working Conference on Mining Software Repositories (MSR)*, Cape Town, May, pp. 181–190.

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

Jamieson, M. J., and Fallah, M. H. 2012. "Agile Quality Management Techniques," *Software Quality Professional* (14:2), pp. 12–21.

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)*, Zurich, pp. 9–13.

Kaner, C., and Bond, W. P. 2004. "Software Engineering Metrics: What Do They Measure and How Do We Know?," in *10th International Software Metrics Symposium*, Chicago, pp. 1–12.

Kitchenham, B. 2004. "Procedures for Performing Systematic Reviews," *Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1.*, Keele University Technical Report, Staffordshire.

Kropp, M., and Meier, A. 2015. "Agile Success Factors - A qualitative study about what makes agile projects successful," Zurich.

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.
Maruping, L. M., and Venkatesh, V. 2009. "A Control Theory Perspective on Agile Methodology Use and Changing User Requirements," *Information Systems Research* (20:3), pp. 377–399.
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.
McKay, J., and Marshall, P. 2001. "The dual imperatives of action research," *Information Technology & People* (14:1), pp. 45–59.
Meso, P., and Jain, R. 2006. "Agile software development: adaptive systems principles and best practices," *Information Systems Management* (530), pp. 19–30.
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.
Myers, M. D., and Newman, M. 2007. "The qualitative interview in IS research: Examining the craft," *Information and Organization* (17:1), pp. 2–26.
Okoli, C., and Schabram, K. 2010. "A Guide to Conducting a Systematic Literature Review of Information Systems Research," *Sprouts Working Papers on Information Systems* (10:26), pp. 1–49.
Oza, N., and Korkala, M. 2012. "Lessons Learned in Implementing Agile Software Development Metrics," in *UK Academy for Information Systems Conference (UKAIS)*, Oxford, pp. 1–12.
Padmini, K. V, Dilum Bandara, H. M. N., and Perera, I. 2015. "Use of software metrics in agile software development process," in *Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, pp. 312–317.
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.
Rauch, N., Kuhn, E., and Friedrich, H. 2008. "Index-based Process and Software Quality Control in Agile Development Projects," in *CompArch2008*, Karlsruhe, pp. 1–4.
Reeves, C. A., and Bednar, D. A. 1994. "Defining Quality: Alternatives and Implications," *Academy of Management Review* (19:3), pp. 419–445.
Ryan, S., and O'Connor, R. V. 2013. "Acquiring and sharing tacit knowledge in software development teams: An empirical study," *Information and Software Technology* (55:9), pp. 1614–1624.
Salo, O., and Abrahamsson, P. 2007. "An iterative improvement process for agile software development," *Software Process: Improvement and Practice* (12:1), pp. 81–100.
Sarigiannidis, L., and Chatzoglou, P. D. 2014. "Quality vs risk: An investigation of their relationship in software development projects," *International Journal of Project Management* (32:6), pp. 1073–1082.
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.
Schultze, U., and Avital, M. 2011. "Designing interviews to generate rich data for information systems research," *Information and Organization* (21:1), pp. 1–16.
Siakas, K. V., and Siakas, E. 2007. "The agile professional culture: A source of agile quality," *Software Process: Improvement and Practice* (12:6), pp. 597–610.
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," *Lecture Notes in Computer Science (LNCS 4589)*, 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.
Tarhan, A., and Yilmaz, S. G. 2014. "Systematic analyses and comparison of development performance and product quality of Incremental Process and Agile Process," *Information and Software Technology* (56:5), pp. 477–494.

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.
Webster, J., and Watson, R. T. 2002. "Analyzing the past to prepare the future," *MIS Quarterly* (26:2), pp. xiii–xxiii.
Weick, K. E., Sutcliffe, K. M., and Obstfeld, D. 2005. "Organizing and the Process of Sensemaking," *Organization Science* (16:4), pp. 409–421.
Whitworth, E. 2008. "Experience Report: The Social Nature of Agile Teams," in *Agile 2008 Conference*, Toronto, pp. 429–435.
Wiese, I. S., Côgo, F. R., Ré, R., Steinmacher, I., and Gerosa, M. A. 2014. "Social metrics included in prediction models on software engineering," in *10th International Conference on Predictive Models in Software Engineering (PROMISE)*, Turin, pp. 72–81.