Agile metrics: How and why yeeeep!

Eetu Kupiainen
Aalto University
Konemiehentie 2
Espoo, Finland
eetu.kupiainen@aalto.fi

G.K.M. Tobin*
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld†
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Lawrence P. Leipuner
Brookhaven Laboratories
Brookhaven National Lab
P.O. Box 5000
lleipuner@researchlabs.org

Sean Fogarty
NASA Ames Research Center
Moffett Field
California 94035
fogartys@amesres.org

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

ABSTRACT

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*agile metrics*

General Terms

Theory

Keywords

ACM proceedings, LaTeX, text tagging

1. INTRODUCTION

Software engineering is at a crossroads as there are new leaner and more agile software development methods appearing next to the traditional software development methods. Software measurement has had it's history for decades but that history has been steered towards the needs of traditional software development methods. What is the role of metrics in agile development?

Software measurement has been seen important for research and there are numerous studies and books regarding it. However, measurement research with agile methods remains scarce. The aim of this research is to fill part of that gap.

The angle for this study is industry focused. We want to understand the current state of agile measurement in industry. Thus, we are more interested in case studies and actual empirical findings than unproven theories or models.

This study will lay out the current state of industrial agile metric papers. Moreover, the study uncovers the reasons for metrics usage as well as highlights actions that the use of metrics can trigger. Practitioners will hopefully benefit from metrics this study brings forth.

This paper is organized as follows: (blaa blaa).

(-historialï£¡hestyminen(en ehkï£¡ ite oo tï£¡stï£¡ hirveen innokas) –lï£¡hti koodimittareista –jossain vaiheessa kiinnostuttiin prosessimittareista –agiilit menetelmï£¡t alkavat yleistyï£¡, miten hoitaa mittaaminen siellï£¡? Kun prosessi muuttuu miten mittaus, perinteiset mittarit raskaita(?). –EBSE:n korostaminen

-tore lï£¡hestyminen – tavalliselle talliaiselle intro sceneen – syyt miks me tehï£¡ï£¡n tï£¡llainen tutkimus – mitï£¡ hyï£¡tyjï£¡ tï£¡llï£¡ tutkimuksella on)

2. BACKGROUND

2.1 Evidence based software engineering

2.2 Measurement

According to Fenton et al.[9] "Measurement is the process by which numbers of symbols are assigned to attributes of entities in the real world in such way as to describe them according to clearly defined rules."

2.3 Previous metric research

There are a few mapping studies on software metrics(tï£¡hï£¡n vois lisï£¡tï£¡ kitch whats 2010:ssï£¡ olevat muut mut en tiii£¡ mitï£¡ lisï£¡arvoa ne tois). [19] says there is a large body of research related to software metrics. However, she highlights that all evidence should be critically appraised so that further studies can be based on good quality evidence. She also reminds researchers to understand the context where metrics are taken from - failure to understand context will probably not provide answers to industry-related questions.

–EBSE –Mittaamisen SR:t –Agile mittaamisen muut tutkimukset vai enemmänkin tutkimuksessa käytettyjen termien ja käsitteiden selittäminen –Aims and research questions

*The secretary disavows any knowledge of this author's actions.

†This author is the one who did all the really hard work.

WOODSTOCK '97 El Paso, Texas USA

Table 1: Paper selection funnel

Phase	Amount of papers
Phase 1	774
Phase 2	163
Phase 3	38

2.4 Aims and research questions

The aim of this paper is to provide preliminary results from a systematic review (SR) on agile metrics. Moreover, we are interested on the industrial use of metrics in agile context.

The performed SR has more research questions but for this paper only the following are considered:

1. Why are metrics used?
2. What actions do the use of metrics trigger?
3. Which metrics are used?

3. REVIEW METHOD

(Tï£¡hï£¡n ehkï£¡ kai tyyliin et vï£¡hï£¡n etitty et lï£¡ytykï£¡ mitï£¡ï£¡n laajaa tutkimusta agile mittareista) (Tï£¡ssï£¡ voisi myï£¡s perustella vielï£¡ SR:n valitsemista menetelmï£¡ksi -Why SR was selected? This is more a ï£¡understanding of problemï£¡ research. Not trying to find solution to specific problem. There is already literature that can be synthesized.)

3.1 Protocol development

Kitchenham's guide for SRs[18] was used as a basis for developing the review protocol. Additionally, other guidelines[38], a lessons learned from SRs [1], a SR case[7] and a SR on SR[20] were used to further understand the challenges and opportunities of SRs.

The protocol was also iterated in weekly meetings with the researchers, as well as in a pilot described in 3.2.5.

3.2 Search and selection process

The strategy for finding primary studies was following:

1. Initial check that there are enough papers to conduct SR
2. Stage 1: Automated search
3. Pilot
4. Stage 2: Include and exclude based on title and abstract
5. Stage 3: Include and exclude based on full text. Conduct data extraction and quality assessment.

Table 1 shows the selection funnel in terms of amount papers.

3.2.1 Search strings

Scopus was used to find the primary documents with automated search. The first search string was:

TITLE-ABS-KEY(software AND (agile OR lean OR "crystal method" OR "crystal clear" OR dsdm OR "dynamic systems development method" OR fdd OR "feature driven development" OR "agile unified process" OR "agile modeling" OR scrumban OR kanban OR scrum OR "extreme programming" OR xp) AND (measur* OR metric OR diagnostic OR monitor*)) AND (LIMIT-TO(SUBJAREA, "COMP")) AND (LIMIT-TO(LANGUAGE, "English"))

It found 512 hits 19.9.2013.

Then I noticed that some previously found key papers were missing from the hits because they were under sub area "Engineering", thus the second search string:

TITLE-ABS-KEY(software AND (agile OR lean OR "crystal method" OR "crystal clear" OR dsdm OR "dynamic systems development method" OR fdd OR "feature driven development" OR "agile unified process" OR "agile modeling" OR scrumban OR kanban OR scrum OR "extreme programming" OR xp) AND (measur* OR metric OR diagnosticOR monitor*)) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(SUBJAREA, "ENGI")) AND (EXCLUDE(SUBJAREA, "COMP") OR EXCLUDE(SUBJAREA, "PHYS") OR EXCLUDE(SUBJAREA,"MATE") OR EXCLUDE(SUBJAREA, "BUSI") OR EXCLUDE(SUBJAREA, "MATH") OR EXCLUDE(SUBJAREA, "ENVI") OR EXCLUDE(SUBJAREA, "EART") OR EXCLUDE(SUBJAREA, "DECI") OREXCLUDE(SUBJ "ENER"))

It found 220 hits 7.11.2013.

Then I found out that previous searches didn't include all XP conferences because they were under sub area "Business", thus the third search:

TITLE-ABS-KEY(software AND (agile OR lean OR "crystal method" OR "crystal clear" OR dsdm OR "dynamic systems development method" OR fdd OR "feature driven development" OR "agile unified process" OR "agile modeling" OR scrumban OR kanban OR scrum OR "extreme programming" OR xp) AND (measur* OR metric OR diagnosticOR monitor*)) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(SUBJAREA, "BUSI")) AND (EXCLUDE(SUBJAREA, "ENGI") OR EXCLUDE(SUBJAREA, "COMP"))

It found 42 hits 10.12.2013

3.2.2 Inclusion criteria

- Papers that present the use and experiences of metrics in an agile industry setting.

3.2.3 Exclusion criteria

- Papers that don't contain empirical data from industry cases.
- Papers that are not in English.
- Papers that don't have agile context. There is evidence of clearly non-agile practices or there is no agile method named. For example, paper mentions agile but case company has only three releases per year.
- Paper is only about one agile practice, which is not related to measuring.
- Papers that don't seem to have any data about metric usage. Also, if there is only a few descriptions of metrics but no other info regarding reasons or usage.
- Papers that have serious issues with grammar or vocabulary and therefore it takes considerable effort understand sentences.

- Papers that refer to another paper where the actual case is discussed.
- Papers that are in academic or semi-academic setting - customer or part of workers are from industry. The reason for this is that it doesn't fully represent industry setting as software development methods are likely enforced by academia.
- Papers where results cannot be separated by setting, for example surveys where there is data both from academia and industry. Similarly exclusion if results cannot be separated by software development method.
- Papers where the setting is not clear. For example only mention of context is "100 junior developers".
- Papers that are full conference proceedings. Individual papers should be already be separated individually.
- Papers that where the measurements are only used for the research. For example author measures which agile practices correlate with success.
- Papers that don't even show measurement usage in a pilot setting. For example method or metric is used to a static industrial data set.
- Papers that are about the same case, from the same author and same research focus, basically the paper format has just changed slightly.

3.2.4 Stage 1 - Automatic search

Scopus was used as the only search engine as it contained the most relevant databases IEEE and ACM. Also, it was able to find Agile- and XP conference papers. Only XP Conference 2013 was search manually because it couldn't be found through Scopus.

3.2.5 Pilot

A pilot study was conducted in order to refine the aim of the research and get familiar with the research method. Moreover, it was possible to modify the method and tools.

The plan was to go through 15 papers from the automatic search:

- 5 by top relevance
- 5 by top citations
- 5 by random

The pilot resulted in changing citation manager tool from Zotero to Jabref. Also, selection stage 2 and 3 were combined. Some codes were generated. The quality assessment checklist was decided on.

3.2.6 Stage 2 - Selection by title and abstract

In stage 2 papers were included and excluded based on their title and abstract. As the quality of abstracts can be poor in computer science[18], full texts were also skimmed through in case of unclear abstracts - especially intro, case description and conclusions was checked briefly. Also, if there were any unclear cases - these were discussed in weekly meetings and a descriptive exclusion rule was created if necessary.

Selection process was also done for a sample of 26 papers by one other researcher - the level of agreement was substantial with Kappa 0.67[21].

3.2.7 Stage 3 - Selection by full text

Stage 3 included multiple activities in one work flow. Selection by full text was done, interesting data was coded and quality assessment was done. Once again, if there were unclear papers, they were discussed in meetings with other researchers. Sample set of 7 papers was also included/excluded by other researcher, with an almost perfect agreement, Kappa 1.0[21].

3.3 Data extraction

Integrated coding was selected for data extraction strategy[4]. It provides focus to research questions but flexibility regarding findings. Deductive coding would have been too restraining and inductive coding might have caused too much bias. Integrated coding made it possible to create a sample list of code categories, namely:

- Why is measurement used?
- How is measurement used?
- Metrics with metric name
- Interesting results

The coding started with reading the full text and marking interesting data with a temporary code. After, reading the full text I would check each data pieces and code again with an appropriate code based on the built understanding. Metric codes were named based on the similarity with the text it was collected from.

We created a rule set for collecting metrics:

- Collect metric only if team or company uses it.
- Don't collect metrics that are only used for the comparison and selection of development methods.
- Don't collect metrics that are used to compare teams.
- Collect metric only if something is said about why it is used or what actions it causes.

Atlas.ti was used to collect the qualitative data(viiteAtlasiin).

To evaluate if the same metrics would be found, another researcher coded metrics from two papers. Capture-recapture method[29] was then used which showed that 90% of metrics were found.

3.4 Quality assessment

Quality assessment form adopted from [7] was used to evaluate the quality of each primary document. Additionally, a relevancy factor was added to the same assessment to describe how useful was the paper for this study. The scale for the relevancy factor is:

- 0 = should be already excluded
- 1 = contains only descriptions of metrics with no additional info
- 2 = some useful information related to metrics
- 3 = A good amount of relevant information regarding metrics and metric usage

3.4.1 Data synthesis

Table 2: Categories for why and how measurement usage

Categories	Sources
Iteration planning	[8][28][3][11][15] [22][12][14]
Iteration tracking	[26][34][22] [6][15][5][10][8][23] [35][36] [13][32][30][28][11]
Motivate and enable team to improve	[35][33][16][5][3] [28][32][34]
Problem identification	[26][35][23][16][32] [22][25][31][24][37]
Pre-release quality	[2][17][17][6][35]
Post-release quality	[25][3][10][32]
Changes in processes or tools	[16][26][24][5][32] [17]

4. RESULTS - WHY AND HOW ARE METRICS USED

This chapter presents the preliminary results from the systematic literature review.

High level reasons for using measurements are listed in Table 2.

4.1 Iteration planning

Many metrics center around iteration planning. Tasks for the next iteration are selected and prioritized with the help of metrics.

Many metrics were focused to help in the prioritization of tasks for the next iteration [11][12][14]. Effort estimation metrics were used scope out features that would not fit to the iteration[8]. Consequently, velocity metrics are used to calculate how many features is the team able to complete in an iteration[28]. Velocity metrics can be also used to improve next iterations estimates[22]. Also, startAndEndDate metric was used to point out interdependent tasks in planning phase[15]. Knowing the teams' effective available hours is also useful when selecting tasks for a iteration[3]. Measuring the completion of tasks enables selecting incomplete tasks to the next iteration[15]. Prioritization of features can also be affected by a metric that measures the amount of revenue a customer is willing to pay for a feature[14]. Higher valued customers get their features done first.

4.2 Iteration tracking

Purpose of iteration tracking is to make sure that the tasks selected for the iteration are completed or that necessary modifications are done to the plan to end the iteration according to schedule. Metrics help in monitoring, identifying problems and predicting end result.

Progress metrics made it transparent to stakeholders how the iteration is progressing. [26][34][22] [6][15][10] [35][36]. Progress metrics included number of completed web pages[15], story completion percentage[35] and velocity metrics[6]. However, using velocity metrics can also have negative effects such as cutting corners in implementing features to maintain velocity with the cost of quality[8]. Risk management was also mentioned often [34][35][6][33]. If it seemed that not all planned tasks could be completed, tasks were cut from the iteration [22][6][23] or extra resources were added [6][23].

When there were problems that needed to be fixed - metrics helped in making decisions to fix them - whether they were short or long term [32][6][26][5]. It was possible to base decisions on data, not only use common sense and experience [33]. Balance of work flow was mentioned as a goal for using metrics in multiple papers [28][25][11][26][6][16]. Crosstraining people to work on multiple disciplines was used to balance the work flow[23][33]. Also, metrics were used focus work on tasks that matter the most, avoid partially done work, task switching[30] and polishing of features[33]. Typically an iteration ends in a release but if too many defects are found the release must be delayed as happened in one case [13].

4.3 Motivate and enable team to improve

(Agile methods emphasize self-empowerment - this was also visible in the used metrics.) This section describes metrics that are used to motivate people and make them improve on their own.

Metrics were used to communicate different data about the project or product to the team members [35][33][28][32][34]. Measurement data motivated teams to act and improve their performance[33][28][3][5][16]. Some examples include fixing the build faster[16][5], fix bugs faster[3] or creating unit tests to declare the completion of a feature[33].

Metrics can be also used to prevent harmful behaviour such as cherry picking tasks that are most interesting. Measuring work in progress and setting WIP limits prevent cherry picking by enforcing only few work items at a time.[23]

4.4 Problem identification

Metrics are often used to identify or avoid problems in processes and work flows. This chapter describes how metrics can be used to spot problems.

There were multiple cases highlighting how measures are used to identify or predict problems so that the problems could be solved or avoided[26][35][22][25][31][24][37].

Sometimes there can be work phases where no value is added for example "waiting for finalization". This type of activity is called waste and can be identified by using lead time.[27]

Creating awareness with defect trend indicator helped to take actions to avoid problems[32]. Some metrics are used continuously while some are only checked in certain intervals(ei oo muistissa viitettï£¡ tï£¡hï£¡n). In some cases there are defined limits in metrics that trigger improvement actions(en muista viitettï£¡ tï£¡hï£¡n). One observed solution to problems was to find the root cause[16][23].

4.5 Pre-release quality

Metrics in the pre-release quality category were used to prevent defects reaching customers and to understand what is the current quality of the product.

Integration fails were a common problem to avoid with metrics[2][17]. Moreover, metrics were used to make sure that the product is sufficiently tested before the next step in the release path[17][6]. Additionally, making sure that the product is ready for further development was mentioned. Also, using metrics to improve pre-release quality was goal in one case[17]. Some metrics force the approach of writing tests before actual code[35].

4.6 Post-release quality

Metrics in post-release quality deal with evaluating the quality after it has been released.

Customer satisfaction, customer responsiveness and quality indicators are seen as attributes of post-release quality. Some metrics include customer input to determine post-release quality[25] [10][3][] while other metrics use pre-release data as predictors of external quality[32] [25][10]. Customer related metrics include for example defects sent by customers[3], change requests from customers[25] and customer's willingness to recommend product to other potential customers[10]. Quality prediction metrics include defect counts[25], maintenance effort[32] and deferred defect counts[10].

4.7 Changes in processes or tools

This chapter lists what kind of changes metrics had for processes and tools.

The successful usage of sprint readiness metric and story flow metric changed company policy to have target values for both metrics as well as monthly reporting of both metrics by all projects[16].

At Ericsson by monitoring the flow of requirements metric they decided to change their implementation flow from push to pull to help them deliver in a more continuous manner. Also, based on the metric they added intermediate release version to have release quality earlier in the development cycle.[25]

Changes to requirements managements were also done based on metrics in an other case at Ericsson[24].

The problem with broken build, and the long times to fix the build gave birth to measurements that monitor and visualize the state of the build and the time to fix it[5][16][17](ei vï£¡lttis oikea kappale). Also, additional code style rules were added to code check-in and build tools so that builds would fail more often and defects would get caught before the release[16][17]. Similarly, testing approaches were changed based on flow metrics[24][32]. ï£¡

5. DISCUSSION

I found it interesting that there were some metrics that reflected agile values, for example metrics that motivated team to act and improve. That could be seen as adhering to fifth principle of agile software: "Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done."

Also, there were hints leaning towards principle "Working software is the primary measure of progress." with a team measuring progress through working code presented to the customer(viite, ja pitï£¡is lisï£¡tï£¡ resultteihinkin ellei jo ole siellï£¡).

Moreover, there were metrics targeted to balance the flow of work. This has some similarities with principle 8 "Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely."

Maybe not that strong evidence but the metrics, which purpose was to identify problems and thus change processes sounded a bit like the last principle "At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly."

Finally, as many metrics focus on tracking and timely completion of the iteration it sounds a lot like principle 3 "Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale."

In general, I think there were many metrics that were targeted for the team - instead of high focus on managerial or upper management reporting metrics. Making metrics visible for the team enables them to independently act and improve without the need of rapid supervision and telling people what to do.

(toinen judu mitï£¡ tï£¡ï£¡llï£¡ voisi olla niin vertailu perinteisiin tai agiilikirjallisuudessa suositeltuihin) (kolmas judu: Koodimittarit oli aika heikosti edustettuna - vain muutama. Joka on sinänsä mielenkiintoista koska tutkimuksissa ne on hyvin edustettuna(kai).)) ()

5.1 Implications for research and practice

5.2 Limitations

Telecommunications sector is widely represented in this study with eight papers from Ericsson. Also, Israeli Air Force was presented in three papers.

Sometimes it was hard to understand which metrics an author was referring when a "why" was described. Moreover, I had to sometimes assume that when author describes the reasons for using a tool, he would be actually talking about the metrics the tool shows.

Whenever a new coding rule was decided it was hard to make sure that all previous primary documents would get the same treatment.

Coding "sense" improved over time so it is possible that some information was not spotted from the primary documents in the beginning of the study.

It is possible that researcher bias could have had an effect on the results. I have positive mindset towards agile methods, as well as towards certain metrics over others.

6. CONCLUSIONS

7. ACKNOWLEDGMENTS

U-QASAR rahoitus?

8. REFERENCES

[1] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4):571–583, 2007.

[2] M. Cataldo and J. Herbsleb. Factors leading to integration failures in global feature-oriented development: An empirical analysis. In *Proceedings - International Conference on Software Engineering*, pages 161–170, Waikiki, Honolulu, HI, 2011.

[3] T.-H. Cheng, S. Jansen, and M. Remmers. Controlling and monitoring agile software development in three dutch product software companies. In *Proceedings of the 2009 ICSE Workshop on Software Development Governance, SDG 2009*, pages 29–35, Vancouver, BC, 2009.

[4] D. Cruzes and T. Dyba. Recommended steps for thematic synthesis in software engineering. In *Empirical Software Engineering and Measurement*

(ESEM), 2011 International Symposium on, pages 275–284, 2011.

[5] P. S. M. dos Santos, A. Varella, C. R. Dantas, and D. B. ao Borges. Visualizing and managing technical debt in agile development: An experience report. In H. Baumeister and B. Weber, editors, *Agile Processes in Software Engineering and Extreme Programming*, volume 149 of *Lecture Notes in Business Information Processing*, pages 121–134. Springer, 2013.

[6] Y. Dubinsky, D. Talby, O. Hazzan, and A. Keren. Agile metrics at the israeli air force. In *Proceedings - AGILE Confernce 2005*, volume 2005, pages 12–19, Denver, CO, 2005.

[7] T. Dybå and T. Dingsøyr. Empirical studies of agile software development: A systematic review. *Information and Software Technology*, 50(9Ű10):833–859, Aug. 2008.

[8] A. Elssamadisy and G. Schalliol. Recognizing and responding to "bad smells" in extreme programming. In *Proceedings - International Conference on Software Engineering*, pages 617–622, Orlando, FL, 2002.

[9] N. E. Fenton and S. L. Pfleeger. *Software metrics: a rigorous and practical approach*. PWS Publishing Co., 1998.

[10] P. Green. Measuring the impact of scrum on product development at adobe systems. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, Koloa, Kauai, HI, 2011.

[11] D. Greening. Enterprise scrum: Scaling scrum to the executive level. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, Koloa, Kauai, HI, 2010.

[12] N. Haugen. An empirical study of using planning poker for user story estimation. In *Proceedings - AGILE Conference, 2006*, volume 2006, pages 23–31, Minneapolis, MN, 2006.

[13] P. Hodgetts. Refactoring the development process: Experiences with the incremental adoption of agile practices. In *Proceedings of the Agile Development Conference, ADC 2004*, pages 106–113, Salt Lake City, UT, 2004.

[14] P. Hodgkins and L. Hohmann. Agile program management: Lessons learned from the verisign managed security services team. In *Proceedings - AGILE 2007*, pages 194–199, Washington, DC, 2007.

[15] N. Hong, J. Yoo, and S. Cha. Customization of scrum methodology for outsourced e-commerce projects. In *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, pages 310–315, Sydney, NSW, 2010.

[16] C. Jakobsen and T. Poppendieck. Lean as a scrum troubleshooter. In *Proceedings - 2011 Agile Conference, Agile 2011*, pages 168–174, Salt Lake City, UT, 2011.

[17] A. Janus, R. Dumke, A. Schmietendorf, and J. Jager. The 3c approach for agile quality assurance. In *Emerging Trends in Software Metrics (WETSoM), 2012 3rd International Workshop on*, pages 9–13, 2012.

[18] B. Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33:2004, 2004.

[19] B. Kitchenham. WhatŠs up with software metrics? Ű a preliminary mapping study. *Journal of Systems and Software*, 83(1):37–51, Jan. 2010.

[20] B. Kitchenham and P. Brereton. A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55(12):2049–2075, 2013.

[21] J. Landis and G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[22] V. Mahnic and N. Zabkar. Measuring progress of scrum-based software projects. *Electronics and Electrical Engineering*, 18(8):73–76, 2012.

[23] P. Middleton, P. Taylor, A. Flaxel, and A. Cookson. Lean principles and techniques for improving the quality and productivity of software development projects: A case study. *International Journal of Productivity and Quality Management*, 2(4):387–403, 2007.

[24] S. b. Mujtaba, R. Feldt, and K. b. Petersen. Waste and lead time reduction in a software product customization process with value stream maps. In *Proceedings of the Australian Software Engineering Conference, ASWEC*, pages 139–148, Auckland, 2010.

[25] K. Petersen and C. Wohlin. The effect of moving from a plan-driven to an incremental software development approach with agile practices: An industrial case study. *Empirical Software Engineering*, 15(6):654–693, 2010.

[26] K. Petersen and C. Wohlin. Measuring the flow in lean software development. *Software - Practice and Experience*, 41(9):975–996, 2011.

[27] K. b. Petersen. A palette of lean indicators to detect waste in software maintenance: A case study. *Lecture Notes in Business Information Processing*, 111 LNBIP:108–122, 2012.

[28] R. Polk. Agile & kanban in coordination. In *Proceedings - 2011 Agile Conference, Agile 2011*, pages 263–268, Salt Lake City, UT, 2011.

[29] G. Seber. *The Estimation of Animal Abundance and Related Parameters*. Blackburn Press, 2002.

[30] M. Seikola, H.-M. Loisa, and A. Jagos. Kanban implementation in a telecom product maintenance. In *Proceedings - 37th EUROMICRO Conference on Software Engineering and Advanced Applications, SEAA 2011*, pages 321–329, Oulu, 2011.

[31] B. Shen and D. Ju. On the measurement of agility in software process. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4470 LNCS:25–36, 2007.

[32] M. Staron, W. Meding, and B. Söderqvist. A method for forecasting defect backlog in large streamline software development projects and its industrial evaluation. *Information and Software Technology*, 52(10):1069–1079, 2010.

[33] D. Talby and Y. Dubinsky. Governance of an agile software project. In *Proceedings of the 2009 ICSE Workshop on Software Development Governance, SDG 2009*, pages 40–45, Vancouver, BC, 2009.

[34] D. Talby, O. Hazzan, Y. Dubinsky, and A. Keren. Reflections on reflection in agile software development. In *Proceedings - AGILE Conference, 2006*, volume

2006, pages 100–110, Minneapolis, MN, 2006.

[35] V. Trapa and S. Rao. T3 - tool for monitoring agile development. In *Proceedings - AGILE Conference, 2006*, volume 2006, pages 243–248, Minneapolis, MN, 2006.

[36] J. Trimble and C. Webster. From traditional, to lean, to agile development: Finding the optimal software engineering cycle. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 4826–4833, Wailea, Maui, HI, 2013.

[37] D. Tudor and G. Walter. Using an agile approach in a large, traditional organization. In *Proceedings - AGILE Conference, 2006*, volume 2006, pages 367–373, Minneapolis, MN, 2006.

[38] J. Webster and R. T. Watson. Analyzing the past to prepare for the future: Writing a. *MIS quarterly*, 26(2), 2002.

APPENDIX

A. BLAA