



LUND UNIVERSITY

Variations on the Evidence-Based Timeline Retrospective Method A Comparison of Two Cases

Bjarnason, Elizabeth; Hess, Anne; Doerr, Joerg; Regnell, Björn

Published in:

39th Euromicro Conference Series on Software Engineering and Advanced Applications

2013

[Link to publication](#)

Citation for published version (APA):

Bjarnason, E., Hess, A., Doerr, J., & Regnell, B. (2013). Variations on the Evidence-Based Timeline Retrospective Method A Comparison of Two Cases. In *39th Euromicro Conference Series on Software Engineering and Advanced Applications* (Vol. 2013, pp. 37-44)

Total number of authors:

4

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Variations on the Evidence-Based Timeline Retrospective Method

A Comparison of Two Cases

Elizabeth Bjarnason¹, Anne Hess², Joerg Doerr², Björn Regnell¹

¹Department of Computer Science
Lund University,
Lund, Sweden
{elizabeth.bjarnason, bjorn.regnell}@cs.lth.se

²Fraunhofer Institute
for Experimental Software Engineering
Kaiserslautern, Germany
{anne.hess, joerg.doerr}@iese.fraunhofer.de

Abstract—Project retrospectives can be a powerful tool for process improvement through obtaining new insights. However pure experience-based reflections may lead to incorrect conclusions. Our method, evidence-based timeline retrospectives (EBTR), mitigates this risk by providing a pre-generated timeline that visualises project history based on evidence rather than relying on subjective opinions and biased memories. Through a comparative study of two cases, a set of variation points has been evaluated. The variation points enable configuring the EBTR method to different contexts and retrospective goals. The results indicate that by selecting certain variations the EBTR method can be configured to support either wide assessments (e.g. the overall impact of a new process) or assessments of a specific process area. For example, through using open or semi-structured discussions, or by varying the applied timeline technique.

Keywords—project retrospectives; process improvement; empirical software engineering

I. INTRODUCTION

Software process engineering is considered fundamental in software engineering [1][2]. Thus, the identification of weaknesses and improvement opportunities of software engineering processes is an important but challenging activity [1]. Several approaches have been proposed that support software process improvements, e.g., based on simulations [3], application lifecycle management solutions, [4], or prioritization techniques [5].

In general, retrospective analysis can be an effective tool for assessing software processes by identifying problems and best practices. Retrospective meetings can support process improvements both directly by identifying weaknesses and improvement strategies, and indirectly through team members gaining new insights and learning concerning best practices [6][7][8][9]. However, retrospectives based solely on participants' experiences of events pose a risk of drawing incorrect conclusions [10] and may become a forum for emotional venting rather than constructive discussions [6][8].

An evidence-based retrospective method was designed to combat this by injecting the retrospective with a pre-generated timeline of visualised project history based on evidence gathered from available systems [9]. This evidence-based timeline retrospective (EBTR) method was previously evaluated for one case (denoted case 1) [9].

In order to further evaluate and explore the EBTR method it was applied to a second case (denoted case 2) and the outcome compared. The aim of this comparative study was to empirically observe the effect of varying the EBTR method over a set of variation points (VPs). In this paper, we report on the influence that each VP may have on (RQ1) new insights and learning; (RQ2) timeline support for meeting; and (RQ3) topics discussed at the retrospective meeting. Though this paper describes the EBTR method, additional details are given in [9].

The remainder of this paper is structured as follows: Section II presents the generic EBTR method. Section III describes the two cases and Section IV the research method. The specifics of the evaluated EBTR variants are described in Section V. The results are presented in Section VI and discussed in Section VII, and we conclude in Section VIII.

II. EVIDENCE-BASED TIMELINE RETROSPECTIVES

Evidence-based timeline retrospectives (EBTRs) inject pre-constructed timelines into retrospective meetings. Project history is visualised in evidence-based timelines (EBTs) by displaying time-stamped evidence of project events from various systems. EBTs can prompt memory and support reflection of past events. At a retrospective meeting multiple roles share their experiences, reflect on events and good practices, and identify improvements. Kerth describes a method where a timeline is produced at the meeting by the participants [11]. Our method enhances on this by providing prepared EBTs, which saves meeting time and provides objective information. In addition, it includes one phase for planning and one for validation to ensure final agreement.

The method was initially designed for assessing RE in a project context though generic enough to be customised for different retrospective goals. The generic method is described in this section (also see [12]), while the EBTR variants for the two cases are described in Section V.

The EBTR method consists of four phases: (1) *planning*, (2) *EBT construction*, (3) *EBTR meeting* with the project team, and finally, (4) *validation* of the outcome. Each phase is described in the following sections.

A. Phase 1: EBTR Planning

The definition of *goals* in this phase enables focusing the EBTR on strategic improvement areas. The main vehicles

for achieving these goals are the EBTs (see phase 2) and a set of *focus questions* (see example in [12]). The focus questions are defined in the planning phase and used to focus EBTR meeting discussions (phase 3) on issues relevant to the EBTR goals.

The EBT aspects, evidence types, and visualisation are defined during the planning phase. The *aspects* to visualise are defined based on the goals. The type and source of *evidence* to collect and suitable *visualisations* are identified. The *projects* to include in the assessment are also selected in this planning phase.

B. Phase 2: EBT Construction

The EBTs are constructed by collecting evidence from various systems, e.g. scope and prioritisation systems, requirements databases, planning tools, defect management systems, etc. Project history is visualised by displaying this evidence along a timeline for each aspect. The visualisation technique can be varied, see Section III for a description for each case.

C. Phase 3: EBTR Meeting

The EBTR meeting is to facilitate group reflection in-line with EBTR goals and was designed according to guidelines for project retrospectives [11] and focus groups [13]. The focus questions (from phase 1) and EBTs (from phase 2) are used to stimulate a discussion.

The meeting participants represent key roles throughout the project life cycle, similar to Collier's retrospective method [6]; ideally 4-8 project members and 1 moderator.

The meeting room is prepared by posting the EBTs in a central location, e.g. on the wall. A whiteboard or flipchart, and pens and post-it notes are needed for capturing information. Seating participants around the EBTs encourages interaction with the EBTs and with each other.

The EBTR goal and EBTs are presented at the meeting. The moderator then leads a discussion based on the focus questions. A set of *prompting questions* suggested by Kerth [11] is available for reinvigorating or redirecting discussions. The participants add clarifications, corrections and additional information to the EBTs, thus, producing updated and jointly agreed EBTs.

The final part of the meeting consists of jointly summarising findings and lessons learned with a set of *sum-up questions* based on the concluding part of Kerth's timeline exercise [11]: things that worked well; what was learnt; what needs improving; what is still puzzling; and what needs to be discussed further.

D. Phase 4: EBTR Validation

In this phase the meeting outcome and conclusions are validated by the retrospective participants reviewing notes and updated EBTs. Further validation can be obtained through additional meetings to agree on an action plan for addressing identified problems and improvements.

III. THE TWO CASES

A. Case 1: Product Development Company

The EBTR method was initially designed for and applied at a company in the telecommunication domain. The company has around 4,000 employees and develops software using an agile development process. All new functionality is defined as features that are prioritised in a product backlog and developed in order of priority. Each feature is developed in a separate feature project that integrates software into software release projects. A feature project life cycle has a lead time of 9 weeks to 2 years and includes handovers between different units and teams; from request through design, development in cross-functional teams, system integration and system testing, and finally customer acceptance. Typically around 200-250 features are integrated into a main software release project.

A feature project involves several roles including product manager, project sponsor, project manager, project architect, developer and tester. The product manager acts as a customer proxy and is responsible for scope decisions. The project sponsor is responsible for ensuring resources. The feature architect is responsible for adhering to architectural strategy and guidelines. The developers and testers iteratively detail requirements in collaboration with the product manager, and develop and verify software accordingly. Finally, the feature project interacts with system-level roles for architecture, integration, and testing.

B. Case 2: Research Project

The EBTR method was applied to a German research project called IBIS [14], which involved two research partners and two company partners, SMEs (small and medium sized enterprises) with ~20 employees. The project aimed at designing a method that enables developers without specific knowledge in usability engineering to systematically design software products that are intuitive to use, creative and innovative. The resulting IBIS method was designed by integrating *image schemas* [15] (recurring cognitive structures and patterns) into a task-oriented requirements engineering process [16]; and it was defined to be easy to integrate into the company partners' software engineering (SE) processes. The usefulness and applicability of the IBIS method was evaluated throughout the research project through comparison of industrial projects conducted at each company's site some using the method.

Different roles were involved in the IBIS project: researchers that developed and evaluated the IBIS method; project managers for each company partner and SE roles defined by the IBIS method and typically included in SE projects at the company's sites. These SE roles comprise product managers and developers. The product managers elicit and specify requirements and evaluate intermediate and final product versions with the customers and end users. The developers design interactions and corresponding UIs, implement and test the software.

IV. RESEARCH METHOD

The main aim of this comparative study was to explore and evaluate variations of the EBTR method by comparing two cases. A number of variation points (VPs) were identified and an EBTR variant was applied to each case (see Figure 1). The outcome for the two cases has been analysed to identify differences potentially caused by the VPs. This comparative study was performed in three main steps: preparations, data collection and data analysis.

A. Preparations

The study was prepared at a number of meetings where the researchers discussed the EBTR method, and designed and planned this study. Previous experience of applying the method (for case 1) was shared and potential VPs were discussed and agreed. For example, the same focus questions were to be used for both cases, but for case 2 the retrospective discussions were to be more structured and limited to these questions. Furthermore, the EBTR meetings were to be longer for case 2 (4 hours vs. 75 minutes) mainly due to availability. The full set of variations points is described in Section V. Characteristics for comparing the selected projects were also discussed and agreed. The reported set is shown in TABLE I.

A separate researcher managed each case and was the contact point between this study and the investigated project(s). This researcher (Bjarnason or Hess) planned and performed the EBTR(s) for their case including constructing the EBTRs.

B. Data Collection

The same data collection protocol was used for both cases. Apart from jointly updated EBTRs, extensive notes were taken at the EBTR meetings. Transcriptions were sent to the participants for validation. Furthermore, the participants' EBTR experiences were gauged by a questionnaire with scale *Not at all*, *Somewhat*, *Fairly much* and *Very much* and a focus group with evaluation questions. Identical questionnaire and evaluation questions were used in both cases (available on-line [12]).

C. Analysis

The final set of VPs and their potential impact were identified at a workshop by the involved researchers. At this

workshop the EBTR variants for each case were presented and the collected data compared. Differences and similarities were discussed and classified as VPs or as effects of a VP.

In order to understand the impact of the variations on EBTR meeting discussions, a *topic analysis* was performed at the workshop on the notes of one meeting for each case. The researcher responsible for the case analysed the notes and identified the discussed topics. These topics were then matched to the focus question topics used at the EBTR meetings and the findings compared between the two cases.

Similarly, the focus group and questionnaire data were compiled and analysed by comparing the results from the two cases. The observed differences were then compared to the VPs and potential connections identified.

V. TWO VARIATIONS OF EBTR METHOD

The EBTR variant for case 1 was applied to three development projects (P₁₁-P₁₃), while the EBTR variant for case 2 was applied to one research project (P₂₁). Project characteristics are shown in TABLE I. The variants differ in the following VPs: (VP1) EBTR goal; (VP2) EBT content and visualisation; (VP3) EBTR meeting participant preparations; (VP4) EBTR meeting length; (VP5) discussion structure at EBTR meeting; and (VP6) EBTR meeting moderator. Each EBTR variant is described below. The relevant VPs are given within parenthesis.

TABLE I. CHARACTERISTICS OF THE INCLUDED PROJECTS.

Project id	Lead time (months)	Project size: developers of total	N:o of roles in project	N:o of EBTR particip.
<i>Case 1</i>				
P ₁₁	28	1 of 4	6	4
P ₁₂	13	1-2 of 13	8	9
P ₁₃	14	4-5 of 13	9	6
<i>Case 2</i>				
P ₂₁	7	4 of 11	4	5

A. EBTR Variant for Case 1

1) Phase 1: EBTR Planning

For case 1, the EBTRs were planned in close collaboration with company representatives and EBTR goals, aspects and evidence were defined and agreed. The main *goal* (VP1₁) was a general assessment of the RE aspects of the company's new development model and what

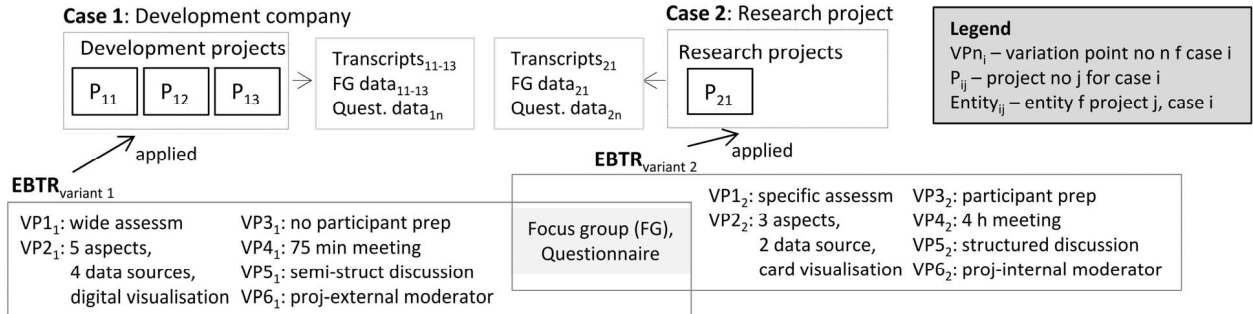


Figure 1. Overview of study setup: one EBTR variant per case. Both variants evaluated through transcription, focus group and questionnaire.

impact it has on project lead time. A secondary goal was to encourage project members to reflect and learn about good requirements practices, and thereby enable improvements in future projects. In line with these goals, evidence was selected to cover a number of *aspects* (VP2₁), namely people, project state, decisions, artefacts and planning.

2) Phase 2: EBT Construction

Evidence for each aspect (VP2₁) was selected to represent high-level events (e.g. project phase) and low-level events (e.g. customer meeting, filing of issue report). The evidence was extracted from available systems for scope, release and project planning, requirements, test and defect management. The extracted data was visualised in one MS Visio timeline per aspect (see Figure 2) and printed on A3 papers (4 in width), see [9] for further details.

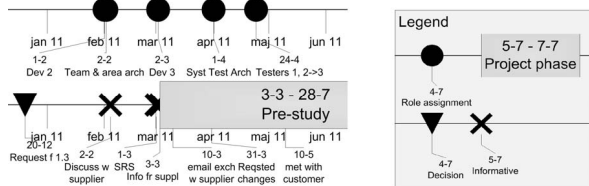


Figure 2. An extract from the EBTs used for case 1.

3) Phase 3: EBTR Meeting

Participants were invited to the meeting without prior information about the EBTR method or any other preparations (VP3₁). The meetings were booked for 75 minutes (VP4₁). Two of the meetings ran over with approx. 15 minutes. A semi-structured discussion format was used (VP5₁). The EBTs and the focus questions were used to encourage open discussions. Spontaneous follow-up questions were used to explore mentioned issues and topics. Despite attempting to include everyone there were junior participants who said nothing or very little, in particular in the largest meeting (P₁₂) and in one with two very strong leadership roles (P₁₁). The moderator was well acquainted with the case (e.g. development process and terminology) but had no prior relationship to the projects (VP6₁).

4) Phase 4: EBTR Validation

Notes of the discussions were sent out to the participants a few days after the meeting together with updated EBTs. Evidence added at the EBTR meeting was marked in the timelines with a separate colour (VP2).

B. EBTR Variant for Case 2

1) Phase 1: EBTR Planning

The main goal for case 2 (VP1₂) was to assess the usefulness and applicability of the IBIS method compared to currently applied methods; to identify problems, ideas for improvements and good practices. Secondary goals were to learn from the experience of working in a research project compared to industrial projects and to identify new ideas for future projects. Three aspects were selected: (VP2₂), namely (i) performed *activities*, (ii) important *events* and (iii) delivered *artefacts* (including planned and actual dates).

EBT visualisation was decided to be done with flip chart paper and coloured cards.

2) Phase 2: EBT Construction

Evidence for each aspect (VP2₂) was collected by the two project managers at each company. This evidence was extracted from data collected throughout the project for evaluating the applicability of the IBIS method. This activity also acted as preparation for the EBTR meeting (VP3₂). The moderator constructed the EBT with the evidence by noting it on cards using a colour scheme to separate between activities, artefacts and events. The cards were arranged along a timeline drawn on two flip-chart papers, see Figure 3. Thus, data for all aspects was represented in one EBT though visually separated by colour.

Prior to the EBTR meeting the EBTR method was briefly presented to most participants at a project meeting (VP3₂).

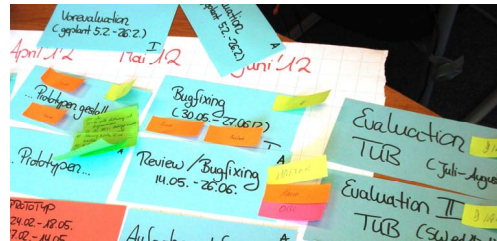


Figure 3. An extract from the EBTs used for case 2.

3) Phase 3: EBTR Meeting

The EBTR meeting was organised by the moderator who was also actively involved throughout the IBIS project (VP6₂). The meeting was opened by a brief introduction and by everyone sharing their expectations (VP3₂). The EBT and the visualisation scheme were then presented, and the topics covered by the focus questions were discussed topic by topic. The discussion was structured as follows (VP5₂): (i) the topic was presented, (ii) the participants reflected individually and noted issues on post-its (10-15 min), (iii) each participant presented their issues and added them to the EBT, and (iv) the presented issues were discussed. The meeting was concluded by a discussion on the sum-up questions (see Section II.C) using the same discussion structure as for the EBTR meeting. The meeting took 4 hours (VP4₂).

4) Phase 4: EBTR Validation

After the EBTR meeting, the outcome of the meeting was consolidated and reviewed by the participants. The final outcome has been published in a project report [14].

VI. RESULTS

The results of this comparative study are here presented according to the three facets of the research questions RQ1-RQ3 (see Section I). For each facet, the results are presented per case based on data gathered through focus groups, questionnaire and topic analysis (see Section IV). The results are related to variation points in Section VII.

All retrospective participants were present at the focus groups where they shared experiences of the retrospective including improvements. The 20 questionnaire respondents represent all roles present at EBTR meetings. For case 1, this was product manager, project manager, line manager, architect, developer and tester. Their experience in current roles varies from 3 months to 10 years (4 years for the majority) and in total ranges from 5 to 27 years (evenly distributed over respondents). For case 2, the following roles were represented: project manager, product manager, developer and company CEO. Their experience in current roles varies from 1 to 15 years, and in total 1 to 16 years.

A. New Insights and Learnings (RQ1)

For case 1, several participants stated at the focus group that they had gained and learnt from the EBTR meeting. One project sponsor said that he now realised that the new company strategy would have had an impact on this project's scoping decisions. One tester gained new insight into the overall process, in particular the early requirements phases and said: 'For me, it is very positive to see the entire picture.' A project manager said that this kind of retrospective could improve and motivate people when starting a new project.

For case 2, the participants stated that they consider EBTR meetings as very useful for reflecting on aspects that went well or could be improved. In this particular case most of the discussed issues were not completely new to the participants due to intensive evaluation activities and frequent discussions at project meetings throughout the IBIS project. However, the EBTR method did support the participants in summarising their experiences. Thus, the participants considered the outcome of the EBTR meeting a very good project result; one that has been delivered to the customer financing the research project.

Comparison of the questionnaire responses for new insights and learning from the two cases revealed some interesting differences (see Figure 4). While the participants for case 1 experienced that they gained *somewhat to fairly much* new insight and learning concerning the big, overall picture (questionnaire 4a), the degree to which participants for case 2 experienced this was *not at all to somewhat*. For good practices (questionnaire 4d), there is also a higher grading for case 1 than for case 2, while needed improvements (questionnaire 4e) are almost identical. These differences could be explained by the fact that in case 1 the

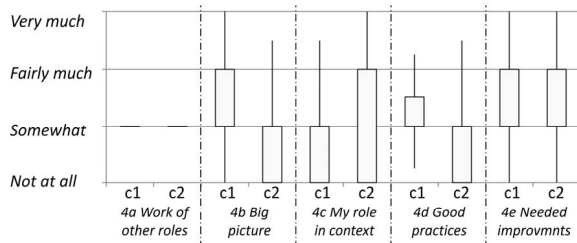


Figure 4. Questionnaire responses for new insights and learning per case (c1 and c2). Boxplots w 25/75 percentiles shown.

projects are part of a very large organisation while for case 2 the particular development projects are more stand-alone and with less 'big picture' to relate to. However, the higher ratings (for 4b and 4d) could be an indication of an effect of a VP, e.g. that the EBTs were more detailed (VP2₁) or that the moderator's lack of prior knowledge of the project (VP6₁) led to explicitly mentioning more contextual factors and practices as opposed to assuming them to be common knowledge.

B. EBT Support for Meeting (RQ2)

For case 1, several participants expressed that compared to experience-based retrospectives the EBTs supported reflection of the entire life cycle. One participant said: 'It would have been harder to discuss the project without the prepared timeline. The graphical presentation makes you think.' A product manager, and some developers and testers appreciated seeing a compilation of the big picture including the phases in which they are not actively involved. Similarly, one participant said that the method supported extending individual perspectives. Furthermore, several participants from different projects said that EBTs support memory recall and that preparing them before the meeting was preferable. One participant said: 'It helps us to remember what happened. It would've been difficult to start talking based on nothing. It's a long time since we did this.'

For case 2, the participants also said that the EBT enabled seeing the big picture and identifying relationships between events. One participant was impressed by being able to see all project activities at a glance and easily become aware of the spent effort and achieved outcome of the project. Thus, the visualisation of project history supported memory recall of certain events and reflection of relevant issues as prompted by the focus questions. Furthermore, the EBT supported the participants in identifying (previously unnoticed) relationships between issues and their consequences through the whole project life cycle. For example, some late scope changes were identified as being caused by a lack of communication at the start of the project. This previously unidentified connection enabled improving the IBIS method to avoid such problems in future.

The participants expressed that at the end of the meeting the EBT was crowded with cards and post-its, making it hard to work with. They proposed preparing the EBT with just flip-chart paper, and use cards/post-its during the meeting.

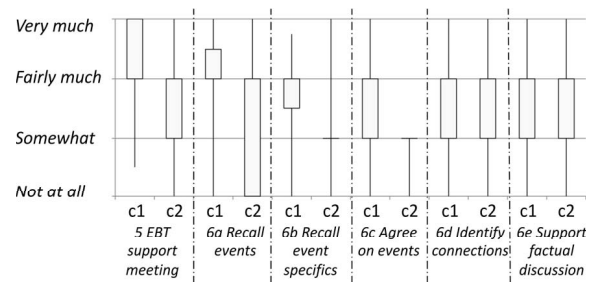


Figure 5. Questionnaire responses for EBT support for meeting per case (c1 and c2). Boxplots with 25/75 percentiles shown.

The questionnaire data concerning EBT support for the meeting (see Figure 5) indicates that the EBT variant used for case 1 provided better support for the meeting, for memory recall and for agreeing on past events than the one used in case 2. However, the EBTs were perceived to provide the same degree of support in both cases for identifying connections between events and supporting a factual discussion.

C. Topic Analysis (RQ3)

Comparison of topics discussed at one EBTR meeting per case (see Section IV.C) reveals that for case 2, all discussed topics could be matched to focus topics. While more topics in total and outside of the focus topics were covered for case 1, see Section VI.C.

TABLE II. NUMBER OF FOCUS AND NON-FOCUS TOPICS DISCUSSED AT SAMPLED EBTR MEETINGS; ONE PER CASE AND FOCUS TOPIC AREA.

	Focus topic area			Non-focus topic area	Sum
	Scope	Communication	Planning		
Case 1	4	4	4	10	22
Case 2	8	2	5	0	15

D. Limitations

Limitations and threats to validity are presented here according to guidelines by Runeson et al. [17].

Construct validity regards how well the research method correlates to the targets research questions. A combination of focus group and questionnaire was used to mitigate the risk of misinterpreting the participants' experience of the EBTR method. Variation points (VPs) were iteratively defined rather than planned from the start. In combination, with varying multiple VPs it is not possible to ensure which variation point causes which effect. However, potential dependencies between VPs have been considered.

Reliability concerns the independence of data and analysis from specific researchers. The risk of researcher bias was addressed with triangulation of meeting notes and cross-analysis of data among the authors. The results were reviewed by researchers not involved in the data collection.

Internal validity concerns whether causal conclusions are warranted and complete. The difference in EBT evidence collection poses a risk. For case 2, EBTR participants collected the evidence and may have introduced a bias, thus limiting retrospective discussions. Also, for case 2 several issues had been previously discussed, thus some insights were not new. For case 1, proponents for the EBTR method were conducting the evaluation. The risk of biasing participants' views was partly mitigated by not focusing on the method, but rather on the specific EBTR goals.

External validity concerns the ability to generalise and transfer findings to other cases. Our aim is not to draw statistically valid conclusions outside the two cases, rather to understand and describe variability aspects in relation to their contexts. Results transferability needs to be assessed by comparing our cases with other cases. To support this we

have characterised the cases and the projects. Furthermore, the results are qualitative in nature, both due to design and to limited amount of data points (shown in Figures 4 and 5).

VII. DISCUSSION

The outcome of applying the EBTR variants to the two cases is compared in this section and the potential effect of each VP (see Section V) is discussed. A summary of our interpretation of the results is also shown in Figure 6.

A. VP1: Retrospective Goals

The width, or focus, of the EBTR goal varied between the two cases, which influenced several other variation points and seems to have affected the outcome. Case 1 had a wide EBTR goal of assessing the RE aspects of their agile development process. While for case 2, the goal was to assess the IBIS project regarding communication, workload between different roles and lessons learnt regarding the IBIS method. The observed differences in amount and focus of discussed topics (see Section VI.C) and extent of new insight into the larger context (see questionnaire 4a, Section VI.A) correspond to the width of the EBTR goal. However, due to the influence that the EBTR goal had on the design of other VPs, we believe that VP1 only has an indirect effect on these factors.

VP1 affected the design of the EBTs (VP2) and the selection of discussion structure (VP5). The aspects and types of evidence for the EBTs were selected in line with the EBTR goal, i.e. for the wider goal of case 1 more aspects and evidence types were selected, while for the more focused goal of case 2 less aspects and evidence types were selected. In addition, the discussion structure was selected to match the goal width, with a more structured discussion for the more focused goal of case 2.

B. VP2: EBT Content and Visualisation

The EBTs used in the two cases varied in the amount of aspects and evidence that were included, and in the applied visualisation technique. For case 1, five different aspects were used and evidence extracted from four different systems by the moderator. While for case 2, three aspects were selected and all evidence was selected from two systems by two participants. This resulted in a larger set of data for case 1 than for case 2. For case 1, the large amounts of data were managed by visualising the evidence in several EBTs using a digitalised format. For case 2, the time stamped data was visualised in one EBT using physical cards. In both cases, the moderators were responsible for visualising the collected evidence in the EBTs.

The range and amount of evidence in the EBTs correspond well to the amount and range of topics discussed during the EBTR meetings (see Section VI.C). Thus indicating that larger and wider sets of data visualised in a clear digitalized way (as for case 1) can lead to discussing a broader range of topics. In contrast, selecting a more focused and limited set of evidence, and visualising this in a

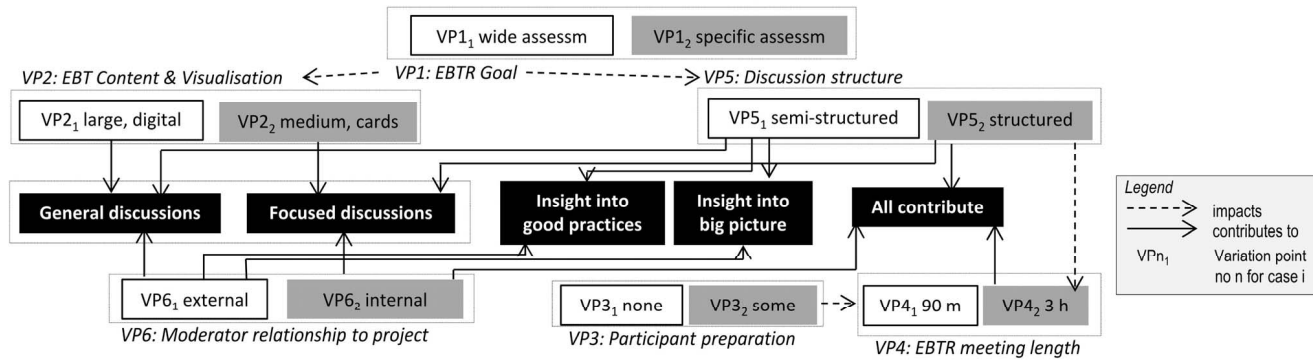


Figure 6 Summary of identified connections between variations points (VP) and effects (black boxes).

simpler way (as for case 2) can result in focusing the EBTR discussions on more specific topics (see Section VI.C). Furthermore, for case 1 the significantly higher degree of participant insight into the bigger picture (see questionnaire 4b in Section VI.A) may also be partly explained by the detailed EBTs used for this case. It is possible that they provide a richer picture of a wider range of events, not limited to current insight.

C. VP3: EBTR Meeting Participant Preparations

The degree of participant preparation was different for the two cases. In case 1, the participants were consciously not prepared due to a goal to design the EBTR method so as to require minimum development resource effort. Instead, the EBTR method was introduced at the beginning of EBTR meeting. For case 2, the EBTR method was introduced to most of the participants at a project meeting. Thereafter the participants agreed to apply the method. Furthermore, for case 2 two participants were also prepared by being involved in the evidence collection.

Participant preparation (i.e. case 2) was expected to enhance the EBTR meeting by strengthening the degree of new insights and the amount of support provided by the EBTs to the meeting. However, the degree of new insights gained from EBTR (questionnaire #4, see Section VI.A) are either similar for both cases, or higher for case 1. Furthermore, the degree of EBT support for EBTR meeting was seen as significantly higher for case 1 rather than case 2 (questionnaire 5, see Section VI.B).

This lack of observable effect may be explained by the large difference in meeting time between the cases. Even if the length of the EBTR meetings could have been reduced (or avoided running over time) by a short preparation, the total meeting time is most likely similar. The decision whether or not to prepare participants beforehand needs to be made case by case depending on the specific situation. For example, for case 1 a 10-minute presentation of the EBT visualisation could have enabled a quicker start of the actual discussions at the EBTR meeting.

D. VP4: EBTR Meeting Length

There was a large variation in the length of the EBTR meetings. For case 1, the meetings were booked for 75

minutes; 2 of 3 meetings ran out of time. For case 2, the meeting was planned for 3 hours, but took approx. 4 hours.

There are no directly observable effects of the differences in meeting length. A longer meeting time could be expected to result in a higher degree of new insights and learning. But, this is not the case. Rather, the results indicate more new insight of the bigger picture and for good practices for case 1, and the same degree of insight for improvements (see questionnaire 4, see Section VI.A). This could partly be explained by the participants in case 2 having a high degree of pre-insight into good practices due to continuous assessments throughout the project.

A positive effect of a longer EBTR meeting is that it allows for more time for discussions and could (if managed correctly) facilitate all participants having a fair opportunity to share and discuss their views. More meeting time also supports selecting a more structure discussion format (this was one of the reasons for not selecting this for case 1).

E. VP5: Discussion Structure at EBTR meeting

Two variations in discussion structure have been evaluated and found to have an effect on the outcome of the EBTR method. For case 1, a semi-structured discussion was moderated, based on the focus questions but not limited to those topics (see Section V.A.3). For case 2, the discussions were more strictly structured according to the focus questions and the participants were given time to individually reflect on each topic before sharing and discussing their views (see Section V.B.3).

The topic analysis shows that for case 1, a larger number and wider range of issues and topics were discussed at the analysed EBTR meeting. In contrast, the majority of the topics discussed for case 2 can be connected to the focus questions (see Section VI.C). Thus, selecting a structured discussion format may lead to more focused discussions and thus more specific findings. A semi-structured format may support exploring a wider area and be suitable for investigating causes and connections between topics.

Furthermore, the higher rating of new insight into the bigger picture and good practices for case 1 (questionnaire 4b and 4d, see Section VI.A), could potentially be partly attributed to the wider discussions resulting from the open discussion format.

Finally, the structured format used in case 2 encouraged all participants to equally share and discuss their views.

F. VP6: EBTR Meeting Moderator

Two different variations concerning the moderator's relationship to the project were evaluated. In both cases, the moderators had good knowledge and insight into the general domain of the projects and previous experience of moderating group meetings. However, for case 1 the moderator had no previous relationship with the projects to which the EBTR was applied. For case 2, the moderator was an active project member.

The more focused set of topics discussed in case 2 (see topic analysis, Section VI.C) may be partly supported by the moderator's existing relationship with the project. This enabled the moderator to support the discussions in identifying potential improvements, though this factor was rated at the same level for case 1 and for case 2 (see questionnaire 4e, Section VI.A). Furthermore, it was easier for the moderator for case 2 to ensure that everyone was included in sharing and discussing their views, which was harder in case 1 since the moderator did not know the name of all the participants.

In contrast, the wider set of topics discussed in case 1 (see Section VI.C) and the higher degree of new insights and good practices among the participants (questionnaire 4b and 4d, see Section VI.A) may be partly supported by the moderator having no previous relationship to the project. Thus, the specific project was new to the moderator. This may have led to the moderator asking and the participants sharing relevant information, which would otherwise have been assumed to be common knowledge and not mentioned.

VIII. CONCLUSIONS AND FUTURE WORK

Project retrospectives can be an effective way for an organisation to assess and continuously improve their development processes. By project members meeting to reflect on project history after project completion, new insights can be gained into good practices, problems and needed improvements. Team reflections can be further supported by injecting facts (evidence) into the retrospective meeting in the form of a pre-constructed visual timeline. A previous evaluation of our evidence-based retrospective (EBTR) method showed that this supports memory recall and factual discussions, and thereby enhancing group reflections around project events.

However, for this to be an effective process improvement tool rather than merely a group bonding activity, the project retrospectives need to be targeted towards strategic goals. Furthermore, the retrospective meetings need to cover topics relevant to these goals.

This comparative study has identified and evaluated six variation points of the EBTR method and their potential effect. The specificity of the EBTR goal is found to impact the retrospective outcome. By setting the variation points, the method can be customised either towards assessing a specific process area or topic, or towards a broader assessment of a

process and its influence on surrounding processes and roles. Future work includes further evaluations of the EBTR method for other cases and other combinations of VPs. In particular, evaluation of different timeline visualisation techniques is an interesting avenue to explore.

ACKNOWLEDGMENT

We thank Dr. Richard Berntsson Svensson for contributing to the data collection for case 1. We also thank the EBTR participants. The study was partly funded by EASE (<http://ease.cs.lth.se>) and by IBIS project (www.ibis-projekt.de) funded by the German Federal Ministry of Education and Research under project number 01IS11017.

REFERENCES

- [1] M.S. Emami, N. Binti Ithnin, O. Ibrahim, "Software process engineering: Strengths, weaknesses, opportunities and threats," *6th Int. Conference on Networked Computing (INC)*, 2010, pp.1-5.
- [2] A. Dikici, O. Turetken, O. Demirors, "A Case Study on Measuring Process Quality: Lessons Learned," *38th EUROMICRO Conference on Softw. Eng. and Adv. Appl. (SEAA)*, pp.294-297, 2012
- [3] R. Cabral Silva Filho, A.R.Cavalcanti da Rocha, "Towards an Approach to Support Software Process Simulation in Small and Medium Enterprises," *36th EUROMICRO Conference on Softw. Eng. and Adv. Appl. (SEAA)*, pp.297-305, 2010
- [4] H. Lacheiner, R. Ramler, "Application Lifecycle Management as Infrastructure for Software Process Improvement and Evolution: Experience and Insights from Industry," *37th EUROMICRO Conference on Softw. Eng. and Adv. Appl. (SEAA)*, pp. 286-293, 2011
- [5] T. Birkholzer, C. Dickmann, J. Vaupel, "A Framework for Systematic Evaluation of Process Improvement Priorities," *37th EUROMICRO Conference on Softw. Eng. and Adv. Appl. (SEAA)*, pp. 294-301, 2011
- [6] B. Collier, T. DeMarco, P. Fearey, "A defined process for project postmortem review, IEEE Software, vol. 13, issue 4, pp. 65-72, 1996
- [7] E. Derby, D. Larsen, *Agile retrospectives: making good teams great!* Pragmatic Bookshelf, 2006.
- [8] M. Drury, K. Conboy, K. Power, "Decision making in agile development: A focus group study of decisions and obstacles", *Agile Conference*, pp. 39-47, 2011.
- [9] E. Bjarnason, R. Berntsson Svensson, B. Regnell, "Evidence-based timelines for project retrospectives—A method for assessing requirements engineering in context", *IEEE 2nd Int. Workshop on Empirical Requirements Engineering (EmpiRE)*, pp. 17-24, 2012.
- [10] M. Jorgensen, D. Sjoberg, "The importance of NOT learning from experience", *Proc. Of European Softw. Process Improvement*, 2000.
- [11] N. Kerth, *Project Retrospectives. A handbook for team reviews*, Dorset House, 2001.
- [12] E. Bjarnason, Research study material on evidence-based timeline retrospective method on-line at: http://serg.cs.lth.se/research/experiment_packages/ebtetro/ (last accessed 2013-06-28)
- [13] C. Robson, *Real world research 2nd ed.*, Blackwell publishing, 2002
- [14] K. Fetzer, A. Hess, K. Lange et al., *Weber Schlussbericht Gestaltung intuitiver Benutzung mit Image Schemata (Final report)*, 2013 (www.ibis-projekt.de)
- [15] J. Hurtienne, K. Weber, L. Blessing, "Prior Experience and Intuitive Use: Image Schemas in User Centred Design", in Langdon, Clarkson, Robinson (Eds.), *Designing Inclusive Futures*. Springer, 2008
- [16] S. Adam, J. Doerr, and M. Eisenbarth, "Lessons learned from best practice-oriented process improvement in Requirements Engineering – A glance into current industrial RE application", *REET09*, 2009
- [17] P. Runeson, M. Höst, A. Rainer, B. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*, Wiley 2012