



LUND UNIVERSITY

V-GQM: a feed-back approach to validation of a GQM study

Olsson, Thomas; Runeson, Per

Published in:

Proceedings of the Seventh International Software Metrics Symposium

DOI:

[10.1109/METRIC.2001.915532](https://doi.org/10.1109/METRIC.2001.915532)

2001

[Link to publication](#)

Citation for published version (APA):

Olsson, T., & Runeson, P. (2001). V-GQM: a feed-back approach to validation of a GQM study. In *Proceedings of the Seventh International Software Metrics Symposium* (pp. 236-245). IEEE - Institute of Electrical and Electronics Engineers Inc.. <https://doi.org/10.1109/METRIC.2001.915532>

Total number of authors:

2

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

V-GQM: A Feed-Back Approach to Validation of a GQM Study

Thomas Olsson Per Runeson
Dept. Communication Systems
Lund University
Box 118,
SE-221 00 Lund, Sweden
[thomas.olsson|per.runeson]@telecom.lth.se

Abstract

The Goal/Question/Metric paradigm is used for empirical studies on software projects. Support is given on how to define and execute a study. However, the support for running several subsequent studies is poor. V-GQM introduces a life-cycle perspective, creating a process, spanning several GQM studies. After the GQM study has been completed, an analysis step of the plan is initiated. The metrics are analysed to investigate if they comply with the plan or has extended it, and also to investigate if the metrics collected answer more questions than posed in the plan. The questions derived from metrics are then used to form the goal for the next GQM study, effectively introducing a feed-back loop. By introducing the bottom-up approach, a structured analysis of the GQM study is possible when constructing several consecutive GQM studies. A case study, using V-GQM, is performed in an industrial setting.

1 Introduction

The Goal/Question/Metric approach (GQM) is a method for performing empirical studies on software projects [1]. The idea is to introduce the goals of a study before doing any data collection. The theory is that only useful data is gathered, leading to more cost-effective studies than if a large amount of data is gathered without a clear purpose or use.

When the data have been collected, an analysis is performed to validate if the collected metrics answers the questions in the study and that the goal of the study is achieved. However, the procedure does not take into account the potential unforeseen benefits of the metrics. For example, in a GQM study of a verification and validation process, the distribution of defects over different phases is measured. This is a process-oriented study used for identifying improvement areas. The metric collected includes information of the

distribution of defects on different parts of the product in addition to the distribution over phases. As a result, a product attribute is measured as well as the process attributes with the same metric, although this was not one of the goals.

Another aspect considered in this paper is when GQM studies are part of several consecutive projects. How should subsequent studies improve and learn from the experience of the previous studies?

This paper introduces V-GQM (Validating Goal/Question/Metric) as a method for analysing a GQM study, after the data have been collected, as a mean for validating the study. The metrics and questions are analysed, based on experience of the collected data, with the purpose of adapting it to the current environment and to attain experience for subsequent studies. The analysis does not only take into account the data collection procedures, but also external factors in the environment, for example organisational changes or technology advances, in the context of the study.

Experience Factory (EF) is a concept for gathering and storing experiences from software projects within an organisation [2]. Implementing an EF is a rigorous and difficult task [7]. Similarly to all process improvements efforts, EF has to be introduced gradually. The methods presented in this paper can be used as a bottom-up approach for implementing parts of an EF. As experiences are documented and studies evolve, GQM is put in a larger perspective, where previous experiences are a central part of designing new studies.

It should be noted that it is not the data collected that is subjects for the analysis. Rather, it is the GQM plan that is analysed. The collected data and derived results are not interesting by themselves, but aspects such as what has been collected, or if it was possible at all to collect the data, are investigated. Hence, it is only after a completed data collection and analysis phase that the principles described here are put into work.

This paper is organised as follows: Section 2 briefly presents related work. Section 3 describes the concepts and constituents of V-GQM. A case study is performed in an industrial setting, using the V-GQM method. This is presented in Section 4. A summary and a discussion on V-GQM are presented in Section 5.

2 Related work

A few studies on validation of GQM studies are published. In the work of Niessink and van Vliet [9] and Birk, van Solingen and Järvinen [3], an emphasis is put on the value of a study, rather than on the data collection.

Niessink and van Vliet argue that studies must be related to overall business and organisational goals. They define a generic process for implementing studies, which takes into account business and organisational goals [9]. The ideas are similar to one of the purposes of this paper, that is, the coupling of the overall goals in the organisation to the goals in the GQM plan. However, there are differences. Most importantly, this work is done in the context of GQM. Also, as Niessink and van Vliet [9] state themselves, the generic process is on a too high level of abstraction to directly support implementation of studies. V-GQM, on the other hand, is a hands-on approach at a practical abstraction level.

In Birk et. al. [3] it is concluded that GQM needs to be extended in the areas of management of goals for several subsequent studies, distinctions made between initial and routine implementation of studies and that tool support is important for the success of a GQM study. The idea of extending GQM to include long-term goals when investigating several, subsequent GQM studies, is the core principle in V-GQM.

Offen and Jeffery present an extension to GQM, denoted M³P [10]. M³P extends GQM by adding a couple of steps before the normal ones in the GQM approach. These are introduced to include organisational and business goals in the goal of a study. This is done before the actual study and does not focus at all on the validation of a study. V-GQM and M³P complement each other, and there is no contradiction to applying both in the same study.

3 The V-GQM method

GQM has proven to be an effective instrument for implementing successful studies [3, 8]. However, GQM is criticised for focusing too much on the data collection and too little on external factors [9], and also for the lack of a structured way of implementing several, consecutive studies [3].

GQM provides a structured approach to implementing a study. A structured approach is also needed to analyse the study as such, with a perspective ranging over more than one GQM study. The analysis approach presented here provides this mechanism.

The proposed approach can be compared to the V-model [11]. Figure 1 shows how a V-model for GQM can look like.

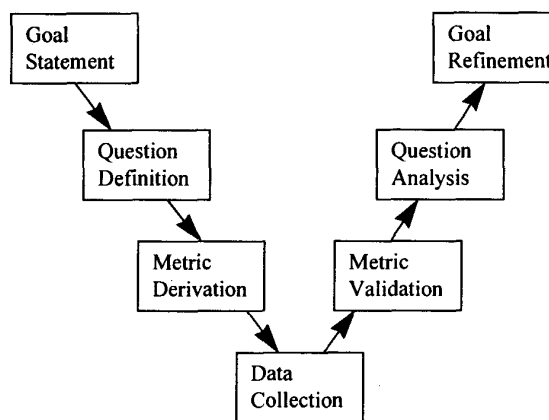


Fig. 1. V-GQM method

First, the goals (requirements) are stated. Second, the questions are defined (high-level design). Third, metrics are derived (low-level design). Lastly, the metrics are collected and the result analysed (implementation). Normally, this is where a GQM study finishes. Instead of stopping here, three more steps are introduced: Validating metrics (unit testing), question analysis (integration testing) and goal refinement (system testing).

In section 3.1, the validation procedure is described for the collected metrics. Section 3.2 present the concept of identifying questions from the newly analysed metrics. Further, section 3.3 presents how the goal of a GQM study can be evolved from the analysis of metrics and questions. Issues about the problems and risks with using this approach are discussed in section 3.4.

Just as within a software project, the actual process for executing a V-GQM study might vary. There can be iterations between the steps in the model, or the study as such might be performed in an iterative manor, as for example in the spiral model [4].

3.1 Metrics validation

The first additional step in V-GQM is the validation of collected metrics. This includes an analysis of what has actu-

ally been collected. The metrics are divided into four different categories:

- Unavailable - The metric could for some reason not be collected.
- Extended - More data is collected than suggested in the GQM plan.
- Generalisable - The collected data can be generalised, i.e. used to answer more than the question raised in the plan.
- Sufficient - The metric answers what it should answer, but not more than it should and can be collected.

As the last category does not require any more analysis, it is not further elaborated.

Extended metrics - In practise, more data is often collected with the motivation that it could easily be done while collecting other data, see Figure 2. The question is why these extensions were not included in the original plan. For example, if effort (person hours) is collected, the lead-time may also be included. The reason may be that it is easily collected at the same time. As a result, more information is collected without it being specified in the GQM plan. This is not necessarily wrong, as issues can be overlooked when writing the plan.

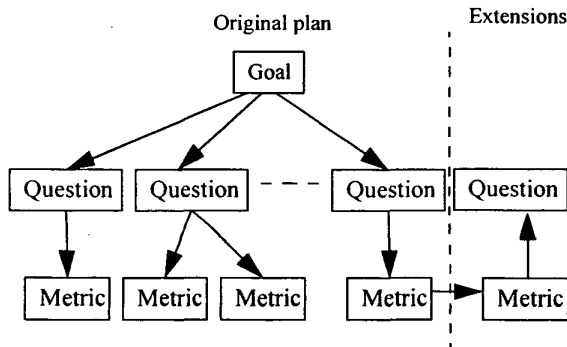


Fig. 2. Extended metrics

Generalisable metrics - Sometimes metrics answer more than is asked for, that is, the answer is generalisable. It is not obvious how to uncover these generalisable metrics without a structured analysis. As the metrics are derived from questions, it can be difficult to look beyond those for more answers than for the questions raised. Figure 3 describes the situation. Consider the following example. The aim of a metric is to collect the distribution of defects over time as the project proceeds with the purpose of identifying where to introduce new verification and validation efforts. However, the metrics collected also informs about the defect density of the product, leading to information on the quality of the product. The former is a process metric, while the latter is a product metric. Effectively, the answers ob-

tained can be generalised to answer other questions than of those in the plan.

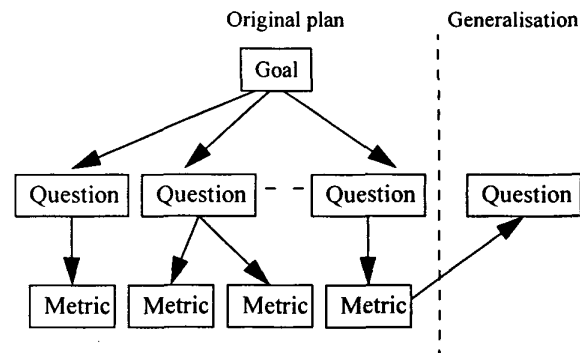


Fig. 3. Generalisable metrics

Unavailable metrics - There might also be certain data that could not be collected, see Figure 4. It is not always obvious on beforehand what data is possible to collect. There are several reasons for why some data can not be collected. It may be too costly to extract the data, sufficient measurements may be lacking from the development or the data may be of poor quality, making it useless.

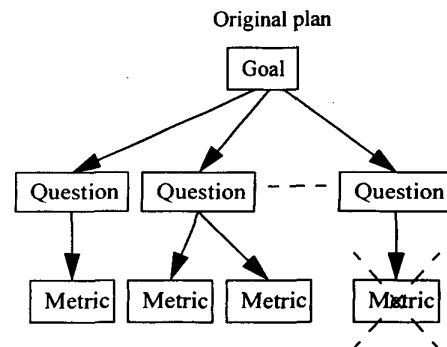


Fig. 4. Unavailable metrics

By applying a structured analysis approach of the metrics, the issues described above can be dealt with and experiences gained. Note that the aim is not to add new metrics to the plan. On the contrary, metrics might very well be discarded if they prove to be redundant or not possible to collect. The procedure aims at optimisation of the study to improve the collection effort as well as cost-benefit relationship.

The procedure for validating the metrics is as follows:

1. Identify metrics that could not be collected and what procedures that have to be introduced to be able to collect the metrics.
2. Identify metrics for which more data has been collected than required by the questions. Analyse what else has

been measured and why the extension has taken place.

3. Identify metrics for which the data answers more questions than posed. Analyse the data to disclose available information.

The order between the steps is not critical. Some might find it more instructive to analyse the questions in the order they are presented in the plan rather than as suggested here. However, it is important to categorise all the questions into one of the four categories.

3.2 Question analysis

When the metrics have been validated the questions are analysed. From the metrics validation, the questions are improved and new questions are derived. Only a very coarse filtering of the questions should be performed. Obvious irrelevant questions should be discarded. However, all other questions should be kept even if the amount is becoming quite large. The elicitation of questions should be done in the next step, goal refinement, where the overall goal is considered and a better overview is retained. If questions are discarded in this step, then they will never reach the goal refinement step and potential benefits may be lost.

One aspect in the analysis to be considered is the effort that a question requires to be answered. This information is used in the next step when the GQM plan for the subsequent GQM study is constructed. By estimating the effort, informed decisions on the extent of a subsequent study can be made. The effort is related to the metrics. However, since one metric might answer several questions, it is important to identify both the effort for a metric to be collected as well as the benefits of collecting it. If the effort is only related to a metric and not the collection of the metric, an incomplete view of the actual effort will lead to an over-estimate. Therefore, the effort should be calculated in the question analysis step, and not in the metrics validation. This information is later used in the goal refinement step.

In the question analysis, actions to resolve issues are identified. For example, if a metric can not be collected this leads to an unanswerable question. In that case it may be necessary to introduce new measurements in the development process to be able to answer the questions in the GQM study. An alternative approach is to remove the question, leading to that less information is collected.

Actions are identified in this step, but the decisions on which actions to take are made in the next step, goal refinement. Again, with the same motivation as for the effort analysis, that is, the overall goal needs to be considered rather than a single question.

3.3 Goal refinement

The last step is also the first step in the subsequent GQM study. From the metrics validation and question analysis, the goal is refined and evolved based on the experience from the previous study. The plan for the next project is drafted based on the previous plan as well as the result from the analysis of the previous study.

There are two sets of influencing factors: First, external factors such as goals of the organisation and changes in the setting. Second, the analysis from the study under analysis, giving input on the data collection as such, concerning availability of data and experience of usefulness. The two sets of factors work together to facilitate not only successful studies, but also usefulness for the organisation in large.

Based on future goals and analysis of past studies, a new GQM plan is derived. The analysis includes elicitation of questions and metrics and also potential introduction of new data collection procedures. With the analysis, informed decisions can be made regarding cost and benefit.

3.4 Risks and problems

As with any other empirical method, care must be taken to keep an objective view and a focus on the applicability of the result. Especially, researcher expectancies can be a problem. However, this is likely to have a greater effect on the first implementation of a GQM plan than on the post-analysis presented here. Still, this factor should not be overlooked.

There is a risk in using the result as a basis, since it might lead to an involvement where the study is formed to derive a certain result. This risk is not in any sense limited to the approach presented in this paper. Nevertheless, this is a risk that needs consideration.

If too much focus is paid to the data collection, the resulting analysis will be more focused than necessary and, perhaps more important, will act prohibitory on evolving and expanding the study. By keeping the long-term goal in mind, this effect can be minimised. Even if some information is difficult to collect, it can still be worth while collecting it, depending on the benefits of the information.

Another risk is inappropriate use of data. For example, if the use of a project-planning tool is concluded to imply a mature planning process a mistake can be made. The implication is not necessary true, since over-zealous belief in CASE tools are not uncommon. If the causal relationships are investigated, this can to a large extent be handled.

4 Case study: ABB Automation Products AB

A GQM study is performed at ABB Automation Products AB (ABB APR) in Malmö, Sweden. The purpose of the study is to establish a baseline of current practise. The V-GQM method presented in section 3 is used on the study as an example on how it can be used in practise.

4.1 Introduction to ABB APR

ABB APR produces software for developing and supervising large-scale industrial control systems. Products range from PID controllers to operator interfaces. The products

are used worldwide by many customers. The systems developed are distributed and of real-time character.

The development organisation is divided into two departments, one developing the operator and maintenance interface for the control systems, the other developing the control development system and core systems. The study presented here is conducted at the latter department. When referring to ABB APR later in this paper, the second of the two departments is referred to.

ABB APR has more than 20 years of experience of software development. ABB APR is close to being at CMM level 2 [6], judged by independent evaluators, not at an official evaluation. The aim is to be at CMM level 2 before the end of year 2000.

Table 1. Question definition

Process definition		Quality focus
Process Conformance	Process Domain Understanding	
1. Which test methods are used by the developers and testers?	9. How well do the testers and developers understand the requirements?	13. What is the quality of the requirements?
2. Which inspection methods are used? Which documents are inspected?	10. How well do the testers and developers understand the overall function?	14. How much time does the people involved in the project spend in each phase?
3. Which group of people are responsible for producing the following test specifications: <ul style="list-style-type: none"> • Unit test cases • Integration test cases • System test cases • Acceptance test cases 	11. How much experience do the testers and developers have with the product? Within the domain?	15. What is the lead-time of the development phases?
4. Which group of people are responsible for performing the following activities: <ul style="list-style-type: none"> • Document reviews • Unit test • Integration test • System test • Acceptance test 	12. What is the general atmosphere towards changes and improvements to the development and testing process?	16. How does the distribution of faults look like over the different phases?
5. Which design methods are used?		17. How many failures (defects still in the product) are observed in the delivered product?
6. What is the relationship between the following documents: <ul style="list-style-type: none"> • Requirements • Design • Test cases 		18. How is the internal delivery quality perceived?
7. What kinds of CASE tools are used, if any?		19. How many people are involved in the following activities: <ul style="list-style-type: none"> • Requirement specification • Development • Testing
8. How is the company organised?		One person might very well be involved in several activities
		20. When are the different documents produced?

4.2 The process improvement initiative

ABB has centrally identified software technology is a key area for strategic development. To support the various software activities it has been decided to invest money in improving the software process within all ABB companies. As a part of this initiative, ABB APR has invested in a three-year project together with LUCAS¹ with the aim at improving the verification and validation (V&V) process.

In addition to the V&V project, there are several other activities aimed at improving the software development process. For example, improving the requirement process and the estimation of project resources.

4.3 The GQM plan

A baseline of current practise is needed before introducing process improvements [12]. The GQM plan used in this case study has the purpose of establishing a baseline and also serves as basis for continuous evaluation of the current practise at ABB APR. The plan is structured as found in, for example, [5, 13].

4.3.1 Goal statement. The purpose of this project is to characterise the V&V process. As it is an initial study, a wide perspective is taken:

Analyse the verification and validation process for the purpose of characterisation with respect to effectiveness and efficiency from the viewpoint of developers, testers and researchers in the context of the Atlas project

The Atlas project is the currently running project for the next release of the software. As this is an initial study, the perspective is broad and it is expected that a number of measurements might have to be introduced to be able to control future process changes.

4.3.2 Question statement. In total 20 questions are defined. The questions are divided into two main areas, process definition and quality focus. Also, process definition is divided into process conformance and process domain understanding [5]. The questions are found in Table 1.

4.3.3 Metrics definition. The last step in the definition of a GQM plan is the metrics definition [12]. The scales

Table 2. Scale description

Scale	Explanation
Descriptive	The answer is put in general, non-quantifiable terms. For example, boundary value analysis is used when constructing unit tests.
Ordinal	The answers can be compared to each other, but have no absolute value. For example, understanding of requirements might be better or worse between two persons. However, it is not possible to quantify the understanding.
Absolute	The answer is quantifiable and has an absolute meaning. For example, number of testers involved in system testing.

used are explained in Table 2 and the categories are described in Table 3 [13].

Table 4 describes the question-metric relationship. For the descriptive metrics, observation of people involved in the development and information collected by key personnel is the main mechanism for collecting data. The information is not subjective or gradeable as the metrics are descriptive. For the metrics concerning ordinal assessment, interviews are performed with people involved in the development. Typically, subjective data is collected and statistical analysis is performed to analyse the result. Lastly, data is extracted from existing reporting system for the absolute metrics. Again, these are not subjective in nature and collected by the authors.

Table 3. Metric categories

Category	Description
Process	The answer concerns the development process as such. For example, which phases are used in the process?
Resource	The answer is a description of how many resources are used. For example, amount of time. This metric also includes people metrics. For example, the experience of a tester.
Product	The answer is a description of the product or environment on which the survey is being conducted. For example, number of faults found in the product after delivery to customer.

1. Lund University Center of Applied Software Research
<http://www.lucas.lth.se>

In total, there are 16 unique metrics defined. There are some questions that require several metrics to be defined, as well as there are some metrics that give answer to several questions.

Table 4. Question-Metric relationship

Metric	Question
Descriptive, Process	1, 2, 3, 4, 5, 6, 7, 8
Ordinal, Resource	9, 10, 11
Descriptive, Resource	12
Ordinal, Product	13, 18, 20
Absolute, Resource	14, 15, 19
Absolute, Product	16, 17

4.4 Metrics validation

The validation of the metrics is done separately in three categories, defined earlier: Unavailable metrics, extended metrics and generalisable metrics. Even though the validation is done per question, it is the metrics and not the questions that are validated.

Table 5. Metrics not possible to collect

Question	Reason
5	Design is to a large extent done implicit or by senior developers only, still mostly without formal methods. This effectively makes this data difficult to collect.
14	The time reports filed are not of sufficient granularity or consistency.
15	The time reporting system does not contain sufficient data. However, a coarse overview is possible from the metrics collected for question 16.
16	Only parts of the metrics could be collected. There exist no classification system. Making classification based on existing data would be an enormous task, as the number of reports is great.
20	As the time reporting lacks details, this classification of documents over time is not possible.

4.4.1 Identify metrics not possible to collect. The question in Table 5 are those for which the metrics could not be collected. As seen in the table, mostly absolute metrics

could not be collected. This indicates that quantitative management is not in place, which is compliant with the current CMM level [6]. Also, question 5, regarding design, cannot be answered, due to the fact that no specific design method is employed at ABB APR.

The metrics missing are: Data on design methods, data on man-hours, data on lead-time, and, partly, defect data.

As this is an initial study, it is not surprising that as many as four metrics out of 16 are unavailable. This is hopefully rectified in the subsequent studies, by introducing new measurements suggested by this V-GQM analysis.

4.4.2 Extended metrics. In Table 6, the metrics by which more is measured are presented, i.e. extended metrics. Three questions concern characterisation of the process at ABB APR and one data on distribution of defects.

For the process-related questions, knowledge was implicit and thereby not included in the questions. As the knowledge is implicit, the actual process used is not known. This information is collected in the process-related questions, leading to a documented V&V and requirement process.

Unexpectedly, the collection of defect data proved to be efficient for collecting phase data. It was assumed before the collection phase that the information on phases would be gathered from time reports. However, as the time reporting is insufficient, this is instead, to some extent, revealed from information on defects.

Table 6. Extended metrics

Question	Explanation
1	The original question aims at identifying test methods, e.g. black-box and mutation testing, coverage. However, a detailed description of the test process is collected, identifying process steps and entry and exit criteria.
9, 13	Both questions concern requirements and as in the previous question information about the requirement process is collected and documented.
16	Even though all of the information can not be collected for this question (see Table 5), other aspects are collected. To collect information on the distribution of defects, an attempt to identify phases by baseline documents is pursued. This results in some information on the different phases and their lead-time in the project.

4.4.3 Generalisable metrics. The generalisable metrics are presented in Table 7. There is no general trend in the four metrics where generalisation is possible.

Question 7 deals with the use of CASE tools. This can to some extent indicate the underlying maturity. For example, if a requirement prioritisation tool is used, this indicates a mature requirement process. A further analysis is found in the question analysis.

Question 12 has the purpose of characterising how motivated people are to change and evolve the working process. Due to the nature of the question, it is natural to ask a question regarding previous improvement suggestions.

Question 16 and 17 relate to the defects and their characterisation. As a lot of information is collected regarding distribution over time and components, this can also be used for analysis of the quality of the software product.

Table 7. Generalisable metrics

Question	Explanation
7	From which tools are being used, information on maturity and capability of the developers can be derived. For example, if a code coverage tool is used, this indicates a certain skill concerning testing. If a project-planning tool is used, then the organisation might have a good maturity concerning data collection.
12	Due to the nature of the question, a direct answer is not possible. Instead, questions regarding other process improvements and the attitude against these are asked. By this, an indication of the success and employment of earlier changes are measured.
16, 17	The two questions collect information concerning defects, before and after delivery. This information can also be used to indicate the quality of the product, not just the efficiency of the V&V process.

4.5 Question analysis

The questions are analysed with the same division of categories as the metrics, that is unanswerable, extended and generalised questions.

4.5.1 Unanswerable questions. In total, there are five questions that cannot be answered, questions 5, 14, 15, 16 and 20. Out of these, three concern time reporting, one concerns defect reporting and one concern design issues.

Question 5 cannot be answered since the design method is implicit and not conducted with any specific method. One way of looking at it is to say that this means that no specific design method is used. Another way would be that as the design is implicit, the question cannot be answered. Either the question is removed leading to reliance on question 7 (see section 4.5.3), or it is kept and the answer is considered to be that no specific design method is used. In either case, very little effort is needed.

Questions 14, 15 and 20 are seeking the answer to various issues related to time. These questions are not possible to answer due to insufficient data. It is possible through existing data, together with information on distribution of defects (question 16), to get an idea of the lead-time, but not the effort (person hours). One alternative is to leave the measurement process as is and effectively remove questions 15 and 20, and also leave question 14 with a somewhat coarse answer. This requires no additional effort for a subsequent study. If, however, a better answer is wanted, the collection procedures have to be improved. Introducing these procedures can be costly and it is likely to take at least a couple of months before reliable data can be collected.

Data about defect distribution and classification is dealt with in question 16. With existing data it is possible to relate defects in time, but not to separate components and it is not possible to classify the defects. The former can fairly easily be introduced into the defect reporting system if, for example, configuration management entities are denoted in review and test logs, as well as in defect reports. The defect classification requires introduction of a classification system and training to obtain consistency. With the current process at ABB APR this is likely to be costly and provide low quality data. Independently of course of action, the question has to be revised.

4.5.2 Extended questions. In the extended category there are four questions where the metric involved has been extended.

In question 1, issues regarding test method are asked. However, information on the test process is also collected. Two mutually exclusive actions can be taken: Either to narrow the question to include only the original sought answer, or to create two or more questions answering both process and method questions. Both actions are cheap to introduce.

Questions 9 and 13 relate to the requirements. As with the previous question, process information is also collected, not just answers concerning the quality and understanding of the requirements. Again, either the answer is made more consistent with the question, or new questions are introduced, both at a low cost for subsequent studies. Both options are viable alternatives.

Last in this category is question 16, dealing with defect distribution. It is possible to define the phases by the baseline of documents, called an artefact driven process. This information is naturally extracted from information on defects. However, it is more reasonable to define this in a separate question. The metrics would be shared among a new question on phases or artefact relationship, with question 15 and 16, and to some extent also question 6 (as the logical relationship between documents are related to, but does not necessarily imply a causal relationship as well).

4.5.3 Generalisable questions. The last category of questions is the generalisable ones. This is the fuzziest of the categories. From the analysis, four questions are identified to be generalisable.

Question 7 investigates the use of CASE tools. Even though more information can be derived from this question, it can be dangerous to make the conclusion that a project-planning tool is used then the organisation has a mature planning process. It would be more reasonable to use the information in this question to get an indication of activities performed in the project that can require a closer analysis from the study. Hence, in the area of design methods, this can be used as an early warning system if a move is made towards introducing a specific design paradigm.

Question 12 is the fuzziest question. It deals with the motivation of the developers and testers to change and improve the development process. To answer this question, thoughts and attitudes toward previous process improvements are discussed with people involved in the project. Thereby, information on earlier improvements is also gathered. It is plausible to complement this question with concrete investigation of success factors and interviews on process changes. However, this is somewhat costly and moves slightly out of focus from the original goal of the GQM study.

Information on defects, before and after delivery, is the object of question 16 and 17. This information on the product and defect reports can also be used as an indicator of quality, which is part of the V&V process. To facilitate an analysis of the quality, one or more questions can be added with little cost as no new data has to be collected (new metrics can be introduced, but it is not necessary). If, however, this is not chosen, no changes have to be done to the questions.

4.6 Goal refinement

The goals of the overall project are kept as in the original goal in the GQM plan, including the broader perspective. As an overall measurement program is not in place at ABB APR, this GQM study include V&V specific issues as well

as more general elements, such as design and requirement procedures.

The purpose of the GQM study is to baseline the current practise. However, as several questions are left unanswered due to insufficient measurements, several process improvements are derived and suggested to ABB APR.

From the analysis of the study, a couple of choices have to be made. In the discussion below, questions are selected, based on the analysis in section 4.5, and improvements identified are put to ABB APR.

It is judged that the extensions to be able to answer questions 14, 15 and 20 would be too costly to introduce in a large scale. There exists some form of reporting to the project manager, concerning tasks estimated to be finished the next couple of weeks. It is possible, without much effort from the developers, to introduce a clearer phase reporting system into the existing reporting. But introducing effort measurements are deemed to be too costly and risky, considering that several other improvements are also introduced. This is effectively complemented with the extension of question 16, as the baseline of documents can effectively be used to identify phases as well. The study is extended with one or more concrete questions concerning phases.

As for question 16, on defect distribution, a classification system is left to the future, but changes to introduce component-related reporting is put to ABB APR. This would generate much value, both concerning the V&V process as well as quality.

Questions 1, 9 and 13 are extended to include specific questions on the V&V and requirement process. By classifying activities in time and method a better view of the entire process is obtained with little more cost to subsequent studies.

Regarding the design method, question 5: There is no need to exclude the question, as the effort for collecting the information is small. Also, the potential risk in overlooking this issue is larger than the added effort for collecting the information. Hence, the question is kept as is.

From the analysis of question 7, it is deemed unnecessary to make any changes. Current practise lacks many of the areas where CASE tools can be used effectively. By maintaining this question as is, no cost is added to performing a future study and still a good indication of the improvement is kept.

One result of the study is that previous process improvement efforts have had varying success and the impact is not clearly known. As a result, question 12 is extended to include separate questions on previous improvement initiatives, as well as more concrete questions on attitude and motivation.

Lastly, even though the maturity is somewhat lagging to introduce quality measurements, this is pursued. With the

introduction of a finer granular defect reporting system, data on quality is easily derived and hence questions are deemed advisable to add in coming studies.

4.7 Case study summary

The GQM study is performed in the context of a three-year research project between ABB APR and LUCAS. In this context, a greater emphasis in the analysis is put on the introduction of new measurements at ABB APR, rather than streamlining the study to current conditions. As this is an initial study, several aspects need improvement. A couple of measurements are suggested to ABB APR as well as new questions to improve the GQM plan. Most notable is that there are several questions left unanswered. By doing the reverse analysis of the GQM plan and result, this is likely to be reduced in the future.

The case study shows how V-GQM can be used in one cycle in a larger context, including several GQM studies.

5 Summary and conclusions

The GQM approach is successfully utilised to implement improvement programs. GQM gives support for definition and collection of metrics, which are set in a certain context and related to certain goals, thereby providing more relevant information. However, there is limited support in the method to validate the GQM approach as such, in particular as a part in a long-term improvement program.

In this paper, a GQM method with validation extensions is presented, the V-GQM method. In the spirit of the V-model for software development, steps for metrics validation, question analysis and goal refinement are added to feed experiences back from one instance of a GQM study to a subsequent study. The method is presented and applied to a case study at ABB Automation Products AB.

The experiences from applying the V-GQM indicate that it provides useful support for validation of a GQM study. Thereby a subsequent study can be made better by a systematic approach to analysing the outcome of a GQM study.

Acknowledgements

The authors are thankful to Johan Nilsson, ABB APR for good co-operation in the study. This work is conducted at LUCAS, Lund University Center of Applied Software Research, which is funded by NUTEK (the Swedish National Board for Industrial Development), Swedish

industry (among them ABB Automation Products AB) and Lund University.

References

- [1] Basili, V.R., Caldiera, C., Rombach, H.D. "Goal Question Metric Paradigm". *Encyclopedia of Software Engineering* (Marciniak, J.J., editor), Volume 1, John Wiley & Sons, pp. 528-532, 1994b.
- [2] Basili, V.R., Caldiera, C., Rombach, H.D. "Experience Factory". *Encyclopedia of Software Engineering* (Marciniak, J.J., editor), Volume 1, John Wiley & Sons, pp. 469-476, 1994b.
- [3] Birk, A., van Solingen, R., Järvinen, J. "Business Impact, Benefit, and Cost of Applying GQM in Industry: An In-Depth, Long-Term Investigation at Schlumberger RPS". *5th International Software Metrics Symposium*, pp 93-96, 1998.
- [4] Boehm, B. "A Spiral Model of Software Development and Enhancement". *IEEE Computer*, pp. 61-72, May, 1988.
- [5] Briand, L., Differding, C., Rombach, D. "Practical Guidelines for Measurement-Based Process Improvement". *Software Process: Improvement and Practise*, 2(4), pp. 253-280, 1996.
- [6] Carnegie Mellon University/Software Engineering Institute. *The Capability Maturity Model: Guidelines for Improving the Software Process*. SEI Series in Software Engineering. Addison-Wesley Publishing Company, 1995.
- [7] Houdek, F., Schneider, K., Wieser, E. "Establishing Experience Factories at Daimler-Benz: An Experience Report". *International Conference on Software Engineering*, pp. 443-447, 1998.
- [8] van Latum, F., van Solingen, R., Oivo, M., Hoisl, B., Rombach, D., Ruhe, G. "Adopting GQM Based Measurements In an Industrial Environment". *IEEE Software*, 15(1), 78-86, 1998.
- [9] Niessink, F., van Vliet, H. "Measurements Should Generate Value, Rather Than Data". *6th International Software Metrics Symposium*, pp. 31-38, 1999.
- [10] Offen, R., Jeffery, R. "Establishing Software Measurement Programs". *IEEE Software*, 14(2), pp. 45-53, 1997.
- [11] Royce, W. W. "Managing the Development of Large Software Systems: Concepts and Techniques". *Proceedings IEEE WESTCON*, Los Angeles, USA, pp. 1-9, 1970.
- [12] van Solingen, R., Berghout, E. *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill Publishing Company, UK, 1999.
- [13] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A. *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Boston, MA, USA, 2000.