

Master Thesis
Software Engineering
March 2012



Evaluating GQM⁺ Strategies Framework for Planning Measurement System

Negin Asghari

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

Internet : www.bth.se/com
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 24 weeks of full time studies.

University advisors:

Kai Petersen, Cigdem Gencel

Department of System and Software Engineering, Blekinge Institute of Technology

External advisor:

Ericsson AB

Multimedia

Ölandsgatan 1, SE-37123 Karlskrona, Sweden

www.ericsson.com

Contact Information:

Author:

Negin Asghari

Address: Bredgatan 16, 37132, Karlskrona, Sweden

E-mail: neginasghari@gmail.com

ABSTRACT

Context. Most organizations are aware of the significance of software measurement programs to help organizations assess and improve the ways they develop software. Measurement plays a vital role in improving software process and products. However, the number of failing measurement programs is high and the reasons are vary. A recent approach for planning measurement programs is GQM+Strategies, which makes an important extension to existing approaches, it links measurements and improvement activities to strategic goals and ways to achieve this goals. However, concrete guides of how to collect the information needed to use GQM+strategies is not provided in the literature yet.

Objectives. The contribution of this research is to propose and assess an elicitation approach (The Goal Strategy Elicitation (GSE) approach) for the information needed to apply GQM+strategies in an organization, which also leads to a partial evaluation of GQM+strategies as such. In this thesis, the initial focus is placed on eliciting the goals and strategies in the most efficient way.

Methods. The primary research approach used is action research, which allows to flexibly assess a new method or technique in an iterative manner, where the feedback of one iteration is taken into the next iteration, thus improving on the method or technique proposed. Complementary to that, we used literature review with the primary focus to position the work, explore GQM+strategies, and to determine which elicitation approach for the support of measurement programs have been proposed.

Results. The Goal Strategy Elicitation (GSE) approach as a tool for eliciting goals and strategies within the software organization to contribute in planning a measurement program has been developed. The iterations showed that the approach of elicitation may not be too structured (e.g. template/notation based), but rather shall support the stakeholders to express their thoughts relatively freely. Hence, the end-result was an interview guide, not based on notations (as in the first iteration), and asking questions in a way that the interviewees are able to express themselves easily without having to e.g. distinguish definitions for goals and strategies.

Conclusions. We conclude that the GSE approach is a strong tool for the software organization to be able to elicit the goals and strategies to support GQM+Strategies. GSE approach evolved in each iteration and the latest iteration together with the guideline is still used within the studied company for eliciting goals and strategies, and the organization acknowledged that they will continue to do so. Moreover, we conclude that there is a need for further empirical validation of the GSE approach in further full-scale industry trials.

Keywords: Goal-based measurement program, GQM, Goal definition, Goals and Strategies elicitation

CONTENTS

ABSTRACT	III#
CONTENTS	IV#
LIST OF FIGURES AND TABLES	VI#
1# INTRODUCTION	1#
2# CHOICE OF RESEARCH METHODOLOGY	3#
3# BACKGROUND AND RELATED WORK	5#
3.1# SOFTWARE MEASUREMENT PROGRAMS IN GENERAL.....	5#
3.1.1# <i>Benefits of Running Measurement Programs</i>	5#
3.1.2# <i>Challenges</i>	5#
3.2# OVERVIEW OF GOAL-BASED MEASUREMENT PROGRAMS.....	6#
3.2.1# <i>GQM</i>	6#
3.2.2# <i>GQIM</i>	8#
3.2.3# <i>Model, Measure, Manage (M³P) Paradigm</i>	10#
3.2.4# <i>MIM</i>	11#
3.2.5# <i>GQM⁺ Strategies</i>	12#
3.3# RELATED WORK ON GQM+STRATEGIES	14#
4# ACTION RESEARCH ON GSE-APPROACH	16#
4.1# ACTION RESEARCH DESIGN	16#
4.1.1# <i>Research Context</i>	16#
4.1.2# <i>Research Process</i>	16#
4.1.3# <i>Design of Elicitation Meetings and Observations</i>	18#
4.2# RESULTS.....	21#
4.2.1# <i>Initial meetings</i>	21#
4.2.2# <i>Archival data</i>	22#
4.2.3# <i>Action Research Iterations</i>	23#
4.2.4# <i>GSE 1.0 – GSE Notation Based Interviews</i>	24#
4.2.5# <i>GSE 2.0 Approach</i>	28#
4.2.6# <i>GSE 3.0 Approach</i>	29#
4.2.7# <i>GSE 4.0 Approach: Released Version</i>	30#
5# DISCUSSION	35#
5.1.1# <i>Main Lessons Learned</i>	35#
5.1.2# <i>Comparison with Related Work</i>	35#
5.1.3# <i>Validity Threats</i>	36#
6# CONCLUSION	38#
6.1# RESEARCH QUESTIONS.....	38#
6.2# FUTURE WORK.....	39#
7# REFERENCES	40#
8# APPENDICES	43#
8.1# INTERVIEW STRUCTURE VERSION I.....	43#
8.2# INTERVIEW STRUCTURE VERSION II.....	46#

LIST OF FIGURES AND TABLES

Figure 1-GQM notation in three levels.....	7#
Figure 2-The GQIM Model	9#
Figure 3-The M ³ P Framework.....	10#
Figure 4-Measurement Information Model (ISO/IEC 15939).....	12#
Figure 5- GQM+ Strategies Method.....	13#
Figure 6- GQM+ Strategies component	14#
Figure 7- Research Design.....	17#
Figure 8 - GSE Iterations.....	23#
Figure 9 - GSE Notation.....	25#
Figure 10-GSE Example.....	25#
Figure 11 - GSE Tree for Customer Responsiveness program.....	27#
Figure 12 -MindMapSkeleton.....	29#
.....	
.	
Table 1- Research Methods Alternatives.....	3#
Table 2-Context Elemetns	16#
Table 3 Triangulation	17#
Table 4 Interview Design	20#
Table 5-GSE Status in each interations	23#
Table 6- GSE Questionnaire.....	28#
Table 7- Rational of Questions in Questionnaire.....	28#

1 INTRODUCTION

In today's competitive software development market, there is an increasing need for having a well-structured measurement program for software organization to understand, control and/or improve the current status of software process/product/resources. A well structured measurement program gives organizations the ability to follow up on goals at different levels of the organization by translating business strategies to the set of operational goals and break them down to respective measurements needed to evaluate them [3][5]. The key success for measurement initiatives is to provide a well-defined plan for executing measurement programs, which can be efficiently, implemented afterwards [23]. A high quality software measurement program should lead to sound, complete, lean, and consistent measurements [24].

Nowadays, software organizations are aware of the importance of measurement programs for sound and fact-based decision-making. However, the success rate of running measurement programs is low with a variety of reasons why measurement programs fail (see e.g. [24] [17] [16][5]). According to the reference just mentioned, the following reasons for failed measurement programs can be found: it is not uncommon for organizations to collect redundant data and irrelevant data that nobody use; collect data that might be useful but people are not aware of it; low commitment for the measurement program among different responsible people within the organization; lack of transparency with regard to the benefit of measurements in their role in software process improvement, good decision making, and software value management.

Researchers believe that any measurement should be done based on the goals of the organization and there are many goal-based measurement programs that have been proposed to contribute to planning measurement programs such as; Goal Question Metric (GQM) [4], Goal Question Indicator Metric (GQIM) [10], and GQM+ Strategies [3][5]. Furthermore, work supports different steps of the goal-based measurement process, such as providing support for metric selection. For example, the framework for Optimum Measure Set Decision (OMSD) is presented in [8][9], which helps organizations select the best set of measures with regard to their goals, and Structured Prioritized Goal Question Metric (SPGQM) [31] using a structured measurement framework supporting prioritization of goals and metrics through tractability.

GQM is a well-known framework, which has served the industry for quite a long time. It is a top-down approach to define measures based on the goals of the organization. GQIM tried to make the GQM approach more comprehensive by providing a linkage between business goals and other measurable goals in software organizations with a more structured and comprehensive model. Recently a new approach emerged, called GQM+Strategies [3][5]. The approach extends the GQM approach by introducing business goals to which measurement and operational goals are linked. With this the GQM-approach gains a long-term perspective and hence the measurement program is more likely to be sustainable. Furthermore, strategies of how to reach the goals are incorporated to be able to follow up on whether a strategy was successful in reaching the business goals. Given the benefit of sustainability of the measurement programs based on GQM+strategies this approach forms the basis of this thesis.

As mentioned earlier, there is e.g. support in selecting metrics through optimization. However, in order to run GQM+Strategies plenty of information has to be collected (regarding business strategies and goals from multiple perspectives). We believe this to be challenging and complex in large organizations. In analogy, e.g. eliciting requirements for a very complex system is also challenging, and a vast amount of research has been

investigating requirements elicitation. Hence, it was checked whether elicitation approaches are available and evaluated in order to run goal based measurement programs in general, and GQM+Strategies in particular. We applied the following search on Google Scholar to figure out whether this research gap has been investigated: (*GQM OR “goal based measurement”*) *AND elicitation*

The majority of papers (over 500) returned were related to requirements engineering, but none was focusing on **how to** elicit information for a measurement program in general, or GQM+Strategies in particular. Hence, as we identified this gap this is the starting point for the research in this thesis.

The aim of this thesis is to aid organizations in eliciting information to be able to run GQM+Strategies in their organization. For this purpose two objectives are to be fulfilled:

- 1) Design an elicitation approach for GQM+Strategies and apply it in organizational context.
- 2) Evaluate the elicitation approach and based on evaluations update the elicitation approach accordingly.

The following research questions are answered in this research:

- *RQ1: What do we know about GQM+Strategies with respect to its practical application?*
The purpose of answering this research question is to learn about the detailed structure and needs of how to support GQM+strategies through the elicitation approach (e.g. what information is required as input).
- *RQ2: How should software organizations elicit goals and strategies as input for GQM+ Strategies?* The answer to this question aims at finding a useful and practically applicable way of eliciting information for GQM+Strategies.

The remainder of the thesis is structured as follows. Section 2 presents the reasons for choosing the research methodologies. Section 3 presents the related work on running measurement programs in general, and GQM+Strategies in particular. Section 4 illustrates the action research used in this thesis, including research design as well as results. Section 5 discusses the results, and Section 6 concludes the thesis.

2 CHOICE OF RESEARCH METHODOLOGY

Research is a set of process to collect, analyze and interpret information to answer to the desired question(s). Research methods are set of tools and techniques used to accomplish the researcher's work (J. Creswell, 2009). Research methods are divided into three groups, namely qualitative, quantitative, and mixed methods. An overview of the different methods is given in Table 2.

Table 1- Research Methods Alternatives

Research method	Description	Selection	Data collected
Literature review/Systematic Literature Review	Literature reviews are distinguished in terms of how systematic they are. Reviews according to [19] aim at aggregating evidence and have a thorough process of finding and aggregating evidence. If the purpose is not to aggregate evidence, a classical literature review can be conducted, which summarizes the articles relevant and mostly aims at positioning own work.	Yes, Classic literature review	Qualitative, Quantitative
Action Research	Action research is done in real world settings and the researcher is taking an active part or role in the research. The researcher introduces an action (e.g. a new method) and during active participation observes what is working or not working. The researcher collaborates during the introduction of the change [12].	Yes	Primarily qualitative, and quantitative
Experiment	In this method the researcher actively influences something to observe the consequences in a controlled environment. They manipulate one or more independent variables and control and measure any changes in other variables [25].	No	Primarily quantitative and use of statistical inference
Case study	Case studies are observational studies focusing on a concrete case (e.g. a company or a project) and do post-hoc observations to understand or explore a real-world setting [33].	No	Primarily qualitative, observational study with no use of inference

We first chose normal literature review, as the main purpose is to position the research and gather and provide the needed background information. We did not use systematic literature review as the initial search showed that there is very limited information regarding research on how to elicit information from stakeholders to run measurement programs. That is, there was no need to aggregate evidence. In the literature review, however, we have a structured approach in identifying and selecting the literature based on forward snowballing.

We chose action research as in our case we propose a solution that is applied in elicitation meetings where the researcher acts as measurement responsible in the organization. With several meetings being held, this allows for iterative improvement, which is part of action research. Case study is more focused on post-hoc observations (e.g. what happens after something has been introduced/used for a while), and is therefore not used. Furthermore, we need real experts that have good knowledge about their information needs as otherwise we cannot test the elicitation approach with respect to their knowledge to be elicited. This also excludes an artificial setting such as in a lab experiment.

3 BACKGROUND AND RELATED WORK

This section shortly motivates the need for measurement programs, elaborates on the benefits and challenges related to running them, and provides an overview of the well-known measurement programs.

3.1 Software Measurement Programs in General

The main purpose of software measurement programs is to provide tangible evidence whether projects meet specified goals or not. Software organizations, like any other organization, depend on measurements to provide the organization with a corporate memory that can be used to control, understand, improve, assess, and evaluate the different aspect of their products and product development lifecycles [34].

3.1.1 Benefits of Running Measurement Programs

Those software organizations that pursue measurement program have a competitive advantage over those not having measurements in place [26]. They can ensure that their organization's product/processes/resources are efficient given that they are able to collect meaningful data and sound measures. Generally, measurement is a way to assess and evaluate software entities against a baseline. Having well defined measurement programs can work as driving force for software process improvement by making informed decisions. It enables software organization to benchmark their performance against other organizations (e.g. comparing lead-times or product quality). Besides that, measurements facilitate communication among stakeholders at different levels of the organization.

3.1.2 Challenges

Despite all the efforts invested on measurement programs during the last years, planning and implementing a well-structured measurement program is still a big challenge for many organizations. Although, software organizations are aware of the significance of measurement program, the success rate of it still is low for various reasons [24][17] [16][5]:

- There are difficulties in software organization in deciding how to collect metrics as well as which metrics should be collected. It is common that many organization collect redundant data, too much unnecessary data or they collect data that nobody uses [5]. It is also common in a big organization that there are some collected data, which might be useful, but people are not aware of existence of them [24]. This underlines the need to have a good way of eliciting information in a good way to make sure we collect data that matters.
- The commitment and buy-in among stakeholders is poor. The benefits of measurement are not transparent and clear across the organization. It happens that stakeholders are not aware of the main purpose of data collection. Therefore, without this transparency, it is difficult to motivate the management group and related stakeholders to support the measurement program continuously and also collect the right data [5].
- We lack structured ways for analyzing data, which can result in unsound strategic decision-making. This creates numerous problems, such as reducing cost effectiveness of measurement programs, as well as increasing metric ambiguity when managers and developers interpret measurements [5].
- There are many templates defined for eliciting required information for measurement program. Basically, these templates and structures defined what to collect rather than how to collect the required data. Having a well defined approach for collecting and

eliciting a right information results in avoiding gathering irrelevant data or overhead in collected data. As mentioned earlier, this is where this thesis aims to contribute.

Many goal-based measurement frameworks have been introduced to overcome these problems. The most well known is Goal Question Metric (GQM), which was introduced by Victor Basili in 1984. GQM has served many software and non-software organization for planning their measurement programs. This approach was followed by many extensions, such as V-GQM [28] and GQIM [10].

The most recent extension of GQM is GQM+Strategies, which tries to address the weaknesses of existing goal-based approaches for planning measurement program, by making them sustainable through the incorporation of long-term goals and relate those to strategies. This approach tries to avoid collecting useless, harmful, and expensive data by providing the linkage among data related to goals at different level of organization. This provides transparency, both for managers and measurement stakeholders. The first version of GQM+ Strategies was published as a white paper. A series of publications was published to illustrate the method's usefulness on hypothetical examples [20] and also in a case study [21], which evaluated this framework in a telecommunication company using blooms taxonomy, the results indicating applicability of the framework in a practical context.

In the following we provide an overview of existing goal-based measurement approaches, which is then followed by related work on GQM+strategies, for which we propose strategies for eliciting the information needed to apply it.

3.2 Overview of Goal-Based Measurement Programs

3.2.1 GQM

Goal Question Metric is a top down approach, which suggested that measurements in software development should be driven by measurement goals. This approach has been developed by Victor Basili and his colleagues in 1984 and has been evolved over a time. GQM has been served as a measurement program in industry for a long time with the purpose of improving the process and quality of software product development. GQM defines a measurement model consisting of three levels as shown in Figure1. Measurement goals represent the goal that should be achieved with the measurement; questions represent the information need (i.e. answers to the questions help to achieve the measurement goal). On the bottom level we have the measures that help in answering the questions.

The GQM approach is mostly used as to set operational goals for software projects. In this approach, each metrics connect indirectly to a goal though one or more question(s), therefore no metrics are defined because of measurement sake [7], a metric has no justification if there is no measurement goal associated to it.

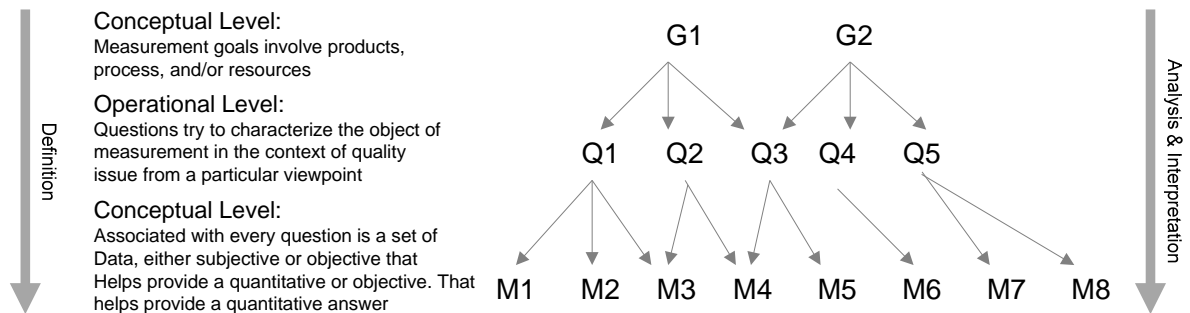


Figure 1-GQM notation in three levels

In the following we summarize the reported strengths of the GQM approach. First of all, measures are based on the goals of the organization, which prevents from collecting irrelevant measures and hence data [13][11]. As an example, GQM has been applied in several organizations such as Motorola, HP, or AT&T and has helped to focus on collecting useful data [1]. Furthermore, the Quality Improvement Paradigm has been used successfully, of which GQM is an integral part and can be, for example, used to support brainstorming activities to identify measures. Furthermore, studies mention cost effectiveness of collecting data when applying GQM to decide on metrics [3][31].

At the same time, the use of GQM is related to some challenges. One critique on GQM is that it is not repeatable and reliable and depends very much on the person using it, e.g. two different persons with the same goal will end up with different questions and metrics [18] (Card, 1993). Furthermore, it is also hard to find out when to stop adding goals and metrics (preferably we would like to measure “everything”) [13]. Change is not managed systematically as well, if the business model of an organization changes its goal changes, and measures thus have to change as well [13]. It is acknowledged that GQM has benefits, however, there is a lack of evidence with regard to the cost of GQM-based measurement programs in comparison to running measurement programs in a different way. The cost can be, for example, increased unnecessarily as GQM is a top-down approach and hence might ignore data already collected at the company [11]. Berander and Jönsson [7] mentioned three challenges of using GQM for planning measurement program. Firstly, there is a risk of metric explosion, which can be a hindrance to an industrial application of a measurement framework. Secondly, GQM might not cover all process perspectives, which are supposed to be measured. Hence, the GQM tree might become incomplete and unbalanced. Thirdly, GQM only considers those perspectives defining the GQM tree, which might hamper the organizational acceptance. Basili et al. [5] elaborate that GQM does not encourage measurement stakeholders to document context criteria, assumptions, and estimated unknowns, which can affect the interpretation of data. Overall, GQM does not explicitly motivate and integrate the measurement of goals at different level of organization (e.g. from strategic to more operational) [5]. That is one of the motivations why GQIM and GQM+Strategies were introduced.

3.2.2 GQIM

Park et al. [29] developed the framework named GQ(I)M, which is similar to GQM, but more comprehensive. GQ(I)M is the acronym for Goal Indicator Measure and the “I” in parenthesis emphasizes the difference between this framework and GQM. GQ(I)M is another goal-driven measurement program, which extends GQM with different steps, for example the incorporation of business goals and actions to be taken to implement the measures. Furthermore, indicators (i.e. definitions of charts and visualizations) play an important role. Figure 4 represents the GQIM notation according to [11]. The GQIM process has 10 steps [29] as described below:

1. Identify business goals
2. Identify what information needs (things you want to know/learn)
3. Identify sub-goals
4. Identify the entities and attributes related to those sub-goals
5. Formalize measurement goals
6. Identify quantifiable question and indicators that are used to achieve measurement goals
7. Identify the data to be collected to construct the indicators, which help to answer the questions
8. Define the usable measures and make these definition operational
9. Identify actions to take for implementing defined measures
10. Provide plan for measure implementation

The first four steps are for defining business goals and structure the organizational objectives that can be used further on as a basis for the measurement goals. This methodology start with business goal of the organization at high level and breaks them down to the sub goals, and then it proceeds to the measures and indicators, which try to achieve those goals. GQIM has developed in Software Engineering institute (SEI).

The GQIM model has an advantage over GQM, contrary to GQM which focus more on project level and measurement goals, GQIM tries to map business objectives to the organization’s information needs and to develop indicators and metrics to monitor the progress in attaining goals [10]. This model is far more comprehensive than GQM by maintaining tractability to overall business goals [10]. This gives the model a more long-term perspective, hence making the measurement program potentially more sustainable. Furthermore, Boyd [10] showed the applicability of GQIM.

However according to [17] GQIM consist of many templates to be filled in and steps in the measurement process, which makes GQIM difficult and exhausting to follow. Furthermore this extra documentation grows over the time and makes managing the measurement program troublesome when the measurement program evolves [17].

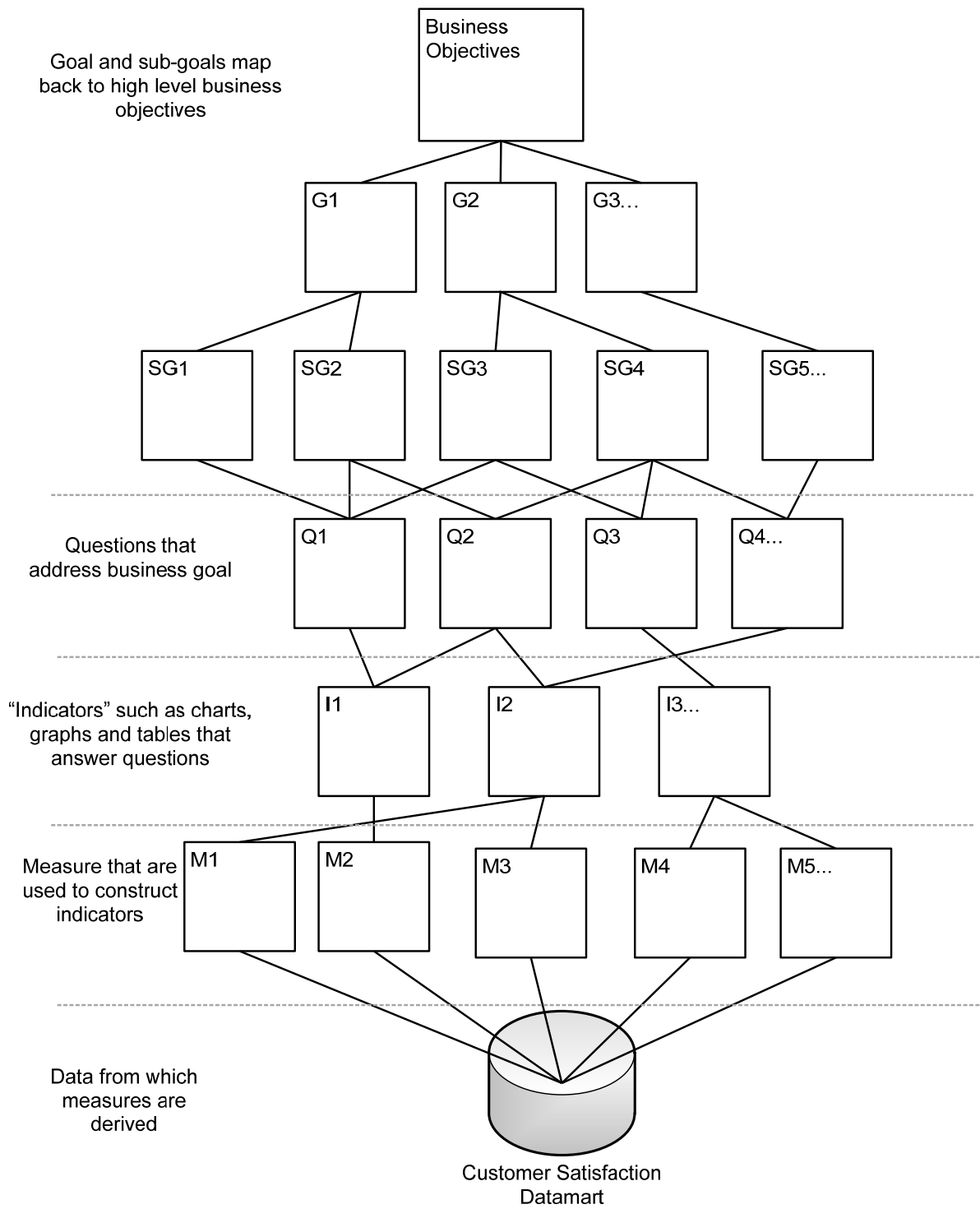
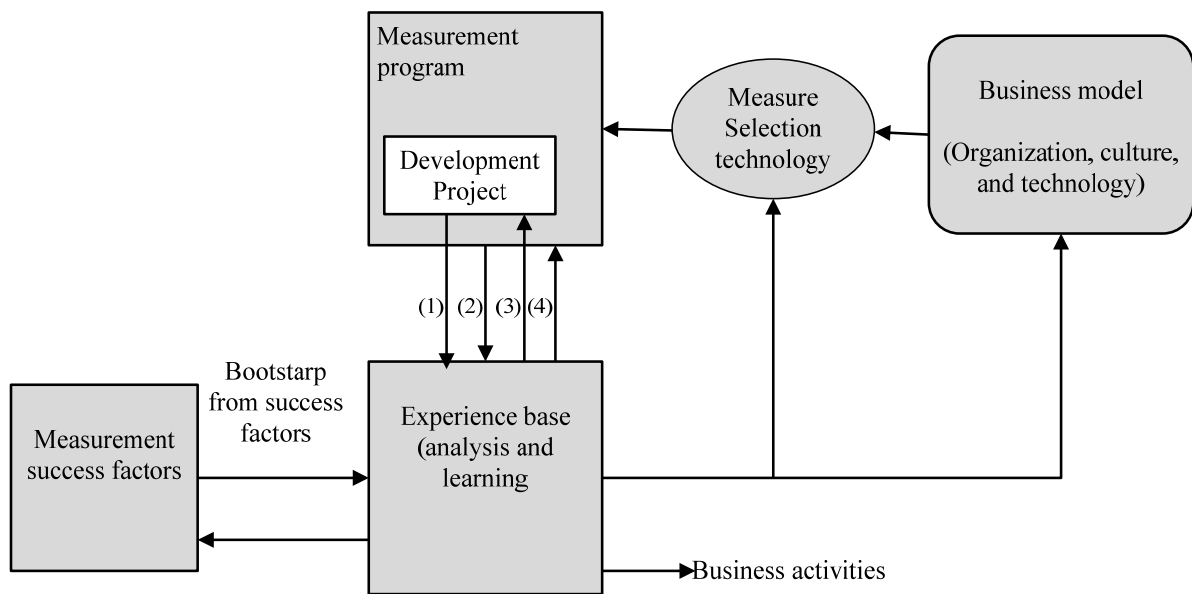


Figure 2-The GQIM Model

3.2.3 Model, Measure, Manage (M³P) Paradigm

Offen and Effery [27] introduced this model in 1997, which is an extension of Quality Improvement/Goal Question Metric. The main focus of M³P is to propose a well-defined linkage among the numerical data and related development and business context by coupling business, organizational and technical issues into a measurement program context. The context here includes e.g. organizational factors, cultural factors, and existing technology. Figure 5 represents the notation of this model. It distinguishes the measurement program focusing on long-term goals, and the project focused improvement program, which focuses on more short term/project goals.



- Key:
- (1) Measurement of measurement program (metameasures)
 - (2) Project-specific data
 - (3) Project improvement (short-term)
 - (4) Program improvement (long-term)

Figure 3-The M³P Framework

In contrary with GQM, this model is focusing on the business goals and measurement implication, which avoid collecting irrelevant data, and tries to ties up the business goals to the measurement goals and respective sub-goals at different level of organization [27]. This helps avoiding collecting data just for the sake of collecting data, and provides traceability among goals at different level (long term view as well as short term view). Besides, M³P model prevents an explosive growth of goals, sub-goals and respective questions and derived measures [27]. This model extends the QIP/GQM and overcome the linkage problem between numerical data and the development and business context.

Like other measurement models, M³P also has some limitation and challenges. According to [27], this model has difficulty in translating the business strategies into goals and strategies. It needs a lot of effort to establish a common understanding among the different involved stakeholders in respective measurement program [27]. Besides, Basili et al. [3] believe that

in this model with goals at different levels being connected to measurements, it is not easy to aggregate measurements consistently for each goal (e.g. how to propagate measurements upwards from the bottom of the goal tree to the top).

3.2.4 MIM

The main purpose of this framework is to help manager in collecting required information to plan, implement and evaluate a measurement program as well as improving software measurement within the project or organizational measurement structure. It does so by providing a standard definition of terminologies with respect to the collection of measurements to fulfill information needs (that correspond to questions in GQM). This model defined a set of processes and activities in which we one adequately specifies the required measurement information, measurements, how the analysis results can be applied and finally how to make a good decision based on the valid results (ISO/IEC 15939). Figure 6 depicts the MIM model.

In the model an entity represents a real-world construct (e.g. source code), which has an attribute (e.g. size). The attribute then can be measured (e.g. through lines of code). Different measures of attributes can be aggregated (e.g. adding up different sizes of different components, or combining size with another measure such as effort as a project attribute to determine productivity). These are then presented as an indicator and an associated interpretation, which then fulfills the information need of the stakeholder (i.e. answers their question).

This model is a standard model, which provides a common terminology to share information about measurement programs [15]. That is, it provides a common language, which improves communication among researchers and practitioners on software measurement issues [14].

One of the challenges is that it does not show the economical aspect of the software measurement e.g. human resources cost [14]. This model has capacity for improvement by adding new features; e.g. depiction of the external source of measurement [14].

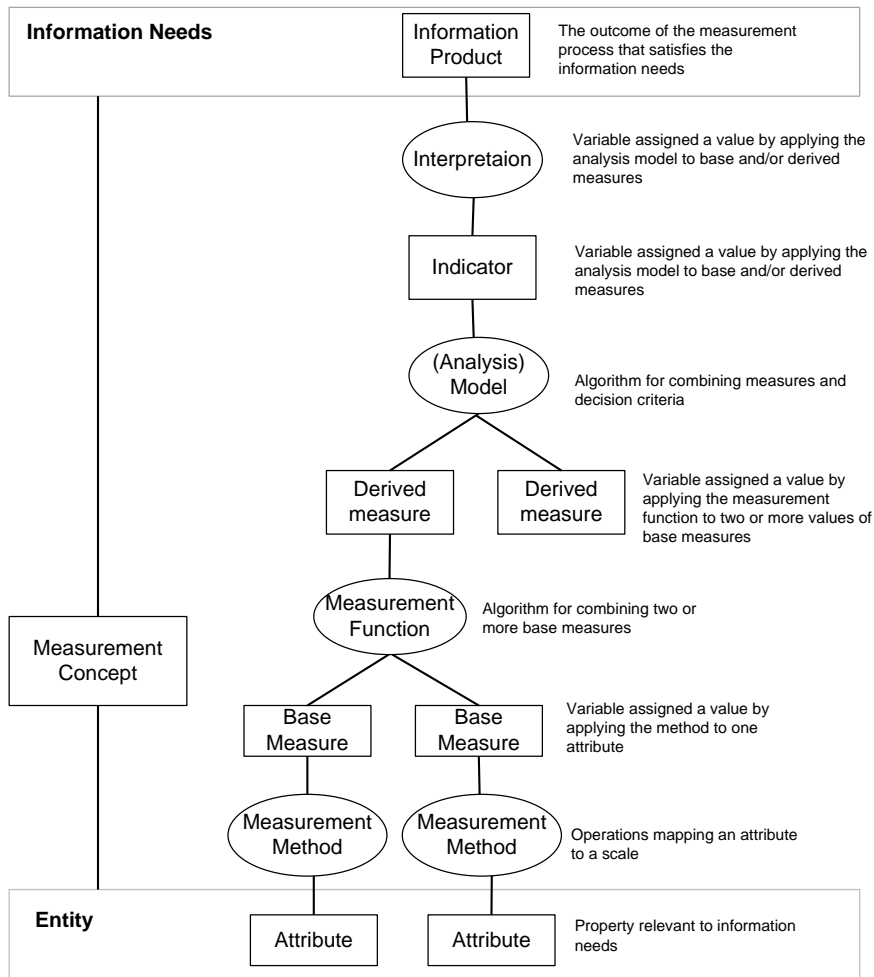


Figure 4-Measurement Information Model (ISO/IEC 15939).

3.2.5 GQM⁺ Strategies

GQM⁺ Strategies framework depicted in Figure 7 is based on the familiar GQM paradigm that adds a capability to it to create measurement programs with alignment between business goals and strategies, software goals, and measurement goals [3][5]. The basic idea of the framework is to explicitly link measurement goals to higher-level goals of an organization and also to goals and strategies of the overall business. GQM has no explicit support for integrating its software measurement model within different level of organization, e.g. more strategic product management level down to project/implementation level. Furthermore, GQIM tried to be more comprehensive in providing linkage among business goals and other measurement goals, but did not structure the goals according to different levels of the organization, and the strategies for the goals have not defined for this method.

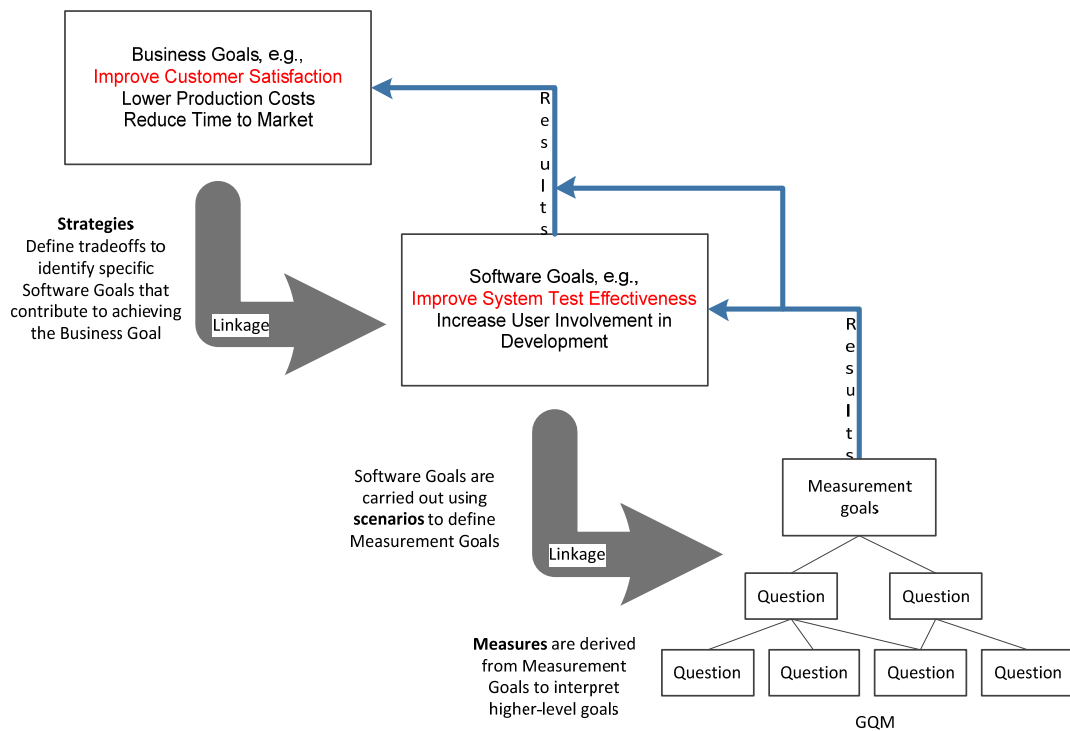


Figure 5- GQM+ Strategies Method

GQM+ Strategies basically developed for software-related organization, although it is assumed that the basic concept can be generalized to establish in an organization-wide measurement program. In this framework different goals and strategies are defined in different level (e.g. business level, software level, project level, etc.). Figure 8 represents the notation of this framework. It has eight conceptual components defined as below [5]:

- *Business goals:* are the goals that organization hopes to achieve during its time operation
- *Context factors:* Characteristics representing the organizational environment and can affect the model and data that is used.
- *Assumptions:* Unknowns, which can affect the interpretation of data (e.g. increase customer satisfaction can result in x% increase in annual revenue)
- *Strategies:* is an activity suggested to partially/totally achieve a respected goal
- *Level i goals:* strategy of each level is a goal of next level, for example strategy of business level is a goal for a software level or goal of i-1 level is part of strategy of level i goal.
- *Interpretation model:* It is a model to depict that whether the goals at all level have been achieved or not
- *GQM+ Strategies element:* Each goal with its respective strategies based on assumption and context factors.
- *GQM graph:* A basic GQM graph for specific GQM+ Strategies elements, which define question, related metrics, and interpretation models.

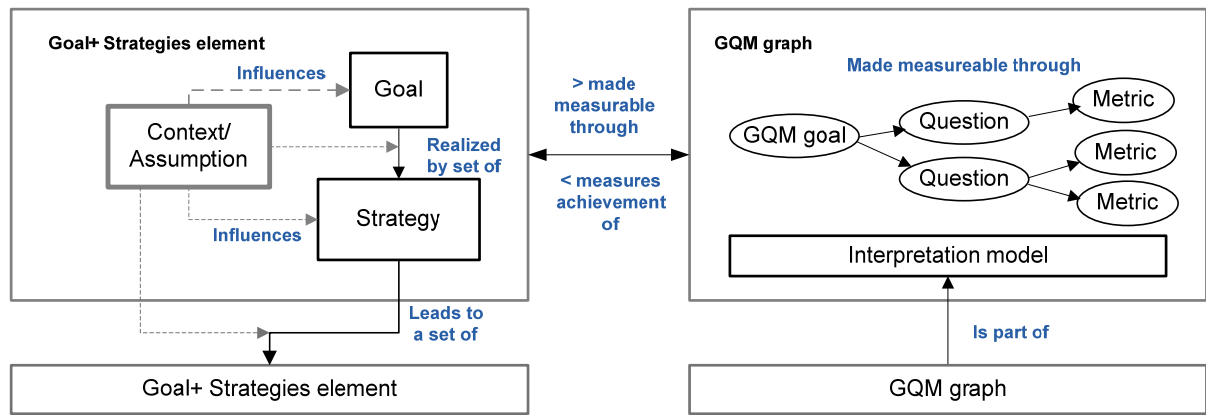


Figure 6- GQM+ Strategies component

The entire model provides an organization with a more structured measurement program by linkage among a different level of organization. And contribute to update the interpretation model based on a rolling up the resulting measurement data at each level.

After having explained GQM+Strategies as a concept, the next subsection presents a detailed literature review of papers on the concept.

3.3 Related Work on GQM+Strategies

After having introduced the well-known goal-based measurement program planning approaches, we thoroughly survey the literature on GQM+Strategies as for this approach we designed our solution. We identified the first paper that introduced GQM+Strategies and with Google scholar looked up all the papers that cited this paper (forward snowballing). This gave in total 27 papers in Google Scholar. From these 27 papers we identified six papers that had GQM+Strategies as their main focus. In addition to the forward snowballing we searched Google Scholar for “GQM+Strategies” in parentheses. This provided two more relevant papers.

Mandic and Basili [20][22] introduce new types of goals besides business goals, namely value goals and integrate them into GQM+Strategies. This should help to understand the value of an organization. Value goals relate to business goals, as they are to check which contribution the achievement of a business goal makes in achieving the value goal. Overall, the idea is based on return on investment, i.e. the return is the value gained, and the investment is the cost in creating the value in the organization. An illustrative example is presented to understand the approach, but no empirical evaluation has been done.

An evaluation of GQM+Strategies in the military training domain, outside of software engineering, is presented in [30]. He elaborates on assumptions; to what extent GQM+Strategies has to be modified in order to be applicable; and the ease of applying the approach (efficiency and its relation to effectiveness). GQM+Strategies defines levels three levels: business level, organizational level, and product level. These levels were translated into more generic ones to be applicable to other domains, namely strategic level, design level, and execution level. Then C applied the approach to military training to evaluate training expenses. The study showed that the approach could be transferred to different domains, but at the same time requires excellent understanding of the concepts to be understandable.

GQM+Strategies was applied in the Japan Aerospace Exploration Agency [18]. Applying the approach and presenting the outcome of the GQM+Strategies activity evaluated the application of GQM+Strategies. The outcome was in the main focus of the work, however, how the outcome was reached and which challenges/support the approach supported was not elaborated. Some lessons learned have been presented, namely: (1) the relationship between different organizational units and organizational levels is clarified; (2) it is important to involve the people of the organization; and (3) when having used the approach, components for goals and strategies should be recorded and be reused to reduce effort in running the measurement program. According to the authors GQM+Strategies also needs to be extended to incorporate priorities of goals.

Mandic et al. [23] evaluated the GQM+Strategies approach in a Finnish company. They used Bloom's taxonomy to evaluate the level of understanding that practitioners have in using the approach. This reaches from level 1 (remember the concepts) up to level 6 (being able to create the best possible GQM+Strategies grid/graph). The levels in between those are level 2 (understanding GQM+Strategies), level 3 (being able to apply GQM+Strategies), level 4 (being able to analyze GQM+Strategies) and level 5 (being able to evaluate GQM+Strategies). People were given training in the approach, which should allow them to achieve level 3. Higher levels require practical experience. In the evaluation of the approach 12 people from the company participated. Overall, the majority of the participants remembered the concepts. For higher levels (2-4) they indicated that they indicated that they feel comfortable in reaching those step working together with an expert supervisor. The practitioners with their current knowledge level considered level 5 and 6 as not being doable.

Mandic and Basili [20] extend the GQM+Strategies approach by contrasting cost and benefit. Therefore, they proposed to associate a budget with each business goal in GQM+Strategies. Overall, the GQM+Strategies graph contains information about goals and their realizations, assumptions to be made in order to achieve a goal, the estimated budget to achieve them, and the actual costs and benefits achieved. Contrasting the actual costs and benefits represents the earned value. No evaluation of the approach has been provided, and no practical experience is reported so far on this approach.

The authors in [32] provide a case study of applying GQM+Strategies in the case of the Japanese Information-Technology Promotion Agency (IPA) with the goal of identifying the degree with which projects are contributing to business goals, and with this information to align the projects. The work was motivated by the observation that alignment of activities with business goals is seldom accomplished. A project alignment matrix is then used to show which strategic goals exist, and which project and its related goals match those strategic goals. In the case study five projects and strategy goals have been identified. The researchers conclude that the approach helped to achieve traceability from strategies to project goals, and hence support to update project goals accordingly.

Mandic and Markku [21] propose a tool that allows capturing context information for GQM+Strategies as well as goals and strategies. The tool is build to support people eliciting information for GQM+Strategies to create the tree/grid for the measurement program. However, the paper does not provide information of how to elicit the information for stakeholders in order to capture them in the tool. In a small survey with 12 participants the researchers found that the participants agreed on the usefulness of the tool in structuring the GQM+Strategies elicitation.

4 ACTION RESEARCH ON GSE-APPROACH

4.1 Action Research Design

The purpose of this study is to develop an approach for eliciting goals and strategies in a complex software organization to plan a measurement program with GQM+Strategies in a complex organization. Complex refers to organizations that have many stakeholders, processes, strategies, rules and basic departments. In order to evaluate the developed approach we conducted action research by using our solution in industry.

4.1.1 Research Context

The case in this study is Ericsson AB, which is one of the major telecommunication companies in the world offering telecommunication services and products to telecommunication operators. Ericsson is certified by ISO 9001:2000. It is market-driven company and a frequently changing market influences its product development approach, which is incremental with the use of agile practices. Ericsson uses balanced scorecard as a performance tool for managing and measuring the software process, product, and recourses. Table 11 represents more detail regarding the context of the study.

Table 2-Context Elementns

Context element	Description
Domain	Telecommunication, data communication, and multimedia solution
Market	Highly dynamic and customized with a market share of 35%
Certification	ISO 9001:2000
Measurement	Balanced Scorecard
Process	Incremental process on the principal level, and agile practices on development level

4.1.2 Research Process

Figure 9 represent the action research we performed in collaboration with Ericsson. The study is designed in an iterative manner. To find a starting point an initial meeting was held with the company to decide on an initial measurement program as a first unit of analysis. In order to understand the context (which is very important as the measurement responsible has to know terminology, structure of organization, products, etc. to understand the information elicited) archival data and documents were studied. Of course, especially with the huge amount of abbreviations and company specific terminology, the researcher had informal meetings with a variety of people whenever there was a need to clarify information from the documentation. After having a good enough understanding of the initial situation the GSE approach was evaluated by applying it in the company. The approach was in the form of an interview, which is further described later. While doing the interview we made observations by taking notes and also collecting feedback from the person interviewed in the end. This feedback was used to update the elicitation approach and then apply it in the next iteration, and so forth.

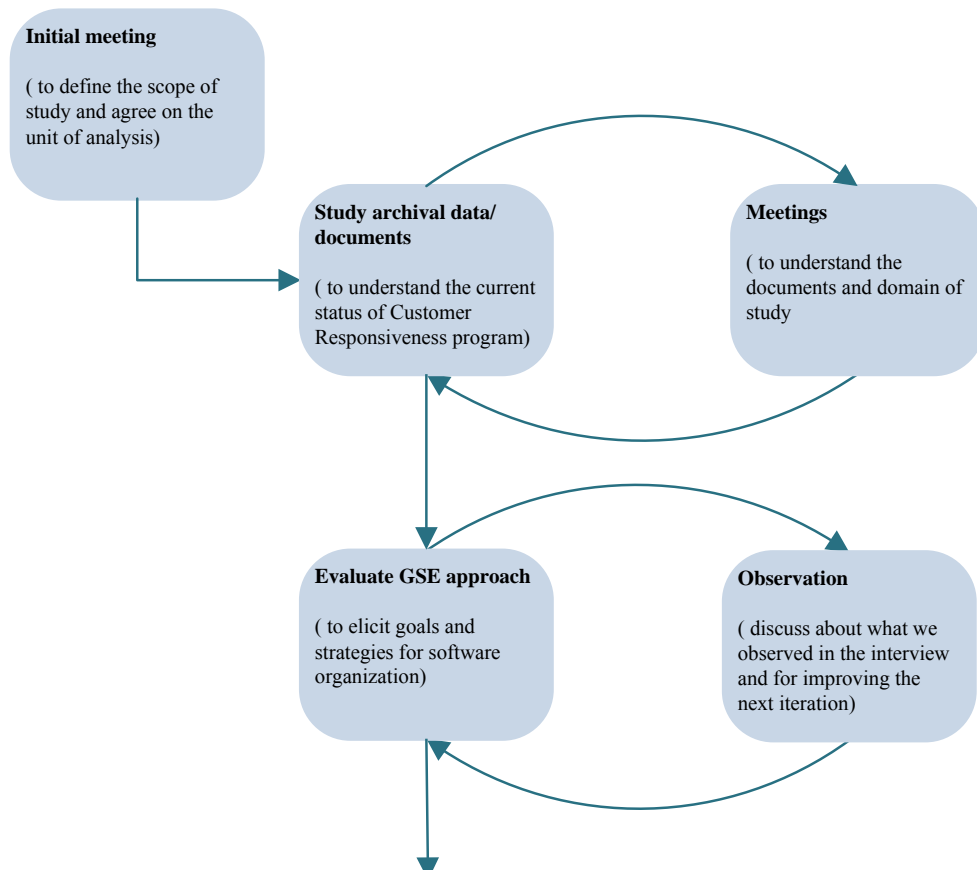


Figure 7- Research Design

We use different sources to collect the required data by following the triangulation approach. The reason we used this approach is because our study relies on qualitative data, which is broader and richer, but less precise than quantitative data. In other words, it is based on interpretation and prone to bias. Therefore, to increase the precision of the study we take different angles of studied objects into consideration. Two different type of triangulation are applied in this study, namely data (source) triangulation and observer triangulation. We have done data source triangulation by study literatures, archival data/documents and interview transcript, and observer triangulation by conducting the interview with three interviewers, the interview leader and designer being the author of the thesis, and the observers being the research supervisors. In Table 12 you can see the definition of different introduced type.

Table 3 Triangulation

Triangulation types	Concept
Data (source) triangulation	When different source of data is used for collecting the same data at different occasions.
Observer triangulation	When more than one person observe the studied object
Methodological triangulation	When a combination of data collection methods used to conduct the study. e.g. qualitative and quantitative method
Theory triangulation	When alternative theories and viewpoints take into consideration

First we start with reviewing literatures regarding goal-based measurement framework for planning the measurement program. There are many templates defined for eliciting required information for measurement program. Basically, these templates and structures defined what to collect rather than how to collect the information. To best of our knowledge, there is a lack of a well-defined approach for collecting and eliciting the right information that helps in avoiding irrelevant data or overhead in data collection (see trial search presented in the introduction). Therefore, we developed the Goal Strategy Elicitation (GSE) approach based on the GQM⁺ Strategies idea and conduct the action research to evaluate it. This approach evolved during the study in an iterative way. This approach would help us to measure the effectiveness of GQM⁺ Strategies idea in planning the measurement program.

Sources of data to achieve this goal is, studying archival data/document of Ericsson and interview stockholders involved in selected unit of analysis; Customer Responsiveness Program, Cost of Poor Quality Program, Performance measurement Program.

In the following we provide some more detail about the different steps shown in the process figure above.

Initial Meetings: To start our study, we conducted five meetings in which we discuss about the project and the main purpose of the project. Besides, we discussed the applicability of the action research study and how the organization can benefit from the result. Another purpose of these initial meetings was to define the scope of study clearly and select the unit of analysis.

Archival Data and Meetings to Understand the Data: There are two internal document repositories within Ericsson, which investigated in this study for finding relevant documents. They are used to extract the required data to understand the defined goals and strategies to create an initial goal tree. All the documents were saved in word or power point format (mostly power point) in the repository. Most of the documents have been written abstractly and were hard to understand; therefore ten meetings (online and face-to face) have been conducted in order to understand the content of documents. Online meeting were necessary in cases where the interviewee was sitting in another branch of the organization located in another city.

Elicitation Meetings and Observation: To elicit the goals and strategies, we conducted six meetings with the measurement stakeholders related to different measurement programs for which we interviewed them. The detailed interview structure is described in following sections. The main purpose of these interviews was to evaluate the GSE approach in an iterative way. During the interviews we used the GSE approach for eliciting goals and strategies and observe the applicability of the approach, then the required actions to improve the approach were taken and applied in coming iteration.

4.1.3 Design of Elicitation Meetings and Observations

In this phase, industrial practitioners are interviewed to evaluate the GSE approach to elicit goals and strategies for planning a goal-based measurement program. It is important to mention that interview is a part of GSE approach. Designing the interview in this guideline has five steps: (1) deciding the themes of the interview (which represents the GSE instrument, as well as the feedback collection approach); (2) designing the interview; (3) selecting the interviewees conducting the interview; (4) transcribing and analyzing the interview; and (5) report the results of the interview. We used the designed interview guidelines to avoid the probable pitfall and also to improve the quality of the collected data. The outcome of this phase answers following research question:

- *RQ2: How should software organizations elicit goals and strategies as input for GQM+ Strategies?* The answer to this question aims at finding a useful and practically applicable way of eliciting information for GQM+Strategies.

(1) **Deciding the themes of the interview:** What we intend to do is to determine the way in which we want to conduct the interview. Thus, we formalized the main purpose of investigation. The purpose of the interview mainly was finding the best way to elicit the goals and strategies of the organization for planning a measurement program. In the Initial meetings, in order to familiarize with the aim of the research, we provided managers of the measurement program with the purpose of the study, and expected outcome of the action research. The actual themes identified are presented in further detail in the following paragraphs.

(2) **Designing the interview:** The quality of data collected from interview depends on a way the interview conducted and the people who are interviewed. There are three alternatives for conducting the interview, which are semi-structured, fully structured, and unstructured interview. We conducted the qualitative semi-structured interviews. In contrary with fully structured interview, semi-structured interview gives us an opportunity to bring up a new question during the interview and follow up on interesting ideas and topics brought up by the interviewee. Unstructured interview do not suit our study because our interview should be conducted with multiple stakeholders, and later on (if e.g. several stakeholders are interviewed) their answers should be comparable. Unstructured interviews have the risk to collect different information in every new interview. The interview themes, as decided in (1), are described in the following paragraphs.

Theme I starts with the introduction regarding the goal of interview/study and how the interviewee can benefit from the outcome of the study. Furthermore, a reason about why the interviewee has been selected and the main purpose of the interview is conveyed. We continue with a short description of what will happen during the interview.

In Theme II, we request interviewee's consent to record the interviews to capture all their views accurately. We also made it clear that everything the interviewee says during the interview will be treated confidentially and will be deleted after transcription and interpretation. Afterwards, the interviewee is asked about their experience and background both in Ericsson and outside Ericsson (if any). Then, the pre-defined questions relevant to goals-strategies elicitation are asked. This part is done in an iterative way and evolved throughout the action research process.

In Theme III, a Likert's scale was used as a survey for evaluation and feedback, and as a starting point for discussion. This survey contained a set of questions regarding Understandability/Usability, Completeness, Correctness, Overhead of information, and Acceptance. Each questions followed by explanations, which were asked to clarify the reason of the rating. We specifically asked about limitations when we noticed that a low rating would be given.

In Theme IV, the interview is continued in absence of interviewee to discuss about what we observed during the interview, this session was also recorded and notes were taken. This helps us to draw a conclusion right after finishing interview and avoid missing or forgetting what we observed. This part was important to decide on the improvements to be done in the following iteration.

The first three parts takes one and half an hour and the last part, Part IV, takes approximately one to two hours. Table 13 shows the division of time assigned to each part.

Table 4 Interview Design

Part #	Time limit
Theme I	30min
Theme II	45min
Theme III	15min
Theme IV	60-120 min

According to the initial meetings, all the interviewees have been selected based on the relevancy and experience they have regarding the unit of analysis, as well as with regard to the interest of the company on which measurement programs to focus on. Furthermore, all the interviewees have an average of 15 years experience working in Ericsson, which indicates that they have a good knowledge of the domain and processes used in the company.

(3) **Selecting the Interviewees and Conducting the Interview:** The interview should start with a briefing of the interviewees about the purpose of interview. We got the support of two managers at Ericsson in deciding on the measurement programs and relevant stakeholders with respect to the interest of the organization. The manager then contacted the interviewees to get their consent. Thereafter, we sent an email to all interviewees to inform them about the purpose of the study. In total we conducted six interviews. The following list introduces the interviewees with respect to their role and experience.

- *Interviewee I:* She has been working in Ericsson for 19 years. He has been working as a Principal Architect for a year, and before that she was a biomedical engineer, building software for hemodialysis. Before that, she was a computer science tutor at London School of computing.
- *Interview II* has worked over a 30 years at Ericsson in different levels of the organization and in different positions, and is now a system manager. She has managed software development and system development and during the last 10 years has been involved mostly in system management. At the moment she is working in the management of research and development at the company, which follows two targets: one is to excel in reuse and the other is improve the system architecture. Literally, she is documenting their blueprint architecture, which represents process steps for facilitating reuse.
- *Interviewee III:* This interviewee is working as a solution Architect She has been working with Ericsson for 18 years in Sweden. She has worked in different positions, such as system integrator, system tester, support engineer, and project manager.
- *Interviewee IV* is a program manager responsible for requirements roadmap and allocation. She has nine years experience working in Ericsson and 20 years experience in other companies like; Anoto, Telia Mobilite, UAB, ELLEMTEL, and Logica svenska AB. She has worked as a developer, tester, product manager, and now she has been working as a program manager for two years.
- *Interview V:* This interviewee is involved in strategic management, and she has been working for Ericsson for 17 years. She has experienced in R&D organization within Ericsson with the focus on radio networks.
- *Interview VI:* She is a system manager. She has worked over twenty years at Ericsson in different levels of organization and in different positions. She has managed software development and system development and during last 10 years she has been involved more in system management. At the moment she is working in the R&D management.

We conducted six interviews, of which three were face-to-face and three of them were making use of videoconference. Further details on the interviews and the feedback collected are provided in the results.

(4) **Transcribing and Analyzing the Interview:** With transcribing the interview, we prepared the interview material for analysis. We used four methods/tools to capture the interview sessions, namely tape recording, taking notes, using Gliffy (online tool to create diagrams), and Mind-map (a tool used to organize information). Interview interpretation is different from person to person, therefore we first write down the interviews, and then rephrasing the text to become more understandable (e.g. removing “ahms”). Finally we had a discussion regarding our interpretation about each interview and get ready for the next one.

We transcribed all four part of interview right after each interview, and for interpretation we used all the transcriptions along with the note we took during the interview, including documentations made in the tools, and photographs taken from e.g. whiteboard drawings and notes.

For those parts in the transcribed text that we could not understand, we contacted the interviewee and asked for more information, either by e-mail or in person.

The interview has been analyzed by noting down observations in our own words where we observed challenges/what was not working so well in the interview. Much of this documentation was also done in the sessions we had after the interviewees. To increase the reliability of this analysis, we had multiple researchers present, reducing the bias in the perception of an individual researcher. This makes the suggestions of improvement for the next iteration more reliable. The transcribed interviews were therefore useful in seeing whether we missed any important problems in our post-meeting. Furthermore, we summarized the feedback collected from Theme III to see what improvements the practitioners see, and also took them into consideration when designing for the next iteration. In order to structure the analysis we assigned codes to the statements that are related to either what was working well in the interview, or what was not working well.

(5) **Reporting the Results:** Reporting the interview is the last part of a long process, which results in new knowledge, which contributes to the development of respective field. After analyzing all the collected data, findings of the study are divided in two groups; first we summarized the content of the interview to provide actual input about the measurement program, and secondly we summarized the experience made with respect to the GSE approach. This serves as input and guide for the people in the organization when they would like to use the approach.

4.2 Results

In the following we document the results of the steps that were presented in 4.1.2 that presented the research process, this includes initial meetings, information retrieved from archival data, and finally the results related to the use of the GSE instrument.

4.2.1 Initial meetings

In the Initial meetings, we agreed on one of the measurement programs, namely “Customer Responsiveness Program” as a starting point to evaluate the approach for eliciting goals, sub-goals and their respective strategies. The customer Responsiveness program has been initiated 2010 for reducing lead-time end-to-end. Lead-time is the period of time needed from receiving a customer request until the product is developed and delivered to the customer. The main rational to start this program was to provide to the customers what they want in a rapid manner. In the follow up meetings we added other measurement programs, namely Cost of Poor Quality and Performance Measurement, to evaluate the GSE approach.

The choice of measurement programs was driven by the interest of the organization, and the priority of planning/gathering information for specific programs.

4.2.2 Archival data

In this section, we studied archival data. Here we highlight some of the challenges when using archival data for the identification of information for measurement program planning. The result of this activity was an initial list of goals and strategies identified from the documents. Overall, the goal was to extract all the goals and strategies defined in documents, and to provide the linkage among them to create one goal/strategy tree as defined by GQM+Strategies. By just relying on documentation we faced the following challenges:

Variation in concept: There were many documents with several revisions, where the last revision has been considered for this study. It was common to encounter the confliction between different concepts like; target, goal, key performance Indicator (KPI), sub KPI, sub goals, strategy, and action. This variety in concept creates difficulty to map the required information to the model. For example if *Developing an Adaptable Product* set as a goal, further on it was mentioned as a strategy or sub KPI in some other documents, which creates confusion in traversing the GQM⁺ Strategies tree. It is good to define a distinctive definition and also tailor these concepts down as much as possible. This is what we took with us when planning the interviews, i.e. we provided definitions of these concepts (goals, strategies, and actions).

Ambiguity in goal/strategy definition: Another challenge, which is really important to discuss about, was ambiguity in defining goals and strategies. When any respective unit set a goal for themselves, by the time it is completely understandable for the people who were involved in the planning of the measurement program. However, later on when new people joined to the unit or when new sub-goals and strategies were added, if these goals were not well defined it created ambiguity in understanding the main purpose of the defined goal. This is not only true for people that are new, but also for people that have been working with the program for a longer time. We noticed this when trying to clarify the concepts in the documents with the practitioners.

Abstract Information: All the documents include many abbreviations and were written in an abstract level, which is likely to be understandable for the owner of the document together with the stakeholders who are involved writing the document. With external people (such as persons defining a measurement program for a development unit) this is a challenge. Although there was an Ericsson wiki, which allowed searching for abbreviations, most of the time there were more than two definitions and in some cases even ten definitions for the specific abbreviation. To facilitate and accelerate the process of understanding, we conducted several online/face-to-face meetings with the owner of the documents in order to clarify abbreviations and in the end actually being able to read the documents.

Missing linkage among documents (goals): As we mentioned earlier, all the information was documented mostly in power point and word format. These documents cannot provide tractability among goals and information needs at different levels of the organization. In other words, documents are information islands where in one document we found strategies, in another we find the goals, and we are not sure if and how these are related. Traceability links are missing. We approached this by making a list of all the goals and strategies, which we extracted from different documents. Then from that list, based on our interpretation, we created a GQM+Strategy goal-tree by trying to find the linkage among them.

Overall, the challenges related to relying on documentation lead to the conclusion that this is not a reliable source for gathering information about measurement programs, given the

above problems. Furthermore, we are not able to capture the interest and perspective of particular key stakeholders. Thus, there is a need for talking to the stakeholders.

4.2.3 Action Research Iterations

The Table 14 is an overview of the iterations needed to achieve a working result and Figure 11 also represents the GSE approach road map together with its description. The table shows that each interview, and that there were three major versions of the GSE approach, while GSE 3.0 stabilized in its design during the final three interviews. The table also shows that three measurement programs have been studied, one regarding customer responsiveness (related to being responsive and able to quickly respond to customer needs), cost of poor quality (being able to reduce cost of poor quality and relate it to the investment that is needed to achieve quality), and performance measurement (assessing the overall performance of the R&D organization).

Table 5-GSE Status in each iterations

Approach	Technique/tool	Status	Measurement Programs
GSE 1.0			
Interview 1	GSE Notation + Semi-Structured Questionnaire	Face-to-face	Customer Responsiveness
Interview 2	GSE Notation + Semi-Structured Questionnaire	Video-conference	Customer Responsiveness
GSE 2.0			
Interview 3	Semi-structured Questionnaire + whiteboard	Face-to-face	Customer Responsiveness
GSE 3.0			
Interview 4	Semi-structured Questionnaire + Mind-Map tool	Video-conference	Customer Responsiveness
Interview 5	Semi-structured Questionnaire + Mind-Map tool	Face-to-face and Video-conference	Cost of Poor Quality
Interview 6	Semi-structured Questionnaire + Mind-Map tool	Face to face and Video-conference	Performance measurement

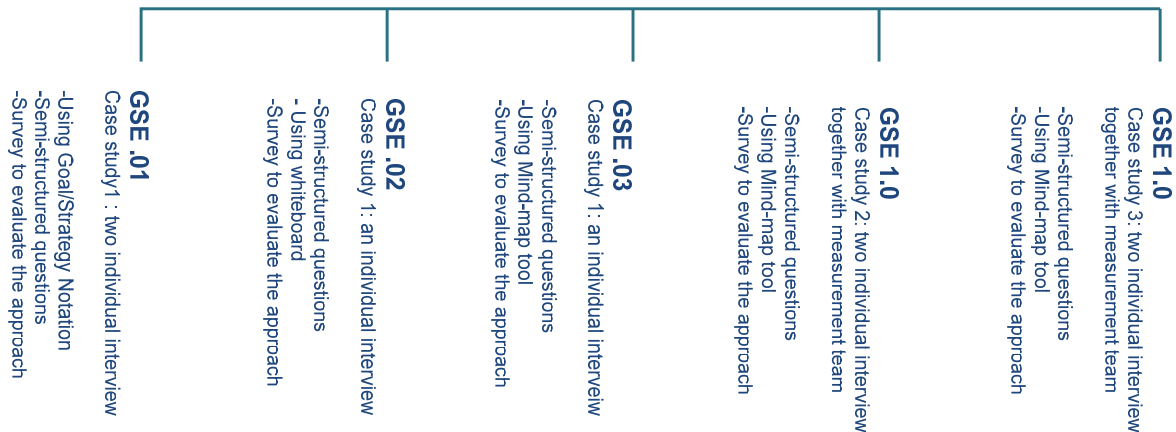


Figure 8 - GSE Iterations

4.2.4 GSE 1.0 – GSE Notation Based Interviews

First, we present the intervention, followed by the results of applying it in Interview 1 and Interview 2.

4.2.4.1 Intervention/Action – Introducing the Notation used for Elicitation

For evaluating the GSE approach we developed a notation depicted in Figure 12. This notation inspired of variability models, in particular the orthogonal variability model [6]. Triangles represent goals, and rectangles present strategies. Goals can be connected to other goals, or to strategies. Strategies can also be connected with each other. Connecting a goal to a strategy shows that this strategy is used to fulfill the goal. There are mandatory strategies (this strategy has to be in place) and optional strategy (this strategy could be used, but does not have to be), which is important to represent alternative strategies. The line connecting goal and strategy has a statement associated with it documenting the rational why the strategy fulfills the goal. Furthermore, links between strategies can indicate that one strategy requires another strategy to be implemented (e.g. introducing a component based architecture might imply organizational changes to be done). We represent these strategies by annotating the links with include and exclude statements.

A cutout of the customer responsiveness program diagram is shown in Figure 13. It shows the overall goal (customer responsiveness) and then shows the different strategies (reduce lead-time, develop an adaptable product, and architecture lift. These are the strategies on business goal level. These then become goals on the more operational level, and then have strategies themselves again. As mentioned before, a goal on one level becomes a strategy on the next level.

During the interviews we brought up an initial tree on a screen and showed it to the interviewee (based on the information we collected from the archival data). Then during the interview we asked the interviewee to add, change, or update the tree by adding or removing goals or strategies, or updating the links between them. Before doing that, we thoroughly explained the terminologies as defined by GQM+Strategies. This very much corresponds to a template/notation driven elicitation approach.



Figure 9 - GSE Notation

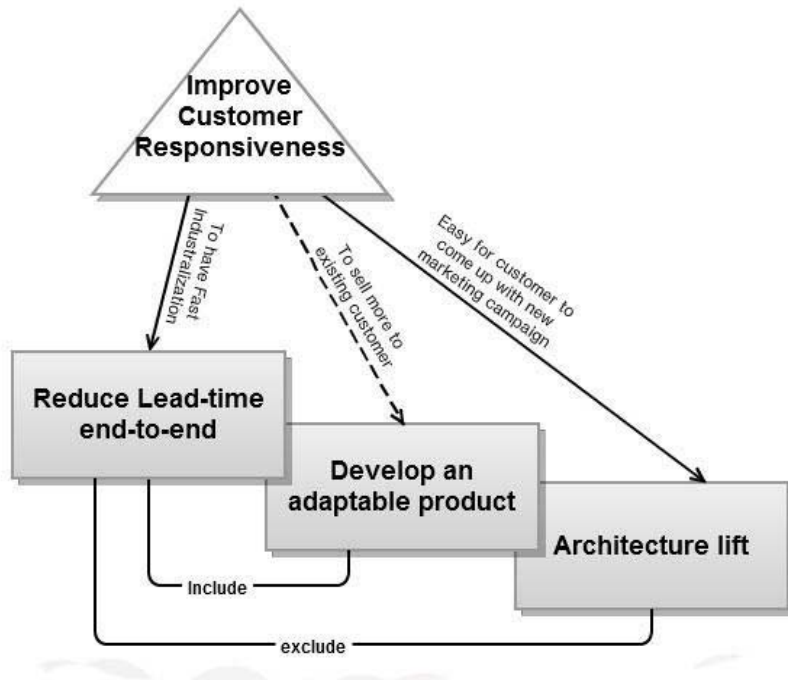


Figure 10-GSE Example

4.2.4.2 Evaluate of GSE Iteration One

We conducted two interviews. After analyzing the interviews similar observations were made for both of them. We found that (1) the Goal and Strategy Concept create confusion; (2) using a notation restricts people expressing themselves and their ideas, and (3) Goal and Strategy become even more vague concepts as we go further down on the tree.

Goal and Strategy concept create confusion: During the first two interviews we found that the interviewees have a problem with the concept of strategy, especially as we break the goal down to sub-goals and go further down the tree of goals and strategies. However, in theory,

goal and strategy have a really clear concept as defined by GQM+Strategies, but when it comes to practice we observed something different. A goal is something that any person, system, organization plans and commits to achieve, and strategy is a plan of action to achieve that goal. One of the problems we faced during the first interview is the interviewee confused the meaning of the goal and strategy and how strategy at each upper level becomes a goal at lower levels. The following phrase are examples of strategy definition that the interviewee defined for himself to understand the situation better:

- Strategy is an action
- Strategy should be implemented
- Strategy is a plan to achieve a goal
- Strategy is a use of specific tools
- For me, strategy cannot have a verb, I'm just thinking though
- What is different between the goal and strategy, we have to check that because they almost the same
- I got lost, it is kind of like an inception, goal and goal and goal, dream and dream and dream and then you get lost

During the one and half hour, the interviewee still tried to understand the goal and strategy concept rather than defining them and interrelate each strategies to related goals. The concept of strategy made interviewee stop from talking, which is contradictory to what we assumed, namely that the visualization of the goal and strategy tree will help the interviewees in updating and adding their information. We found that different people perceived the concept differently depending on their own definitions they had in mind (e.g. from their organizational context). That is, they could not just forget those. We also observed during the interview that they checked the definitions several times to remind themselves. As a consequence, communication took a long time, as the use of the notation was not intuitive.

Notation Restricted people expressing themselves: We observed that the definition of a notation with the concepts goal and strategy hindered interviewees in expressing themselves, telling us what they wanted to say. The main reason was that it took time for them to map their thoughts within the notation structure, and might even miss expressing some important information due to that. Overall, we can conclude that interviewees cannot talk and express their own idea and struggle to adapt to the notation all the way down to define the goals and strategies with these restrictions.

Goal and strategy become vague in a lower level: During the first two interviews, we used the GSE notation to keep track of goal, strategy, and rationale behind the relation of each goal and strategy. On the highest level this might still be intuitive, but when there are too many translations of goals into strategies, the confusion increases as the tree increases. Figure 14 represent a cutout of GSE tree that is the result of first two interviews. This figure shows that, as it goes to the lower level of the tree, the goals and strategies faded and are not clear anymore. However, this notation was developed to increase the level of transparency, but did not succeed in doing so. The interviewees themselves in the lower level (e.g. level 3 and level 4) were not sure that, they are talking about goals or strategy. The last level in figure shows that goal “Improve the quality of API¹ and SDK² and three other strategies located at the same level. The reason might be that the different branch of tree do not grow simultaneously and give a balanced three. Therefore, the goal on one branch can be a strategy for another branch, which became counterintuitive to the interviewees.

¹ Application Programming Interface

² Software Development Kit

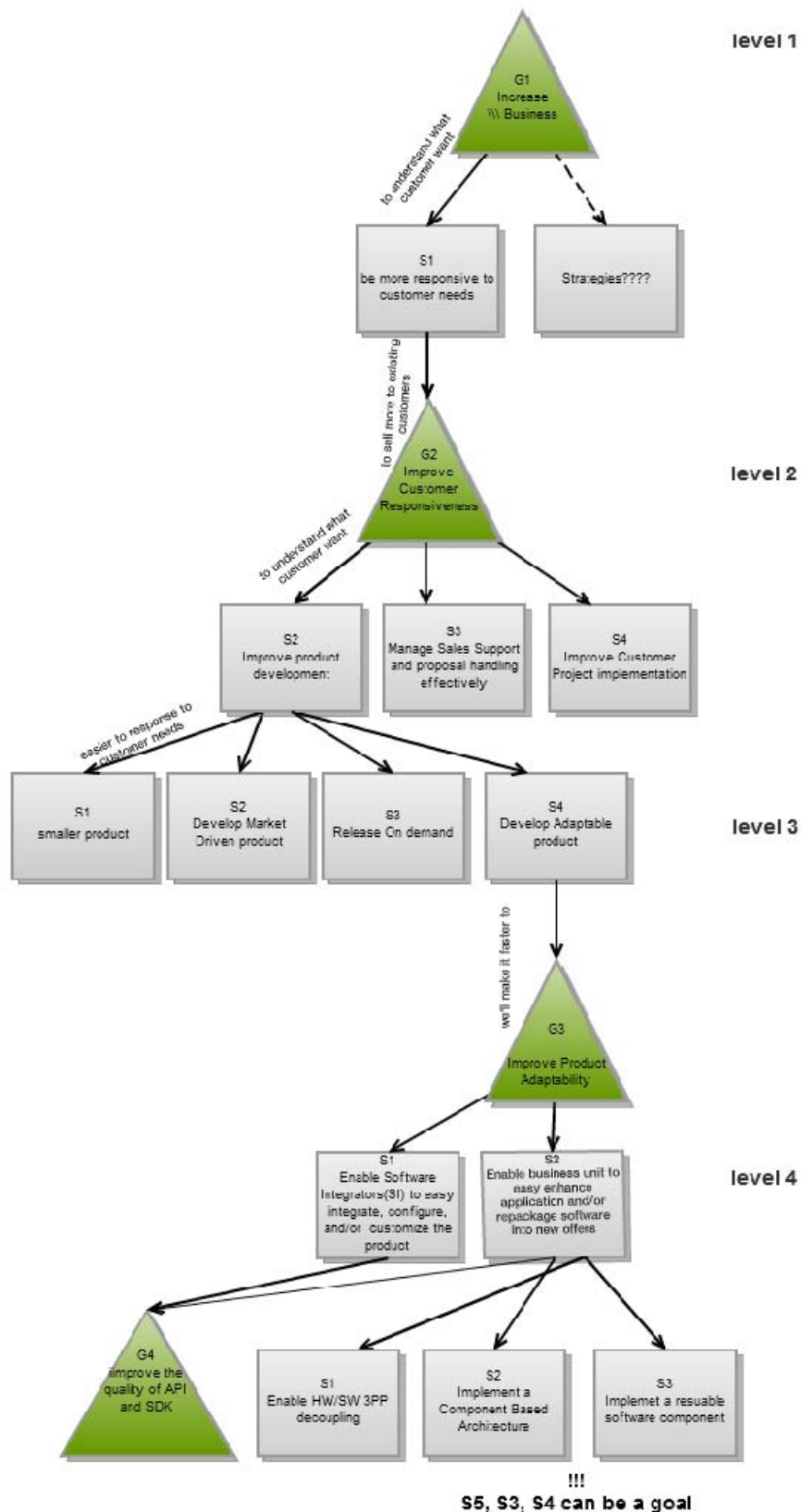


Figure 11 - GSE Tree for Customer Responsiveness program

Communication took long time as the use of notation was not intuitive: A consequence of the approach not being counterintuitive is that it takes very long time to elicit information and to modify/update the goal-strategy tree. Even though it has obvious advantages as having traceability in place right away, and we being able to store it or put it into a tool, it leads to frustration on the side of the interviewee, and hampers the motivation of the interviewee to take part in the process of measurement definition.

As a consequence of this first attempt we redesigned our approach to elicit the goals and strategies.

4.2.5 GSE 2.0 Approach

4.2.5.1 Intervention/Action – Semi-Structured Interview Guide using White-Board

The failure of the first version of the GSE approach led to the next version, which was designed with the intention to allow the practitioners to express their ideas more freely. At the same time we linked the questions we ask to rationales to make sure that we capture the goals and strategies. The questions in Table 15 are the ones used in interview three and four.

Table 6- GSE Questionnaire

#	Questions for goal/strategy elicitation
1	What is Customer Responsiveness Program?
2	Why was it initiated?
3	The Customer Responsiveness program is successful when: a) b) c)
4	Who can influence the success of a, b, c, d, e, f...
5	What do you do to achieve a, b, c?
6	In these activities that you just mentioned, which ones are working out good today?
7	Which ones have the highest improvement potential?
8	Do you have a plan to improve them? How?

Table 16 explains the rationales behind each question. These questions were designed to capture the goals and strategies, and also the relevant stakeholders involved. We designed the questions so that practitioners talk freely and naturally without having to think about the concepts of strategies and goals. For example, instead of asking: “what goals do you want to achieve” we asked: “The Customer Responsiveness program is successful when:”. The interviewee then has to complete the sentence.

Table 7- Rational of Questions in Questionnaire

#	Rationale of Questions
1	It gets the person talking freely. It is also a good start to understand the program and context.
2	It provides the rationale for the existence of the program
3	It gives hints for information needs and even measurement needed
4	It identifies other measurement stakeholders who involved
5	It identifies key actions to be evaluated by the measurements to understand, control and improve
6	It obtains information about current state/baseline
7	It obtains information about desired improvement areas
8	It identifies improvement actions, and also give information about the effect of these actions through measurement

Besides having the interview guide we also defined an interview process. The interviews required three different roles, namely moderator, observer, and scribe.

- *Moderation:* The moderator is the person who is leading the interview, i.e. giving the introduction, and asking the questions from the interview guide.
- *Support:* One person is responsible for listening and keeping track of follow-up questions needed. That person should be given the opportunity to ask these questions every now and then (moderator should be aware and make sure the support person gets that time).
- *Documentation:* One person captures the information provided by the interviewee. In this version of the interview we used a white-board to capture the information.

4.2.5.2 Evaluation of GSE Iteration Two

The interviewee felt more comfortable: We used semi-structured interview instead of the GSE notation to elicit the goals and strategies. We observed that the interviewees talk freely and fluently, and also did not have to ask for clarification of the questions, the answers indicate that the questions were easily understandable. We can conclude that the interviewees talk more freely and answer the questions without being constrained by concepts and/or notation. The content of the transcription showed that the interviewees are much more focused on the content of the actual measurement program.

More information captured faster: Interviewees have more freedom to talk, which helped us to capture the flow of the thoughts and the connection between thoughts in a very short time.

Whiteboard enabled a good communication: In the second approach, we used the whiteboard to map the interviewee’s thoughts on the board and illustrate what we understood from his/her answers. Capturing the information on a white-board allowed us to summarize how we understood the interviewee, as he/she would see all the information captured during the interviewee. This process was iterative and the interviewee was also correcting the things we documented by pointing out misunderstandings. The drawback of using the white-board was that we had to capture the information very fast, leading to poor handwriting, and at the same time the information was not so well structured.

This led to the next version of the interview with two main changes. First, the person responsible for documentation would now document everything in a mind-map tool that has the structure of the interview as a starting point. Second, the person documenting would pace the interview and ask people to stop in order to summarize.

4.2.6 GSE 3.0 Approach

4.2.6.1 Intervention/Action – Semi-Structured Interview Guide Using Mind-Map

MindMap is a graphical method/tool for taking notes. It is a diagram used to represent words, task, activities, or other items, which are centralized around a key word or area. Figure 15 depicts a skeleton of MindMap used for our measurement program.

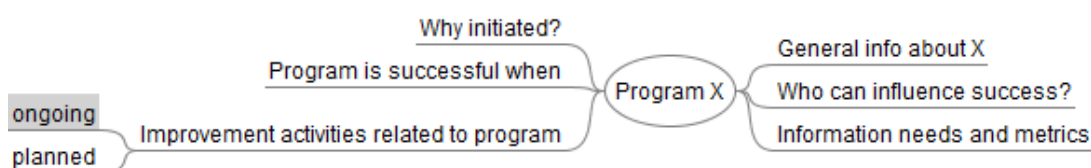


Figure 12 -MindMapSkeleton

The word in the center (Program X) can be exchanged for any measurement program ongoing, or large decision to be made. As the mind-map can be easily represented on the computer screen, it is particularly useful as it facilitates distributed elicitation when measurement stakeholders are distributed.

4.2.6.2 Evaluate GSE Iteration Three

In this version, we used the mind-map tool instead of whiteboard to facilitate the documentation and recording of information. As mentioned earlier, in contrast to using whiteboards, there was no worry about losing or missing information. To avoid any bias and confirm the applicability of the last version of GSE approach we applied GSE version 3.0 to two other different programs: “Cost of Poor Quality” and “Performance Measurement”. Figure 16 represent the outcome of GSE approach from Cost of Poor Quality program as an example. After the interview we showed the results to the interviewee to confirm the captured information. They confirmed the accuracy of the content and found this approach useful and easy to use.

As we said earlier, this approach has been evolved during the study and still needs to evolve further. Therefore, the measurement stakeholders who want to use this approach can modify the approach and release a new version to fulfill their requirements. In this study, GSE version 3.0 was the last release of this approach. However, we noticed some additional extension:

- We noticed that practitioners often mentioned information about measurements; hence we added a question regarding measurements collected today.
- In addition, measurements are already elicited by asking the interviewees what information they need to be successful with respect to the successes mentioned in the interview. Thereby, in discussion with the practitioners we elaborated on the issue that measurements might drive wrong behavior. For example, when having a measure on lead-time, quality might be compromised to perform better on the lead-time measure. This hence leads to sub-optimization.

Overall, this led to the released version of the approach that his now used at the company. The released version as described in the following section is the same as the manual for the practitioners in the company who use the approach.

4.2.7 GSE 4.0 Approach: Released Version

Part I: Introduction

1. Some general preparations and tips for the interview: Install MindMap tool (e.g. FreeMind) to your computer(s) for capturing what the interviewee is saying. Important here: The interviewee should be able to see the MindMap and follow what the interviewers understand from his/her comments.

It is important that the person taking the notes is given time to capture the information, he/she determines the pace of the interview. As a start, the person taking the notes has a skeleton of the mind-map with branches indicating the main information to be captured. Here, it is important that the person taking the notes is free to extend the mind-map according to his/her interpretation.

Different people might have different structures, but similar information should be found in the MindMap. The requirement on the MindMap is that other persons should be able to understand it, that means we do not just identify keywords, but the relationship between information should be captured. See, for example, the MindMap example provided in this

document: See e.g. branch “*What do you do to achieve different success measures*” followed by a short description of what is done. Here, it would not be good to just say “*Actions*” instead, in short: use short phrases when possible, not single keywords.

Give people time and let them talk freely and follow their thoughts/connection between thoughts. Make sure that you progress with the interview questions, but do not “correct” the opinions of the interviewee, as you are after their opinion. However, you are free to ask follow-up questions if you need clarification.

People prefer to be interviewed individually, when possible conduct the interviews with one person at a time.

2. Roles in the interviews: Here we describe three different roles that the measurement analysts supposed to take during the interview:

Moderation: The moderator is the person who is leading the interview, i.e. giving the introduction, asking the questions from the questionnaire.

Support: One person is responsible for listening and keeping track of follow-up questions needed. That person should be given the opportunity to ask these questions every now and then (moderator should be aware and make sure the support person gets that time).

Documentation: One person captures the information provided by the interviewee using the Mind-Map tool. For that the person has an empty skeleton of the MindMap with different branches for the questions, so that these can be directly filled in. The person doing the documentation should be given time to summarize the answers to the questions. Below are two examples, one Mind-Map skeleton done before the interview, and one completed Mind-Map output during an interview with a stakeholder.

3. Context: Provide the following information to the interviewee: Purpose of the interview and the process of the interview as described below.

3.1 Purpose of the Interview: Describe the goal of the interview (e.g. to get a better understanding of perspective X), where X is, for example, cost of poor quality program, customer responsiveness program, or any other perspective.

Reason for selecting the interviewee (e.g. as interviewee is working with X program he/she is considered an important measurement stakeholder).

3.2 What will happen during the Interview? Tell interviewee that you provide a number of open ended questions that you will ask to be able to cover the information needs of stakeholders in order to identify useful measures to aid in achieving good quality measurement program in Ericsson (e.g., measuring only what is needed and nothing else, aid in managing projects in a better way, driving the right behavior and so forth).

Tell interviewee that there are no right/wrong answers and that you are looking for his/her opinions, and that the interviewee is always free to ask questions for clarification, or not to answer a specific question.

Part 2: Questionnaire

The Questionnaire has 2 part: a) Questions related to Interviewee background and b) Questions for eliciting goals and information needs.

Questions to collect Interviewee Information

Name of the interviewee:

What is your current role in the organization? (*need to know at which level of the organization you are working, for which group of persons you are a representative*)

How long have you been working at Ericsson? (*gives some hints about domain knowledge, process knowledge, etc.*)

What areas have you been in before and how long have you been working in? (*gives some hints about domain knowledge, process knowledge, and that interviewee might be influenced by several perspectives, not just the current one*)

Question to be used during Goals/Strategies and Information Needs Elicitation Phase

Q1: What is X Program? (*gets the person talking freely, good start to understand the program and the context*)

Q2: Why was it initiated? (*provides the rationale for the existence of the program*)

Q3: The Program X is successful when:

- a)
- b)
- c)
- d)

(*gives hints for information needs and even measurements needed*):

Q4: What information do you need to follow up the success of the program?
(*makes information needs and success indicators explicit for the whole (program)*)

Q5: Who can influence the success of a, b, c, d, ... (*identifies other measurement stakeholders*)

Q6: What do you do to achieve a, b, c, d,... ? (*identifies key actions to be evaluated by the measurements to understand, control, and improve*)

Q7: In these activities that you just mentioned, which ones are working out good today?
(*obtains information about current state/baseline*)

Q8: Which ones have the highest improvement potential? (*Obtain info about desired improvement areas*)

Q9: Do you have a plan to improve them? (*identifies improvement actions, would like to know the effect of these actions through measurements*)

Q10: What information do you need for achieving a, b, c? (*identifies information needs and success indicators for concrete improvement ideas/actions*)

Q10.1 – What behavior will the measurements drive? (*figures out whether the measurements will really drive the desired improvement actions, or any other kind of behavior that is not leading to any benefit for Ericsson*)

Q10.2 – How do you think the measurement could be tricked? (*achieves robust and accepted measurements that show the “reality”*)

Q11: What are you collecting today (measurements)? (*identifies current state of measurement – in which phase are you understanding, controlling, improving*)

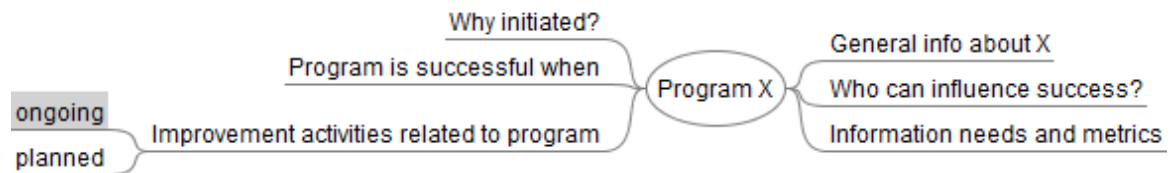


Figure1 -The Skeleton for Goals and Strategies Elicitation



Figure 16. An Example MindMap output for Cost of Poor Quality

5 DISCUSSION

There are some goal-based measurement models introduced in literature. The most well known ones of these measurement models are: Goal Question Metric, Goal Question Indicator Metric, Model Measure Manage, Measurement Information Model, and GQM+ Strategies. All these measurement models aim to introduce a model for planning measurement based on the goal of organization. These measurement models have some benefits and challenges mentioned before. It is important for large organizations to focus their measurements based on their more long-term/strategic business goals, and to relate them to strategies. In particular, the studied organization had several measurement programs running while the thesis was done. All of these programs were complex and there was a requirement by the organization to relate them to strategic goals. Hence, GQM+Strategies was chosen as a baseline approach for which the elicitation instrument was developed.

5.1.1 Main Lessons Learned

We conducted several iterations of developing the GSE approach for eliciting information to plan GQM+Strategies based measurement programs. Two main lessons have been learned:

Lesson 1: Do not restrict the interviewee through notations and templates. This lesson traces back to evidence we found during the action research. In first iteration of GSE, interviewees could not transfer their information and map it to the GSE notation, and we observed that, during the interview they tried more to understand the concepts and notation rather than provide required information for eliciting goals and strategies. For example, the interviewee from the beginning to the end of interview tried to rephrase the concept of strategy and map his knowledge to the GSE notation. We experienced that it was working much better to let the interviewees express their ideas more freely while having an interview guide that makes sure that the relevant information is captured.

Lesson2: With respect to timing it is important to use concepts and terminology that is familiar to the stakeholders, and get them to answer in natural terms. For example, instead of asking ‘what are the strategies/actions for this specific goal? ‘We asked what you do to achieve this goal?’. Another example was to ask the interviewee to complete sentences, e.g. “X is successful when...” to find goals and also measures.

Overall, this research indicates to recommend to practitioners to use semi-structured interviews without using the terms goals and strategies, but find more natural ways of asking the questions, as presented in our solution. Templates and notations might be confusing.

This research might have implications beyond the context of measurement programs. For example, in goal-based requirements engineering similar questions could be answered to learn what the goals of using an application are, what the strategies of deployment and use are, and how the success of the deployment of the application should be determined (quality measures). In particular, existing work on requirements elicitation focused on interviews could benefit.

5.1.2 Comparison with Related Work

In the related work on GQM+Strategies the issue of understanding GQM+Strategies [30] and the need for support when using the approach was raised [23]. Every participant had to be trained (also the stakeholders that are to provide the information for the content of the

measurement program who are not experts, e.g. normal developers) [23]. In large organizations one might have to talk to a lot of stakeholders and would require organization-wide training. Template-based and notation-based elicitation approaches would require excellent knowledge of the concepts, and even with the understanding of goals and strategies there might be problems to map the real-world situations to them correctly as indicated by this research. However, with an interview approach that elicits the information, and at the same time asks the questions in a way that the interviewee does not have to map to goals/strategies/context and so forth, much training effort could be potentially saved. Then only the measurement responsible has to have the competence to map the interview results to the GQM+Strategies concepts and represent them as a graph. The Mind-Map is a good complement, as it comprises detailed statements of the interviewees, as overall the template information seems to be abstract otherwise.

5.1.3 Validity Threats

Threats to validity might affect the reliability of the outcome of the study [33]. In the following we discuss validity threats of relevance for our investigation.

These threats can affect the outcome of the study with respect to cause/effect without the knowledge of the researcher, and is referred to as internal validity threats [33].

Bias in Understanding of Archival data: We used archival data to draw a sample life-cycle tree for one goal as an input to evaluate our GSE approach for eliciting goals and strategies. We could bias the result while interpret the documents. Documents contain abbreviation and terminology in the company might mean different things and deviate from the understanding of the researcher. Furthermore, the level of abstraction of the documents (if it is high) can leave room for interpretation. To avoid any bias from the researcher's side we studied the documentation individually. After that we shared our understanding with the owners of the documents to clarify if our understanding was correct.

Bias in analyzing qualitative data during interview: Interview was part of our GSE approach to elicit the goals and strategies. To evaluate the developed approach, we analyzed the transcribed interviews and the notes capturing our observations. There was a possibility to forget what we observed during the interview; although we took notes there is still a chance of missing data. To avoid this threat, we had one to two hour(s) discussions about what we observed right after finishing the interview. The discussions were recorded and transcribed as part of the interview.

Misunderstanding the extracted data from interview: During the interview there was risk of the researchers misunderstanding the interviewee. However the mind-map/whiteboard/GSE notation used as a tool to represent the interviewee's thoughts to them for transparency and consistency of captured information. But still there is a risk of misunderstanding information. To overcome this threat conducted member checking of the results with each of interviewee to verify our interpretation. Furthermore, after we transcribed the whole interview, in a case of ambiguity in any part of the transcription, we return back to the interviewee to clarify it and make it as precise as possible.

Trust: Interviewees might be biased or hold back information due to a lack of trust towards the researchers. One of the supervisors was present during the interviews and is employed at the case company, which potentially reduces this threat, but might not mitigate it. We further tried to reduce this threat by explaining the purpose of the interview, and guaranteeing that any recorded material will be deleted. Another reason for the reduction of this threat is that company representatives helped in contacting the interviews, and that it was seen as an initiative in relation to the ongoing measurement programs.

Limitation in Generalizability due to the Study of one Company: We evaluated the GSE approach by conducting action research study in a single organization, which can cause a diversity threat. To reduce this threat, we used the GSE approach on three different programs; Customer Responsiveness Program, Cost of Poor Quality Program, and Performance measurement Program. These programs have been running on different department, and on different levels of the organization. Besides, we overall conducted six interviews to evaluate the GSE approach due to the time limitation from both sides. This amount of interview might not be enough to achieve generalizability. However, we found that the interviews using the approach without notation was working very well in Interview 3, 4, 5, and 6. Hence, with regard to the overall research aim of evaluating the approach no further interviews were required. Further interviews might now lead to minor variations only, such as adding another question, or making a minor update in the phrasing of a question.

Roles and responsibility of interviewee: In our study, we mostly interviewed top level managers at each program, which can be a potential threat to the result. For example, top manager mostly focus on business level and at the least state software development/operational level. They are rarely involved at project level. Thus, the results of the GQM+Strategies activity might be biased towards more business oriented goals rather than project goals. However, we believe that the concepts/notation would be equally confusing to those stakeholders, and that our proposition is that the role might not have changed the general structure of the interview that was released to the company.

Applicability of the approach by internal measurement stakeholder: The GSE approach needs three persons for conducting the interview; moderator, supporter, and documenter. There was a threat that method and technique we used for the GSE approach mapped to our specified ability. To assure that internal measurement stakeholders also are able to use the approach, we conducted an interview in attendance of five other measurement stakeholders. They participated in the interview as listeners to learn how to use the developed approach. After that we conducted another interview in which three of stakeholders who were listener in the previous interview took the roles of us and use the approach themselves. They found the approach easy to use, and hence this indicates that the approach is applicable for usage by practitioners, and not just the persons who developed the approach.

6 CONCLUSION

In this study we applied GQM+Strategies and used a novel elicitation approach in order to capture information needed to plan a measurement program using GQM+Strategies. For that purpose, we used action research. We evaluated our GSE approach in three iterations by interviewing relevant measurement stakeholders. We applied our approach to three different measurement program; Customer Responsiveness program, Cost of Poor Quality program, Performance Measurement program. At the end we provide the organization with the GSE approach and guideline to how they can use it.

In the following we provide the answers to our research questions.

6.1 Research Questions

RQ1: What do we know about GQM+Strategies with respect to its practical application?

GQM+Strategies is a novel approach introduced in 2007, hence there is no extensive evaluation available yet. The approach has been applied, and in the application completed GQM+Strategies results have been demonstrated. However, little evidence on actual feedback by the practitioners is presented, and very few challenges of using the approach. The main challenges mentioned were related to the knowledge needed to apply the approach. That implied that if we would want to use it organization-wide everyone participating in metric definition would have to be trained. This supported the need for having an elicitation approach that can be moderated by a measurement program responsible, but does not require measurement specific knowledge of the stakeholder providing input to measurement program planning.

RQ2: How should software organizations elicit goals and strategies as input for GQM+ Strategies?

We developed GSE approach based on the GQM⁺ Strategies idea to elicit goals and strategies of software organization. GSE notation, which is inspired of variability models and was used in the first few interviews, was not working well for goal/strategy elicitation. We found that the goals and strategies concept created confusion. Although the meaning of goal and strategies are clear, the interviewees found it difficult to map their real world experience to these concepts. Goals and strategies became vague/ambiguous on lower levels when breaking them down to the sub-goals and sub-strategies. Furthermore, the notation restricted people expressing themselves. They tried to map their information to the notation and that hampered them talking freely, which made the elicitation process time consuming and resulted in capturing less information, because the notation was not intuitive. Overall, we conclude that notation/template based elicitation hampers the elicitation process when using GQM+Strategies.

This GSE approach evolved in three iterations and was evaluated in three measurement program; Customer Responsiveness program, Cost of Poor Quality program, and Performance Measurement program. We found that the GSE approach in later iterations was more intuitive and feasible for eliciting goals and strategies within software organization as the interview now was designed in a way that it was asking for goals and strategies, but rather asked the questions in a way that the interviewee does not require the definition of goal and strategy. Instead we, for example, asked for when program X is successful to find the goals, and asked for how the goal is achieved in order to find the strategies. Using visualization tools like MindMap enabled good communication and understanding between

interviewers and interviewees, because it makes the communication clear for both listener and speaker by visualizing a level understandability between them during the conversation. The GSE approach together with guideline is the proposed solution for usage in GQM+Strategies in eliciting goals and strategies in software organization. The organization decided to adopt the approach in future measurement program planning activities. This underlines the usefulness and applicability of the solution proposed and evaluated in this thesis.

6.2 Future Work

Software measurement program is a wide area, which can be improved from different angles. In our study we focused on eliciting the required information to plan measurement programs in a more transparency and structured way for GQM+Strategies.

In order to understand the quality of the elicited information based on our approach more long-term studies are needed. As we also mentioned, the approach was used with business stakeholders. In order to increase the generalizability of the results operational stakeholders need to be interviewed as well.

Furthermore, in general there is a need to conduct more empirical research on GQM+Strategies as the approach so far was evaluated in very few cases. The focus of the evaluations should be on hindrances in using the approach to assure that studies lead to further improvement and support of GQM+Strategies. In related work it was e.g. mentioned that there is a need to incorporate goal prioritization as not everything can be measured due to cost, and not all measures are equally important to the organization due to different importance of goals.

7 REFERENCES

- [1] Baldassarre, M. T., Caivano, D., & Visaggio, G. (2003). Comprehensibility and efficiency of multiview framework for measurement plan design. *Proceedings of the International Symposium on Empirical Software Engineering, (IESE 2003)*, 89–98
- [2] Basili, V., & Green, S. (1994). Software process evolution at the SEL. *IEEE Software*, 11(4), 58–66
- [3] Basili, V., Heidrich, J., & Lindvall, M., Regardie M., Rombach H. D., Seaman C. B., Trendovic A., (2007). Bridging the Gap between Business Strategy and Software Development. *Proceedings of the International Conference on Information Systems (ICIS 2007)*, 25
- [4] Basili, Victor R, Caldiera, G., & Rombach, H. D. (1996). The goal question metric approach. *Encyclopedia of Software Engineering, John Wiley & Sons*, 1994.
- [5] Basili, Victor R, Lindvall, M., Regardie, M., Seaman, C., Heidrich, J., Münch, J., Rombach, D., et al. (2010). Linking Software Development and Business Strategy through Measurement. *IEEE Computer*. 43(4):57-65
- [6] Benavides, F., & Ruiz-Cortés, A. (2009). Feature Model to Orthogonal Variability Model Transformations. A First Step. *Actas de los Talleres de las Jornadas de Ing del Software*, 3(2), 81–90.
- [7] Berander, P., & Jönsson, P. (2006). A goal question metric based approach for efficient measurement framework definition. *Proceedings of the 2006 ACM/IEEE international symposium on International symposium on empirical software engineering (IESE 2006)*, 316
- [8] Bhatti, A. M., & Abdullah, H. M. (2009). Master thesis: Deciding on Optimum Set of Measures in Software Organizations. *Master thesis*, (Blekinge Institute Höskolan).
- [9] Bhatti, A. M., Abdullah, H. M., & Gencel, C. (2009). A Model for Selecting an Optimum Set of Measures in Software Organizations. *Proceedings of the 16th European Conference on Software Process Improvement (EuroSPI 2009)*, 44–56.
- [10] Boyd, A. (2002). The goals, questions, indicators, measures (GQIM) approach to the measurement of customer satisfaction with e-commerce Web sites. *Aslib Proceedings*, 54(3), 177–187.
- [11] Boyd, A. J. (2005). The evolution of goal-based information modelling: literature review. *Aslib Proceedings*, 57(6), 523–538.
- [12] Brydon-Miller, M., Greenwood, D., & Maguire, P. (2003). Why Action Research? *Action Research*, 1(1), 9–28
- [13] Card, D. (2003). Integrating practical software measurement and the balanced scoreboard. *Proceedings 27th Annual International Computer Software and Applications Conference. (COMPAC 2003)*. 362–363
- [14] Card, D. N. (2003). Practical Software measurement. *Proceedings of the 25th International Conference on Software Engineering (ICSE 2003)*, 738–739.
- [15] Garcia, F., Bertoa, M., Calero, C., Vallecillo, A., Ruiz, F., Piattini, M., & Genero, M. (2006). Towards a consistent terminology for software measurement. *Information and Software Technology*, 48(8), 631–644.

- [16] Gopal, a., Krishnan, M. S., Mukhopadhyay, T., & Goldenson, D. R. (2002). Measurement Programs in Software Development: Determinants of Success. *IEEE Transactions on Software Engineering*, 28(9), 863–875.
- [17] Goethert, W., Hayes, W., (2001). Experiences in Implementing Measurement Programs. *Technical Report TR:CMU/SEI-2001-TN-026*
- [18] Kaneko T., Katahira M., Miyamoto Y., Kowalczyk M.: Application of GQM+Strategies® in the Japanese Space Industry. *IWSM/Mensura 2011*: 221-226
- [19] Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. *Keele University EBSE Technical Report EBSE200701* (Vol. 45, p. 65). Keele University and University of Durham.
- [20] Mandic, V., & Basili, V. (2010). Utilizing GQM + Strategies for Business Value Analysis An Approach for Evaluating Business Goals. *Proceedings of the 2010 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement ESEM 10 (2010)* (p. 1).
- [21] Mandic, V., & Markku, O. (2010). SAS: A Tool for the GQM + Strategies Grid Derivation Process. *Lecture Notes in Computer Science, 6156/2010*, 291–305.
- [22] Mandic, V., & Basili, V. (2010). An Approach for Evaluating Business Goals, Technical Report TR TR-TOL-2010-2802, University of Oulu, Department of Information Processing Science
- [23] Mandic, V., Harjumaa, L., Markkula, J., & Oivo, M. (2010). Early empirical assessment of the practical value of GQM+ Strategies. *Proceedings of the International Conference on Software Process (ICSP 2010)*, 14–25.
- [24] Mendic, M. G., Basili, V. R. V. R., & Mendonsa, M. G. (2000). Validation of an approach for improving existing measurement frameworks. *IEEE Transactions on Software Engineering*, 26(6), 484–499
- [25] Nielsen, K. A. (1995). The Experiment as Research Method in Sociology. *Dansk Sociologi*, 6(4), 72–92.
- [26] Niessink, F., & Vliet, H. V. (1998). Towards Mature Measurement Programs. *Proceedings of the 2nd Conference on Software Maintenance and Reengineering (CSMR 1998)*: 82–88.
- [27] Offen, R. J., & Jeffery, R. (1997). Establishing software measurement programs. *IEEE Software*, 14(2), 45–53.
- [28] Olsson, T., & Runeson, P. (2001). V-GQM: a feed-back approach to validation of a GQM study. *Proceedings Seventh International Software Metrics Symposium (METRIC 2001)*, 236–245.
- [29] Park, R. E., & Goethert, W. B. (1996). *Goal-Driven Software Measurement — A Guidebook*, Carnegie Mellon University, SEI
- [30] Sarcia, S. Is GQM+Strategies really applicable as is to non-software development domains? *Proceedings of the International Symposium on Empirical Software Engineering and Measurement (ESEM 2012)*
- [31] Tahir, T. (2009). A Structured Goal Based Measurement Framework Enabling Traceability and Prioritization. *Proceedings of the 6th International Conference on Emerging Technologies (ICET 2010)*, 282-286.
- [32] Trendowicz A., Heidrich J., Shintani K. (2011): Aligning Software Projects with Business Objectives. *IWSM/Mensura 2011*: 142-150

- [33] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2000). *Experimentation in software engineering: an introduction*. Kluwer Academic Publishers.
- [34] van Solingen, R., & Berghout, E. (2001). Integrating goal-oriented measurement in industrial software engineering: industrial experiences with and additions to the Goal/Question/Metric method (GQM). *Proceedings Seventh International Software Metrics Symposium (METRIC 2001)*, 246–258.

8 APPENDICES

8.1 Interview Structure Version I

Study Plan

This interview is planned to be conducted in 3 parts: The introduction part, the case study part and the evaluation part.

Purpose of interviewing

We planned 6 interviews to evaluate the 1st iteration of GSE approach, which includes the process description, modeling notation and questionnaire to be used for Goals and Strategies Elicitation phase. For measurements, we need to identify organizational goals and strategies to support decision making on what to measure in order to determine if we achieve our goals, and if we have chosen the right strategies.

Based on your input we will make updates and refinements. For that purpose we chose customer responsiveness program/product adaptability as a start/pilot case study and you have been selected as one of the experts with the help of PO Olsson. We selected you as you are involved in Ericsson initiative to define an Ericsson wide measurement strategy, and we would like to get your opinion and viewpoints.

What will happen during the Interview?

We have already come up with a set of initial goals and strategies for customer responsiveness program/product adaptability based on a number of documents and powerpoints that we collected in Ericsson. Here, we would like to:

- Elicit further the goals and strategies using the process description, modeling notation and questionnaire
- Collect feedback on the usability, completeness, correctness of our approach to be used in Ericsson.

Important: There is no right or wrong answers; we are looking for your opinions/viewpoints.

Confidentiality Statement

Ask for recording. Explain reason for recording: aid in interpretation (removes an interpretation step, otherwise we interpret by taking notes, then interpreting the notes, will be able to capture your viewpoint much better when recording).

Highlight: Recording will be treated confidentially, recordings will be deleted.

Tell them about member-checking (to get feedback about interpretation/summary)

Part 2: Action Research (45 Min.)

Questions to collect Interviewee Information

Name of the interviewee:

Acceptance:

I would like this approach of identifying goals and strategies to be adopted at Ericsson.

○ ○ ○ ○ ○ ○ ○
Completely Completely
Agree x Disagree

Explanation of rating (why you rated in this way)

8.2 Interview Structure Version II

This Guideline is composed of 2 parts: a) the Introduction part, which explains the process to be used when conducting the interviews with the measurement stakeholders and b) the main Interview Questionnaire (and associated MindMap skeleton) that is to be used for eliciting the goals and information needs in the organization. (ISO/IEC 15939 Standard was used as a base for the terminology used in this Guideline).

Study Plan: This interview is planned to be conducted in 3 parts: The introduction part, the case study part and the evaluation part.

Purpose of interviewing: We planned 6 interviews to evaluate the 1st iteration of GQM-Lean, which includes the process description, modeling notation and questionnaire to be used for Goals and Strategies Elicitation phase. For measurements, we need to identify organizational goals and strategies to support decision making on what to measure in order to determine if we achieve our goals, and if we have chosen the right strategies.

Based on your input we will make updates and refinements. For that purpose we chose customer responsiveness program/product adaptability as a start/pilot case study and you have been selected as one of the experts with the help of PO Olsson. We selected you as you are involved in Ericsson initiative to define an Ericsson wide measurement strategy, and we would like to get your opinion and viewpoints.

According to the first two interviews, we decided to continue with the 2nd Iteration for GQM-Lean. In this iteration we skip the life cycle goal tree and start from the top level goal. Therefore we let the interviewee to come down and define the goals and strategies himself.

What will happen during the Interview? There are set of Questions to first define the definition of the top level goal. And then with the help of interviewee goal break down structure will happen.

Important: There is no right or wrong answers; we are looking for your opinions/viewpoints.

Part 2: Action Research (45 Min.)

Goal: We would like to measure customer responsiveness and hence need to understand how it is structured.

Confidentiality Statement: We would like to request your consent to record our interview to ensure that we capture your views accurately. The recording will be treated confidentially and deleted after transcription and interpretation. Please let us know if an audio recording is acceptable to you, otherwise we will just take written notes of our conversation.

Questions to collect Interviewee Information

Name of the interviewee:

What is your current role in the organization?

How long have you been working at Ericsson?

What areas have you been in before and how long have you been working in?

Experience before coming to Ericsson (study, work at other companies etc.):

Question to be used during Goals and strategies Elicitation Phase

Interview Questionnaire for 3rd and 4th Interviewees – 2nd Iteration for GQM-Lean

Q1: What is customer responsiveness program?

Q2: Why was it initiated?

Q3: The customer responsiveness program is successful when:

d)

e)

f)

g)

Q4: Who can influence the success of a, b, c, d, e, f...

Q5: What do you do to achieve a, b, c?

Q6: In these activities that you just mentioned, which ones are working out good today?

Q7: Which ones have the highest improvement potential?

Q8: How can we improve them?

Part 3: Evaluation and feedback (30 Min.)

3.1 Understandability/Usability:

a. The interview questions were easy to understand.

Completely
Agree

Completely
Disagree

Explanation of rating (why you rated in this way)

3.2 Completeness:

Some important types of information (e.g. rationale) have not been captured during the interview

Completely
Agree

Completely
Disagree

