Empirical Software Engineering manuscript No.
(will be inserted by the editor)

An Elicitation Instrument for Operationalising GQM$^+$Strategies (GQM$^+$S-EI)

Kai Petersen · Cigdem Gencel · Negin Asghari · Stefanie Betz

Received: date / Accepted: date

Abstract Context: A recent approach for measurement program planning, GQM$^+$Strategies, provides an important extension to existing approaches linking measurements and improvement activities to strategic goals and ways to achieve these goals. There is a need for instruments aiding in eliciting information from stakeholders to use GQM$^+$Strategies. The success of GQM$^+$Strategies highly depends on accurately identifying goals, strategies and information needs from stakeholders.

Objective: The research aims at providing an instrument (called GQM$^+$S-EI), aiding practitioners to accurately elicit information needed by GQM$^+$Strategies (capturing goals, strategies and information needs).

Method: The research included two phases. In the first phase, using action research method, the GQM$^+$S-EI was designed in three iterations in Ericsson AB. Thereafter, a case study was conducted to evaluate whether the information elicited with the designed instrument following the defined process was accurate and complete.

Kai Petersen (corresponding author)
Blekinge Institute of Technology
Karlskrona, Sweden
E-mail: kai.petersen@bth.se

Cigdem Gencel
Free University of Bolzano-Bozen,
39100 Bolzano-Bozen, Italy
E-mail: cigdem.gencel@unibz.it

Negin Asghari
Ericsson AB
Kista, Sweden
E-mail: negin.asghari@ericsson.com

Stefanie Betz
Blekinge Institute of Technology
Karlskrona, Sweden
E-mail: stefanie.betz@bth.se

Results: We identified that the industry requires elicitation instruments that are capable to elicit information from stakeholders, not having to know about the concepts (e.g. goals and strategies). The case study results showed that our proposed instrument is capable of accurately and completely capturing the needed information from the stakeholders.

Conclusions: We conclude that GQM$^+$S-EI can be used for accurately and completely eliciting the information needed by goal driven measurement frameworks. The instrument has been successfully transferred to Ericsson AB for measurement program planning.

Keywords Goal Driven Measurement · Goal Elicitation · Strategy Elicitation · Information Need · GQM$^+$Strategies · Action Research · Case Study

1 Introduction

In today's competitive software development market, there is an increasing need for having a well-structured measurement program for software organizations to understand, control and improve the current status of software processes, products and 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 [Basili et al., 2007, Basili et al., 2010]. The key success for measurement initiatives is to provide a well-defined plan for executing measurement programs, which can be efficiently implemented afterwards [Mandic et al., 2010b]. As [Mendonça and Basili, 2000] point out, a high quality software measurement program should lead to sound, complete, lean, and consistent measurements.

A number of frameworks, models and standards (cf. [Basili et al., 2010, Tahir and Jafar, 2011]) have been developed to support software organizations in planning their measurement programs. The main idea of many approaches is to provide a set of metrics that might be relevant for a specific application domain, process area, or quality aspect. Links to goals or links to information needs are typically only vaguely provided if at all. Mechanisms to creating such links are usually missing. However, there are a few exceptions, that aim at deriving measures from measurement goals (such as Goal Question Metric [Basili and Weiss, 1984, Caldiera and Rombach, 1994]) or connecting measures to organizational goals (such as Balanced Score Card [Kaplan et al., 1996]).

The empirical evidence collected in different contexts have shown that goal-driven measurement approaches are rigorous, adaptable and flexible [Boyd, 2005]. Among these, some of the well-known ones are Goal Question Metric (GQM) [Basili and Weiss, 1984, Caldiera and Rombach, 1994], Balanced Score Card [Kaplan et al., 1996], Goals Questions Indicators Measures (GQ(I)M) [Boyd, 2002], Model Measure Manage Paradigm (M3P) [Offen and Jeffery, 1997], ISO/IEC

15939 standard on software measurement process [ISO/IEC15939, 2002] and Decision Maker Model [McGarry et al., 2002].

The Goal-Question-Metric (GQM) served the industry for quite a long time with the purpose of improving the process and quality of software product development [Ardimento et al. 2006]. GQM suggested that measurements in software development should be driven by measurement goals in a top down fashion. GQM defines a measurement model consisting of multiple levels: goals, questions and metrics. Measurement goals represent what should be achieved with the measurement. Questions represent the information needs of the stakeholders (i.e. answers to the questions help to achieve the measurement goal), and the metrics help in answering the questions. In this approach, each metric connects indirectly to a goal through one or more question(s), therefore no metrics are defined which do not have any justification related to a measurement goal.

GQM has been extended and complemented over time. Offen and Jeffery [Offen and Jeffery, 1997] introduced the M^{3}P which proposed 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. M^{3}P hence incorporates business goals, and emphasizes linking numerical results to them similar to GQ(I)M [Boyd, 2002, Park et al., 1996]. GQ(I)M is the acronym for Goal Question Indicator Measure and the "I" in parenthesis emphasizes the difference between this framework and GQM. GQ(I)M 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. In comparison to GQM, GQ(I)M incorporates business goals in addition to project goals, hence incorporating the business perspective [Boyd, 2002]. It was specifically designed for the CMM/CMMI family of process improvement models [Park et al., 1996].

Recently, [Basili et al., 2007, Basili et al., 2010] introduced GQM$^+$Strategies to explicitly link organizational goals across multiple levels (e.g. business, software, and process). This model extends the GQM 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 for 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, we chose this approach to form the basis for this research.

On the other hand, in order to run goal-driven measurement frameworks, plenty of information has to be elicited; including goals on multiple levels, strategies, and information needs. In particular, capturing all this information from the perspective of stakeholders accurately is essential, given that they will use the information as a decision parameter for metrics selection. The amount of information to be captured also increases when the scale of the organization is large (e.g. in terms of number of systems, or the number of people or

projects involved). The different perspectives (e.g. roles in the organization) then have to agree on the accuracy of the captured information to achieve a common understanding, and to decide what is important (e.g. through walk-throughs or discussion). Alternatives for prioritizing the importance of goals and strategies and how to use that for selecting software metrics is discussed in [Graham, 2009].

However, the existing goal-driven approaches neither explicitly define how to, nor provide supporting guidelines for the information elicitation process. [Kilpi, 2001] briefly mentioned the overhead costs due to time consuming goals setting sessions and negotiations in Nokia. Boyd [Boyd, 2005], referring to Wilson et al. [Wilson et al., 2002], claimed that it is difficult to elicit information on business goals and strategies in an organisation as stakeholders were not always fully aware of, or able to articulate their goals. The need to arrive at a common understanding and agreement on what is important further adds to that challenge.

We checked whether elicitation approaches are available and were empirically evaluated in order to capture information needed for goal-driven measurement through a pilot search. The goal was to determine whether there exists any study where the main focus was on how to design an elicitation instrument for goal-based measurement program planning. The terms used were *"(GQM OR "goal based measurement") AND elicitation"*. The majority of the papers (over 500) returned were related to requirements engineering, but none was empirically evaluating and comparing approaches to elicit information for a measurement program in general, and GQM$^+$Strategies in particular. In [Park et al., 1996], some guidelines are provided for eliciting information when applying goal-driven measurement process. However, no empirical evidence was provided, which the guidelines were built upon. Hence, we identified the following research gap: For goal-based measurement program planning we know what to collect, but have little empirical evidence for how to best collect the required information.

With the purpose of aiding organizations in eliciting information needed for GQM$^+$Strategies based measurement program planning, we fulfilled the following objectives:

1. Design an elicitation instrument for GQM$^+$Strategies in close collaboration with industry and practically apply it.
2. Empirically evaluate the approach with respect to its ability to accurately capture information for GQM$^+$Strategies from practitioners.

The remainder of this paper is structured as follows: Section 2 provides background on GQM$^+$Strategies and the studies conducted on the topic. Thereafter, Section 3 presents the related work on information elicitation approaches in general, and in the context of goal-based measurement. Section 4 provides the research methodology used in this study. Section 5 presents the results, followed by a discussion of the results and their implications in Section 6. Section 7 concludes the paper.

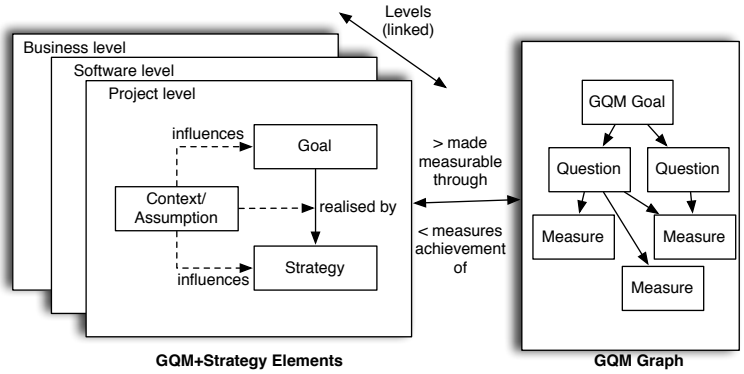

Fig. 1: GQM+Strategies Concepts based on [Basili et al., 2007]

2 Background

2.1 GQM+Strategies Structure and Concepts

GQM+Strategies [Basili et al., 2007,Basili et al., 2010] is the most recent advancement in goal-driven measurement program planning. Figure 1 illustrates the structure of GQM+Strategies. The key elements of GQM+Strategies are goals, strategies, and assumptions.

Goals, strategies, and assumptions/contextual information is collected and linked with each other. These are collected on multiple levels, e.g. business level, software level, and project level and linked with each other across levels.

To avoid a disconnection between the business perspective and operational measures the GQM+Strategies elements are linked to GQM graphs to document how they could be quantified to determine the achievement of the goals documented in GQM+Strategies elements. For instance, one could learn whether a new strategy used was leading an improvement with respect to the desired business goals. For that it is important to determine an interpretation model, which states when a goal has been achieved.

[Münch and Heidrich, 2008] provide definitions for the concepts used in GQM+Strategies. The definitions from [Münch and Heidrich, 2008] are provided in Table 1.

2.2 Studies on GQM+Strategies

Mandic et al. [Mandic et al., 2010b] 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),

Table 1: Action Research Interviewers

Business goal	Definition	Example
Business goals	Goals the organization wishes to accomplish	Market shares to be achieved, revenue goals, and other benchmarking goals
i-th level goals	Lower-level goals that are inherited from higher-level goals (level i-1)	Lvl i (business goal) improve market has strategy (improve quality), leads to lvl i-1 goal (reduce customer trouble reports by X%) on software level
Strategy	Approaches to achieve the goals	Improve product quality, increase efficiency, etc.
Assumptions	Estimated unknowns affecting the interpretation of data in relation to linkage between elements	Assumptions how different goals influence each other
Context factors	Environmental factors representing the organizational environment	Experience in measurement programs, means for automated data collection
GQM$^+$Strategies Element	Single goal and its strategies, including context and assumptions	Left box in Figure 1
GQM-Graph	Single GQM goal and related questions, metrics, and interpretation models	Right box in Figure 1
GQM$^+$Strategies Grid	Collection of GQM$^+$Strategies elements, GQM-Graphs and all links	Figure 5

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 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 et al. [Mandic et al., 2010a,Mandic and Basili, 2010] extended the GQM$^+$Strategies by introducing the value goals with the intent of analysing business goals. To this end, they integrated business value analysis (BVA), which provided a coupling of cost-benefit and risk analysis (value goals) with operationally measurable business goals. 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 was presented to understand the approach, but no empirical evaluation has been done yet.

Mandic and Oivo [Mandic and Oivo, 2010] proposed a tool that allows capturing context information for GQM$^+$Strategies as well as goals and strategies. The tool was built to support information elicitation for the GQM$^+$Strategies when creating the tree/grid for the measurement program. However, the paper

does not provide information of how to elicit the information from 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.

An evaluation of GQM+Strategies in the military training domain was presented in [Sarcià, 2010]. In this study, the authors elaborated the extent the GQM+Strategies has to be modified in order to be applicable. GQM+Strategies defines multiple levels as 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. In addition, they evaluated the ease of applying this approach in terms of efficiency and effectiveness by applying it 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.

GQM+Strategies has been applied in the Japan Aerospace Exploration Agency [Kaneko et al., 2011]. The outcome was in the main focus of the work. However, how the outcome was reached and which challenges/support the approach supported were not elaborated. Some lessons learned have been presented, namely: (1) clarify the relationship between different organizational units and organizational levels; (2) involve the people in the organization; and (3) when having used the approach, record components for goals and strategies and reuse them to reduce effort in running the measurement program. According to the authors, GQM+Strategies also needs to be extended to incorporate priorities of goals.

The authors in [Trendowicz et al., 2011] 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.

[Jürgen Münch, 2013b] conducted a recent case study reflecting on their experiences when using GQM+Strategies. The study lasted for five months and was structured along several workshops. The main drivers for elicitation in this case were workshops and planning meetings held with company representatives. The study highlights several challenges: 1) Initial training is needed given that stakeholders do not all have prior knowledge of the GQM+Strategies concepts; 2) it is challenging to find a suitable entry point to start the investigation; and 3) understanding the organizational structure is a challenge given that roles are not always clearly defined. Key recommendations were given as: having representatives from multiple levels of the organization to partic-

ipate; using the concept of saturation to determine stability, and using goal evaluation patterns which appeared as very useful.

[Jürgen Münch, 2013a] conducted another investigation at the same company as in [Jürgen Münch, 2013b] with the goal of determining goal harmonization and alignment. The study presents revisions of goals for revisions of the GQM+Strategies grid. The findings indicated that during workshops only little new information was added to the grid. Hence, overall the evidence indicated that GQM+Strategies helps in alignment and harmonization of the goals of an organisation. Other positive side-effects mentioned was increased transparency.

3 Related Work

The area with most research on elicitation in software engineering is associated with the requirements engineering activity. Two systematic reviews [Davis et al., 2006,Dieste and Juzgado, 2011] have been conducted to aggregate evidence to identify which elicitation approach has been most effective.

In the systematic review by Davis et al. [Davis et al., 2006], interviews have been found to be most effective in terms of amount of information gathered. Based on vote counting (counting number of studies in favor or against a specific outcome), they found that structured interviews appear to be more effective than unstructured interviews, and that unstructured interviews are more effective than the think aloud technique. Another finding was that visual aids do not aid discovering new information, but help in focusing the discussion.

The review by [Dieste and Juzgado, 2011] analyzed studies on interview techniques which were categorized into interviewing, laddering, sorting, and protocol review techniques. The findings of this review also support that interviews appear to be the most effective technique for elicitation. The least effective technique was protocol analysis.

Park et al. [Park et al., 1996] suggested using structured brainstorming, the Nominal Group Technique [Scholtes et al., 2003] and interviews for eliciting information when applying goal driven measurement.

As part of the GQM approach, GQM abstraction sheets are a means to conduct interviews to elicit GQM-based information. An abstraction sheet selects a unit of study, purpose, quality focus, and context for one specific goal. It is answered from a viewpoint. Four quadrants are to be filled in with the interviewee, namely quality focus (measured properties), variation factors (influencing factors on measured properties in the quality focus), baseline hypotheses (current situation with respect to measured properties), and variation hypotheses (effect variation factors will have on the properties in the quality focus) [Briand et al., 1996]. The approach has been successfully used in industry as part of the overall GQM-approach (cf. [Van Latum et al., 1998]). The abstraction sheet is valuable in eliciting information building the GQM graph

as part of GQM+Strategies. It has been evaluated as part of GQM, but has not been the main focus of an empirical study.

Contribution to related work: There are a number of publications on the GQM+Strategies, while only few represent industrial case studies (see Section 2). The existing case studies did not provide a guide that a measurement program responsible could take to communicate and elicit information from relevant stakeholders beyond the presentation of the definition and concepts, as well as the final result (GQM+Strategies grid). The grid itself is the outcome of the elicitation (i.e. communicating with the stakeholders, studying documentation, etc.).

There exist templates to fill in GQM+Strategies related information. Furthermore, the GQM+Strategies grid as an instrumentation instrument could be used for elicitation. In this study, the traditional instruments (template/GQM-grid) have been utilized first with an initial assumption that they would be easy and intuitive to populate with information. However, it became evident that a new solution was needed, which does not rely on substantial experience on measurement concepts when interviewing practitioners (see Section 5.1.1).

This research adds to the current body of knowledge on the GQM+Strategies by iteratively designing an elicitation instrument in industry and providing reflections for each design iteration for alternative elicitation approaches. The designed instrument is then used in a case study focusing on a specific measurement program, making a novel contribution in empirically evaluating the quality of the elicited information with respect to accuracy and completeness in industry.

4 Research Methodology

This research is divided into two phases building upon each other: i) Phase I - Design of the Elicitation Instrument and ii) Phase II - Evaluation of the Elicitation Instrument. Figure 2 provides an overview of the research methodology used in this study.

In the first phase, we used action research [Brydon-Miller et al., 2003] to iteratively design the elicitation instrument for the GQM+Strategies (called GQM+S-EI) in close collaboration with the case company (see Section 4.2 for details). We took the role of measurement program planners for one particular improvement program at the company. The research question (RQ1) for this phase of the study was:

- *RQ1: 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 from stakeholders. The initial design goal of GQM+S-EI was to capture the main structural elements (goals, strategies, links between them) from practitioners to make the construction of the GQM+Strategies grid as presented in [Basili et al., 2010]. The

GQM$^+$Strategies grid is the outcome of the elicitation activity, but can also serve as an instrument for elicitation. We used a notation-based grid structure as the starting point for elicitation, and made changes iteratively using action research to improve. Each iteration provides insights of which strategies of elicitation are applicable. Having defined a stable and well working solution for elicitation, we used the instrument in the second phase.

In the second phase, we conducted a case study [Runeson and Höst, 2009] on a different improvement program to empirically evaluate the GQM$^+$S-EI in eliciting information accurately and completely from the stakeholders (see Section 4.3 for details). The research question was as follows:

- *RQ2: How accurate and complete is GQM$^+$ S-EI in capturing GQM$^+$ Strategies relevant information from stakeholders?*

The answer to this question provides evidence for the accuracy and completeness in capturing information from stakeholders, which are the main intention of the elicitation approach.

Another important related issue to eliciting information is to decide on the right goals and strategies for a measurement program to be successful (e.g. how to prioritize elicited goals, and how to choose metrics accordingly). However, this is another step in the overall measurement program and is not elaborated here. Information about this important issue can be found in, for example, [Gencel et al., 2013].

4.1 Research Context

This study was conducted as part of a research project (GQM-Lean: Establishing a Lean and Sustainable Software Measurement Program funded by Ericsson Software Research (ESR)). The research context is characterized by the company and the measurement program being studied.

The industry partner for this study was Ericsson AB, which is one of the major telecommunication companies in the world offering telecommunication services and products to telecom operators. Ericsson is certified by ISO 9001:2000. It is a market-driven company and the frequently changing market influences its product development approach, which is incremental with the use of agile practices. Details on the process followed in the company can be found in [Petersen and Wohlin, 2009a,Petersen and Wohlin, 2010a]. Ericsson uses balanced scorecard as a performance tool for managing and measuring the software process, products, and resources. Table 2 represents more detail regarding the context of the study. The same company was investigated in both Phase I and Phase II.

Ericsson AB has been using ISO 15939 to structure their measurements, aimed at defining common base measures that will fulfil the most important information needs (10 to 20). Further requirements formulated were the possibility to benchmark externally, strengthen organizational alignment, and satisfying the information needs of multiple stakeholders. Many measures were

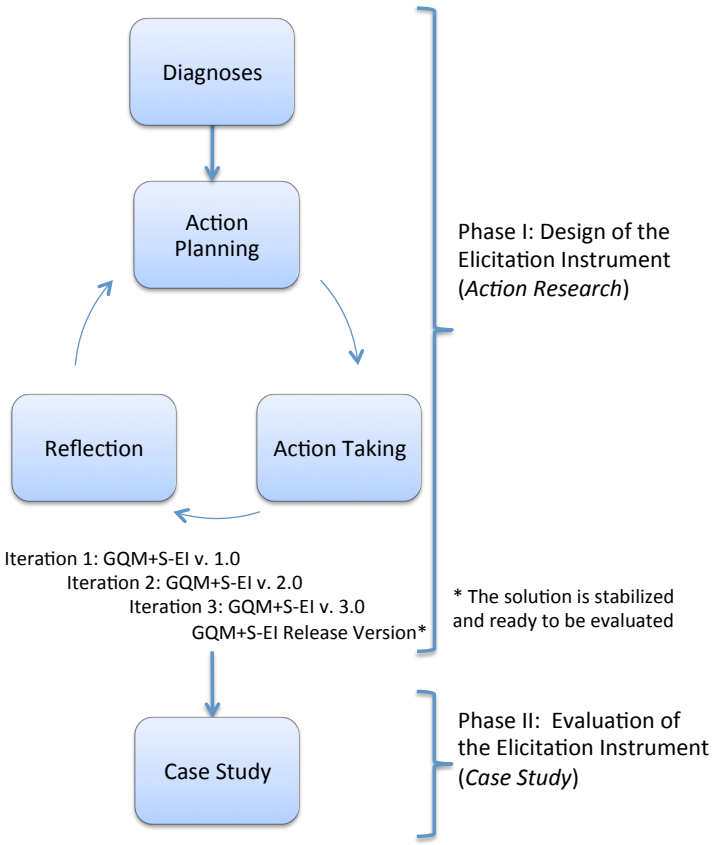

Fig. 2: Research Methodology

collected manually and had to be aggregated from different databases. Hence, the measures identified the need to support measurement programs running currently by automation to reduce the workload, increase efficiency, and to make better analysis of the collected data. Recently, Ericsson AB built dashboards that are connected to collected data that has been captured automatically (e.g. for lead-time and throughput related measures). No GQM-based representation has been used in the studied case prior to this study.

Table 2: Context Characterization

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

4.2 Phase I: Design of the Elicitation Instrument

4.2.1 Units of analysis

Improvement programs are initiatives to improve and are to be supported by measurement programs. In our initial meetings in Ericsson AB, we agreed to develop GQM$^+$ S-EI in a currently ongoing improvement program already planned, namely 'Customer Responsiveness Program'. The Customer Responsiveness program has been initiated in 2010 for reducing lead-time end-to-end in creating customizations requested by the customers in-between releases (cf. [Mujtaba et al., 2011]). The reason for requesting them is that customers would not like to wait for a new release before receiving a feature they desire. Having developed a wide range of customizations, the company decides before a release, which of the customizations developed should become part of the next release delivered to all customers. Furthermore, we also made interviews in two of the other ongoing improvement programs, one focusing on evaluating the 'Cost of Poor Quality' and the other on 'R&D Performance' evaluation. The choice of improvement programs was driven by the interest of the organization, and the priority of planning and gathering information for specific programs.

4.2.2 Overview of the method

We used action research method (action research [Brydon-Miller et al., 2003]) for designing the elicitation instrument. The reason for choosing this method was that the company committed to provide the required resources and continuously collaborate with us to iteratively design the elicitation instrument to be later on transferred to Ercisson AB. Action research method follows the steps of diagnosis, action planning, action taking, and reflection. The action is an intervention introduced by the researcher to induce a positive change in the company. Examples of action research applied in the software engineering context can be found in [Baca and Petersen, 2013,Iversen et al., 2004].

Phase I in Figure 2 provides an overview of the research process used in this phase of the research. We started with the diagnosis step in order to understand the challenges in industry in relation to existing approaches and decide on the scope of the research. Furthermore, we investigated the

improvement program the case company chose to work on when designing the elicitation instrument. We studied the process documentation in order to get familiar with the terminology used in the company. Whenever there was a need, we held informal meetings with the practitioners in the company to clarify answers to our questions. Having a good understanding, we continued with the action planning and action taking steps designing and testing an initial version of GQM$^+$S-EI. Thereafter, we reflected on the results and identified improvement points GQM$^+$S-EI for the next iterations. The instrument was finalised and released after three iterations.

4.2.3 Roles and their interaction

The action research team consisted of three persons who acted as responsible for measurement program planning. In this role, they elicited and aggregated information related to measurement program planning from stakeholders. Their initial tasks were defined on a high level beforehand, as is shown in Table 3.

Table 3: Action Research Interviewers

Name	Role	Description
Negin Asghari	Lead interviewer	Guides the interview by following the guidelines provided by GQM$^+$S-EI.
Cigdem Gencel	Support	Asks follow-up questions and follows interesting lines of questioning.
Kai Petersen	Scribe	Documents the information provided while using GQM$^+$S-EI and provides summaries in-between.

Six interviews were needed to achieve a stable and well working version of GQM$^+$S-EI. The practitioner profiles of the interviewees are provided in Table 4. 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 interviewees are very experienced and know the organization well, and all have been with the company for more than 10 years. The manager then contacted the interviewees to get their consent. Thereafter, we sent an invitation email to all interviewees to inform them about the purpose of the investigation.

4.2.4 Data collection

In the action research process understanding the processes, products, practices, and corporate terminology is essential to be able to act in the role of a measurement program planner. To acquire the knowledge, archival data (in particular process descriptions, power-point presentations, and measurement reports) have been studied. In case information was not understandable to us, our contact persons identified experts who could help in that situation. Most

Table 4: Action Research Interviewees

Interviewee ID	Current Role	Description
AI1	Architect	This interviewee has been working in Ericsson for 19 years. She 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.
AI2	System manager	This interviewee has worked over 30 years at Ericsson in different levels of the organization and in different positions. 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: excel in reuse and improve the system architecture. As a system manager, she is documenting their blueprint architecture, which represents process steps for facilitating reuse.
AI3	Solution architect	This interviewee 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.
AI4	Requirements engineer	This interviewee is a requirements engineer responsible for requirements road-map and allocation. She has nine years experience working in Ericsson and 20 years experience in other companies like; Anoto, Telia Mobile, UAB, ELLEMTEL, and Logica Svenska AB. She has worked as a developer, tester, and product manager before.
AI5	Strategic management	This interviewee is involved in strategic management, and she has been working for Ericsson for 17 years. She has experienced in RD organization within Ericsson with the focus on radio networks.
AI6	R&D Manager	This interviewee 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 RD management.

of the documents have been written abstractly and were hard to understand for people not yet familiar with the measurement programs. Therefore, ten meetings (online and face-to face) have been conducted in order to understand the content of the documents. Online meetings were necessary in cases where the interviewee was sitting in another branch of the organization located in another city.

In total, six interviews have been conducted to design three different versions of GQM$^+$S-EI (see Section 5.1). The data collection consisted of four steps:

– Step 1. Introduction: The purpose has been introduced to the practitioner being interviewed. The introduction has evolved during the action research

iterations, and can be found in Section 5.1.1, Section 5.1.2 , Section 5.1.3 and Section 5.1.4 for GQM$^+$S-EI v.1.0, v.2.0, v.3.0 and release version, respectively.
- Step 2. Application of the GQM$^+$S-EI instrument to elicit information. The first version of the elicitation instrument is described in Section 5.1.1. The versions 2.0 and 3.0 are presented in Section 5.1.2, Table 8. The release version of GQM$^+$S-EI is shown in Tables 9 and 10 in Section 5.1.4.
- Step 3. Feedback collection: We gathered feedback from the interviewees regarding GQM$^+$S-EI. During Step 3, the interviewees have been asked regarding their reflection and discussion on the following aspects for all interviews using opinion questions. They rated their level of agreement with the statements below on a 7-point likert scale of 'Completely agree' to 'Completely disagree'. They were also asked to comment on their answers. The ratings and the comments were then used as a starting point for a more open discussion of why they gave that rating, which led to concrete actions suggested for improvements.
  - Understandability/usability: The interview questions to be used as part of GQM$^+$S-EI were easy to understand.
  - Completeness: Some important types of information (e.g. rationale) have not been captured.
  - Correctness: We have been asked the right questions to be able to capture the intention of process related to the improvement program.
  - Overhead: We have been asked for irrelevant information, which brings unnecessary overhead.
  - Acceptance: I would like this elicitation approach (interview guide) to be adopted at Ericsson.
- Step 4. Research team reflection: The research team (see Table 3) immediately reflected what happened during the use of GQM$^+$S-EI, and discussed further improvements.

All data collection steps were recorded on tape and then transcribed. In case we needed clarification, we contacted the interviewees. Then, we presented the interview results to the practitioners to evaluate their accuracy (member checking [Yin, 2011]).

In parallel to the interviews, steering group meetings with customer representatives (including our contact persons at the company) have been held. During the meetings the progress has been reported, and the changes to the solution presented. This was an additional source for discussing, reflecting, and identifying improvements to GQM$^+$S-EI. Five steering group meetings have been held during the action research.

4.2.5 Data analysis

During the analysis phase, we used the transcribed text, our notes taken during the interview sessions and the answers to questionnaire that we used to get feedback on each iteration (Step 3). The interviews were analyzed by noting

down observations in our own words, where we observed challenges coming to light given that an aspect of the instrument was not working well during interview sessions. Furthermore, we summarized the feedback collected from Step 3 to take into consideration which improvements the practitioners see when designing the new version of the interview. Much of this documentation was also done in Step 4.

To increase the reliability of this analysis, we had multiple researchers present, reducing the bias in the interpretation of an individual researcher (observer triangulation [Runeson et al., 2012]). In order to structure the analysis, we assigned tags to the statements (one tag indicating where things were going well and one where improvements are needed) and associated them to text fragments during the interview (Step 2) and during the research team reflection (Step 3). Based on the analysis we highlighted noteworthy observations as narratives and provided reflections that led changes to the instrument. [Colombo, 2003] suggested narrative analysis as a meaningful way to analyze action research results.

4.2.6 Threats to validity

[Petersen and Gencel, 2013] provided a reflection of validity threat categorizations for software engineering and proposed to discuss four types of validity threats: 1) descriptive validity (ability to describe what we observe objectively and truthfully), 2) theoretical validity (concerns controllability and whether the measures used capture what they intend to capture), 3) generalizability (the degree of generalizability internally (within groups, communities, or a company) and externally (across groups, communities, and companies)), and 4) interpretive validity (whether the conclusions/inferences are reasonably drawn from the data objectively).

Descriptive validity: To avoid the risk that events during the action research are not accurately captured, two actions have been taken. First, all interviews were recorded and transcribed, which removes a step of interpretation if we would have only relied on notes. Second, our summaries of what occurred during the interviews have been verified with the interviewees through member checking. Hence, this threat is considered being under control.

Theoretical validity: A confounding factor between the setups is the background of the interviewees, as well as in some cases we were co-located and in others not. Regarding the background, all interviewees were very experienced and had a good understanding of the improvement programs involved, hence being able to provide the information requested. The video-conference system was advanced given that the company is using global software development; hence it allowed to see the other interviewers, and provided the feature of screen sharing. Overall, the set-up was similar to the face-to-face situation. Learning effect is a threat in action research, given that with every iteration learning occurs. In each interview session, new participants who were not exposed to the questions earlier, were involved to avoid the learning effect. A potential threat could be that the interview team learns and executes

the interviews with greater proficiency. Another threat could have been that the interviewees did not disclose all information as the interviews could have been be perceived as outsiders. At the time of the study, a collaborative research project has been established, with the first author being employed at the company and not being a superior to any of the interviewees. This aided in establishing trust. Furthermore, the authors had an office and were hence embedded in the corporate environment.

Generalizability: One risk is that the outcome depends on the context in which it was applied, i.e. the company and the measurement program. We applied the elicitation approach to support measurement program planning for different improvement programs (customer responsiveness, cost of poor quality, performance measurement) and found that the approach stabilized and the principle design was working well in the three programs. This supports internal generalizability. There exists a threat to external generalizability given that the approach has been evaluated in a single company. We described the context and provided references to detailed process description to support future aggregations of evidence (e.g. cross-case comparisons) to support external generalizability.

Interpretive validity: The interpretive validity concern the conclusions we draw from the observations made in each action research iteration. As these rely on interpretation, as suggested in the context of literature reviews, multiple researchers were involved in the interpretation step. The interpretations were further cross-checked and reflected in the reference group meetings to reduce the threat of interpretive validity.

4.3 Phase II: Evaluation of the Elicitation Instrument

The purpose of this phase was to empirically evaluate whether the designed instrument could completely and accurately capture GQM+Strategies relevant information from stakeholders.

4.3.1 Unit of analysis

Improvement program: During the case study, we studied another improvement program. This improvement program for which a measurement program was planned is concerned with the continuous transformation of Ericsson AB to being a lean organization according to the principles of lean manufacturing and software engineering [Poppendieck, 2003]. The program includes different business units including software development as well as hardware development. The initiative is hence distributed across multiple development units. The measurement program is the basis for creating measurement dashboards to continuously evaluate and improve with respect to being a lean software organization.

4.3.2 Overview of the method

The research method used is a single case study with one unit of analysis [Runeson and Höst, 2009]. The primary means of data collection were interviews in combination with observation.

4.3.3 Roles and their interaction

The team for empirical evaluation of the instrument during the case study consisted of one researcher (Kai Petersen) and two measurement analysts at the case company. Their experiences are shortly described in Table 5. Both analysts were also involved as contact persons during the action research, and as participants in the steering group meeting. They observed the use of the approach during the action research (Phase I). Furthermore, they received a presentation of the tutorial prior to conducting the case study (see Section 5.1.4). Having them being part of the interviewer team is an important part of creating a realistic situation (see e.g. [Ivarsson and Gorschek, 2011] for a discussion about facilitating practical relevance from a subject's perspective).

In this phase, another researcher, Stefanie Betz, was also involved after the interview sessions in: 1) reviewing the grid to make sure the translation of mind map to GQM+Strategies grid was correct, and 2) documenting the outcome of the member checking (walkthrough meeting) to assure accurate recording of requested changes by the practitioners.

Table 5: Case Study Interviewers

Name	Role	Experience
Measurement Analyst 1	Lead Interviewer	The lead interviewer at the time of the interviews was the head of product development excellence, a unit at the company which is concerned with process improvement, methods, tools, and measurements. He worked with cross-organizational and global improvement programs. Before that, he headed product development units, was a line manager, and a designer. He has been with Ericsson since 1995.
Measurement Analyst 2	Support	The support interviewer is a business intelligence product manager, and had repsonsibilities as a system design driver, unit manager, quality manager, as well as designer, tester and trainer. He has been with Ericsson since 1992.

In total, eight interview sessions with the improvement program stakeholders were performed in order to evaluate the instrument by applying it to elicit the required information. Two stakeholders participated in the Interview 7. The interviewees have been selected by the measurement analysts. They were from different development sites and business units. However, they all were involved in the improvement program and shared the main ideas and intentions as described in Sections 4.3.1. Table 6 provides an overview of the interviews, their roles, and their experience. Interviewees have been chosen to capture the

Table 6: Case Study Interviewees

Interviewee ID	Current Role	Description
CI1	Portfolio manager	Responsible for strategic product management. Manager since the 1990s, has been IT and unit manager, and is in the role of a portfolio manager since 2009
CI2	Driver lean and agile transformation	Currently driver (formally responsible) for lean and agile transformation. Worked in R&D, head of technology for 2 years, head of product management for six years.
CI3	Portfolio manager	At Ericsson since 1997, previous roles were unit manager, program manager, and department manager
CI4	Program and project manager	At Ericsson since 1991, previously in management positions in different organisations at Ericsson
CI5	Project manager	In the role of a project manager since 1999 at Ericsson.
CI6	Head portfolio and technology mgt.	At Ericsson for over 18 years, had roles as project manager, line manager, technology manager.
CI7	Interviewee 1: Responsible for operations and innovation/ Interviewee 2: Responsible for integration and verification Efficiency	Interviewee 1: 3-4 years responsible for operations and innovation, previous background in testing, test coordination, and test manager. Interviewee 2: Responsible for integration and verification efficiency, working with operational improvement, previous experience as test manager and line manager, 36 years at Ericsson.
CI8	Change driver for agile and lean, agile coach	At Ericsson for 15 years, headed different system management departments

operational perspective (project, e.g. CI4, CI5 in Table 6), strategic product perspective (Portfolio and program managers, e.g. CI1, CI3, CI6), and process improvement perspective (e.g. CI2, CI8). Choosing multiple perspectives has been important to not achieve saturation by just covering a very limited number of perspectives.

4.3.4 Data collection

The research team used the final release of GQM$^+$S-EI designed in Phase 1 for eliciting information from different stakeholders. During the elicitation process, the GQM$^+$S-EI comprises use of mind-maps (see Section 5.1.4) and GQM$^+$Strategies grid (see Fig 5).

The main source for determining the completeness and accuracy of the elicitation using the instrument is the GQM$^+$Strategies grid. The outcomes of the elicitation session using the mind-maps are documented in the grid, which provided the basis to determine newly added information with each additional

interview. Hence, the interviews using the elicitation approach form the basis of data collection for the study.

The final outcome (GQM$^+$Strategies) was reviewed in a walkthrough meeting, during which member checking (accuracy) was conducted. Prior to the meeting, each member received their own mind-maps by e-mail to make eventual corrections. The lead interviewer (see Table 5) guided the participants of the meeting through the GQM$^+$Strategies grid step by step. Prior to doing this, the first interviewer presented his slides and piloted his walkthrough with the other members of the interview team being there to provide feedback.

During the walkthrough, the participants of the meeting could comment on the grid. The scribe took meeting minutes to document all comments and changes requested. In total, four stakeholders not previously involved in the interviews, but concerned with the improvement program, were present as additional references to give further possibilities to add views to the grid. These included two program managers, the corporate responsible for improving business flow across business units, and a consultant for product development.

4.3.5 Data analysis

After completing all eight interview sessions, the scribe constructed the GQM$^+$ Strategies grid (see Fig 5). This was done by marking goals, strategies, information needs, and metrics in different colours. From the data collected in the first interview session, an initial version of the grid was constructed. Thereafter, the elements of the second interview were added, and so forth. If an element was already present, it was not added again. This way, it would become visible what was new in the GQM$^+$Strategies grid after each additional session.

After having constructed the grid, the last author of this paper checked whether mapping of the information collected in the mind map to GQM$^+$ Strategies grid was correct. Then, the scribe sent the grid to the interviewers and incorporated their feedback. Documenting the outcome of the member checking during the walkthrough meeting, the last author also checked the correctness and completeness of the recording of requested changes by the practitioners.

In this study, we used two measures to evaluate the results of the case study, and hence the GQM$^+$S-EI for accuracy and completeness. The first measure used was saturation (quantitative) to establish completeness, which means that with additional interview sessions to elicit information, less new data is obtained [Guest et al., 2006]. Saturation is recognized as a good measure for qualitative research. [Morse, 1994] stated *"saturation is the key to excellent qualitative work"*. In this study, we measured the saturation by counting the newly identified GQM$^+$Strategies elements with each additional interview session, as having been applied in software engineering research (see e.g. [Ali et al., 2012]).

The second measure used to determine the accuracy of the elicited information was done by member checking. The common approach to determine

the accuracy of qualitative findings is to use member checking [Yin, 2011], which has been conducted through the walkthrough activity with the interviewees and additional stakeholders determining whether their views are also covered. The number of change requests in terms of, for example, reformulating, adding, or deleting GQM$^+$Strategies elements has been used as a measure. The member checking is also an additional means to check for completeness. That is, if many new GQM$^+$Strategies elements would be suggested during the walkthrough, important information would have been missed.

4.3.6 Threats to validity

Descriptive validity: In this phase, we were not given consent to record the interviews due to confidentiality reasons. There was a risk that the GQM$^+$Strategies grid is inaccurately defined by the researcher, which would lead to inaccurate reflections based on the elicited information with respect to accuracy and completeness. Therefore, the first author created the grid from the interviews, and then the lead interviewer and support interviewer from the case study, who are also measurement analysts, reviewed the results. Furthermore, the last author of this paper checked whether mapping of the information collected in the mind map to GQM$^+$Strategies grid as well as the recording of requested changes by the practitioners were correct. These reduced the risk of the data collected in the grid being dependent on the researchers.

Theoretical validity: A lack of diversity in the background and roles of the participants could have led to a result where no new goals or strategies were identified early on. To avoid this threat, people from different business units and perspectives have been interviewed. Hence, this factor appeared to have a limited impact on the result. Furthermore, there was a threat having the researchers who were the developers of the approach drove the elicitation, as this constitutes an unrealistic situation (see [Ivarsson and Gorschek, 2011]). Hence, roles were changed from the action research to the case study, the roles of the lead and support interviews now being taken by corporate measurement analysts.

Generalizability: The measurement program is of large scale involving different business units. Hence, the results with respect to distribution between number of GQM$^+$Strategies elements, and the outcome with respect to completeness and accuracy might differ in other contexts (e.g. in the number of interview sessions needed to reach saturation), such as process improvement in smaller scales.

Interpretive validity: The interpretation was based on well-defined and widely used measures in qualitative research and requirements engineering. Hence, using established approaches in this study (saturation as a measure of completeness, and member checking) reduced the risk in drawing non-reasonable conclusions from the data.

5 Results

In this section, the results of Phase I and Phase II following the research methodology described in the previous section are presented.

5.1 Phase I: Design of the Elicitation Instrument

Table 7 shows the three major versions of GQM$^+$S-EI that were developed. Given that not all stakeholders were co-located at the development site where we had our work places, few interviews had to be conducted through video conference. The company, for that purpose, provided a video conference system where the participants can see each other, and also can share their screens. The majority of interviews were conducted in the customer responsiveness measurement program. Version 3.0 has been tried in three different improvement programs before the final release which was evaluated in the case study.

Table 7: Action Research Iterations Overview

GQM$^+$S-EI version	Interviewee	Instrument/process	Location	Program
1.0	IA1	Notation based + semi structured questionnaire	Face-to-face	Customer responsiveness
1.0	IA2	Notation based + semi structured questionnaire	Video conference	Customer responsiveness
2.0	IA3	Semi-structured questionnaire + white-board	Face-to-face	Customer responsiveness
3.0	IA4	Semi-structured questionnaire + mind-map tool	Video conference	Customer responsiveness
3.0	IA5	Semi-structured questionnaire + mind-map tool	Video + Face-to-face	Cost of poor quality
3.0	IA6	Semi-structured questionnaire + mind-map tool	Video + Face-to-face	Performance measurement

5.1.1 Iteration 1: GQM$^+$S-EI v. 1.0

In each of the iterations, we followed the activities of action research: 1)Diagnosis, 2)Reflection, 3)Action planning and 4)Action taking. Below, we describe each for the first iteration.

Diagnosis: During this activity, our focus was on establishing the current situation at the company. At the time of starting this research, Ericsson AB had already started using ISO/IEC 15939 to define measurements in a consistent and repeatable way. The company was in the process of defining a standard measurement process to be used in the organization. At the same time, the company already has a long history of establishing and using measurements, which was documented in various research studies such as in [Damm et al., 2006, Petersen and Wohlin, 2011, Petersen and Wohlin, 2010b].

Reflection: In the diagnosis stage, our aim was to explore with which elicitation approach to start with considering the current situation of the case company. The first idea was using existing templates (e.g. templates provided in the tutorial material [Münch and Heidrich, 2008] developed at Fraunhofer IESE, and having practitioners fill them in. However, one concern was that they might loose sight of the big picture regarding the overall structure given that the information was added in a document using tables, and not visually connecting the GQM+Strategies elements, such as in the GQM+Strategies grid. We also sent the template to one of the persons in the measurement program, which was not filled in due to the amount of information requested. Filling in the template also requires excellent understanding of GQM+Strategies [Sarcià, 2010]. It was also shown that, even with training, only 3 of 6 levels in the Bloom's Taxonomy tailored to GQM+Strategies could be obtained [Mandic et al., 2010b] (see also related work in Section 3). Therefore, we decided that the template itself would serve as an excellent checklist for completeness of the elicitation outcome, but not as the elicitation instrument itself.

Our first design hence was based on a self-developed notation to represent the GQM+Strategies grid to the stakeholder during the elicitation process, which was combined with a semi-structured questionnaire as discussed further below.

Action planning: The problem of planning a measurement program is similar to configuring a customizable system during product configuration in software product lines [van der Linden et al., 2007]. In software product line engineering, alternative features can be selected for a product. In software measurement program planning alternative measures, strategies, and goals can be selected. Furthermore, there might be dependencies between strategies and goals, as is the case between features in product lines. Hence, we adapted a notation used in product lines [Bühne et al., 2005] that allowed us to clearly distinguish goals and strategies from each other.

An example (cut-out), which we prepared based on the study of existing documentation about the Customer Responsiveness program, is shown in Figure 3. Goals are represented as triangles, and strategies as boxes to clearly distinguish the two elements. On the highest level, the strategic goal, which can have several associated strategies, was found (in this case strategy S1 is to run a customer responsiveness program). This strategy, on a lower level, becomes a goal, which generates further strategies, etc. Two types of links connect goals and strategies. Mandatory links indicate that the strategy has to be implemented in order to achieve the goal, while optional strategies could be implemented, e.g. one might choose one optional strategy over another one. At the links, a rationale is written why the strategy aids in fulfilling the goal, adding rationales to make the notation easily understandable. Furthermore, links exist between strategies: <include> links indicate that in order to implement one strategy another has to be implemented as well; <exclude> links indicate that when one strategy is implemented, it is in conflict with another strategy. In case there is uncertainty with respect to the links (e.g. whether

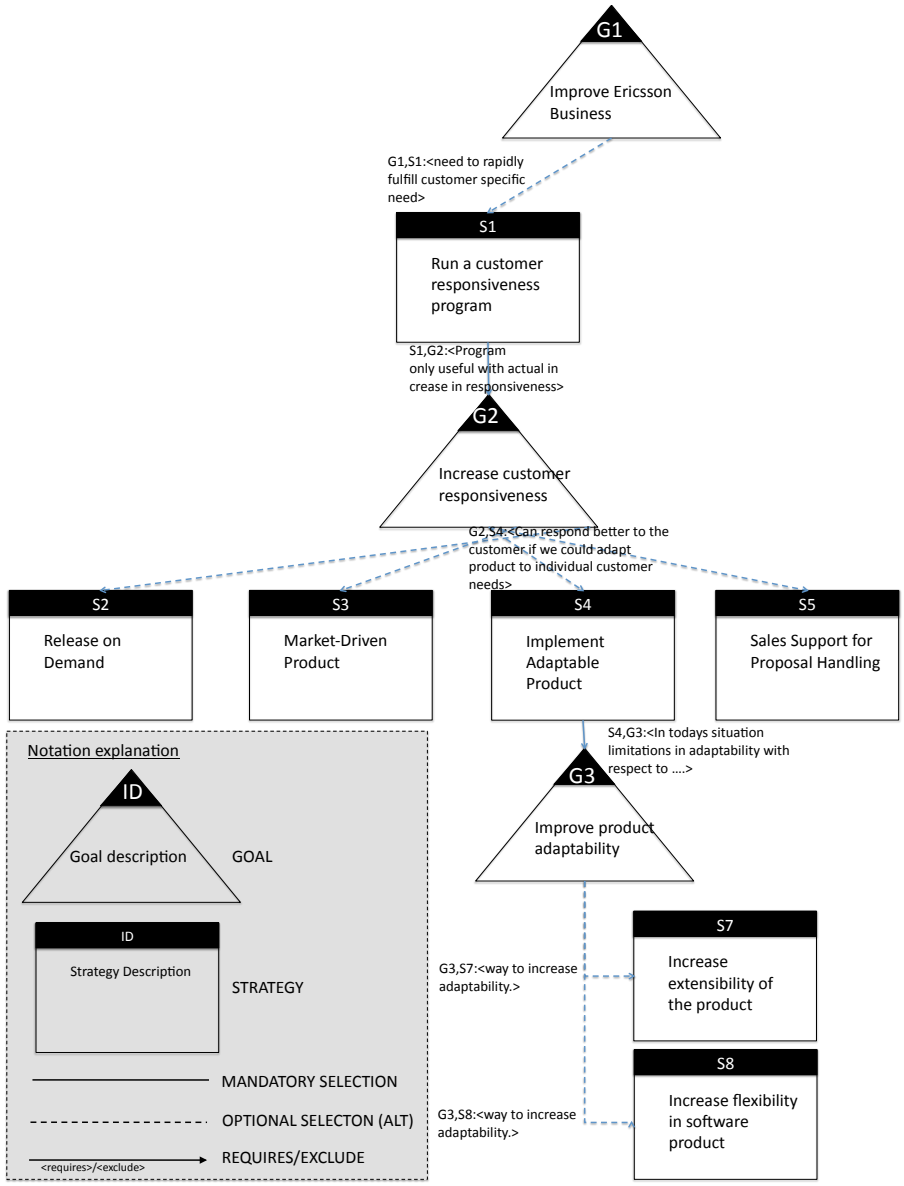

Fig. 3: GQM$^+$S-EI Notation

two strategies are really conflicting) they present assumptions and in the case of certainty context elements.

We used a semi-structured questionnaire to guide the interview as previous reviews [Davis et al., 2006, Dieste and Juzgado, 2011, Park et al., 1996] found interviews that were structured to be most effective technique for elicitation.

The semi-structured questionnaire used complementary to the notation was structured as follows:

- Introduction: We provided an initial skeleton with few goals and strategies already elicited from documentation (see Figure 3). We communicated that we would like to elicit further goals and strategies, and also that we would like to collect feedback about the elicitation approach used.
- Subject information collection: The interviewees were asked for their role in the organization, how long they have been at Ericsson, which areas (and for how long) they have been working in before, and their practical experience and education before joining Ericsson.
- Notation introduction: The notation (as explained above) and the purpose of GQM$^+$Strategies are introduced.
- Application of GQM$^+$S-EI v. 1.0: The elicitation questions asked include:
  1. Are the goals, strategies, the rationals and the relations between them well defined? Please comment whether there are: (a) any extra goals and strategies that are needed?; (b) Are there any goals/strategies that need to be deleted or replaced?
  2. What are the magnitude and time-frame for the goals?
  3. Is there any information you think is important/missing for our purpose that you would like to add?
- Feedback collection: We collect feedback on the elicitation approach, as described earlier in the previous section.

During the process, the interviewees could follow from the model all updates based on the responses given by them as the changes were made directly during the interview sessions.

Action taking: Applying the approach as described above, the interviewer team had the roles as described in Table 4. During the first two interview sessions, the participants were confused by the concept of strategy, especially as we broke the goal down to sub-goals and went further down the tree of goals and strategies. One of the problems we faced during the first interview was that 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 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 am just thinking through.. What is different between a 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 one and a half hours, the interviewee was still trying to understand the goal and strategy concepts rather than defining them and interrelating to each other. The concept of strategy made interviewee stop from talking about the actual answers to the questions. 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 the concepts. As a consequence, communication took a long time, as the use of the concepts was not intuitive.

Reflection: The result of the first two interviews was surprising. In our mind, goal and strategy were clearly defined concepts. Two possible reasons that created confusion might be: (1) given that a goal on one level becomes a strategy on the next, and so forth appears to be confusing to non-experts in GQM$^+$Strategies. (2) Organizations have their own terminology and a meaning for goals and strategies, which are the definitions practitioners have in their minds. As a consequence, for our next design iteration, we aimed to design GQM$^+$S-EI v. 2.0 eliciting for strategies and goals without using these terms.

5.1.2 Iteration 2: GQM$^+$S-EI v. 2.0

Action planning: We formulated open questions in the interview to allow the interviewees to explain their ideas more freely. For example, instead of asking: *"What goals do you want to achieve with the improvement program?"* we asked: *"The improvement program is successful when: a)..., b)..., c)...; please complete the sentence"*. At the same time we linked the questions to rationales to make the purpose of the question traceable in the context of GQM$^+$Strategies. Table 8 provides an overview of the questions and rationales.

Table 8: Questions used in GQM$^+$S-EI v. 2.0

ID	Question	Rationale
1	What is <X> (X is, e.g. improvement program) about?	It gets the person talking freely. It is also a good start to understand the *context*.
2	Why was it initiated/used?	It provides the rationale for the existence of the program, further information about *context* and *assumptions*.
3	<X> is successful when: a)..., b)..., c)...; please complete the sentence.	This gives hints for *goals*, information needs and even measurement needed.
4	Who can influence the success of a, b, c,...	Identifies the relevant *stakeholders* when deploying and evaluating strategies.
5	What do you do to achieve a, b, c,...?	It identifies current *strategies* to be evaluated by the measurements to understand, control and improve.
6	In these activities that you just mentioned, which ones are working out good today?	Identifies *strategies* that work well and could be spread in the organization
7	Which ones have the highest improvement potential?	Identifies *strategies* that need to be improved (e.g. by changing, removing, or adding new strategies).
8	Do you have a plan to improve them? How?	Identifies further *strategies* that should be employed in the future, or changes to the existing strategies.

The process was followed according to Table 4, with Negin Asghari leading the interview, Cigdem Gencel being responsible for listening, interrupting and asking follow-up questions, and Kai Petersen documenting the outcome of the interview. In this iteration, the outcome was documented on a white-board, given that the interviewee was in the same room as the interviewers.

Action taking: Following the process above, the following observations have been made. The interviewees talked freely and fluently, and also did not have to ask for clarification of the questions, the answers indicated that the questions were easily understandable. The content of the transcription showed that the interviewee was much more focused on the content of the actual questions asked. In the second approach, we used the white-board 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 (s)he 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. We had to capture the information very fast, which led to poor handwriting, and at the same time the information was not so well structured.

Reflection: We concluded that the interviewee talked more freely and answered the questions without being constrained by the enforced concepts and/or notation. It was beneficial to capture information on a white-board during the interview as it allowed the interviewee to understand how we interpreted the information. Hence, we concluded that the interview documentation should be visible to the interviewee during the session. This also allowed to conduct member checking at 'real-time', which was at the same time problematic to keep the pace.

5.1.3 Iteration 3: GQM$^+$S-EI v. 3.0

Action planning: In the new version, the same questionnaire was used as we did not identify any reason for a change. However, we decided to change the role of the scribe when documenting the information in order to pace the interview, stopping it and making summaries. Another change was that the sessions could also take place using video conferencing. Hence, instead of using the white-board that becomes difficult to manage during a long interview, a mind-map tool was planned to be used on a shared screen.

Action taking: To avoid any bias and confirm the applicability of the last version of GQM$^+$S-EI we applied GQM$^+$S-EI v. 3.0 to two other different programs: "Cost of Poor Quality" and 'Performance Measurement'. During the interview for 'Cost of Poor Quality' the interviewee was available online. Three observers were present during the interview to learn about how to conduct it. The observers then took the roles of the measurement analysis (see Table 4) during the interview for the 'Performance Measurement'. In particular, measurement analyst 1 in Table 5 took the role of the lead interviewee and measurement analyst 2 took the role of the scribe. Also, Kai Petersen had

the role of the scribe to later compare notes. The other researchers had the support role and made observations.

One noteworthy observation was that during the interviews the participants often referred to measures they were already using. Furthermore, during the interviews the practitioners elaborated on the issue that measurements might drive wrong behavior, and how they could be tricked. 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. After the interviews we showed the results to the interviewees to confirm the captured information was correct. They confirmed the accuracy of the content and found this approach useful and easy to use. Comparing the notes of the two scribes, one version was keyword based, while the other version was using more detailed phrases.

Reflection: In contrast to using white-boards, managing information collected through mind-map appeared to be much easier and convenient for both the analysts. For example, there was no worry about losing or missing information. Furthermore, it was possible to navigate through the information and make it easier for the interviewee to check the collected data. This also added in readability and allowed to structure and provide links between GQM$^+$Strategies elements. Hence, even when conducting the interview co-located, we suggested the use of a mind-map tool. Regarding taking of notes during the interviews, we recommended usage of phrases instead of short notes, as after a longer period of time the mind-maps could still be read and understandable by persons not being present at the interviews. As we noticed that practitioners talked about the current measures collected, we introduced a question related to that. Furthermore, we introduced a question to the behavior that is driven by using the measurements, and how they could be tricked. This was motivated by the observation that the practitioners elaborated on possibilities of driving wrong behaviors through measurements (see Table 10). These led us to the GQM$^+$S-EI release version that has been evaluated later on in the case study.

5.1.4 GQM$^+$S-EI Release

For the release version, we developed a elicitation guideline to be used by the measurement analysts in the company. Below, we provide the content of this tutorial.

Part I: Introduction: In this part, the following information is given to the measurement analyst:

1. *Some general preparations and tips for making interviews with stakeholders:* Here we describe what is needed prior to the interview, and some additional hints for preparation.
   - Install MindMap tool (e.g. FreeMind) to your computer(s) for capturing what the interviewee is saying. 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, that is, 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.
- It is useful to have a prepared basic structure of the MindMap ready where branches are created for the different questions of the interview in Table 10. During the documentation process it is important to document the information in a way that it could be understood by others, i.e. phrases or sentences should be used when documenting what the interviewee is saying. Only documenting keywords from elaborate explanations should be avoided. An example MindMap is presented in Appendix A.

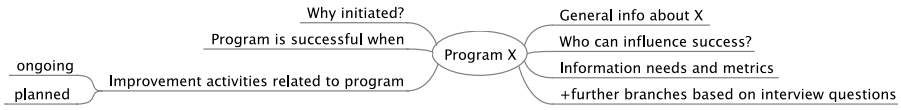

Fig. 4: Mind-Map Skeleton

- 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: The support person is responsible for listening and keeping track of follow-up questions needed. The support 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: The documenter is responsible for capturing and recording the information provided by the interviewee using the mind-map tool. The documenter starts with an empty skeleton of the MindMap with different branches for the questions so that these can be directly filled in. The documenter should be given enough time to summarize the answers to the questions.
3. *Context:* Provide the following information to the interviewee: Purpose of the interview and the process of the interview as described below.

- 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.
- 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 parts: a) Questions related to Interviewee background (Table 9), and b) Questions to capture GQM$^+$-Strategies related information (Table 10).

Table 9: Gathering Information about the Stakeholder

ID	Question	Rationale
1	Name and contact information	Allows for source traceability and collecting further feedback
2	How long have you been working in this organization?	Gives some hints about domain knowledge, process knowledge, etc.
3	What areas have you been in before and how long have you been working in?	Gives some hints about domain knowledge, process knowledge, other perspectives, etc.

5.2 Phase II: Evaluation of the Elicitation Instrument

The individual interviews are integrated into a single grid representing the different viewpoints of the interviewees. The interview structure that is represented in the mind-map is structured along GQM$^+$Strategies elements and allows to explicitly link goals and strategies. For example, when we asked *"<X> is successful when: a)... b)... c)"*, for a, b, and c, we also ask how they are achieved (strategies).

During the actual process of elicitation the documenter creates sub-branches representing different levels of GQM$^+$Strategies. This is done either by directly asking for a refinement (e.g. what does this goal comprise of), or by interpretation. Given that the interviewees can observe the documentation process directly, they can correct the interpretation in case it is wrong. Having several

Table 10: Questions to capture GQM+Strategies related information

ID	Question	Rationale
1	What is <X> (X is, e.g. improvement program) about?	It gets the person talking freely. It is also a good start to understand the *context*.
2	Why was it initiated/used?	It provides the rationale for the existence of the program, further information about *context* and *assumptions*.
3	<X> is successful when: a)..., b)..., c)...; please complete the sentence.	This gives hints for *goals*, information needs and even measurement needed.
4	What information is needed to evaluate success?	Makes required information needs, indicators, and measures explicit.
5	Who can influence the success of a, b, c,...	Identifies the relevant *stakeholders* when deploying and evaluating strategies.
6	What do you do to achieve a, b, c,...?	It identifies current *strategies* to be evaluated by the measurements to understand, control and improve.
7	In these activities that you just mentioned, which ones are working out good today?	Identifies *strategies* that work well and could be spread in the organization
8	Which ones have the highest improvement potential?	Identifies *strategies* that need to be improved (e.g. by changing, removing, or adding new strategies).
9	Do you have a plan to improve them? How?	Identifies new *strategies* that should be employed in the future, or changes to the existing strategies.
10	What information do you need to evaluate whether proposed improvement strategies are successful?	Identifies additional information needs and success indicators for concrete improvement ideas/actions)
11	What are you collecting today (measurements)?	Identifies current state of measurement
12	What behavior will the measurements (currently collected/planned to collect) drive?	Figures out whether the measures will really drive the desired improvement actions, or any other kind of behavior not leading to a benefit.
13	How could the measurements be tricked?	Helps achieve robust and accepted measurements that show "reality".

mind-maps their information has to be integrated. The first interview is used as a starting point. From that interview the initial grid is created based on the mind-map by incrementally adding the elements based on the questions in Table 10. We started with the answers to question 3. (giving the goals), followed by 6., 7., and 8. (giving the strategies). Then we added in the information needs of question 4., 10., and the metrics (question 11.). In the next step the second interview is taken and all new information is added one by one to the grid of interviewee 1, and so forth. We color coded the new information added for each interview, and it is also recommended to keep traceability from each element to the interviewees raising it as this aids in negation.

Table 11 shows the number of GQM+Strategies elements that have been collected during the eight interviews. The absolute number, and the % of all elements collected are also shown. In total, 98 GQM+Strategies elements have been identified.

Table 11: Case Study Interviewees

Element	No. identified	Percentage
Number of business goals	1	1
Strategies related to business goals	4	4
Lower level goals	4	4
Strategies related to lower level goals	18	19
Measurement goals	19	20
Information needs (questions)	29	29
Metrics	23	23
Total	98	100

5.2.1 Completeness

Figure 5 provides a structural overview of the grid derived. Due to confidentiality reasons of the case company, the complete list of goals and strategies are not mentioned. Different colors indicate the additional information provided by the interviews. At the bottom right corner of the Figure the sequence of interviews in relation to the colors is shown (e.g. white color is the first interview, yellow color the second interview, etc.). Diamonds represent strategies, and boxes represent goals on different levels, as well as questions and metrics. What is immediately noticeable in the grid is that the business goal, lower level goals, and strategies have been largely covered already after two interviews. Only to derive the measurement goals, questions, and metrics further interviews were required.

To make the added information for each type of GQM$^+$Strategies more explicit, Figure 6 provides the cumulative number of each GQM$^+$Strategies element as listed in Table 11. The x-axis shows the sequence of interviews. The y-axis shows the cumulative number of GQM$^+$Strategies elements according to type. It is visible that the saturation occurs after the Interview 6 for all element types, with only very few elements being added by the last interviewee in relation to the overall number collected. Furthermore, the higher in the GQM$^+$Strategies hierarchy, the earlier the saturation occurs. For business goals, strategies related to business goals, and lower level goals saturation occurred after the second interview. Similar observations apply to strategies related to lower level goals and measurement goals with only few additions in comparison to what has been added during Interview 1 and Interview 2. Relatively many new items were added for information needs and metrics until Interview 6.

Overall, the result indicates saturation, and hence provides a positive indication that the elicitation approach is completely capturing GQM$^+$Strategies information from the stakeholders.

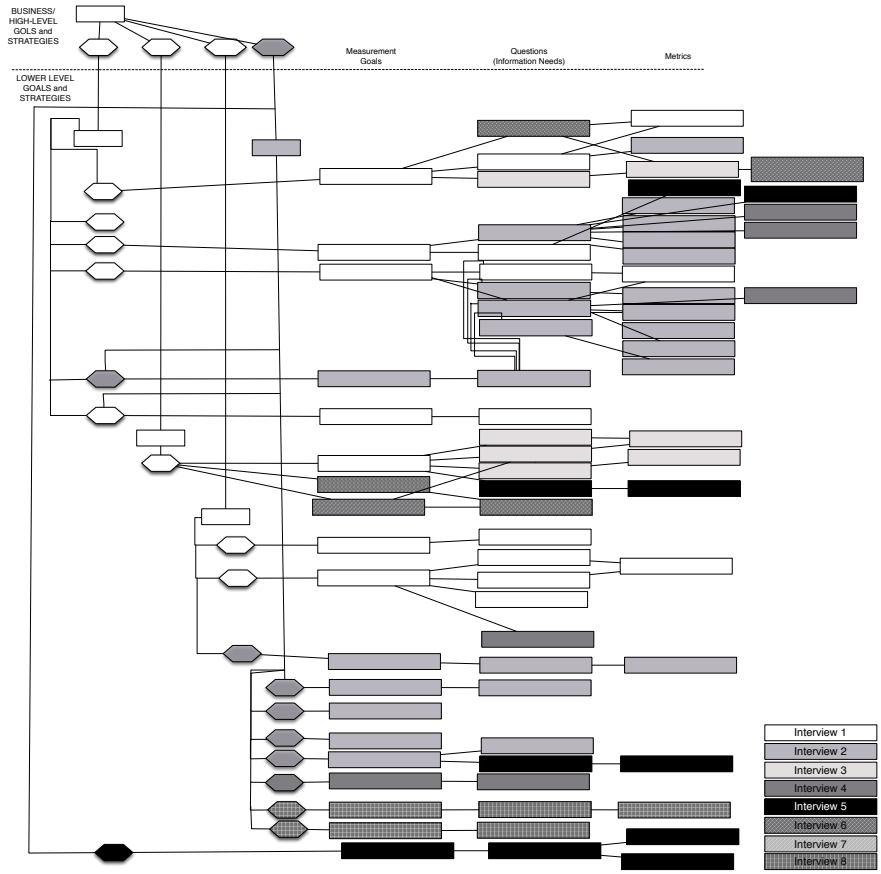

Fig. 5: GQM+Strategies Grid

5.2.2 Accuracy

Overall, five changes have been requested by the practitioners. All changes requested were renaming (one goal and four strategies were requested to be renamed). The following changes were made:

- the strategy "Improve practices and processes" was reformulated to "Continuously improve processes & methods".
- the strategy "Reduce cost" was reformulated to "Achieve cost accuracy".
- the strategy "Make realistic commitments" was reformulated to "Manage uncertainties".
- the strategy "Deliver quality" was reformulated to "Deliver accuracy and quality".

The essence of the strategy has been preserved, and hence no changes in the structure were needed for the GQM+Strategies elements. As an exam-

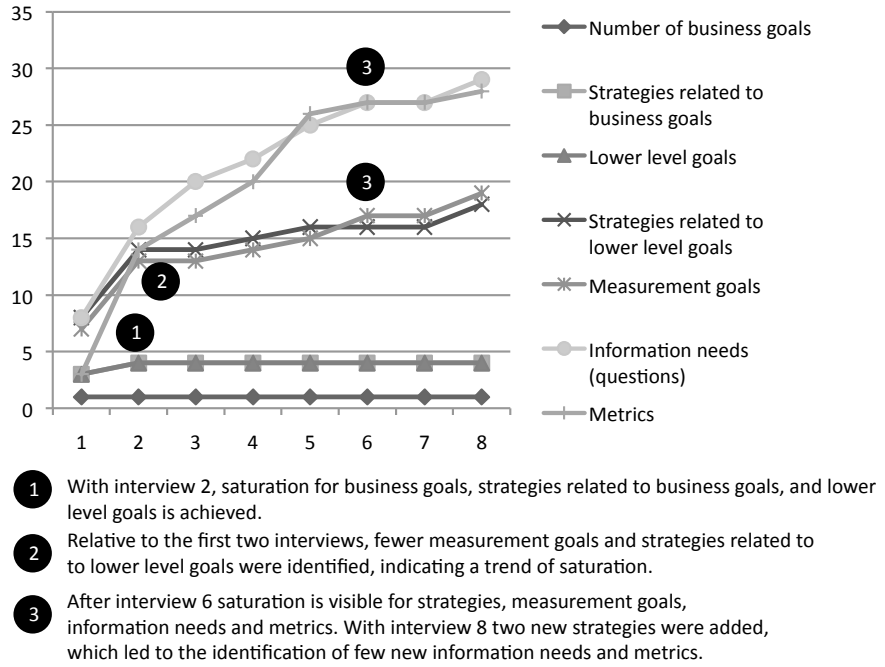

Fig. 6: Saturation in Interviews

ple, "make realistic commitments" means to be dealing with uncertainties and estimating them (e.g. risks) in a good way.

A noteworthy change was the reformulation of the business goal from being related to being profitable to value creation, which was is in-line with the lean idea, focusing all activities on value [Poppendieck, 2003]. That is, profit is seen as a monetary value, while in lean the value terminology is wider (e.g. value for the customer, internal value related to efficiency, and so forth). The term was found to subsume the GQM$^+$Strategies below the business goal well. Similar observations are reported in [Mandic et al., 2010a].

Possible changes are to move (restructure), add, change (e.g. reformulate), or delete elements (Move-Add-Change-Delete [Finke and Hartmann, 2011]). Only changes in the form of reformulations were requested (change). The main change was the reformulation of the business goal. No modifications in structure, additions, or deletions of GQM$^+$Strategies elements were requested. Overall, this supports the accuracy of the interpretations.

5.2.3 Observed side effects

These are observations that have not been incorporated in the research questions earlier, but are noteworthy to mention.

Learning from each other: As pointed out by the Ericsson representatives participating in the review team, capturing the strategies (those that work well

and those that do not) across business unit boundaries allowed to learn from other units. In particular, for a strategy not working well today, another unit might provide a strategy that could help to improve the situation. Hence, the elicitation approach facilitates knowledge transfer relevant for concrete process improvement actions, not just for measurement program planning. This also leads to an aligned understanding of goals and strategies the organization is performing, which is essential in further negotiation processes.

Reflections on cost: The Swedish funding agency KKHög [1] calculates costs for practitioners participating in research projects with 800 SEK per hour (approx. 120 USD or 90 Euro). Eight interviews have been conducted with three interviewers with seven interviews individually, and one interview having two interviewees. Hence, the pure conduct of the interviews with the core team cost (8* (3+7)+2) *800 SEK = 65600 SEK. In addition, the walkthrough meeting with 13 people participating created a cost of 83200, overall creating a cost 148800 SEK. Overall, this is a low cost seen from the improvement program and measurement stakeholders point of view. On the other hand, substantial effort was required by the measurement analysts. However, we did not record the time spent working on analysis and reflection, hence we are not able to provide a number.

6 Discussion

Prior to using GQM+S-EI we investigated company documentation to gain an understanding of the company. This part was essential as part of the research (document analysis being an important part of triangulation), but should be considered as an important part of the elicitation approach itself.

During the iterative design of the elicitation instrument, different alternatives have been explored. The template-based approach is sequential in adding GQM+Strategies element and does not provide an overall picture connecting the elements. The notation based approach directly focused the interviewees on the elements (goals, strategies, etc.) and their meaning. In particular, it was confusing when a strategy on one level becomes a goal on the next level, etc. Hence, there was a need to elicit information to identify the elements without using the GQM+Strategies terminology for the identified reasons, and at the same time provide a structural overview. For this purpose, mind-mapping served best. During the interviews, further information was added sequentially. While the process was initially focused on strategies and goals, it was expanded to incorporate information needs, and metrics as well. Overall, this indicates that, it is essential to allow the practitioners to talk freely in their own terminology during elicitation, and later have the measurement analysts, who are the experts in the concepts, conduct the mapping. During the process of structuring the grid, notation based approaches appear to be most valuable.

During the case study, we found out that our elicitation instrument was capable of accurately and completely eliciting goals and strategies. To determine

[1] http://www.kks.se/

the completeness and accuracy of information elicited we used saturation and member checking as measures. We saw that, higher level goals and strategies were elicited early on with very few interviews. More interviews were needed to reach a saturation state at the lower levels. Overall, this indicates that the elicitation instrument can quickly lead to a comprehensive picture. This result has been obtained by interviewing people from different business units, and also different roles and perspectives that took part in the same improvement program. Thus, in a smaller scale saturation might be reached sooner. An interesting research direction for the future is to investigate how many interviews are needed in different contexts (scale, domain, or processes, see e.g. [Petersen and Wohlin, 2009b]) as this provides interesting insight when estimating the effort for measurement program planning. The result of such a study might vary also depending on the persons involved. When starting with persons on the project level high level goals might be missing early on. In case no high level goals are elicited, we would propose to group existing goals and suggest high level goals (e.g. measurement experts might have experience allowing them to do that), and negotiate them with the measurement stakeholders involved. The case study was also conducted in a situation where no GQM$^+$Strategies grid has been available prior to the case study. For a case with an existing grid we suggest to conduct the interviews as described and complement the existing grid with newly obtained information (see Section 4.3.5 for a description of how to systematically add information). When looking at how much new information is added with new interviews it will become visible whether the existing grid still reflects the situation of the company.

The interview design was geared towards GQM$^+$Strategies, containing questions explicitly focused on strategies. However, it also subsumes information required by other measurement programs (see related work in Section 3). Hence, the solution presented is not just a GQM$^+$Strategies solution, but a solution for all goal-based measurement program planning approaches. More information was collected, but this was not considered an issue as the actual elicitation from the stakeholders appears to be cost-effective.

The cost-effectiveness is also positively influenced by the way that the instrument is designed, i.e. only the measurement analysts need training in the concepts of GQM$^+$Strategies by asking for the information using more intuitive questions allowing to talk practitioners freely (see Section 5.1.4). However, the measurement analysts need to have very good knowledge as they have to transfer the information from the interview into the GQM$^+$Strategies grid, which requires the accurate identification of GQM$^+$Strategies elements.

When having elicited all information there is a need to prioritize and take decisions which metrics to collect. For that purpose, it is important to assure traceability back to the interviewees, which can be directly done in the GQM$^+$Strategies grid by color coding it according to sources (see Figure 5).

We also would like to highlight that it was essential to have multiple people involved during the interviews as well as when interpreting and structuring the GQM$^+$Strategies grid. As recommended by [Keele, 2007] it is useful to involve multiple persons in tasks where interpretation is required to increase the

quality, and hence trust in the end-result. Member checking based on a walkthrough was used to check whether the instrument accurately and completely captured the information for the set of people interviewed. The group should agree on a version of the grid; hence walkthroughs could become part of the elicitation activity in practical usage also (not just for research purposes) to achieve a shared understanding. In the next step, the practitioners have to ask questions such as: "Which goals are important?", "Which strategies are best suited to achieve the goals?", and "Which metrics (based on the previous two answers) should be collected?". Having provided an instrument to accurate instrumentation we recommend that negotiation and review should be a continuous process when adopting the instrument.

7 Conclusion

This research focused on developing an elicitation instrument to completely and accurately identify GQM+Strategies elements from measurement stakeholders, and then from that build GQM+Strategies grid. The research was conducted in two phases. In the first phase, the elicitation instrument was iteratively developed using action research. In the second phase, the final version of the instrument was evaluated with respect to its ability to accurately and completely identify GQM+Strategies elements from stakeholders. Two research questions have been investigated in these two phases:

Action research question (RQ1): How should software organizations elicit goals and strategies as input for GQM+ Strategies? After having studied documentation to familiarize with the improvement programs and company's terminology, six interviews were needed to identify a stable and workable elicitation instrument. Here are the main findings:

- Semi-structured interviews using mind-mapping is preferred over templates and notations for elicitation purposes. Templates require deep knowledge about measurement terminology and information is added sequentially in tables hindering to get an overall overview. Notations also require knowledge about the measurement concepts. Questions not using the terminology when eliciting (e.g. goals and strategies) fared better as practitioners focus on providing insights of the improvement program in their domain during the interviews rather than struggling with the terminology.
- Notations and templates are important from the measurement analysts' point of view when translating the mind-maps into GQM+Strategies elements and grids.
- Multiple roles are required for the interviews when applying the elicitation instrument. The interview leader moderates the interview and asks the questions, the support interviewer interrupts and asks follow-up questions, and the scribe documents the results in the mind-map tool visible on a screen to the interviewee. That way a "real-time" member-checking is also conducted.

– While the interview initially focused on eliciting and structuring goals and strategies, it was expanded to incorporate all elements needed to construct the GQM$^+$Strategies grid.

Case study research question (RQ2): How accurate and complete is GQM$^+$S-EI in capturing GQM$^+$Strategies relevant information from stakeholders? Saturation (measure of completeness) occurred after the sixth of eight interviews, while saturation occurred earlier on for the high level GQM$^+$Strategies elements. This indicates that the elicitation approach achieves good results in completeness. With respect to accuracy (member checking was used to measure) only few changes were requested, which were primarily related to reformulations, no structural changes or additions/deletions were needed. The main change was to reformulate the business goal to better reflect the value focus of lean software development (which was the focus of the measurement program). Hence, we conclude that the designed elicitation instrument accurately reflected the views of the stakeholders.

A number of future research studies would be of interest. First, the elicitation instrument should be investigated in different contexts (e.g. small and medium sized scales). This allows gaining an understanding of how much effort is required for gathering the data in measurement programs in different contexts (e.g. with small scale saturation might occur much earlier). Another interesting direction is how to structure the elicited information. For instance, for elicitation we proposed to utilize product line concepts, which was not applicable at that stage. However, to be able to analyze a measurement program with respect to consistency, product line concepts could be useful in later stages.

Acknowledgements This work has been conducted as part of an Ericsson Software Research (ESR) funded project; GQM-Lean- Establishing a Lean and Sustainable Software Measurement Program. The work also funded partially by ELLIIT, the Strategic Area for ICT research, funded by the Swedish Government. We also would like to thank the measurement analysts who have worked with us in close collaboration, and all the interviewees that participated in this research.

References

[Ali et al., 2012] Ali, N. B., Petersen, K., and Mäntylä, M. (2012). Testing highly complex system of systems: an industrial case study. In *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM 2012)*, pages 211–220.

[Ardimento et al. 2006] Ardimento, P., Baldassare, M. T., Cimitile, M., and Visaggio, G. (2006). Assessing multiview framework (MF) comprehensibility and efficiency: A replicated experiment. *Information and Software Technology*, 48(5):313–322.

[Baca and Petersen, 2013] Baca, D. and Petersen, K. (2013). Countermeasure graphs for software security risk assessment: An action research. *Journal of Systems and Software*, 86(9):2411–2428.

[Basili et al., 2007] Basili, V. R., Heidrich, J., Lindvall, M., Münch, J., Regardie, M., Rombach, H. D., Seaman, C. B., and Trendowicz, A. (2007). Bridging the gap between business strategy and software development. In *Proceedings of the International Conference on Information Systems (ICIS 2007)*, page 25.

[Basili et al., 2010] Basili, V. R., Lindvall, M., Regardie, M., Seaman, C. B., Heidrich, J., Münch, J., Rombach, H. D., and Trendowicz, A. (2010). Linking software development and business strategy through measurement. *IEEE Computer*, 43(4):57–65.

[Basili and Weiss, 1984] Basili, V. R. and Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Trans. Software Eng.*, 10(6):728–738.

[Boyd, 2002] Boyd, A. (2002). The goals, questions, indicators, measures (gqim) approach to the measurement of customer satisfaction with e-commerce web sites. In *Aslib Proceedings*, volume 54, pages 177–187. MCB UP Ltd.

[Boyd, 2005] Boyd, A. J. (2005). The evolution of goal-based information modelling: literature review. In *Aslib proceedings*, volume 57, pages 523–538. Emerald Group Publishing Limited.

[Briand et al., 1996] Briand, L. C., Differding, C. M., and Rombach, H. D. (1996). Practical guidelines for measurement-based process improvement. *Software Process Improvement and Practice*, 2(4):253–280.

[Brydon-Miller et al., 2003] Brydon-Miller, M., Greenwood, D., and Maguire, P. (2003). Why action research? *Action research*, 1(1):9–28.

[Bühne et al., 2005] Bühne, S., Lauenroth, K., and Pohl, K. (2005). Modelling requirements variability across product lines. In *Proceedings of the 13th IEEE International Conference on Requirements Engineering (RE 2005)*, pages 41–52.

[Caldiera and Rombach, 1994] Caldiera, V. R. B. G. and Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of software engineering*, 2(1994):528–532.

[Colombo, 2003] Colombo, M. (2003). Reflexivity and narratives in action research: A discursive approach. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, volume 4.

[Damm et al., 2006] Damm, L.-O., Lundberg, L., and Wohlin, C. (2006). Faults-slip-through - a concept for measuring the efficiency of the test process. *Software Process: Improvement and Practice*, 11(1):47–59.

[Davis et al., 2006] Davis, A. M., Tubío, Ó. D., Hickey, A. M., Juzgado, N. J., and Moreno, A. M. (2006). Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review. In *Proceedings of the 14th IEEE International Conference on Requirements Engineering (RE 2006)*, pages 176–185.

[Dieste and Juzgado, 2011] Dieste, O. and Juzgado, N. J. (2011). Systematic review and aggregation of empirical studies on elicitation techniques. *IEEE Trans. Software Eng.*, 37(2):283–304.

[Finke and Hartmann, 2011] Finke, J. S. and Hartmann, D. (2011). *Implementing Cisco Unified Communications Manager, Part 1 (CIPT1) Foundation Learning Guide:(CCNP Voice CIPT1 642-447)*. Cisco Press.

[Gencel et al., 2013] Gencel, C., Petersen, K., Mughal, A. A., and Iqbal, M. I. (2013). A decision support framework for metrics selection in goal-based measurement programs: GQM-DSFMS. *Journal of Systems and Software, in print*.

[Graham, 2009] Graham, D. (2009). *Foundations of software testing : ISTQB certification*. Course Technology Cengage Learning, Hampshire, rev. ed. edition.

[Guest et al., 2006] Guest, G., Bunce, A., and Johnson, L. (2006). How many interviews are enough? an experiment with data saturation and variability. *Field methods*, 18(1):59–82.

[ISO/IEC15939, 2002] ISO/IEC15939 (2002). Iso/iec 15939 international standard 1st edition 2002: Software engineering - software measurement process. reference number iso/iec 15939:2002(e)(2002).

[Ivarsson and Gorschek, 2011] Ivarsson, M. and Gorschek, T. (2011). A method for evaluating rigor and industrial relevance of technology evaluations. *Empirical Software Engineering*, 16(3):365–395.

[Iversen et al., 2004] Iversen, J. H., Mathiassen, L., and Nielsen, P. A. (2004). Managing risk in software process improvement: an action research approach. *Mis Quarterly*, 28(3):395–433.

[Jürgen Münch, 2013a] Jürgen Münch, Fabian Fagerholm, P. K. M. P. J. P. (2013a). The effects of gqm+strategies on organizational alignment. In *Proceedings of the DASMA Software Metric Congress (MetriKon 2013)*, Santander, Spain.

[Jürgen Münch, 2013b] Jürgen Münch, Fabian Fagerholm, P. K. M. P. J. P. (2013b). Experiences and insights from applying gqm+strategies in a systems product development

organisation. In *Proceedings of the 39th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA 2013)*, Kaiserslautern, Germany.

[Kaneko et al., 2011] Kaneko, T., Katahira, M., Miyamoto, Y., and Kowalczyk, M. (2011). Application of gqm+strategies® in the japanese space industry. In *IJoint Conf of 21st Int'l Workshop on Software Measurement and the 6th Int'l Conference on Software Process and Product Measurement (IWSM/Mensura 2011)*, pages 221–226.

[Kaplan et al., 1996] Kaplan, R., Kaplan, R. S., and Norton, D. P. (1996). *The balanced scorecard: translating strategy into action.* Harvard Business Press.

[Keele, 2007] Keele, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, EBSE Technical Report EBSE-2007-01.

[Kilpi, 2001] Kilpi, T. (2001). Implementing a software metrics program at nokia. *IEEE Software*, 18(6):72–77.

[Mandic and Basili, 2010] Mandic, V. and Basili, V. (2010). An approach for evaluating business goals. Technical report, Technical Report TR TR-TOL-2010-2802, University of Oulu, Department of Information Processing Science.

[Mandic et al., 2010a] Mandic, V., Basili, V. R., Harjumaa, L., Oivo, M., and Markkula, J. (2010a). Utilizing gqm+strategies for business value analysis: an approach for evaluating business goals. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement (ESEM 2010)*.

[Mandic et al., 2010b] Mandic, V., Harjumaa, L., Markkula, J., and Oivo, M. (2010b). Early empirical assessment of the practical value of gqm$^+$strategies. In *Proceedings of the International Conference on Software Process (ICSP 2010)*, pages 14–25.

[Mandic and Oivo, 2010] Mandic, V. and Oivo, M. (2010). Sas: A tool for the gqm$^+$strategies grid derivation process. In *11th International Conference on Product-Focused Software Process Improvement (PROFES 2010)*, pages 291–305.

[McGarry et al., 2002] McGarry, J., Card, D., Jones, C., Layman, B., Clark, E., Dean, J., and Hall, F. (2002). *Practical software measurement: objective information for decision makers.* Addison-Wesley Boston.

[Mendonça and Basili, 2000] Mendonça, M. G. and Basili, V. R. (2000). Validation on an approach for improving existing measurement frameworks. *IEEE Trans. Software Eng.*, 26(6):484–499.

[Morse, 1994] Morse, J. M. (1994). Designing funded qualitative research.

[Mujtaba et al., 2011] Mujtaba, S., Feldt, R., and Petersen, K. (2011). Analyzing strategy and processes for product customization in large-scale industrial settings. In *Proceedings of the 37th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA 2011)*, pages 369–373.

[Münch and Heidrich, 2008] Münch, J. and Heidrich, J. (2008). Aligning business strategies with software measurement - exercise handouts.

[Offen and Jeffery, 1997] Offen, R. J. and Jeffery, D. R. (1997). Establishing software measurement programs. *IEEE Software*, 14(2):45–53.

[Park et al., 1996] Park, R. E., Goethert, W. B., and Florac, W. A. (1996). Goal-driven software measurement?a guidebook (cmu/sei-96-hb-002, ada313946). pittsburgh, pa.: Software engineering institute.

[Petersen and Gencel, 2013] Petersen, K. and Gencel, C. (2013). Worldviews, research methods, and their relationship to validity in empirical software engineering research. In *Proceedings of the Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement (IWSM-Mensura 2013)*.

[Petersen and Wohlin, 2009a] Petersen, K. and Wohlin, C. (2009a). A comparison of issues and advantages in agile and incremental development between state of the art and an industrial case. *Journal of Systems and Software*, 82(9):1479–1490.

[Petersen and Wohlin, 2009b] Petersen, K. and Wohlin, C. (2009b). Context in industrial software engineering research. In *Proceedings of the Third International Symposium on Empirical Software Engineering and Measurement (ESEM 2009)*, pages 401–404.

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

[Petersen and Wohlin, 2010b] Petersen, K. and Wohlin, C. (2010b). Software process improvement through the lean measurement (spi-leam) method. *Journal of Systems and Software*, 83(7):1275–1287.

[Petersen and Wohlin, 2011] Petersen, K. and Wohlin, C. (2011). Measuring the flow in lean software development. *Softw., Pract. Exper.*, 41(9):975–996.

[Poppendieck, 2003] Poppendieck, M. (2003). *Lean software development: an agile toolkit.* Addison-Wesley Professional.

[Runeson and Höst, 2009] Runeson, P. and Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164.

[Runeson et al., 2012] Runeson, P., Host, M., Rainer, A., and Regnell, B. (2012). *Case study research in software engineering: Guidelines and examples.* Wiley. com.

[Sarcià, 2010] Sarcià, S. A. (2010). Is gqm+strategies really applicable as is to non-software development domains? In *roceedings of the International Symposium on Empirical Software Engineering and Measurement (ESEM 2010)*.

[Scholtes et al., 2003] Scholtes, P. R., Joiner, B. L., and Streibel, B. J. (2003). *The team handbook.* Oriel Incorporated.

[Tahir and Jafar, 2011] Tahir, T. and Jafar, A. (2011). A systematic review on software measurement programs. In *Proceedings of Frontiers of Information Technology (FIT 2011)*, pages 39–44. IEEE.

[Trendowicz et al., 2011] Trendowicz, A., Heidrich, J., and Shintani, K. (2011). Aligning software projects with business objectives. In *Joint Conf of 21st Int'l Workshop on Software Measurement and the 6th Int'l Conference on Software Process and Product Measurement (IWSM/Mensura 2011)*, pages 142–150.

[van der Linden et al., 2007] van der Linden, F. J., Schmid, K., and Rommes, E. (2007). *Software product lines in action.* Springer.

[Van Latum et al., 1998] Van Latum, F., Van Solingen, R., Oivo, M., Hoisl, B., Rombach, D., and Ruhe, G. (1998). Adopting gqm based measurement in an industrial environment. *Software, IEEE*, 15(1):78–86.

[Wilson et al., 2002] Wilson, H., Daniel, E., and McDonald, M. (2002). Factors for success in customer relationship management (crm) systems. *Journal of marketing management*, 18(1-2):193–219.

[Yin, 2011] Yin, R. K. (2011). *Applications of case study research.* Sage.

A Mind-Map Example

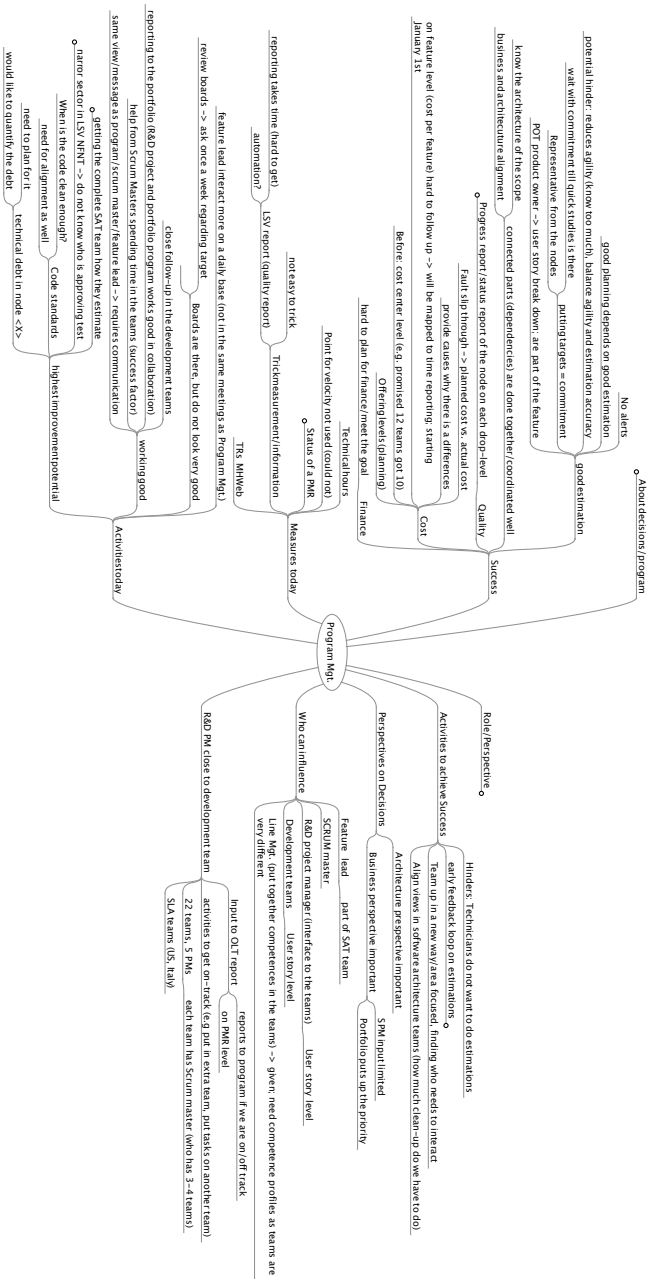

Fig. 7: Mind-Map Example

Kai Petersen is a senior researcher at Blekinge Institute of Technology (BTH), Sweden. He received his PhD from BTH in 2010. His research focuses on software processes, software metrics, lean and agile software development, quality assurance, and software security in close collaboration with industry partners. Kai has authored over 30 articles in international journals and conferences, and has been the industry chair of REFSQ 2013.

Cigdem Gencel is a senior researcher at the Faculty of Computer Science of the Free University of Bolzano, Italy. She received her PhD from the Middle East Technical University in Turkey in 2005. Her research focus is providing novel solutions to real industrial challenges, particularly in the areas of software size and effort estimating, software measurement, software project management, requirements elicitation methods and process improvement. She is a member of the International Advisory Council of COSMIC and a member of the COSMIC Measurement Practices Committee.

Negin Asghari is a Software Developer at Ericsson - primarily at the Call Session Control Function, which is one of the IP Multimedia System-nodes. Before that she was working as a researcher in the same company, being passionate about software process improvement and lean software development, she spent most of her time on a project for Improving Customer Responsiveness. She has got a Bachelor in Computer Engineering from the University of Mashad, Iran and a Master of Science in Software Engineering from Blekinge Institute of Technology (Sweden). She is also a certified Microsoft System Administrator with two years working experience.

Stefanie Betz received her Diploma in Information Engineering and Management and her PhD from the Karlsruhe University, Germany. She is currently a research fellow in the Software Engineering Lab at Blekinge Institute of Technology, Sweden. Before, she was a research fellow at the Institute of Applied Informatics and Formal DescriptionMethods (AIFB), Faculty of Economics and Business Engineering, Karlsruhe Institute for Technology, Germany. She spent the autumn 2009 as an invited visiting researcher at the Enterprise Simulation Laboratory SimLab, Faculty of Information and Natural Sciences, Department of Industrial Engineering and Management, Aalto University Finland. Her research interests are in (global) software engineering and business process management.