TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Angewandte Softwaretechnik

Collaboration-based Metrics for Project Management Support in Global Software Engineering

Christian Lescher

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Johann Schlichter

Prüfer der Dissertation:

1. Univ.-Prof. Bernd Brügge, Ph.D.
2. Assoc. Prof. Daniela Damian, Ph.D.,
   University of Victoria, Victoria / Canada

Die Dissertation wurde am 24.05.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 30.07.2011 angenommen.

Abstract

Global Software Engineering (GSE) has brought new challenges to development teams: Geographic separation, working in different time zones and different cultures and languages make the collaboration of team members more difficult and can lead to project delays, cost overruns and quality problems. While project management metrics to monitor cost, time and quality characteristics of a development project, are state of the practice today, these metrics often reveal problems too late. The hypothesis of this dissertation is that collaboration-based metrics can help to detect and address problems in GSE projects early on. To validate our hypothesis we constructed a set of collaboration-based project management metrics for GSE, combining knowledge from social psychology and software measurement, and tested them in practice. We applied our GSE metrics in an industrial case study of a real-world crisis project as well as a set of GSE student projects, which gave anecdotal evidence that our GSE metrics are helpful to identify and resolve collaboration problems in GSE projects. The GSE metrics provided accurate and useful information to the project teams, while the effort for data collection was acceptable.

Kurzfassung

Globale Softwareentwicklung (GSE) birgt neue Herausforderungen für Entwicklungsteams: Geografische Verteilung, das Arbeiten in verschiedenen Zeitzonen sowie verschiedene Kulturen und Sprachen erschweren die Zusammenarbeit der Teammitglieder und können zum Terminverzug der Projekte, Kostenüberschreitungen und Qualitätsproblemen führen. Während Projektmanagement-Metriken zur Überwachung von Kosten-, Zeit- und Qualitätsaspekten eines Entwicklungsprojekts heute in der Praxis etabliert sind, so sind Probleme anhand dieser Metriken doch oft zu spät zu erkennen. Die Hypothese dieser Dissertation ist, dass auf der Messung von Zusammenarbeit basierende Metriken helfen können, Probleme in GSE-Projekten frühzeitig zu erkennen und zu adressieren. Zur Validierung unserer Hypothese haben wir einen Satz von kollaborationsbasierten Projektmanagement-Metriken konstruiert, wobei wir Wissen aus der Sozialpsychologie mit Software-Metriken kombiniert haben, und in der Praxis getestet. Wir haben unsere GSE-Metriken in einer industriellen Fallstudie eines realen Krisenprojekts sowie in mehreren GSE-Studentenprojekten angewandt, die anekdotisch erwiesen haben, dass unsere GSE-Metriken hilfreich sind, um Probleme in der Zusammenarbeit in GSE-Projekten zu identifizieren und zu lösen. Die GSE-Metriken lieferten korrekte und nützliche Informationen für die Projektteams bei vertretbarem Aufwand für die Datenerhebung.

Acknowledgments

I want to thank my dissertation advisor, Prof. Bernd Brügge, Ph.D. for his great support in my research. With his rich experience and brilliant ideas he gave me thought-provoking impulses without which this dissertation would not have been possible. I am also grateful to my second advisor, Prof. Daniela Damian, Ph.D. for her strong support and her feedback – starting from the doctoral symposium at ICGSE'09 to the finalization of my thesis.

I am grateful to all those involved in the case studies and experiments of my dissertation: First of all I would like to thank my Siemens colleagues for the great cooperation on the industrial case study. I thank the supervisors and participants of the NEREID projects, in particular Prof. Dr. Florian Matthes, Christopher Schulz and Christian Neubert. I want to thank the coaches and participants of the DOLLI4 project and the supervisors Helmut Naughton and Harald Stangl. I thank the students of my master seminar on Global Software Engineering in summer term 2010 for participating in the experiments.

I would like to thank all those who provided feedback on my thesis topic, including the panel and participants of the doctoral symposium at ICGSE'09 and RE'09, my colleagues at the Chair of Applied Software Engineering and Timo Wolf.

I am grateful to Monika Markl and Helma Schneider for the organizational support. I would also like to thank Jennifer Schiller who encouraged me in my decision to write this dissertation and introduced me to the team at TUM and for her good advice for the preparation of my first university seminar. I want to thank Prof. James D. Herbsleb, Ph.D. and Matthew Bass for the good cooperation in the past; it was a great opportunity to learn from their valuable experience in Global Software Engineering.

Finally, I would like to thank my family for their patience and support during the countless hours I spent on weekends, evenings and sometimes also early mornings for writing this dissertation.

Contents

Conventions

We capitalize identifiers such as class names (e.g. GSE Metrics), indicator names (e.g. Team Cohesion) and the term Global Software Engineering. We use American English (except in literal quotes).

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, they are used only for identification and explanation without intent to infringe.

This document contains colored figures and should be printed in color.

Chapter 1

Preface: The Need for Early Warning Indicators

Problems in globally distributed projects often become evident only very late in the project – during integration of parts developed at different sites, or even worse at the customer site. One such undesirable and consequential situation occurred when the new Terminal 5 of the airport London Heathrow was opened. On the first day of operation, several hundred flights were delayed or canceled, thousands of passengers slept at the airport and more than 10,000 bags were lost somewhere in the baggage handling system [KL08].

1.1 The Opening of Heathrow Terminal 5: A Case Study

Everything started off so well. The building was finished on time, and Queen Elizabeth II officially opened Terminal 5 (T5) in a ceremony on March 14, 2008. The airport opened for passengers on March 27, 2008, under the responsibility of British Airport Authority (BAA) as the owner and operator of the airport and British Airways (BA) as the sole airline occupying Terminal 5. The parties involved in the T5 project were so self-confident that they published their success stories already prior to the opening. In their article *Testing the Heathrow Terminal 5 Baggage Handling System – Before It Is Built* of March 2007 Roger Derksen, Huub van der Wouden and Paul Heath described the comprehensive testing concept of T5 with various levels of emulation and on-site testing, and concluded: “All of these benefits are helping BAA deliver a fully operational baggage system for Terminal 5, on time and on budget.” [DvdWH07]

Everything in T5 was thought to be foreseen, as described in the book *Heathrow's Terminal 5: History in the Making* by Sharon Doherty, a “HR and organisational effectiveness director for Heathrow airport and Terminal 5”, who “was accountable for the approach to people management and organisational change on Terminal 5" [Doh08]:

> The operational readiness team has three opening events. The Queen will cut the ribbon prior to the opening date, there is an event on the actual day of the opening and then the Saturday afterwards the Olympic torch will pass through T5 and the National Lottery will be hosted in the building. In preparation over three years, a joint BAA and BA team has worked to ensure that all of the people, processes and systems will be ready and working in the new facility on day one. [Doh08, p.7]

The 'observed behavior' on the day of the opening quite differed from the 'specified behavior' mentioned above. What was intended to be an object of national pride, happened to be – in the wording of BA Chief executive Willie Walsh – a "national embarrassment for years" [lon08]. Indeed the bad reputation seems to have a lasting effect, as London Heathrow Airport was recently voted in a survey of 14,500 frequent flyers to be the "worst airport in the world" [tha09].

Here's a chronology of the events on T5's first day of operation as it was reported in the news [KL08]:

March 27, 2008, 2:00 AM. It is still quiet in the Heathrow Terminal 5 building. Only some security guards walk around.

3:00 AM. The first employees arrive. More than expected are coming by car. They all have maps. However, due to the new environment and missing direction signs they have difficulties in finding the employee parking lots.

3:15 AM. The new baggage handling system doesn't start. Without the system, the baggage handlers don't have work plans, as in the new system these are provided electronically on PDA devices.

4:00 AM. The first baggage and check-in counters open. According to the plan, there should be more counters opened, however, staff is missing.

4:15 AM. The waiting lines grow. It was assumed that 80 percent of the passengers would check-in via Internet or at the self-service check-in machines, but in reality about half of the people go to the check-in counters.

4:30 AM. The employee parking lots are fully occupied. Employees are now redirected to alternative locations at the airport. There they need to wait up to half an hour for shuttle buses to the terminal, as the bus operators haven't sufficient knowledge of the place.

4:40 AM. The problem of the missing employees impacts the whole terminal. The security check for staff cannot open additional lanes, as the security inspectors are still on the way. Additionally, the security checks take longer as the inspectors are not familiar with the system yet and there are a few malfunctions.

4:42 AM. The first plane arrives earlier than expected. Because there is staff missing at the passport check, long waiting lines occur.

5:30 AM. The lines at the check-in counters and the security check grow further. Some of the counters have to be closed due to technical problems.

6:00 AM. The first flights are ready for boarding. But without the baggage handling system, no baggage comes on board.

6:30 AM. In the meantime, all security checks are working, although slower than expected. The baggage handling system still doesn't work properly. The first seven airplanes start without baggage.

7:00 AM. The baggage handling system is restarted. So far, about thousand bags piled up to be processed.

7:30 AM. Due to the long waiting lines many passengers arrive too late at their gate. BA cancels the first flights. But the passengers cannot change their bookings at the gates or the check-in counters. This is only possible at the 26 service counters, of which only two are open.

9:00 AM. An unexpected software error in the baggage handling system transfers the baggage of passengers with connecting flights not to the high rack storage, from where they should go to the connecting flights, but to the baggage claim area. From there they need to be taken manually back to the baggage store. Due to a system crash the identifications of many bags are lost, therefore their destination needs to be registered again. The pile of baggage grows further up.

11:00 AM. A few hundred environmentalists invade the building, around them policemen line up with machine guns. The cleaning staff doesn't come through anymore, the toilets become dirty.

2:00 PM. In the baggage hall more than 5,000 bags have piled up. Baggage is transported unsorted to other terminals.

3:00 PM. It comes to a fight between 30 baggage handlers, where security staff needs to intervene. BA is canceling more and more flights.

4:30 PM. The baggage handling system has reached its capacity limits. BA now allows only carry-on baggage.

10:00 PM. Many passengers sleep on the stiff and uncomfortable chairs in the terminal building. Later in the night, employees distribute blankets, sleeping bags and bottles of water.

11:00 PM. The engineers need to shut down the baggage handling system for reasons of precaution. BA is already canceling flights for the upcoming days.

1.2 Heathrow Terminal 5: Looking for the Root-Causes

What were the root-causes of the problems in the T5 project? As it is widely known, historical events are usually characterized by multiple root-causes (*multi-causality*) which are influenced by each other (*interdependencies*). The same holds for failure of large-scale development projects: There is no single root-cause, but multiple causes, which altogether originate the consequential situation.

As for Heathrow Terminal 5, public investigations followed. The Transport Committee of the British House of Commons published the results along with oral and written evidence of the public hearings on October 22, 2008 [Hou08].

In a nutshell, the investigations revealed three major problem areas:

- Insufficient staff familiarization
- Software errors in the baggage handling system
- Launching operation in a big bang approach

Insufficient staff familiarization. While there was an extensive plan for staff familiarization in the so-called *operational readiness phase* which started half a year prior to the opening of T5, "it is clear that on the first days staff did not have the degree of familiarity and confidence to operate to the planned times in what were new surroundings and in many cases using different equipment." [Hou08, Ev55] In addition, staff training was executed poorly, as described by Mr. Iggy Vaid on behalf of British Airways employees: People were taken to a hotel and shown film or slides. They were also put on a coach to give them an overview of the site, but what was missing was hands on training. For baggage, it was still a building site, therefore on-site training was impossible [Hou08, p.5].

Software errors in the baggage handling system. The baggage handling consists of two systems: the *baggage handling system* which screens and sorts the bags, and the *baggage reconciliation system*, which matches bags to checked-in passengers to ensure that all baggage on the aircraft is accompanied by a passenger. At the start of operation on March 27, there was a problem with the log-in codes for the staff who operate the baggage reconciliation system, as barcode passwords had been generated incorrectly [Hou08, Ev43]. Additionally, about a quarter (26 percent) of users repeatedly entered wrong log-in codes on the opening day causing the system to lock the user out after three attempts; a fast track password resetting process was implemented later on to solve this problem [Hou08, Ev46]. Due to a software error, some bags were incorrectly flagged as not being accompanied by passengers. Another software error caused the system to have problems recognizing data fed in from non-BA flights, which affected 2,900 bags not being transfered to their connecting flights [Hou08, Ev43]. The first–last bag identification in the Reclaim Hall was not working due to software interface issues, which resulted in delay in the reallocation of reclaim belts and late arrival of baggage. Software filters that were put in place for trials of the system prior to the opening in order to prevent transfer bag messages to be generated during testing were not removed before live operation. This lead to transfer bags not being recognized properly and therefore automatically sent to the system's storage facility for manual sorting. A software configuration was incorrectly set which stopped the feed of data from the baggage handling system to the baggage reconciliation system, which caused bags not to be ready for loading. Another software error caused bags from BA flights at Terminal 1 and 4 not being recognized by the T5 system and therefore being held for manual intervention and recoding. Finally, the baggage handling system became overloaded and gridlocked [Hou08, Ev54].

Launching operation in a big bang approach. From the Transport Committee's report it becomes clear, that many of the problems could have been avoided, if T5 was not started by a 'big bang' move, where 380 flights – seventy percent of the BA's total number of flights at all Heathrow terminals – were moved to T5 on the first day [Hou08, Ev6,Ev52]. Instead, the airline should have transfered the flights from terminal 1 and 4 over several weeks in incremental steps in order to be able to react to problems [hea08]. This is a direct analogy with waterfall development processes versus iterative development approaches. While in iterative development problems become obvious and can be solved early on, in a waterfall process problems often occur at the very end of development.

1.3 Early Warning Indications

Could the problems in the Heathrow T5 project have been detected earlier than on the day of its opening? From the committee report it is obvious that employees knew about problems before the opening. However, the existing organizational structure and the prevailing culture did not encourage or even prevented open feedback. Mr Iggy Vaid, on behalf of BA employees, argued in the public hearings:

> **Mr Vaid:** That is another problem of two-way feedback especially with British Airways. I hate to say that about my own airline, but culturally the existing management structure is one where you cannot tell the emperor that he has no clothes; you have to say his clothes are beautiful. No supervisor or person can tell his or her boss that the system will not work. If you do you are not a team player; you are sidelined, so for that reason you say that it works and the emperor has beautiful clothes [Hou08, Ev24].

Another indication that could hint at latent problems was the impaired communication between BA and BAA during the last half year of the project. Asked about the top reasons for what went wrong with the opening of T5, Colin Matthews, Chief Executive of BAA, answered:

> **Mr Matthews:** The first is that however well the airport operator and the airline operator, BA, are working it is also vital that the two are absolutely integrated and together. I think that during the construction of Terminal 5 that appeared to be the case. Around about or just prior to the opening of T5 it seems that that togetherness deteriorated. It is that togetherness that allows you to cope with the issues that arise on the day. Speaking personally, I think that was the biggest thing. [Hou08, Ev31]

and later:

> **Mr Hollobone (Transport Committee):** If you had your time again what one thing would you do differently?
>
> **Mr Matthews:** I would focus resolutely and determinedly on keeping British Airways and BAA in the same room tightly together. That has been my focus since joining on April 1 and it will continue to be my focus under any set of circumstances. I repeat that however good each one is on its own it needs to be really tightly co-ordinated and built together into a single team for success. That would be the single thing that personally I have invested in and will continue to invest in. [Hou08, Ev31]

Also the BAA memorandum confirms that the communication between BA and BAA deteriorated during the operational readiness phase:

> During the construction phase, I consider the joint working between BAA and BA to have been good. However, after September 2007 (the so-called "operational readiness" phase), the relationship worked less well. [Hou08, Ev46]

Finally, after the opening disaster the problems were solved by establishing a joint BA/BAA crisis management team at terminal level and by improving the communication between BA and BAA [Hou08, p.10].

1.4 Implications for Project Management

To sum up, the problems with Heathrow T5 have been in the intangible parts – software and people – and they have not been caught before the opening disaster. To avoid such problems which occur late in a project or – as it happened in the case of T5 – even at the customer site, early warning indicators are needed. It seems that collection of anonymous feedback and watching the communication and 'togetherness' between BA and BAA could have revealed problems much earlier.

Chapter 2

Introduction

Global Software Engineering (GSE) – software engineering in teams which are globally distributed, characterized by geographic separation, different time zones and cultural differences – has become a mainstream trend in industry today. While companies expect to benefit from a better cost position, access to talent regardless of location, and proximity to markets or customers, Global Software Engineering also introduces big challenges: Collaboration[1] and communication become more complicated than in a co-located project where all team members are located in immediate vicinity [Her07]. Difficulties of knowing who to contact about what, of initiating contact, and of communicating effectively across sites are typical effects. It's all the more important since software engineering is of a communication-intensive nature [BD10, p.111] and communication has high impact on project success: In fact, studies showed that the intensity of project-internal communication is positively correlated with project success [WR08]. Other studies confirmed that miscommunication is one major reason for issues such as cost overruns, quality problems and project delays [HLRP99].

Collaboration difficulties are not unique to globally distributed projects only, where team members are located in different countries. Similar effects are visible in any larger project which goes beyond one team room. In a study of engineering organizations, Allen found that, when engineers' offices were about 30 meters or more away, the frequency of communication dropped to nearly the same low level as when people are located in offices at different sites or countries [All84]. While geographic, temporal and cultural distance in general impede collaboration, they are no adequate indicator of collaboration difficulties. It can be the case that the collaboration between two team members which are thousands of miles apart is better than between one of these team members and his boss located in the office next door. Another measurement is needed referring to the *collaboration distance* between two team members, considering aspects such as communication frequency and relationship of team members.

[1] We use the term *collaboration* in the sense of "to work jointly with others or together especially in an intellectual endeavor" [MW11e].

Existing approaches – software measurement, social psychology, and communication metrics – have drawbacks when it comes to supporting project management[2] in detecting problems in GSE projects[3] early.

Software measurement provides support for operational decisions in project management as well as estimation and prediction. It is the basis for specifying and achieving objectives [ED07, p.2], in accordance with the saying 'You can't control what you can't measure' [DM82]. Metrics for project management support have been in place for decades now. They are used by project managers to observe the status of their project, including cost (e.g. actual and planned cost), time (e.g. milestone trend analysis) and quality aspects (e.g. defect rates). Metrics can support project managers in their decisions and help them to perform corrective actions [ED07, MCJ+01]. However, existing metrics measure rather the symptoms – exceeded cost, missed milestones or poor quality – than the root causes of problems in a GSE project.

The impact of trust on effective team collaboration has been studied by social psychologists. They have done comprehensive research by studying social groups and the way in which personality, attitudes, motivations and behavior of the individual influence and are influenced by them [MW11b]. However, psychologists are more interested in either fundamental research (e.g. how do people behave in a certain situation?) or structural approaches (e.g. study of organizational forms) or the area of human resource, e.g. competence models (required skills and abilities), but not in software project management. The kind of measurements introduced in this dissertation are well established in psychology. In our opinion, there is still a lack of cross-discipline collaboration between social psychology and Global Software Engineering research. An example for the value of cross-discipline cooperation is the area of usability engineering which is concerned with designing coherent user interfaces and has shown how useful methods from psychology can be for software engineering. Another example is the utilization of the social network analysis approach in GSE.

Analyzing communication data turned out to be a helpful approach for understanding behavior of globally distributed development teams. Bruegge and Dutoit showed that metrics on communication artifacts can be used to gain significant insight into the development process that produced them [DB98]. A common approach for analysis of communication data developed by Moreno [Mor34] in the 1930s is the so-called *sociometry* or *social network analysis*: Based on communication data the communication relationships between team members are analyzed. The resulting network can be displayed as a graph, and analyzed applying meth-

[2] According to the Project Management Institute, *project management* is the "application of knowledge, skills and techniques to execute projects effectively and efficiently", where a *project* is "a temporary group activity designed to produce a unique product, service or result" [Pro11].

[3] By *GSE project* we refer to a project whose team members are partitioned to at least two sites in different countries with different native languages (cf. definitions in Section 4.4).

ods of graph theory. Damian and her colleagues conducted social network analysis in various contexts, e.g. analyzing communication relationships in so-called requirement-dependency social networks, i.e. social networks of teams, which work on requirements that are dependent on each other [MDSS08]. Wolf et al. reported that the prediction of build failures based on social network analysis was possible under certain conditions [WSDN09]. Current GSE research focuses on social network analysis for examination of completed projects (post mortem analysis), but not on social network analysis for project management. Furthermore, the relationship aspect is not considered. Our explanation why GSE research currently doesn't focus on relationship metrics is that computer scientists seem to prefer 'objective' measurements, but not subjective judgments of GSE team members. While Moreno used sociometry for exploration of social structures via measurement of sympathies and antipathies of group members, communication metrics applied in GSE usually refer to 'objective' data, e.g. from a version control or change management system. However, the expressive power of 'objective' measurements is limited. Change management data for example adheres very much to the development process and does not reveal team conflicts.

The research question addressed in this dissertation is *How can collaboration problems in GSE projects be detected early on before they have serious impact?* The hypothesis of this dissertation is that collaboration-based measurement indicators can help to detect and address collaboration problems in GSE projects. Our research goal is to introduce a collaboration-based measurement model[4] – a set of metrics measuring collaboration inside a project team – to improve decision support in GSE project management.

In our research, we deal with the following aspects:

Analysis of Problem Domain. To thoroughly understand the problem domain of GSE, we analyze related work and conduct case studies as well as experiments. The study of related work includes literature on GSE, software measurement, communication metrics as well as an interdisciplinary view into social psychology.

Definition of Collaboration-based Project Management Metrics for GSE. We define a model of collaboration-based metrics to support GSE project management. The metrics are constructed using an interdisciplinary approach, utilizing knowledge from social psychology as well as software measurement expertise from computer science.

[4]We use the term *measurement model* to refer to a set of metrics representing an abstraction of attributes of the objects under observation, where a *metric* is the quantification of an attribute obtained from collected data.

Application and Validation of GSE Metrics. We apply our GSE metrics to selected projects to gain experience with them and evaluate their accuracy and usefulness as well as the effort for data collection and innovation grade.

The following aspects are *not* in the scope of this dissertation:

Collaboration vs. Project Performance. In our research, we anticipate that improving collaboration implies improving the project performance. The analysis of the relationship between improvement of collaboration and cost, time, quality behavior is not in the scope of this work.

Representative Field Study. The application of the collaboration-based measurement model is limited to a few projects. A broad, representative field study as well as a statistical analysis of results, which would have been possible with a large set of projects, is beyond the scope of this dissertation.

The approach of the collaboration-based project management metrics introduced in this dissertation differs significantly from the existing work. To our knowledge, collaboration-based project management metrics are a new field in GSE research. Existing studies on communication metrics in GSE are usually post mortem analyses, i.e. a project is analyzed after completion, without the objective to improve project management (see Table 3.1 in Section 3). Another new aspect of this dissertation is the interdisciplinary approach of social psychology and computer science.

This document is structured as follows: Chapter 1 motivates the need for early warning indicators by a case study of the Heathrow Terminal 5 opening disaster. Chapter 2 gives an introduction to the topic of collaboration-based metrics for project management support in GSE. In Chapter 3 we present the research process and the research methodology applied in this dissertation. Chapter 4 describes characteristic challenges in GSE. Chapter 5 deals with an interdisciplinary view of GSE teams: the social psychology of distributed teams. Chapter 6 describes the role of communication in software engineering and introduces the concepts of software measurement and communication metrics. In Chapter 7 we define a model of collaboration-based project management metrics for GSE. Chapter 8 describes the first application of our GSE metrics in an industrial case study. Chapter 9 deals with the validation of the GSE measurement model in two additional case studies. Chapter 10 presents a summary of this dissertation as well as future directions.

Chapter 3

Research Process

In order to introduce a collaboration-based measurement model for project management in GSE projects, we took a three step approach. It consists of the phases *Understand Problem Domain*, *Construct Measurement Model* and *Validate Measurement Model*, as depicted in Figure 3.1. We followed the principles of formative evaluation as opposed to summative evaluation. *Formative* evaluation is concerned with forming a program or technology (in our case a measurement model), "by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on" [Toc06]. The primary goal of the application is to develop and improve the program or technology. An example could be the development of a new university course, where formative evaluation is conducted "with a small group of people to 'test run' various aspects of instructional materials" [Blo07]. *Summative* evaluations on the other hand focus on examining the effects or outcomes of a program or technology after its introduction, "they summarize it by describing what happens subsequent to delivery of the program or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes" [Toc06]. The primary goal of the application is to observe the outcome after delivery. In the example of the new university course, summative evaluation answers the question "did the learners learn what they were supposed to learn?" after the university course has been rolled out [Blo07]. In a quote of Robert Stakes the difference between formative and summative evaluation is put at its most succinct: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." [Blo07]

With respect to our research goal to introduce a collaboration-based measurement model to improve decision support in GSE project management, the formative evaluation refers to the development of the GSE metrics, while summative evaluations would apply proven GSE metrics to a broad set of GSE projects and evaluate the outcomes.

Year	Total Papers	Communication	Communication or Collaboration Metrics	For Project Management
2006	25	24	2	0
2007	29	25	1	0
2008	29	27	4	0
2009	46	45	4	0
2010	42	40	1	0
Total	171	161	12	0

Table 3.1: Systematic literature review of ICGSE publications: Number of papers referring to communication, papers on communication or collaboration metrics, and papers which deal with the application of communication or collaboration metrics for project management purposes

Project management metrics for GSE are still in their infancy. We conducted a systematic literature review of all publications of the IEEE International Conference on Global Software Engineering (2006–2010). We analyzed each paper under the following aspects: a) Does the paper refer to 'communication'? b) Does the paper describe or use communication or collaboration metrics of any kind? (e.g. social network analysis, or other approaches of measuring communication), and c) Are communication or collaboration metrics applied for project management purposes? Table 3.1 summarizes the result. As one can see from the figures, almost all papers (in total 94%) refer to the term communication, while only few of them (7%) actually deal with communication or collaboration metrics. However, we found no evidence of usage of communication resp. collaboration metrics for project management in any of the ICGSE publications. The lack of related work on collaboration-based project management metrics in GSE reinforces the decision to follow a formative approach to construct GSE metrics.

In the following we outline the three steps of our research process (see Figure 3.1).

Understand Problem Domain. The objective of the first phase was to get a profound understanding of characteristic challenges in GSE. We put strong emphasis on a thorough and comprehensive analysis, as a good understanding of the problem domain is the foundation for a purposeful definition of a GSE measurement model. Based on *literature research*, we synthesized related work (including an interdisciplinary view and own work which was published earlier). We used *experiments* to learn more about the effects of distributed teams. We conducted selected experiments in student groups, other experiments are cited from literature. Furthermore, we conducted *case studies* of GSE projects, in particular utilizing *semi-structured interviews*.

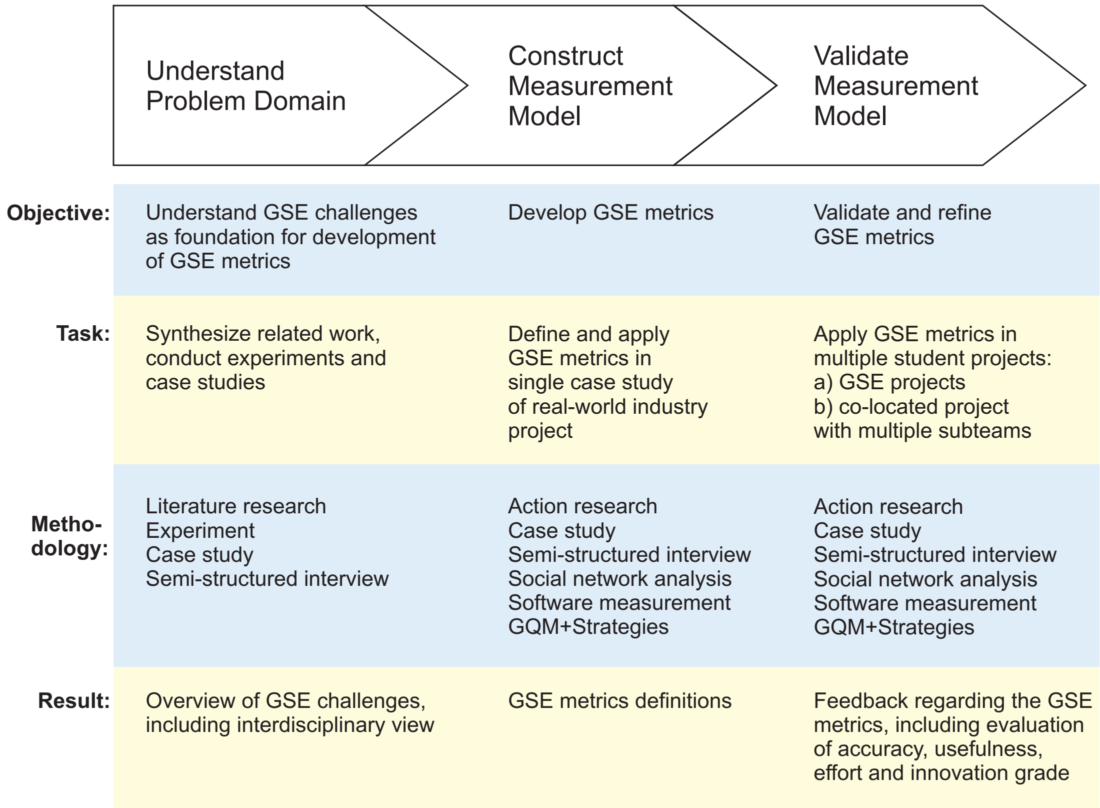


Figure 3.1: Overview of research process and methodologies

Construct Measurement Model. Focus of the second phase was the development of collaboration-based GSE metrics. We defined and applied our GSE metrics in a *single case study of a real-world industry project*, i.e. in realistic context, utilizing *action research*. Phase 2 resulted in a set of GSE metric definitions. We adapted the *GQM$^+$Strategies* approach introduced by Basili, Rombach et al. [BLR$^+$10] which is an extension of the well-established Goal Question Metric (GQM) paradigm [BGR94] for goal-oriented definition of metrics. Other techniques to be mentioned here are: *semi-structured interviews* which we used during our investigations on the industry project, as well as *software measurement* and *social network analysis* which were important foundations for our GSE metrics.

Validate Measurement Model. After the GSE metrics were defined and applied for the first time, a validation phase followed. The same research techniques were used as in the phase before: Action research, case studies, semi-structured interviews, social network analysis, software measurement and GQM$^+$Strategies. However, as opposed to the previous phase, focus was now on application of the GSE measurement model to multiple projects to gain broader experience. We

Experiment	Purpose	Type
NASA space game	Investigate decision making in distributed teams	student experiment
The Derdians	Create culture awareness	student experiment
Quizmaster game	Illustrate actor-observer divergence	student experiment
Robbers Cave	Investigate intergroup relations and conflict	cited from literature
Shipping business game	Investigate effect of reciprocity	cited from literature
Decision alternatives	Illustrate traps in decision making due to human perception	cited from literature

Table 3.2: Overview of experiments

decided to look for suitable student projects as our objects of research, because we assumed that they are more open to experimentation, e.g. in frequent application of the GSE metrics, and allowed for examining many projects because the administration efforts (e.g. negotiation of non disclosure agreements) were considerably lower than in industry projects. We applied the GSE metrics to GSE student projects and one co-located student project with multiple subteams to be able to compare results in both settings. Phase 3 resulted in a feedback regarding the defined GSE metrics, including an evaluation of accuracy, usefulness, effort and innovation grade.

3.1 Experiments

We used a set of experiments to gain knowledge on collaboration in GSE teams. An overview is presented in Table 3.2. We conducted the experiments NASA space game, The Derdians and Quizmaster game inside a GSE master seminar in June 2010. It was a three days block seminar with a group of thirteen international students. The purpose of the NASA space game was to investigate decision making in distributed teams. Three separate teams were created, two distributed teams (team is split across two different rooms) and one co-located team (all team members stay in the same room). A team task was specified and the time needed to solve the task as well as the quality of the outcome was measured. Creating culture awareness was the objective of the experiment The Derdians. Finally, the Quizmaster game was used to demonstrate the actor-observer divergence, an effect from social psychology. Other experiments cited from social psychology include the Robbers Cave experiment, a comprehensive field study on intergroup relations and intergroup conflict, a shipping business game which illustrates the effect of reciprocity in human behavior, and experiments on decision alternatives which reveal traps in decision making due to human perception.

3.2 Case Studies

Besides experiments, case studies were a main element in our research. In the Understand Problem Domain phase, we used case studies to learn more about challenges in GSE projects, in particular through semi-structured interviews. In the Construct Measurement Model and Validate Measurement Model phase we utilized case studies to develop and evaluate our GSE metrics. The main research methodology we used in the case studies during development of our GSE measurement model was *action research*. Action research is based on multiple learning cycles, in which knowledge is iteratively applied and refined. The research takes place in real-world situations and aims to solve real problems [O'B01]. According to the action research model of Kemmis [KM88], each cycle consists of the following process steps: 1) *Plan:* a strategic plan is developed based on the current knowledge, 2) *Act:* the strategic plan is implemented (the *action*), 3) *Observe:* Outcomes are evaluated, and 4) *Reflect:* A critical reflection on the results and the process applied is conducted in order to prepare for the next cycle. We defined and applied our GSE metrics in an industrial case study and validated it in two additional case studies. From the projects' point of view, it was a process of organization development, with an input phase, a transformation phase and an output phase. In the input phase, the current status of the organization and related problems were determined. The transformation phase addressed the identified issues by taking improvement actions, and the output phase reflected the status after the improvement. In our research perspective, the development of GSE metrics based on formative evaluation was the main focus. The action research process allowed for learning in realistic context and incorporating feedback from practical applications to ensure the effectiveness of the defined measurement model.

3.3 Criteria for Selection of Projects

In the following we give an overview of the criteria for project selection in the phases Construct Measurement Model and Validate Measurement Model.

3.3.1 Project Selection for Construction Phase

As mentioned above, we looked for a real-world industry project to construct and apply the collaboration-based GSE metrics. Criteria for the selection of the project were as follows:

- **Industrial project.** In order to learn in a realistic context, a real-world software engineering project from the software and systems development industry should be selected.
- **Globally distributed project.** The project under examination should be globally distributed, with multiple development sites in different countries, multiple native languages and cultural backgrounds, and including a low-cost site.
- **Medium size project.** The project should be of a medium size, i.e. about 30–100 team members, because a too small project would deliver less insights, and a too large project would be difficult to handle for a first time application of the new metrics.
- **Risk project.** We particularly looked for a project with a high risk resp. a crisis project.
- **Willingness to support research.** The project leader should be willing to support experimental research, willing to collect and evaluate data, and open minded to applying the new set of metrics for the first time.

3.3.2 Project Selection for Validation Phase

The objective of the validation phase was to apply the defined GSE metrics in multiple projects to gain broader experience with the defined GSE measurement model and validate its effectiveness. While industry projects are preferred for the validation, our experience is that industry projects – especially in Germany – are significantly more restrictive with respect to legal issues and data collection (e.g. data collection and analysis for groups smaller than 5 persons is not allowed) and at the same time involve a high administration effort with respect to the legal side. We therefore decided to look for student projects, both globally distributed and a co-located project with multiple subteams. Criteria for the selection of the projects were the following:

- **Student projects.** Multiple student projects should be selected in order to gain broader experience in application of the GSE measurement model. We chose student projects, because they were assumed to be more open to experimentation and allowed for examining more projects because the administration efforts are considerably lower than in industry cooperations (see above).
- **Global distribution/multiple subteams.** We looked for student projects in software engineering, which operate in a globally distributed setting, with multiple sites in different countries, multiple native languages and cultural backgrounds, and different time zones. On the other hand, we also looked

for a co-located student project of at least 20 persons with multiple subteams. We expected that examining both GSE projects and a co-located project with multiple subteams would allow us to compare the co-located and distributed setting and see to which extent effects of distribution also occur between subteams of the same project.

- **Willingness to support research.** The project leader should be willing to support experimental research, willing to collect and evaluate data, and open minded to applying the new set of metrics.

3.4 Validation Criteria

A central question in the design of the validation process was: How do we know, if our approach was successful? To be able to answer this question, we established the following validation criteria:

- **Accuracy.** One important aspect of the validation is the accuracy of the metrics, i.e. *how well do the GSE metrics reflect the reality in the project team?* Without accurate data, no meaningful analysis of the team collaboration is possible. The accuracy aspect is particularly important because the metrics represent the subjective perception of the respondents.

- **Usefulness.** Besides the accuracy of data, it is important to see the usefulness of the metrics, i.e. *how useful are the GSE metrics for project management?* Even if the metrics are accurate, they might not be of any particular use.

- **Effort.** Since the metrics are designed to be used throughout a GSE project and applied regularly, they might not take away too much time of the team members. The effort required for data collection and analysis is a crucial aspect, i.e. *is the effort for data collection and analysis acceptable?*

- **Innovation grade.** Finally, also the innovation grade of the proposed approach is an important aspect, because the solution might be accurate and useful, but not new, i.e. *is the collaboration analysis approach new in GSE project management?*

The research process described above was a valuable foundation for developing and validating our GSE measurement model. In the following three chapters, we deal with the major concepts utilized in the first phase of our research: First, we discuss challenges from Global Software Engineering perspective. Second, we take an interdisciplinary view on the social psychology of distributed teams. Third, we look into the concepts of software measurement and communication metrics.

Chapter 4

The Challenges of Global Software Engineering

In this chapter, we give an overview of related work and summarize characteristic challenges which globally distributed development teams are facing.

4.1 The Key Challenge: Collaboration over Distance

Global Software Engineering has brought new challenges to development teams: Geographic separation, working in different time zones and different cultures and languages make the collaboration of team members difficult [HM03, BHL07]. Key challenges are collaboration and communication over distance [Her07]. Difficulties of knowing who to contact about what, of initiating contact, and of communicating effectively across sites are typical effects. Miscommunication – the "failure to communicate clearly" [MW11j] – is one major reason for issues such as cost overruns, quality problems and project delays [HLRP99]. One of the major problem areas is requirements engineering, as handling of requirements and requirements changes are particularly difficult in a globally distributed setting; they are of a communication-intensive nature [DZ03, Her07, Dam07], and requirements changes occur frequently: The traditional rule of thumb for requirements allocated for implementation is to expect them to change by 1 to 3 percent per month [Ebe06].

4.2 Delay in GSE

A key effect of GSE is delay. To illustrate this observation, we would like to report on the results of a controlled experiment conducted with a group of thirteen students within a GSE master seminar in June 2010. The experiment was designed as a group exercise with the learning objective for the students to experience conducting a development-related task via electronic communication media. For this purpose we chose a prioritization task, which is an activity that occurs in practically every development project, e.g. when discussing and assigning priorities of requirements or change request. In order to achieve comparable results, we chose the NASA space game [KK09, NAS, Wik11]. It is a well-known exercise in psychology and management education, but it was not known to any of the thirteen students. The objective in the NASA space game for the team is to prioritize 15 items according to their importance in an emergency situation on the moon. The instructions of the NASA space game are as follows:

> Your spaceship has just crashed on the moon. You were scheduled to rendezvous with a mother ship 200 miles away on the lighted surface of the moon, but the rough landing has ruined your ship and destroyed all the equipment on board except for 15 items. Your crew's survival depends on reaching the mother ship, so you must choose the most critical items available for the 200-mile trip. Your task is to rank the 15 items in terms of their importance for survival. Place a number 1 by the most important item, number 2 by the second most important, and so on, through number 15, the least important. [NAS]

In order to avoid unwanted effects due to language issues as the GSE master course was a mixed course of German and international students, the list of items was provided in both languages – English and German (see Table 4.1).

The goal of the experiment was to analyze group behavior and results of co-located vs. distributed teams. The hypothesis was that co-located teams show better result than distributed teams. One important reason for choosing the NASA space game was that there exists a sample solution against which an absolute evaluation of the results is possible (see Figure 4.1). To analyze team behavior in various settings, three groups were formed: two distributed teams (team B and C) and one co-located team (team A). The physical distribution into two sites was achieved by splitting team B and C into two rooms. The available time for solving the prioritization task was communicated to be 45 minutes.

For the distributed teams (team B and C) the task information was also distributed: We gave team B and C the instructions plus half of the list (items A–H) at site 1 and the other half (items I–O) of the list (without instructions) at site 2. Due to its co-location at one site, team A was given the instructions and the whole list of items to be prioritized. The three teams varied in the type of communication media they were allowed to use. The co-located team A used personal communication. Team B was allowed to use e-mail communication only. Team C was allowed to use any electronic communication media such as video conferencing, e-mail or chat. In fact, team C decided to use skype chat.

During the experiment, the time was measured from beginning of the task until completion (hand-in of results). It turned out that the co-located team was

Item	Description (English)	Description (German)
A	Box of matches	Streichhölzer
B	Food concentrate	Lebensmittelkonzentrat
C	50 feet of nylon rope	Fünfzig Fuß Nylonseil
D	Parachute silk	Fallschirmseide
E	Solar-powered portable heating unit	Tragbares Heizgerät
F	Two .45caliber pistols	Zwei .45 Kal. Pistolen
G	One case of dehydrated milk	Trockenmilch
H	Two 100-pound tanks of oxygen	Zwei 100-Pfund-Tanks Sauerstoff
I	Stellar map (of the moon's constellations)	Stellar-Atlas (Mondkonstellation)
J	Self-inflating life raft	Sich selbst aufblasendes Lebensrettungsfloß
K	Magnetic compass	Magnetkompass
L	5 gallons of water	Fünf Gallonen Wasser
M	Signal flares	Signalleuchtkugeln
N	First-aid kit containing injection needles	Erste-Hilfe-Koffer mit Injektionsnadeln
O	Solar-powered FM receiver-transmitter	Mit Sonnenenergie angetriebener UKW-Sender/Empfänger

Table 4.1: NASA space game: List of items [KK09, NAS, Wik11]

significantly faster: While team A needed 24 minutes to complete the task, team B finished in 53 minutes. Team C even needed 62 minutes. This is equivalent to a delay of factor 2.2 in case of team B and factor 2.6 in case of team C (see Figure 4.3). These results actually match well with experiences reported in literature: Herbsleb and Mockus found in their studies that distributed work items take about 2.5 times longer to complete than similar items where all the work is co-located [HM03].

To assess the quality of the teams' results, the item priorities assigned by the teams were compared to the sample solution published by NASA. The deviation was calculated as the sum of the deviations per item with respect to the rank number. The lower the deviation value, the better the prioritization results. A zero deviation would be a perfect match with NASA's sample solution. The resulting deviations are shown in Figure 4.3. With a deviation of 26 (team A) and a deviation of 28 (team C), team A and C achieved the best results. Team B's result showed a deviation of 37.

After the experiment, the experiences made during the exercise were discussed with the students. Team B and C reported that they had certain problems due to a bad WLAN connection. In addition, for team C it took about ten minutes

Item	Description	NASA rank	Explanation
A	Box of matches	15	No oxygen to sustain flame, virtually worthless
B	Food concentrate	4	Efficient means of supplying energy requirements
C	50 feet of nylon rope	6	Useful in scaling cliffs, tying injured together
D	Parachute silk	8	Protection from sun's rays
E	Solar-powered portable heating unit	13	Not needed unless on dark side
F	Two .45-caliber pistols	11	Possible means of self-propulsion
G	One case of dehydrated milk	12	Bulkier duplication of food concentrate
H	Two 100-pound tanks of oxygen	1	Most pressing survival need
I	Stellar map (of the moon's constellations)	3	Primary means of navigation
J	Self-inflating life raft	9	CO2 bottle in military raft may be used for propulsion
K	Magnetic compass	14	Magnetic field on moon is not polarized; worthless for navigation
L	5 gallons of water	2	Replacement for tremendous liquid loss on lighted side
M	Signal flares	10	Distress signal when mother ship is sighted
N	First-aid kit containing injection needles	7	Needles for vitamins, medicines, etc., will fit special aperture in NASA space suits
O	Solar-powered FM receiver-transmitter	5	For communication with mother ship; but FM requires line-of-sight transmission and short ranges

Figure 4.1: NASA space game: Sample solution developed by NASA experts [KK09, NAS, Wik11]

Item	Team A co-located	Team B e-mail	Team C skype chat	NASA	Delta A	Delta B	Delta C
A	14	15	14	15	1	0	1
B	3	6	3	4	1	2	1
C	7	5	7	6	1	1	1
D	11	9	12	8	3	1	4
E	12	14	4	13	1	1	9
F	13	10	13	11	2	1	2
G	9	3	11	12	3	9	1
H	1	1	1	1	0	0	0
I	8	4	5	3	5	1	2
J	10	11	8	9	1	2	1
K	15	5	15	14	1	9	1
L	2	2	2	2	0	0	0
M	5	8	10	10	5	2	0
N	6	7	6	7	1	0	1
O	4	13	9	5	1	8	4
Deviation from NASA solution:					**26**	**37**	**28**

Figure 4.2: NASA space game: Ranks assigned by team A, B and C and deviation from NASA sample solution

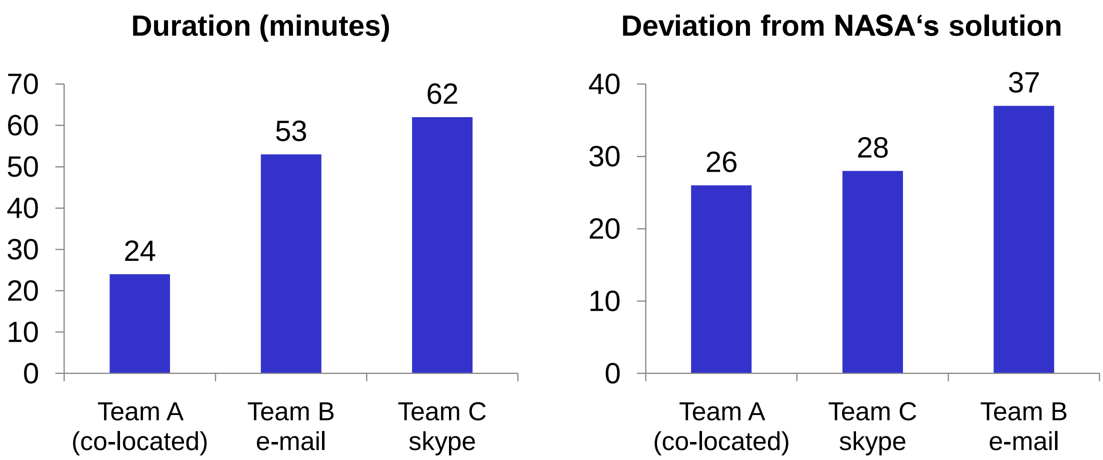


Figure 4.3: NASA space game: Duration and deviation from NASA's sample solution per team

to set up the chat connection, which explains why team C needed 9 minutes more than team B to complete the task although they had a more direct communication medium. However, this effort paid off with respect to the better result achieved, which seems to be a consequence of the more instant communication via chat compared to e-mail. Team B reported that they had problems with overlapping e-mails, i.e. they sent multiple e-mails to the part of the team at the other site before the previous ones were answered. Team C reported that their subteam at site 1 at the beginning didn't recognize that site 2 had no task description. For example, they mentioned the discussion around the compass: Site 2 suggested "we take the compass", which was answered by site 1 with "but it doesn't work on the moon". The reaction of site 2 was: "So what?", as they didn't know the instructions and the context of the task. Both team B and C reported integration difficulties: The available lists were prioritized locally first. It took a high coordination effort for them to agree on a joint list.[1]

4.3 GSE Challenges Observed in International Student Projects

In order to better understand the problem domain of GSE, we analyzed three international student projects in GSE education. The projects were carried out in the so-called *Network of Engineering univeRsities Educating in Intercultural Design* (NEREID) university course, which was originally initiated by Prof. Laurini from INSA de Lyon, France; now multiple international universities are participating [MNS+11]. In the NEREID course, groups of six students from three different sites collaborate on the development of a specific software system. The international student projects we examine here were conducted in winter term 2009/2010. The students were from five different universities, speaking three different native languages:

- Institut National des Sciences Appliquées de Lyon, France
- Tecnológico de Monterrey campus Puebla, Mexico
- Technische Universität München, Germany
- Universidad Tecnica Federico Santa Maria, Chile
- University of Applied Sciences Esslingen, Germany

[1]The experiment turned out to be a great learning opportunity for the student groups and received excellent feedback from the students. It therefore can be recommended as exercise that can be included in future GSE education.

To capture the experiences and lessons learned, we conducted semi-structured interviews with the student teams in Germany, after the projects were completed and evaluated.

4.3.1 Project 1: University Foreign Relations Map

The project 1 aimed at developing a web application to graphically display partner university relationships. The application should make it easier for students to find partner universities and compare possible exchange partners. The scope included displaying a university relations map as overview of partner universities, basic administration capabilities to edit university relationship data as well as provisioning of an online forum for communication between students and uploading experience reports.

The project team was distributed across three sites with two students from each site: Garching (Germany), Lyon (France) and Esslingen (Germany). Esslingen was the initiator of the project. Project 1 had a project leader from the customer site who was appointed already before the project was started and had a close contact with the supervising professor. The worksplit was defined by the project leader, the development tasks were separated by subsystems: The responsibility for the web page was in Esslingen, the online forum was assigned to Lyon and the database development was the task of the team in Garching. The requirements for the project were provided by the supervising professor in Esslingen who also acted as the customer. All communication with the customer was handled by the project leader in Esslingen; there was no direct communication between the team in Garching and the customer. For project communication, weekly skype meetings (voice only) were initially scheduled. However, due to different reasons, the project team was never fully present at these meetings. Later, e-mail communication was therefore used instead. The team in Garching experienced long delays to receive answers to their e-mails, sometimes up to one week. Another issue of distribution was the access to the web server in Esslingen, as the Garching team had no direct access to the server and always needed assistance by the team in Esslingen.

Another problem occurred with the Lyon site, which was responsible for the online forum. According to the Garching team, there was low involvement from the Lyon team and they were most of the time arguing why they could not proceed. The Garching team kept sending e-mails to the project leader and asked for an escalation. However, the problem was not solved. At the very end of the project, the Lyon team posted rudimentary code for an online forum as their contribution, but it was not integrated with the overall system. The Garching team felt angry about their colleagues in Lyon.

The team in Garching reported in the interview that they saw there were different motivations of the sites involved. In particular the team in Esslingen treated the project lightly. When there were discussions about problems with the project leader, he referred to clarifying all issues with his professor in a personal and informal way. According to the Garching team, there was no formal final presentation in Esslingen, but only an informal meeting with the professor, and the project report of the Garching team was used instead of writing an own report.

4.3.2 Project 2: Picture-based Itineraries

The task of project 2 was to develop a picture-based navigation system for pedestrians. While existing navigation systems usually use street names and a map to explain the route, many pedestrian ways do not have street names. Therefore the goal of this project was to create a system which uses pictures augmented with arrows to explain the route, e.g. a view of a crossroad including an arrow to indicate where to go next.

The project team involved students from three countries – Germany, France and Chile – with two students at each site. The supervising professor of the French site played the role of the customer. The team felt there was a great latitude with respect to the requirements, as only high level requirements were given. The team split their work according to subsystems to minimize the need for communication: the server part was taken by the French team, the client part was chosen by the German team, and the navigation arrows were assigned to the Chilean team. In contrast to project 1, this project had no explicitly appointed project leader. (While a team member from France suggested himself in his first e-mail as project leader, the German team answered the decision should be deferred to the first meeting, where the team members get introduced to each other. Later at the meeting, no decision was taken.) As the project task came from France, in the first meeting the French team gave an overview of the high level requirements. As many aspects of the task were still underspecified, the French team took over the responsibility to further clarify the requirements and provide a specification.

Communication was arranged in weekly meetings and exchange of e-mails. The purpose of the weekly meetings was to exchange the current status and to make decisions. While the first meetings had no pre-defined agenda, the meetings became more structured over time. Skype's chat was used as communication medium, as the available network bandwidth was not sufficient for voice-over-IP or video communication. The team reported that they made good experiences using chat, it even helped them to maintain a certain order in communication.

From the beginning of the project, there were problems in the cooperation with the Chilean team. The students from Chile did not attend the first meeting, as they "totally forgot the meeting date". This lead to additional effort, because the information provided during the first meeting and the decisions that had already been made needed to be re-discussed with the Chilean team. The German team did not escalate the problem in order not to discredit their colleagues, only some e-mails were exchanged within the team. There was no contribution from the Chilean team until two days before the final presentation: A code delivery was provided on December 20th, which was not integrated and therefore not working at the final presentation (no arrows visible). Also there were problems with the time synchronization between the three sites. While the German team had to give its final presentation and demonstration of the system on December 22nd, the French team went already on holiday on December 19th. Fortunately the French team could run the server during their holiday from the students' hall of residence, so that the server needed for the demonstration was still available.

4.3.3 Project 3: Cadaster System

The project 3 focused on the development of a system using geographical data of the cadaster system of Puebla (Mexico) to map articles to specific geographic locations. Users should be able to visualize, create and modify articles related to a specific reference point on the map.

The project team consisted of three students from Mexico, two from France and one student from Germany. The topic was defined by the supervising professor from Mexico. A high level requirements specification was already existent at the beginning of the project. Project 3 had a project leader from Mexico who was designated before the project started. The team members jointly organized the distribution of work. They decided to have a functional worksplit across locations: One member of the French team was responsible for organizing meetings (e.g. invitation and agenda) and the documentation (e.g. meeting minutes). Mexico was accountable for the conversation and clarifications with respect to the requirements. The student from Germany took over the role of the technical implementation and integration as lead developer. Additionally, each site was in charge for a part of the system. Furthermore, for each site the team defined a so-called team manager who acted as the single point of contact in case of problems or bug reports with the part of the system, for which the site was responsible. The team had weekly meetings, a video conference meeting for organizational issues and status reports and a skype chat for technical clarifications. When required, the frequency of meetings was increased. The supervising professor from Mexico who was in the customer role occasionally participated in the video conferences. Overall the collaboration went well, although it was noticed that there are cultural differences between the countries: the team from Mexico was perceived to be more relaxed, while the French colleagues sometimes expressed their concerns when progress was not as fast as they wanted. The team members had to bear the time difference in mind, which was 7 hours, but it was not considered a serious problem as the team members were able to agree on meeting time slots, where all sites were available. Sometimes it was challenging to find a time slot for the video conference, because it was difficult for both the French and the Mexican team to get a video conference room due to limited resources. One major issue

Project	Task	Site Structure	Work Split	Communication Media
1	Develop a web application to graphically display partner university relationships	Germany (Garching):2; France:2; Germany (Esslingen):2	Subsystem worksplit	Skype (voice), e-mail
2	Develop a picture-based navigation system for pedestrians	Germany:2; France:2; Chile:2	Subsystem worksplit	Skype (chat), e-mail
3	Develop a system to map articles to specific geographic locations in the cadaster system of Puebla (Mexico)	Mexico:3; France:2; Germany:1	Functional worksplit: organization, requirements, implementation	Video conference, skype (chat)

Table 4.2: NEREID project overview

occurred with respect to understanding of the requirements: The Europeans and the Mexican had different imaginations of a cadaster system: While the students in France and Germany assumed that they need to implement a cadaster system with legal registers – as known in Europe –, the Mexican expected to build a system with a simple geographical map. As the team members detected that they have different domain knowledge, they were able to resolve this issue.

4.3.4 Discussion

Table 4.2 gives an overview of the characteristics of the three student projects. While all of the three projects had a similar site structure, they differed with respect to the work split which the teams had chosen to use. While project 1 and 2 split their work according to subsystems of the system to be developed, project 3 used a functional work split: organization, requirements, and implementation. The worksplit by subsystems has the advantage of minimizing the need for cross-site communication because each site can work independently on one subsystem. However, according to to Berenbach [Ber06] there is a risk of insufficient coordination between the development sites and difficulties in integration of subsystems. The functional worksplit on the other hand introduces communication difficulties between the requirements and the implementation subteam. Also the communication media used differed across the three teams: Project 1

Challenge	Projects
Time coordination, availability	1,2,3
Requirements, domain expertise	2,3
IT infrastructure	1,2
Different motivations per site	1,2
Cultural differences	3
Communication delays	1
Escalation	1,2
Integration of subsystems	1,2

Table 4.3: GSE challenges reported in the interviews

used voice over IP via skype and later e-mail, project 2 decided to use e-mail and skype chat, and project 3 communicated via video conferencing and skype chat.

In Table 4.3 the challenges identified in the three student projects are summarized. Altogether, the problems encountered by the student teams very well reflect real-life issues, in particular communication problems as well as technical, organizational and people aspects [BHL09]. Due to the geographic and organizational separation, the students perceived themselves being in local subteams instead of one global team. In critical situations this resulted in a characteristic “us” versus “them” attitude. However, all of the NEREID projects were finally able to present good results and made an important learning experience. The students of project 1 and 2 stated that the modular architecture and separation by subsystems has proven well, as it reduces complexity and communication needed across sites. The student teams also mentioned that a clearly defined escalation path would help them to deal with problems which are outside of their responsibility. The current experience had shown that the students are reluctant to run down their colleagues in front of their supervising professor.

Both student teams interviewed emphasized that the project was a very valuable learning experience for them, those were aspects which “cannot be learned by listening to lectures only”.

4.4 A Meta Model for GSE

In Chapter 2 we introduced GSE as software engineering in teams which are globally distributed. GSE typically involves *offshoring*, “the practice by which companies outsource processes, chiefly IT-based services, across large distances to other parts of the world, often low-wage countries” [Sch04]. Figure 4.4 which

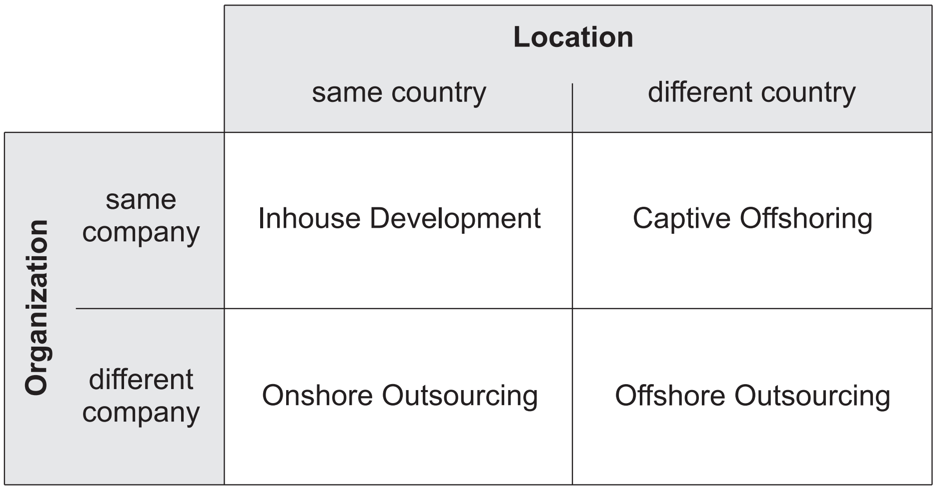


Figure 4.4: Offshoring types, adapted from Schaaf [Sch04, p.3]

we adapted from Schaaf [Sch04, p.3] shows different types of offshoring[2]. It is structured along two dimensions: location – same or different country – and organization – same or different company. If development is conducted within the same country and the same company, we call it *inhouse development*. Development inside the same company, but involving different countries is called *captive offshoring*. The dimension of organization introduces the differentiation between 'make' (same company) or 'buy' (different company), called *outsourcing*. If outsourcing is carried out inside the same country, we name it *onshore outsourcing*. If outsourcing pertains to a different country it is called *offshore outsourcing*.

We use the term *group* for "a number of individuals assembled together or having some unifying relationship" [MW11f] and the term *team* in the sense of "a number of persons associated together in work or activity" [MW11g]. A *project team* is a team which is designed to produce a unique product, for example a software-based system. We use the term *distributed project team* for a project team whose team members are partitioned to different locations, e.g. two different buildings. More specifically, a *GSE project team* is a distributed project team where the locations of the team members include at least two sites in different countries with different native languages. A *GSE project* is a project which is conducted by a GSE project team. Distributed project teams and GSE project teams are facing similar collaboration challenges. In fact, Sangwan and his colleagues define GSE "as software development that uses teams from multiple geographic locations" [SBM$^+$06, p.3], regardless of the distance between the locations: "These teams could be within one country [...] or on other sides of the world" [SBM$^+$06, p.3].

[2]Additionally, in some publications the term *nearshoring* is used which refers to offshoring in close geographic proximity, e.g. a neighboring country.

Figure 4.4 views GSE along two dimensions – location and organization. However, in GSE literature multiple other models can be found. We systematize existing approaches in GSE from literature into a meta model for Global Software Engineering. In the following, we briefly describe the elements of this meta model as well as related challenges and heuristics from related work. We use the term *challenge* to refer to a characteristic which makes Global Software Engineering difficult. Furthermore, a *heuristic* is defined as an approach to either reduce the need to collaborate (i.e. reduce dependencies) or to increase the ability to collaborate effectively. Please note that we give examples of typical challenges and heuristics for GSE, but we do not strive for completeness here. A graphical illustration of the meta model for Global Software Engineering can be found in Figure 4.5. The meta model is organized according to a taxonomy introduced by Rumbaugh and colleagues [RBP+91, p.17]. The focus of Rumbaugh's work was on object-oriented modeling and design for software engineering, i.e. system-related models. However, the taxonomy is transferable also to other types of models, e.g. people-related models. In recent publications, the taxonomy is also called Structure–Behavior–Function framework [GK04, GRV09]. The meta model relates GSE models to three categories (the columns in Figure 4.5):

- **Structure (Object Model).** This category deals with the structure of objects, i.e. components of the objects and connections among them [RBP+91, p.17][GK04, GRV09]. The object in the GSE meta model are either the people involved in software engineering or the system under development.
- **Behavior (Dynamic Model).** This type of models "describes those aspects of a system concerned with time and the sequencing of operations" [RBP+91, p.18], it refers to dynamic behavior, e.g. a sequence of states and transitions between them [GRV09].
- **Function (Functional Model).** Functional models describe the teleology of the objects (i.e. what it is for) [GK04] resp. functions characterized by preconditions and postconditions [GRV09].

The GSE meta model distinguishes between people-related models and system-related models (the rows in Figure 4.5). People-related models are concerned with the team members of a development team, while system-related models deal with the technical system under development.

4.4.1 People-related Models

People-related models deal with the team members of a development team. The people-related models contained in the GSE meta model are: Location Model, Organization Model, Competence Model, Culture Model, IT Infrastructure Model, Collaboration Model and Process Model.

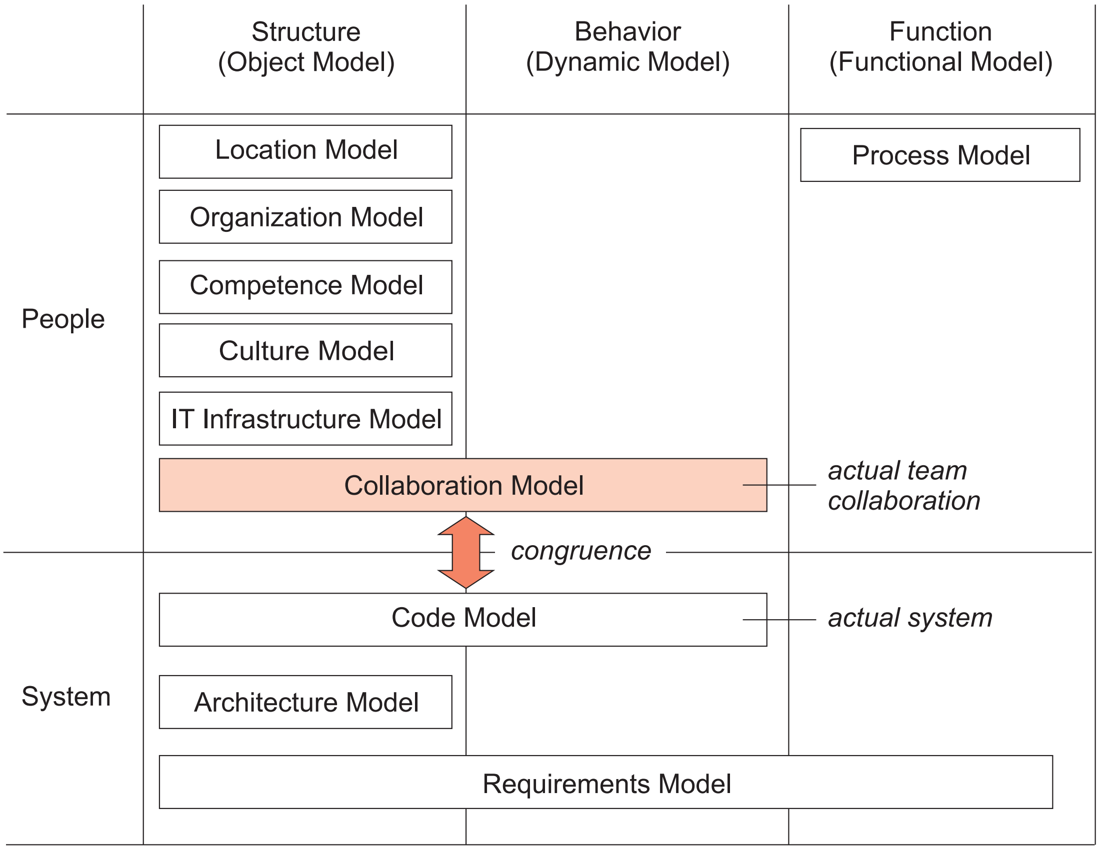


Figure 4.5: Models of Global Software Engineering

Location Model

The Location Model determines the geographic distribution of a GSE team, i.e. the development site, building and room where the team members are located.

The GSE-related challenges concerning the location model are:

- **Geographic distance.** Distance and communication are negatively correlated. This was shown by Thomas Allen in a study of over 500 individuals in seven organizations over a duration of six months. Allen's study revealed that communication drops significantly when engineers are more than about 30 meters away from each other [All84]. The interesting fact is that even when teams are dispersed to multiple adjacent buildings, or just different floors of the same building, their communication drops to such an extent that the impact on communication is comparable to globally distributed teams [Car99, p.44].

- **Degree of distribution.** The number of sites determines the degree of distribution, i.e. the more sites are involved, the more challenging becomes GSE [BHL09].

- **Different time zones.** In association with the geographic distribution of a GSE team, also time zone differences (temporal distance) are challenging in GSE [Car99, BHL09]. Working in different time zones impacts reachability across sites and thus the ability to communicate.

The central heuristic addressing the location aspect is the following:

- **Adjust location model.** This heuristic aims at increasing the ability to collaborate by changing the location model, e.g. by co-locating team members. According to Carmel, reasons for *not* decentralizing a team usually include more control, less duplication of effort and wasted effort, and a better ability to maintain a corporate culture [Car99, p.44]. However, in many cases the possibility to change the location model to co-location are very limited and is beyond the power of the project manager in charge. Best practices related to the co-location of team members include cross-site delegations of team members for e.g. half a year, or to co-locate key members of the team for a certain activity e.g. training at the beginning of the project) [Les10].

Organization Model

The Organization Model is concerned with organizational units, e.g. companies or company departments, and their responsibilities, involving concepts like reporting structures, decision making rules, and communication structures [BD10, p.118], but also organization-specific goals and incentive structures.

Challenges:

- **Coordination of suppliers.** In a globally distributed setting, the coordination of suppliers in the sense of onshore or offshore outsourcing is particularly challenging, e.g. with respect to legal agreements, the communication of requirements or transferring essential domain knowledge to the suppliers.

- **Unclear escalation path.** In globally distributed projects, it is often not clear to whom issues related to cross-site collaboration can be escalated [HL09].

- **Divergent incentive structures.** GSE projects often need to deal with different incentive structures of the involved organizations. For example, a subcontractor organization usually follows different goals than a unit of the same company. Also project priorities may vary, e.g. because the current project is less important to one site than other (possibly local) projects [BHL09].

Heuristics:

- **Supplier management.** The whole area of supplier management is concerned with effective handling of supplier relationships. An introduction to supplier management is out of the scope of this dissertation. We refer to existing literature for more information, e.g. Sangwan, chapter 10 [SBM+06].

- **Establish clear escalation path.** To allow for escalation to address critical issues, a unique escalation path, spanning the involved locations and organizations, needs to be defined and communicated to all team members [HL09].

- **Establish common goal.** To avoid that the organizations involved in GSE projects pursue different objectives, establish common goals and harmonize incentive structures across organizations [BHL09].

Competence Model

A Competence Model is concerned with the capability of the workforce, which can be defined as "the level of knowledge, skills, and process abilities available for performing an organization's business activities." [CHM09, p.5]

Challenges:

- **Domain expertise.** In GSE projects it occurs frequently that team members of specific sites don't have sufficient expertise in the relevant domains of the product under development [BHL09], which increases the need for communication and the risk of making wrong assumptions with respect to the requirements. If for example a medical product is developed, a basic understanding of the context of e.g. a hospital workflow in which the product will be used is crucial.

- **Technology expertise.** Team members need sufficient expertise on the technologies required for the project. This involves for example programming languages, libraries, protocols, middleware, etc. [BHL09].

- **Incompatible language skills.** Inadequate language skills can lead to inefficient communication and misunderstandings [BHL09, HL09].

- **Experience on GSE projects.** If parts of the team lack experience in working in a globally distributed development team, it constitutes an additional challenge [BHL09].

Heuristics:

- **Training.** By providing adequate training for technology, domain knowledge or language, the skills of the team members can be improved [HL09].
- **Team selection.** Team members can be systematically selected for their suitability for collaboration in a GSE project. *Note: This heuristic affects the Organization Model.*

Culture Model

The Culture Model is dealing with the aspect of culture in a distributed team, where culture refers to "the integrated pattern of human knowledge, belief, and behavior that depends upon the capacity for learning and transmitting knowledge to succeeding generations" [MW11i]. GSE team members have to deal with cultural differences. This affects multiple types of culture: Obvious is the *national culture* of the country of origin of the team members. But besides there are also other types, such as *corporate culture*, associated with the company or subsidiary [Car99, p.60].

Challenges:

- **Cultural differences.** Cultural differences (also: *cultural distance*) can lead to misunderstandings and ineffective collaboration [Car99, p.57][BHL09] [HL09].

Heuristics:

- **Create culture awareness.** One typical heuristic is to create awareness for cultural differences within a team [HL09], see Section 4.6.

Carmel is referring to the work of Geert Hofstede [HH04] and Edward T. Hall for the fundamentals of cultural differences. An interesting aspect discussed by Carmel is whether *professional culture*, which developers share across nations and companies, dominates national culture. Carmel presents studies in favor and against that 'software culture' – the professional culture prevailing in the software engineering domain – can be stronger than national culture [Car99, p.73–77]. For instance, Hofstede found significant national cultural differences, although all participants shared the same professional culture. Finally, it becomes clear that differences in national or corporate culture are a major challenge also when people share the same professional culture.

IT Infrastructure Model

The IT Infrastructure Model is concerned with communication and development tools that are used for a GSE project. Typical communication media include e-mail, phone, wiki, chat; typical development tools are for example a multi-site version control system or a change management system.

Challenges:

- **Divergent development environment.** Different development tools can hamper cross-site collaboration [BHL09]. For example, different debugging tools can make it difficult to reproduce a bug reported by another development site.
- **Unreliable or inadequate IT/telecom infrastructure.** Frequent network outages, bandwidth limitations, or restrictions on network usage impact the ability to collaborate [BHL09].

Heuristics:

- **Align development environment.** Development environments should be aligned to each other to avoid incompatibilities and to allow for a smooth collaboration [BHL09, HL09].
- **Collaboration technology.** Make use of adequate collaboration technology to provide good support for globally distributed collaboration [Car99, p.89,92,114], e.g. by making video conferencing and application sharing available.

Process Model

The Process Model is concerned with the development processes used in GSE projects. It is a functional model in the sense of defined activities with preconditions and postconditions; with respect to deliverables and milestones it can also be seen as structural model.

Characteristic GSE challenges related to the Process Model are:

- **Divergent development processes.** Frequently, the sites involved in a GSE project either use different development processes or interpret the development process in different ways [BHL09, HL09].
- **Define a suitable development process**. The development process needs to be defined in such a way that it helps to ensure the necessary cross-site communication and synchronization [HL09].

Heuristics:

- **Harmonize development processes.** Collaboration can be improved by harmonizing development processes and to establish them as a ‘common language’ across sites [HL09, Car99].
- **Use iterative or agile processes.** By using iterative or agile processes, frequent cross-site communication is furthered [PDL09]. Small increments ensure that problems become visible early on [Eck10].

Collaboration Model

The Collaboration Model represents communication structures and behavior of the team members in a GSE project. This model is the focus of our GSE metrics approach.

Typical challenges include:

- **Impeded communication and awareness.** Distribution impedes communication and awareness. Typical challenges are difficulties of knowing who to contact about what, of initiating contact, and of communicating effectively across sites [Her07].
- **Loss of communication richness.** According to Carmel, GSE causes a loss of communication richness and therefore impacts the ability of the team to collaborate effectively [Car99, p.48]. GSE teams use different kinds of communication media, e.g. video conferencing, telephone, or e-mail. They differ with respect to the level of communication richness, where *rich communication* is defined as two-way interaction involving more than one sensory channel, e.g. personal communication. Only a certain part of the message we communicate is in the explicit text we transmit; a substantial portion is nonverbal and implicit, communicated via body language such as gestures, facial expressions, and postures (see also Chapter 6). Intensive problem solving, design, or conceptual collaboration works best using a rich communication channel [Car99, p.48].
- **Team cohesion.** Successful teams are cohesive: Team cohesion – “the act or state of sticking together tightly” [MW11h] – leads to enhanced motivation, increased morale, greater productivity, harder work, more open communication, and higher job satisfaction compared to non-cohesive groups [Car99, p.52]. Team cohesion is more difficult to achieve for a globally distributed team; distance is an impediment to building relationships of trust [Car99, p.53]. A team needs to go through the stages of *forming*, *storming*, and *norming* until it reaches maturity and effective performance in the *performing* stage, as conceptualized in the stage model of team developmental

maturity by Tuckman [Car99, p.54]. In the first phase, the forming stage, the team gets to know each other and clarifies roles, tasks and objectives. Conflicts arise in the storming phase, different leaders evolve and there are different individual goals. When the team moves on to the norming stage, it forms norms, roles and protocols for working together. Finally the team reaches the performing stage, where it is performing well, has a common goal and conflicts are handled constructively. Distributed teams take longer to move through the early stages due to distribution [Car99, p.54].

- **Coordination breakdown.** GSE causes a breakdown of control and coordination mechanisms which were effective in co-located projects. Carmel defines *control* as the process of adhering to goals or policies or standards, and *coordination* as the act of integrating each task and organizational unit so that it contributes to the overall objective [Car99, p.45]. Coordination and control concepts imply a balance between formal mechanisms (e.g. regular meetings) and informal mechanisms (e.g. a chat in the hallway). This balance is distorted in distributed teams. A study by Perry et al. showed that co-located developers spend 75 minutes per day in unplanned, informal, interpersonal interactions, which are not part of the scheduled meetings [Car99, p.47]. This kind of coordination is not possible in a globally distributed setting.

- **Divergent management practices.** In addition to the coordination breakdown described above, incompatibilities between management practices at the various sites involved in a GSE project can also cause difficulties, for example monitoring of overall progress becomes difficult, if the involved sites don't monitor progress in the same way [BHL09].

Typical heuristics related to collaboration are the following:

- **Team building.** Team building measures are used to build trust and improve team cohesion, e.g. a joint kick-off meeting [Car99, p.169] or an icebreaker [KK09].

- **Project retrospectives.** A project retrospective is a team retreat where issues in team collaboration are discussed and reflected. It usually involves a facilitator external to the team who leads the retrospective, which has the advantage that he or she is in a neutral position and thus can build a safe environment for the participants of the retrospective. The focus of a project retrospective is on learning, not fault-finding. All participants of a retrospective need to "truly believe that everyone did the best job he or she could" (see Kerth's prime directive for retrospectives, [Ker01, p.7]) Project retrospectives can be held after completion of a project or during a project. The idea of project retrospectives is also embedded in the principles of agile development (see agile manifesto): “At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.” [BBvB+01]. A project retrospective consists of three phases: a *readying* phase where the team is prepared for the retrospective and trust is built, a phase concerned with reflection of the *past* to understand what actually happened and a *future* phase with the objective to jointly decide on what to keep and change to improve collaboration in the remainder of the project resp. the next project. Kerth provides a comprehensive handbook for how to conduct project retrospectives in software engineering, including proven team exercises for the readying, past and future phase [Ker01].

4.4.2 System-related Models

System-related models are concerned with the technical system under development. The system-related models in the GSE meta model are: Code Model, Architecture Model and Requirements Model.

Code Model

The Code Model is the representation of the actual system under development, i.e. the source code.

Challenges:

- **Integration difficulties.** When subsystems are developed by different sites, often there can be problems with integration of these subsystems.

- **Quality problems.** Another characteristic challenge are quality problems with parts of the system.

Heuristics:

- **Continuous integration.** Use continuous integration approaches to ensure that the various subsystems fit together well. *Note: This heuristic involves the Process Model.*

- **Code quality management.** Implement code quality management measures, e.g. static code checking, to ensure a certain code quality.

Architecture Model

The Architecture Model refers to the top level design of the system under development, which implies essential decisions with respect to the system's structure.

Challenges:

- **Architectural change.** Changes in the system architecture usually have significant impact on multiple development sites involved in a GSE project. Architectural changes imply a high communication need [BHL09].

- **Uncertain component interfaces.** Component interfaces often affect multiple sites when the development of subsystems is distributed. If component interfaces are not well defined or need to be changed, this causes a high need for communication [BHL09].

- **Uncertain allocation of functionality.** If the allocation of functionality is not clearly defined, it imposes a challenge on the GSE project. Significant communication is necessary to achieve a clarification of which functionality is to be developed where [BHL09].

Heuristics:

- **Low coupling.** The term coupling is defined as "the strength of the dependencies between two subsystems or two classes. Low coupling results in subsystems that can be modified with minimal impact on other subsystems." [BD10, p.762] Low geographic coupling reduces cross-site dependencies.

- **High cohesion.** Cohesion means "the strength of dependencies within a subsystem or a class. High coherence is desirable as it keeps related classes together so they can be modified consistently." [BD10, p.760] High cohesion is preferable, especially in GSE: Objects to be developed are assigned to the same development site.

Requirements Model

The Requirements Model contains the requirements, i.e. the desired functionality, non-functional properties and any constraints of the system to be developed.

Challenges:

- **Uncertain requirements.** When requirements are not clear or changing, it is likely to affect multiple development sites and thus create a high need for communication [BHL09].

- **Stringent non-functional requirements.** Non-functional requirements, e.g. performance, security, reliability, availability affect all parts of the system and therefore all development sites [BHL09].

- **Central components or cross-cutting features.** Central components such as an alarm handler or features which affect multiple parts of the system developed by different sites introduce cross-site dependencies and imply a high need for communication [BHL09].

Heuristics:

- **Requirements engineering.** The whole field of requirements engineering is dedicated to improving definition and management of requirements and thus addressing the challenges in the Requirements Model, e.g. by improving requirements elicitation, ensuring a high quality of requirements documents, or approaches such as model driven requirements engineering. An introduction to requirements engineering is not in the scope of this dissertation. For more information, we refer to existing literature, e.g. Davis [Dav93] and Wiegers [Wie03].

According to Conway's Law, "A design effort should be organized according to the need for communication", because "organizations which design systems are constrained to produce designs which are copies of the communication structures of these organizations" [Con68]. In other words: There is a congruence between the people and the system level of the GSE meta model, in particular between the Collaboration Model and the Code Model.

4.5 Organizational Patterns in GSE

Herbsleb presented a study on organizational patterns in GSE and related challenges [Her06], affecting the Location, Organization and Architecture Model mentioned above. The study was based on more than 150 interviews at 14 sites on 3 continents, analyzing location and organization structures that evolved in different product lines at Lucent. The organizational structures found can be categorized along three dimensions: Subsystem, Process and Release. Herbsleb used a three-dimensional cube representation to illustrate the respective organizational structures. As we found the organizational patterns and the graphical representation very valuable for understanding communication needs and tradeoffs in structuring GSE projects, we want to give a brief overview in this section. We also use Herbsleb's cube representation, which has been redrawn and slightly simplified (see Figure 4.6). In the following, we walk through various ways how to slice the cube. In the figures, the work split between two different development sites are shown – site 1 and site 2 –, which are depicted by two different colors.

Separation by Subsystems. In the organizational pattern *Separation by Subsystems*, responsibility of sites is split according to subsystems of the system under development. Figure 4.6 gives an example: While site 1 is in charge of

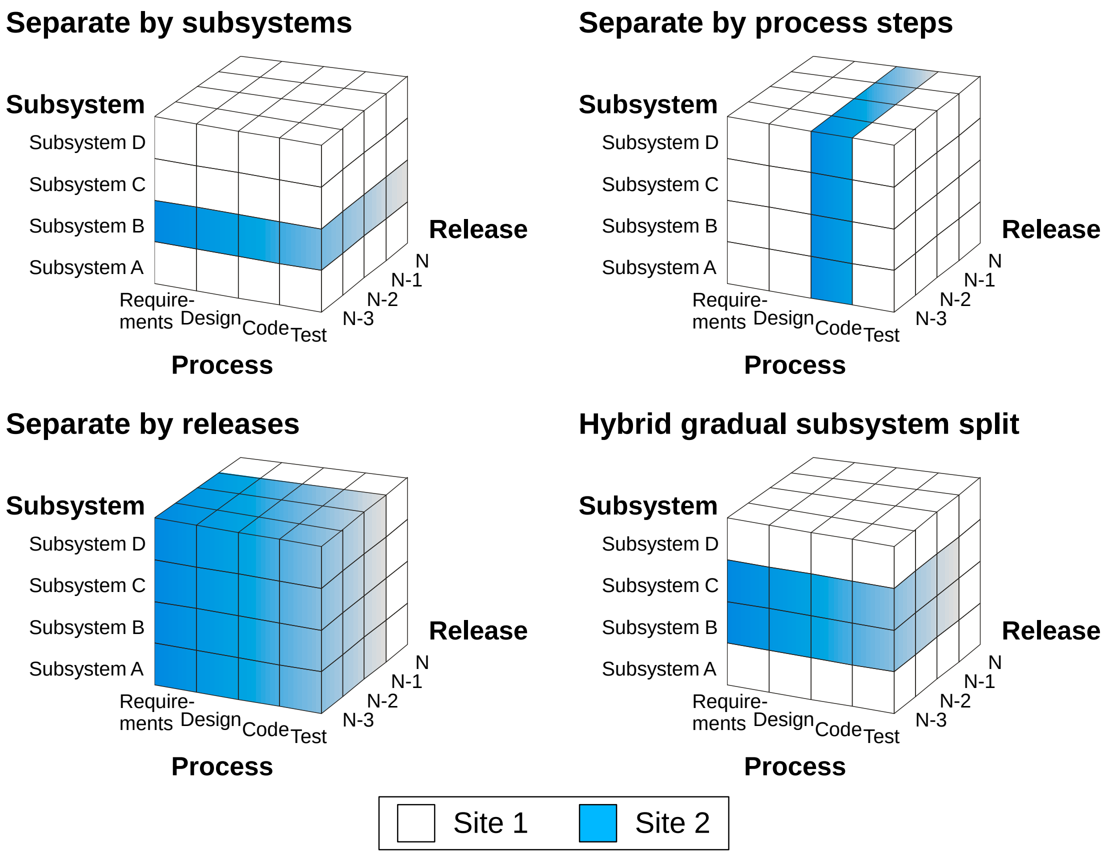


Figure 4.6: Organizational patterns in GSE according to Herbsleb [Her06]

subsystem A, C, and D, site 2 is responsible for development of subsystem B. Both site 1 and site 2 take care of all development activities related to their subsystems, i.e. requirements, design, code and test. The advantage of this pattern is that site 1 and site 2 can develop their respective subsystems relatively independent from other sites, because dependencies across sites are minimized. The development sites might use different development processes, the establishment of synchronization milestones are sufficient. The Separation by Subsystems pattern is usually chosen in case of modular systems with clear and stable interfaces. A main challenge of this organizational pattern is integration, as the subsystems which need to be integrated were developed independently before. Furthermore, the implementation of central components, cross-cutting features (such as uniform alarm handling by all components) and non-functional requirements (such as performance or security) is challenging in this setting.

Separation by Process Steps. In the second pattern, *Separation by Process Steps*, the work between sites is split along the development process. In the example in Figure 4.6, site 1 is responsible for requirements, design and test, while coding is done at site 2. This pattern allows leveraging time zones of site 1 and site 2 to speed up the test and fix cycle; when the working day at site 2 ends, site 1 can take over the development results and test them 'over night'. Additionally, the Separation by Process Steps pattern is often used to gain experience with GSE, using the so-called *extended workbench* approach as a first step. The major challenge of this pattern is the high cross-site communication required. This is particularly problematic when fixing defects, because problem detection and rework occur at different sites. Furthermore, this setting is usually inflexible to changing plans, because resources e.g. at the coding site are usually allocated to multiple projects and therefore availability of resources is restricted.

Separation by Releases. The third dimension is the release of the system under development. In the pattern *Separation by Releases*, the current and previous releases are taken care of by different sites. The example in Figure 4.6 shows e.g. that site 1 has the responsibility to develop the current release (original developers), while site 2 is maintaining all the previous releases (maintainers). Thus the current release is still developed within one site, reducing the need for cross-site communication. Another advantage is that site 2 can learn about the whole system, while being under less pressure than in the development of the current release. The Separation by Releases pattern normally works best, if previous releases will remain stable, typically with no new functionality and few bug fixes only. It is usually used if the current release is much more critical to business than old releases, and the organization wants to expose the new site to the system. Challenges include the required cross-site communication between maintainers and original developers, which was observed to slow down both. Another disadvantage is that maintainers need to learn about the whole system at once. There is a high effort for synchronizing bug fixes. A particular critical challenge is that maintainers are likely to feel disrespected as they do 'less valuable work' compared to the original developers.

Gradual Subsystem Split. Also a combination of the above organizational patterns is possible This is reflected by the *Gradual Subsystem Split* pattern, where responsibility for part of the system is moved, but not for the current release. In the example in Figure 4.6, site 2 has the responsibility for development of the previous releases of subsystem B and C, while site 1 takes care of the other subsystems A and D as well as all the whole current release. This is usually an interim step to take over responsibility for part of the system and used if the new site is inexperienced with the system, but the long term goal is subsystem responsibility at the new site.

In addition to the aforementioned organizational patterns, in his study Herbsleb also observed a split between development (including requirements, design, code and test) and customization, e.g. country-specific adaptations, which shall not be further explained here as it is basically a specific case of the pattern Separation by Process Steps. The organizational patterns observed at Lucent are mainly related to waterfall development processes. However, they also provide valuable insights for iterative-incremental and agile software development.

4.6 Culture

Cultural differences can easily lead to misunderstandings and problems in collaboration, when development sites with different cultural backgrounds are supposed to work together. We don't focus on culture in the definition of GSE project management metrics, because culture is rather static in a development team, unless the team structure is changed, as opposed to communication or team cohesion. However, culture can be the root cause of communication problems, and we consider culture an essential topic in GSE. Therefore we give an introduction on the topic of culture and provide an overview of the cultural dimensions measured by Hofstede in the remainder of this chapter.

4.6.1 Culture as 'Software of the Mind'

According to Hofstede, culture is "the collective programming of the mind that distinguishes the members of one group or category of people from another", influencing people's behavior and thoughts [HH04, p.4]. Besides *human nature* which is universal to all people and *personality* which is individual for each person, culture determines the way we act, think and feel. Culture is learned (not inherited) and specific to a group or category, for example a nation or a company. Furthermore, culture doesn't change quickly, as it reproduces itself through transmission from generation to generation. While the visible *practices – ritual, heroes* and *symbols* – change more frequently, the *values* behind are very stable [HH04, p.9].

4.6.2 Creating Awareness for Cultural Differences

To introduce the topic of culture to students and to create awareness for cultural differences, we use an exercise called "The Derdians" – an experiment simulating the meeting of two cultures: A team of engineers goes to an imaginary country "Derdia" in order to teach the population of Derdia how to build a bridge [GHKdJCG+00, p.62]. It was conducted at our master seminar on Global Software Engineering in June 2010. The thirteen students were allowed to choose between two roles: the engineers (group of 4 persons) and the Derdians (group of 9 persons). After the students selected their role, the engineers were taken to a separate room, where they received a sheet with instructions and were asked to start with their preparations[3]. According to the instruction sheet, the engineers' task is to bring the innovation of a bridge to Derdia, which is a country with mountains and valleys, but no bridges. The bridge has to be built in Derdia, as the Derdians have to learn all construction stages to be able to build future bridges on their own. The bridge is symbolized by a paper bridge of approx. 80 cm length, and has to be stable at the end of the exercise, i.e. it "should support the weight of the scissors and glue used in its construction" [GHKdJCG+00, p.65]. Available materials include paper resp. cardboard, glue, scissors, ruler and pencils. According to the rules of the exercise, each piece has to be drawn by pencil and ruler and then cut out with scissors. The instructions also inform about the time line: The overall time for preparation before going to Derdia is 40 minutes, the time available for teaching the Derdians to build the bridge is 25 minutes. In the middle of their preparation work (approx. after 25 minutes; the engineers should wait until they are picked up by the the instructor), two members of the engineers' team are allowed to make contact for 3 minutes with the Derdians, followed by another 10 minutes for analysis of their observations and finishing preparation.

Meanwhile also the Derdians received written instructions and prepared for their role. What the engineers didn't know was the kind of culture prevailing in Derdia. The Derdians had specific habits in their culture with respect to social behavior, communication and work [GHKdJCG+00, p.63]:

- **Social behavior.** The Derdians always touch each other when communicating. Not being in contact while talking is considered very rude.
- **Greetings.** Derdians greet with a kiss on the shoulder: the first person starts with a kiss on the right shoulder of the counterpart, the other responds with a kiss on the left shoulder.
- **Insult.** If a Derdian is insulted by not being greeted or not touched while communicating, he/she starts shouting loudly about it.
- **Yes/No.** Derdians always say 'yes', even when they mean 'no', but then accompanying the 'yes' with empathic nodding.
- **Work behavior.** Tools are gender-specific: Scissors are used by men only, while pencil and ruler are female tools. Glue can be used by persons of both genders.
- **Foreigners.** A Derdian man will never get in contact with another man unless he is introduced by a woman (no matter if Derdian or not).

[3] The full instruction sheets can be found in the *Intercultural Learning T-Kit No. 4* by Gillert et al. [GHKdJCG+00, pp.62–65]

Figure 4.7: The 'Derdians' in action: construction and final load test of the bridge

At the master seminar the meeting of the two cultures of the engineers and the Derdians brought new insights to the students about dealing with cultural differences. It occurred that during the first contact, the two members of the engineers' team examined the available building material and tools only, not paying any attention to the people of Derdia whom they were supposed to collaborate with. When the actual build phase started, the engineers experienced kind of a culture shock. However, they figured out the cultural rules quickly and were able to adapt to them. Figure 4.7 shows the Derdians and engineers constructing the bridge and the final load test, which was passed successfully. The students mentioned that the exercise was a lot of fun to them and at the same time an eye-opener for cultural differences. Although the culture of the Derdians was imaginary, they have seen a lot of parallels in meetings between different cultures in real life.

4.6.3 Cultural Dimensions

Hofstede has done a ground-breaking study in measuring cultural differences. He conducted a broad survey inside IBM and collected data from IBM employees in more than 50 countries. Based on the data collected, Hofstede analyzed cultural differences between the countries involved. The study was replicated by others later on, which delivered additional data.

Hofstede identified five cultural dimensions – Power Distance, Individualism, Masculinity, Uncertainty Avoidance and Long-Term Orientation – which we describe in the following.

Power Distance. The first cultural dimension, Power Distance, is "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally" [HH04, p.46]. Countries with a low Power Distance Index (PDI) treat everyone as equal as possible, while a high PDI indicates an unequal distribution of power. In countries with a high power distance, a basic virtue is respect and obedience towards authorities; in turn the superior takes care and supports the inferiors. At work bosses are supposed to be benevolent autocrats, and inferiors pay the superior ones respect and expect to be told what to do. High PDI usually means strong vertical hierarchies as well as wide gaps in salary between top and bottom of the organization.

Individualism. Hofstede's second dimension is concerned with how much people of a culture are seen as individuals (individualism) as opposed to group members (collectivism). According to Hofstede, "Collectivism [...] pertains to societies in which people from birth onward are integrated into strong, cohesive in-groups which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty" [HH04, p.76], while "Individualism pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family" [HH04, p.76]. In cultures scoring high on Individualism (IDV), everyone is supposed to think of oneself as 'I' rather than 'we', and encouraged to voice one's own opinion. At work, management means management of individuals rather than groups. In contrast, in a collectivistic culture, harmony inside the group is important, as an individual is strongly dependent on his or her group. In the workplace, bosses choose and treat employee's according to the group they belong to.

Masculinity. The third cultural dimension refers to the distribution of emotional roles between the genders, masculinity as opposed to femininity. In a masculine culture, "emotional gender roles are clearly distinct: men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be more modest, tender, and concerned with the quality of life" [HH04, p.120]. The opposite, a feminine culture is present "when emotional gender roles overlap: both men and women are supposed to be modest, tender, and concerned with the quality of life" [HH04, p.120]. In cultures with a high masculinity index (MAS), one lives in order to work, job choices are based on career opportunities, management makes decisions in an assertive way. Conflicts are carrier out based on the principle 'let the strongest win'. In contrast, in feminine cultures (low MAS) people work in order to live. Job choices are based on personal interest rather than status and career opportunities. Free time is preferred over more income. Management tries to handle conflicts by negotiation and compromise instead of fighting. Virtues are modesty and the ability to get along with everybody.

Uncertainty Avoidance. Hofstede's fourth cultural dimension, Uncertainty Avoidance, is "the extent to which the members of a culture feel threatened by ambiguous or unknown situations" [HH04, p.167]. In societies with a high Uncertainty Avoidance Index (UAI) ambiguity is not tolerated, and firm rules are established in great detail of what is right and what is wrong. At work, employees tend to stay longer with a company. They focus on facts and on substance in a task and are better at implementing things because of their great discipline. A low UAI on the other hand implies a high tolerance for ambiguity and uncertainty. There is no need for many rules, and rules are questioned if they have no obvious meaning. In the workplace, employees will change the employer frequently. They focus on common sense, and their free thinking enables them to be better at inventing things.

Long-Term Orientation. The fifth cultural dimension was discovered first in an Asian study, the so-called Chinese Value Survey [HH04, p.29] and adapted later by Hofstede. It distinguishes long-term and short-term oriented cultures. "Long-term oriented societies foster pragmatic virtues oriented towards future rewards, in particular saving, persistence, and adapting to changing circumstances. Short-term oriented societies foster virtues related to the past and present such

Country/Region	PDI	IDV	MAS	UAI	LTO
Arab Countries	80	38	53	68	n/a
China	80	20	66	30	118
France	68	71	43	86	39
Germany	35	67	66	65	31
Great Britain	35	89	66	35	25
India	77	48	56	40	61
Japan	54	46	95	92	80
Russia	93	39	36	95	n/a
Sweden	31	71	5	29	33
USA	40	91	62	46	29

Table 4.4: Cultural dimensions of selected countries according to Hofstede: Power Distance (PDI), Individualism (IDV), Masculinity (MAS), Uncertainty Avoidance (UAI) and Long-Term Orientation (LTO) [HH04]

as national pride, respect for tradition, preservation of 'face', and fulfilling social obligations." [Hof11] This dimension is measured in the Long-Term Orientation Index (LTO). In business, long-term investments are valued more than quick returns. Achieving and maintaining a stable market position has priority over immediate results. Tradition is seen as impediment to innovation and is therefore de-emphasized. Caring for both social and business relationships is considered essential. In cultures with a low LTO, respect for tradition and social rules are important. In business, there is less innovation because of the need to preserve tradition. Immediate profits are more important than savings or cutbacks for a long-term stable market position.

Table 4.4 shows a set of selected countries with their values in Hofstede's five cultural dimensions.

4.7 Summary

In Chapter 2 we stated that geographic, temporal and cultural distance in general impede collaboration, but are no adequate indicator of collaboration difficulties. We claimed that a new measurement is needed referring to *collaboration distance*. With the GSE challenges presented in this chapter, we identified aspects relevant for collaboration distance.

Chapter 5

Social Psychology of Distributed Teams

This chapter deals with an interdisciplinary view of GSE teams: the social psychology of distributed teams.

5.1 Intergroup Relations

It is repeated in GSE literature that team cohesion and trust are important success factors, because distribution impairs the relationship between teams (see also Chapter 4, where we discussed challenges such as lack of trust, lack of informal communication and loss of teamness). For example, Herbsleb and Mockus state in their studies on globally distributed development teams:

> There is substantial evidence that distant colleagues feel less 'teamness,' experience some conflict in work styles and report that their distant colleagues are less likely to help out when workloads are especially heavy. The lack of 'teamness' presumably reflects the fact that distance interfered with the usual stages by which individuals become coherent groups or teams. [HM03]

To better understand the reasons behind these effects between distributed teams, in the following we take an interdisciplinary view, utilizing existing research from the field of social psychology.

5.1.1 The Robbers Cave Experiment

Sherif and colleagues [She88] investigated group behavior and intergroup relations when individuals having no established relationships are brought together. This study was also recently cited in software engineering context, see keynote 'Who do you trust?' by Linda Rising at AGILE'08. The *Robbers Cave Experiment* is seen as one of the most successful field experiments on intergroup relations ever conducted [Bro86, p.535]. Sherif et al. were able to analyze the behavior in group formation, the effects of competition between groups, and ways of conflict resolution, all in realistic context. The subjects – twenty-two eleven year old boys – were taken to a summer camp for three weeks. They were completely unaware that an experiment was going on with them. The researchers all took roles in the camp (mainly counselors) and used their breaks during the day and evenings to make notes on their observations. To backup the findings, the observation method was combined with other methods, such as filling in questionnaires with sociometric data and other laboratory-type methods which were embedded into the lifelike situation at the camp. Instead of telling the subjects what to do in the experiment, the researchers placed them in demanding problem situations and observed their behavior.

The Robbers Cave Experiment is an experiment in social psychology. In social psychology, two disciplines are combined: psychology and sociology. *Psychology* is "the science of mind and behavior" [MW11a]. On psychological level, subject of the analysis is the *individual* and its psychological functioning with respect to aspects such as motives, judging, perceiving, learning, etc. [She88, p.5] On the other hand, *sociology* refers to "the science of society, social institutions and social relationships" [MW11c], subject of analysis is the *social group*. Sociology deals with concepts such as the social organization, institutions, value systems, etc. [She88, p.5] Consequently, social psychology is "the study of the manner in which the personality, attitudes, motivations, and behavior of the individual influence and are influenced by social groups" [MW11b]. Sherif et al. emphasized the importance of the two level concept – the psychology of the individual and the sociology of the group – for their research, allowing for a deeper understanding and comparing the findings obtained on different levels with each other [She88, p.6].

To distinguish between groups, Sherif et al. use the terms ingroup and outgroup: According to Sherif, a group is defined as "social unit that consists of a number of individuals who, at a given time, stand in more or less definite interdependent status and role relationships with one another, and that explicitly or implicitly posseses a set of values or norms regulating the behavior of individual members, at least in matters of consequence to the group" [She88, p.26]. From the point of view of individuals belonging to a specific group, the defined social unit of their own group is referred to as *ingroup*. The social units of which they are not a part are called *outgroups* [She88, p.26].

General Design of the Study

The focus of the study was on intergroup relations. "As an experiment in social psychology, the study undertook to trace over a period the formation and functioning of negative and positive attitudes, as a consequence of experimentally introduced situations, of members of one group toward another group and its members" [She88, p. 24]. The Robbers Cave Experiment consisted of three successive stages, each of approx. one week duration: 1) ingroup formation, 2) friction phase, and 3) integration phase. The purpose of stage 1 was the formation of two separate ingroups of subjects that previously have not been acquainted with one another. Stage 2 introduced situations of competition to produce friction between the two groups. Finally, the objective of stage 3 was to try out approaches of conflict resolution to reduce tension between the groups.

Selection of Subjects and the Experimental Site

High effort was put into a careful selection of experimental subjects. The goal was to have two balanced groups. Subjects had to be "well-adjusted boys of the same age and educational level, from similar sociocultural backgrounds, and with no unusual features in their personal backgrounds insofar", i.e. no problem cases or isolates [She88, p.53]. Furthermore, they must not be acquainted with one another before the experiment. In this way it was ensured to rule out effects from existing relationships and previous experience of unusual degrees of frustration at home or school. In a comprehensive selection process, researchers first approached multiple schools in Oklahoma City. Their first contact was to principals of appropriate schools. After showing a letter of recommendation from higher school authorities, the effort was briefly described as experimental camp of the University of Oklahoma with the purpose to study interaction within and between groups [She88, p.55]. In order to avoid that principals and teachers would recommend their favorite boys, the researcher explained that he or she would like to go to the school yard and pick candidates from first hand observation. At the playground, the researcher watched for about five to ten candidates suitable for the experiment, and asked the playground supervisor for the name of the boys and more characteristics. Next, the researcher consulted the homeroom teacher to get further information on each boy [She88, p.56]. This procedure usually resulted in five or six potential subjects per school. In the manner described, about 200 potential subjects were identified.

In the next step, the parents of about 50 boys that seemed most suitable were interviewed. The study was described to them accurately, but not exhaustively, as a three-week summer camp under adult supervisions to see how the boys would work together. Parents were not permitted to visit their boys during the three weeks period, however, boys were allowed to leave early if they wanted to go home. Finally, 22 subjects were selected for the experiment. They were assigned by researchers to two comparable groups.

The experimental site was chosen according to three criteria: 1) the site had to provide isolation from the outside world to avoid external influences, 2) separate facilities for two groups were needed to keep them separate during ingroup formation, and 3) the camp and the surrounding area had to allow flexibility in choosing and planning ingroup and intergroup problem situations. After inspection of several camps, a Boy Scouts of America camp surrounded by Robbers Cave State Park was chosen, located in a densely wooded area in the Sans Bois Mountains of southeastern Oklahoma, which fulfilled the desired criteria [She88, p.59].

Stage 1: Ingroup formation

The aim of stage 1 was the formation of two independent ingroups. The two groups were transported separately by bus to their own facilities at the Boy Scouts of America camp surrounded by Robbers Cave State Park. In this stage, the groups didn't know of each other. Group activities included hiking, swimming, preparing meals from bulk ingredients, joint tent pitching, and a group treasure hunt. A variety of problem situations arose that required cooperation, such as transportation of boats and equipment, planning and executing hikes, improving swimming places, organizing campfires, or preparing meals in the woods with bulk ingredients [She88, p.86]. In both groups, a leadership structure and a status hierarchy evolved. Both groups developed own norms regulating the behavior in relations with one another. In one group, a norm of being 'tough' and not complaining about injuries was established, while at the same time cursing a lot was the way to behave. In the other group swimming nude became the norm, and after two boys left the camp due to homesickness, it was taboo to show feelings of homesickness. Both groups gave themselves an own group identity – the 'Rattlers' and the 'Eagles' – and stenciled T-shirts, caps and a flag with it. Near the end of stage 1, the Rattlers and the Eagles became aware of the presence of the other group: Hearing distant voices, or finding paper cups at a hideout where they had left none, leading to resentful discussions that 'outsiders' had been there [She88, p.71]. Immediately the ingroup-outgroup language of 'us versus them' was observable. "They better not be in our swimming hole." [She88, p.94] "Those nigger campers" [She88, p.95] The Rattlers and the Eagles expressed that they wanted to challenge the other group of boys.

Stage 2: Friction phase

Stage 2 introduced situations of competition to produce negative attitudes toward the outgroup. For this purpose, a tournament between the Rattlers and the Eagles was announced. It consisted of ten sport events (3x baseball, 3x tug-of-war, 1x touch football, and 3x tent pitching). Additionally, there were five events judged by staff members (3x cabin inspections, 1x skits and songs, 1x treasure hunt). The scores from the latter events were used to ensure a neck-to-neck competition between the two groups until the end of the tournament. Prizes of great appeal value to the boys were offered as reward: The group with the highest cumulative score in the contests would win a trophy, and each member of the winning group would receive a medal and a four-bladed knife [She88, p.98].

When the two groups came into contact for the first time, they verbally harassed each other [She88, p.98]. During the first two days of the tournament, the norm of 'good sportsmanship' prevailed, but gave way soon to "increased name calling, hurling invectives, and derogation of the outgroup" [She88, p.101]. After the Eagles had lost the first baseball game and the first tug-of-war, they took the Rattlers' flag, burned it and hung the scorched remnant back up [She88, p.109]. Now war began: The next morning, when the Rattlers had finished breakfast and discovered their burned flag, they became furious. The Rattlers went to the Eagles and asked if they burned it, and when they admitted, a fierce fight started. Staff had to intervene. Later on, at the second tug-of-war, the Eagles used the strategy to sit down and dig in their feet. The Rattlers were totally exhausted when after 7 minutes of tugging strenuously they sat down and dug in, too. In the following the Eagles were able to pull the tired Rattlers across the line [She88, p.111]. The Rattlers perceived this strategy as totally unfair and, later at night when the Eagles were sleeping, they made a commando raid on the Eagle's cabin, turning beds over and ripping mosquito netting. Most Eagles wanted to retaliate that night, but the researchers prevented the act of retaliation when they heard from the Eagles that rocks should be used [She88, p.112]. The following day, the Eagles had breakfast first, and when the Rattlers were in the mess hall for breakfast after them, the Eagles armed themselves with sticks and bats and carried out a retaliation raid on the Rattlers' cabin. Back in their own cabin, they prepared new weapons for a possible return raid by the Rattlers – socks filled with rocks. For safety reasons, the counselors intervened [She88, p.112]. Finally the Eagles won the tournament (through researchers manipulation in plotting the routes of the final treasure hunt), received their prizes and went for a celebratory swim. During that time, the Rattlers raided the Eagles' cabin, "messing up beds, piling personal gear in the middle of the cabin, settings boats at the dock loose and absconding with the prize knives and medals" [She88, p.114]. Another fight followed, and staff decided to stop it to avoid physical injury [She88, p.115].

At the end of stage 2, hostility was at its peak and the members of the two groups didn't want to have anything to do with the members of the other group under any circumstances [She88, p.115]. After the end of the tournament and one additional day dedicated to ingroup activities, the subjects were asked to fill in a questionnaire. It was explained to them that "they were being asked to do so to help the administration find out what they thought of their new acquaintances and how they were enjoying camp" [She88, p.137]. They were asked for stereotype ratings (such as brave, tough, friendly, sneaky, smart alecs, stinkers) of their own and the other group as well as sociometric ratings of friendship preferences. About 93 percent of the friendship choices referred to ingroup members [She88, p.122]. Ingroup members were seen as brave, tough and friendly, while almost all outgroup members were evaluated as sneaky, stinkers and smart alecs.

Stage 3: Integration phase

After strong friction between the two groups was achieved in stage 2, researchers now could proceed to stage 3, the integration phase. Goal of this stage was to explore the effectiveness of two ways of conflict resolution: 1) non-competitive contact situations and 2) the introduction of situations presenting superordinate goals that could only be achieved jointly by the two groups. Altogether, seven contact situations as non-competitive team building measures were arranged, such as watching a movie together, joint shooting of firecrackers on Forth of July (Independence Day), and eating lunch together. The mere contact situations resulted in jeering, catcalls and insulting remarks, there were no signs of intermingling between the groups, and the joint meals actually turned into food fights – the Rattlers and the Eagles "throwing rolls, napkins rolled into a ball, mashed potatoes and so forth" on each other [She88, p.157]. None of the contact situations contributed to reducing friction between the two groups, nor all of the contact situations together.

As it became obvious that mere contact situations were going to be insufficient in reducing intergroup friction, the researchers started with the introduction of superordinate goals. The following main hypothesis was tested:

> Hypothesis 2 (Stage 3)
>
> When groups in a state of friction are brought into contact under conditions embodying superordinate goals, the attainment of which is compelling but which cannot be achieved by the efforts of one group alone, the groups will tend to cooperate toward the common goal. [She88, p.48]

The first superordinate goal pertained to drinking water. All of the water of both groups in the camp came from the same reservoir. Staff manipulated the drinking water system in such as way that they closed the valve of the tank and stuffed the faucet at the end of the tank with pieces of sacking, so that the water supply was cut off at the tank; only the water that was already in the main pipe supplied the camp for a few more hours. When no more water came through at the boys' camps, staff informed the Rattlers and the Eagles that the problem could be a leakage in the pipe from the tank and that four groups would be necessary to inspect four segments of the pipe. To avoid that the subjects would potentially blame the problem on camp staff, the boys had been told that "in the past, on occasion, vandals had tinkered with the water system, causing difficulties" [She88, p.162]. While all boys were ready to help, the volunteers for each of the four segments were either all Rattlers or all Eagles. After the inspection of the pipe didn't reveal any problem, the four groups all met at the tank. Most of the boys were thirsty, "some of the Rattlers still had a little water in their canteens, but the Eagles did not even have their canteens with them" [She88, p.163]. When they found the clogged faucet, they jointly worked for 45 minutes on cleaning it. After the problem was solved and the water came through, "the boys rejoiced in common" [She88, p.165]. The Eagles who had no water bottles with them were permitted to drink first. During and after this problem situation, there was for the first time a good-natured intermingling of Rattlers and Eagles. However, there was another food fight at dinner that night.

Similar situations with superordinate goals followed. During a Campout at Cedar Lake, the truck that should bring food for both groups was manipulated by staff so that it did not start. Both groups jointly had a tug-of-war against the truck to pull it and get it started, which they achieved altogether. The effect was intermingling of group members and friendly interaction between Rattlers and Eagles [She88, p.172]. In other situations, the boys had to cooperate to rent a movie which they all wanted to see, but could not be afforded by one group alone, or collaborate on preparing their meals from bulk ingredients. It took some time and multiple problem situations embodying superordinate goals until the group lines finally disappeared.

At the end of stage 3, stereotype ratings were repeated. While at the end of the friction phase, ratings of outgroup members were very unfavorable, now by the end of the integration phase, ratings of outgroup members "were preponderantly favorable in both groups" [She88, p.194]. Also friendship choices were now much more favorable toward the outgroup.

Finally, all boys insisted on going back on the same bus, they sat down in a mixed group. During a refreshment stop, the Rattlers used their prize money from one of the contents to buy malted milks for all boys [She88, p.187]. The cumulative effects of the situations with superordinate goals resulted in eliminating tension between the groups and thus the above hypothesis was confirmed.

5.1.2 Relevance for GSE

The results of the Robbers Cave Experiment are valid for groups in general and also highly relevant for GSE teams. Stage 2 gave a vivid example that frame conditions of competition between teams lead to intergroup conflict and social distance. In particular, perceived injustice and inequity yielded in an escalation of the situation. Also in GSE projects, there are multiple sources of competition between subteams that are located at different sites: Sites often have divergent incentive structures, and local goals which are not in line with the overall project. For example, in a collaboration of two sites, one site might have other projects which are of higher priority and therefore put only limited effort in the GSE project in question. Also the transfer of responsibility to an offshore location can cause fear in team members at the original site to lose their job, i.e. they see the offshore location as a competitor and therefore are reluctant to provide full support to the other site. Furthermore, perceived injustice and inequity often play an important role in conflicts in GSE projects. Problems in the progress of the overall project are then blamed to the colleagues at the distant site, because "they don't fulfill their duties properly", e.g. because their deliveries are not of the expected quality, even though the source of the problem might be in the way of collaboration itself (e.g. how requirements are communicated). Frequently, 'us versus them' situations can be observed in GSE between subteams at different sites.

Stage 3 of the Robbers Cave Experiment showed that mere contact situations are insufficient to reduce tension in intergroup conflicts, while the introduction of superordinate goals were effective to resolve conflict and transformed the Rattlers and Eagles to 'one team'. This fact should be taken into account in GSE as well. For team building purposes, it is not sufficient to have pure contact situations only, but it is essential to establish a common superordinate goal "the attainment of which is compelling but which cannot be achieved by the efforts of one group alone" [She88, p.48].

5.2 Perception and Reality: The Influence of Human Factors

Social psychology can help to understand the influence of human factors on the collaboration of distributed teams. Knowledge about general effects in human behavior is obtained by conducting and analyzing controlled experiments. A controlled experiment "is one in which there is strict control exercised on variables, with typically one or more variables being changed or 'manipulated' while all other variables remain the same" [EC11]. In social psychology, controlled experiments are used to study human behavior. Experimental psychologists explore human behavior in different contexts and use measurement and observation methods for analysis. In controlled experiments, a situation is created and the behavior of a person in the given situation is measured. The person is considered a black box, as no direct insight into the mind of the affected person is available. The results provide valuable information about human behavior.

Typical characteristics of experiments in social psychology are the following:

- Simple target-oriented experiments with generalizable results

- Situation is carefully prepared, other persons than the subject under observation are often confederates

- Behavior is measured and analyzed for correlations with variations of context

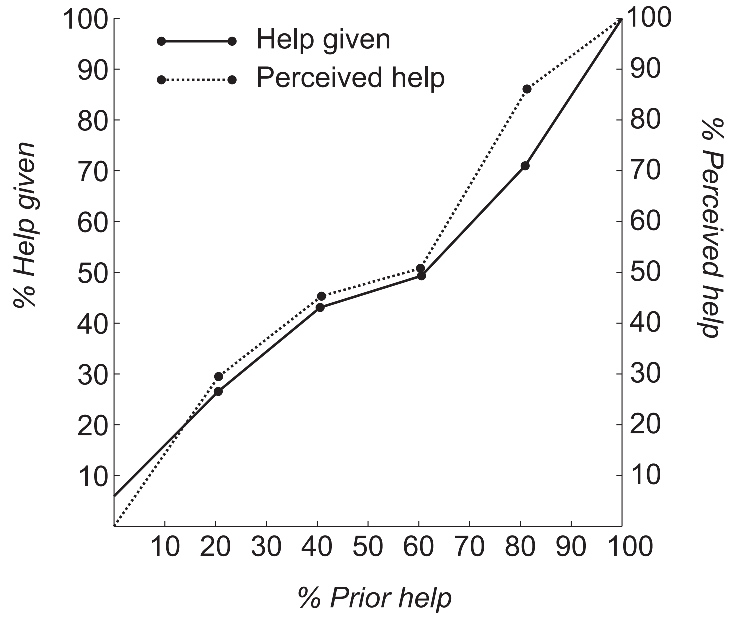


Figure 5.1: Reciprocated help and degree of perceived help as a function of degree of prior help [Bro86, p.61]

- A control experiment[1] is used to compare behavior in a 'neutral' situation

Teamwork heavily relies on helping each other, but readiness to help others depends significantly on human factors such as trust and personal relationship, mood and help received previously. In a simple experiment conducted by Isen and Levin [Bro86] the effect of mood on the willingness to help was analyzed. At a time where coin-box telephones were still the rule, pay phones in a shopping mall were prepared in such a way that people could find a coin left by somebody before in the pay telephone slot. Finding or not finding a coin was the independent variable in the experiment. The fortune of finding a coin created a good mood with the subject. When leaving from the telephone, the subject was walking behind a young woman dropping a folder full of papers and thus was given the option to help or not to help. The young woman was a confederate of the experimenters, and the course of actions had been thoroughly planned [Bro86, p.57]. The behavior of the subjects was measured: 14 of 15 (93%) of the subjects who found a coin helped the young woman by picking up the dropped papers, while only 2 of 26 (8%) of the subjects who didn't find a coin helped. There is a strong evidence that putting people in a good mood causes them to be more helpful.

[1]A control experiment is "an experiment in which the subjects are treated as in a parallel experiment except for omission of the procedure or agent under test and which is used as a standard of comparison in judging experimental effect" [MW11d].

Besides mood, *reciprocity* – the principle in human behavior of responding to a positive action with another positive action, and responding to a negative action with another negative one – plays an important role in the question of people being helpful. Help given is strongly correlated with previous help received. In an experiment by Wilke and Lanzetta [Bro86, p.60] a shipping business game was used to analyze help given depending on prior help received from the other side. Figure 5.1 illustrates the results of this experiment.

5.2.1 Perception Biases

As human beings we are prone to biases in our perception of other people. In this section we briefly describe three such biases: the actor-observer bias, the self-presentation advantage and ingroup–outgroup perception.

Actor-Observer Bias

An effect that is well known in social psychology is the so-called *actor-observer bias*: Acting persons (actors) see factors from outside – i.e. the situation they are in – as reason for their actions, while observers attribute the actors' behavior to the character of the actor [Bro86, p.136, p.174]. This holds for individuals as well as for in- and outgroup perception. Usually the actor is right, because he or she has more data available, i.e. he or she knows the own thoughts and considerations that lead to the action. The actor-observer bias is an effect that is present in everyday life, also in GSE: While the team members at one development site are under pressure and know that it was due to the situation, for example a technical problem, that they could not complete a given task on time, their distant colleagues are likely to attribute it to personal traits ("they were lazy").

Self-Presentation Advantage

Another perception bias is the inherent self-presentation advantage of teachers and speakers: Because of the role and situation they are in, they are perceived as more knowledgeable than others. The quizmaster game, an experiment from social psychology, illustrates this self-presentation bias. Brown replicated this experiment every year with students at Stanford university and stated that it was one of his favorite experiments, because "it always works" [Bro86, p.181]. We decided to use the quizmaster game as warm up exercise in the GSE master seminar in June 2010.

Students were asked to form 2-person teams. Within each team, roles were randomly assigned: one student became quizmaster, the other took the role of the contestant. Quizmasters then got 15 minutes time to compose ten questions,

My role:

I was ❑ quizmaster ❑ contestant

Number of questions answered correctly: ______

Please rate **your own level of general knowledge** on a scale from 0 (very poor) to 100 (excellent), where 50 is the average TUM student:

My own level of general knowledge: ______

Please rate **your team member's level of general knowledge** on a scale from 0 (very poor) to 100 (excellent), where 50 is the average TUM student:

My team member's level of general knowledge: ______

Figure 5.2: Questionnaire filled in by contestants and quizmasters at the end of the exercise

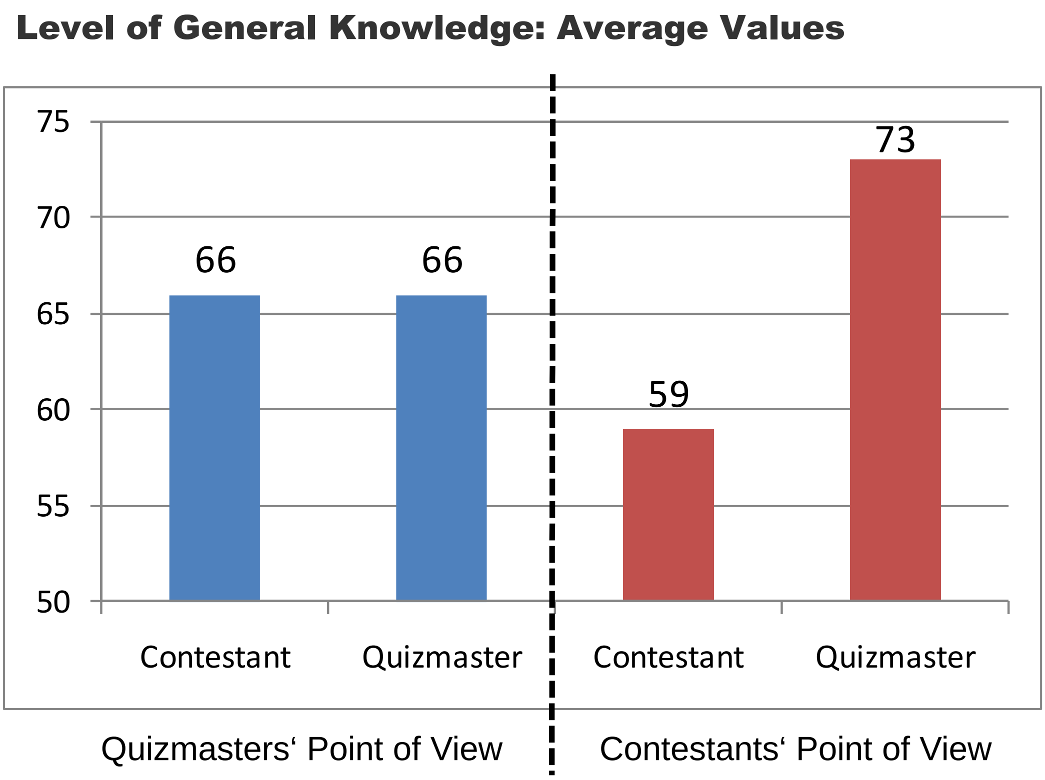


Figure 5.3: Results of the quizmaster exercise, clustered by role

which should be challenging to answer but not unfair or impossible. The contestants had 15 minutes of free time until their quizmaster was ready, and then subsequently were asked to answer the ten questions. At the end of the exercise, all students were asked to fill in a short questionnaire to rate the level of general knowledge of themselves and of the other person in their team (see Figure 5.2). The results have been evaluated in front of the audience (entering the values into a spreadsheet that had been prepared before). Figure 5.3 shows the ratings of general knowledge obtained. While from quizmasters' point of view, contestants and quizmasters are at the same level of general knowledge, contestants think that quizmasters have a higher level of general knowledge than contestants. The explanation of this effect is that contestants treat the brief samples of knowledge displayed in the game as equally representative of the general knowledge of the two participants. Because the quizmasters are controlling the topic and the questions, they are perceived to be more knowledgeable [Bro86, p.181].

Ingroup and Outgroup Perception

In social psychology it has been shown that "when people are assigned to a group, any group, they immediately, automatically, and almost reflexively think of that group as better than the alternative, an out-group for them, and do so basically because they are motivated to achieve and maintain a positive self-image" [Bro86, p.551]. It has been proven in various experiments that it is sufficient to be assigned to an arbitrary group – on a random basis – to cause ingroup favoritism. Close personal contact is usually stronger than other group memberships like religion or nation. Lufty N. Diab repeated the Robbers Cave procedure in 1963 with eleven-year-old boys in Beirut, each of the two groups (they called themselves 'Blue Ghosts' and 'Red Genies') consisted of five Muslims and four Christians [Ber08, p.178]. The experiment showed that assignment to the two separated groups (with no previously established relationships) and interacting together in the same place tightly together was stronger than any religious denomination, as none of the fights were along religious lines.

5.2.2 Hidden Traps in Decision Making

Even when dealing with facts and figures, people are far less objective than they think they are. Hammond, Keeney and Raiffa [KRH06] describe psychological traps that are particularly likely to undermine objective decisions. A vivid example for how perception influences decision making is the *framing trap*: Consider the following problem which in an experiment was posed to a group of insurance professionals:

> You are a marine property adjuster charged with minimizing the loss of cargo on three insured barges that sank yesterday off the coast of

> Alaska. Each barge holds $200,000 worth of cargo, which will be lost if not salvaged within 72 hours. The owner of a local marine salvage company gives you two options, both of which will cost the same. Which plan would you choose?
>
> Plan A will save the cargo of one of the three barges, worth $200,000.
>
> Plan B has a 1/3 probability of saving the cargo on all three barges, worth $600,000, but has a 2/3 probability of saving nothing. [KRH06, p.7]

In a second study, a group was asked to choose between:

> Plan C will result in the loss of two of the three cargoes, worth $400,000.
>
> Plan D has a 2/3 probability of losing all three cargoes, worth $600,000 but a 1/3 probability of losing no cargo. [KRH06, p.7]

Note that plan A and B describe the same facts as plan C and D. The difference is that the situation is framed in different ways: in terms of gains and in terms of losses. The interesting evidence is: In case of the first experiment, 71% of the participants chose plan A, which is less risky. In the second experiment, 80% of the participants chose plan D. The results demonstrate that people are risk averse when a problem is presented in terms of gains but risk seeking when a problem deals with avoiding losses [KRH06].

Further known traps are the *anchoring trap* (initial impressions or data anchor subsequent judgments), the *status-quo trap* (people are reluctant to change), the *sunk-cost trap* (make choices that justify past choices, even if the past choices turned out to be wrong), *confirming evidence trap* (interpreting information in a way that supports the existing viewpoint while avoiding information that contradicts it), the *overconfidence trap* (tendency to be overconfident about accuracy of estimates and forecasts), the *prudence trap* (estimates are adjusted to be 'on the safe side'), the *recallability trap* (inability to recall events in a balanced way distorts decision making, e.g. dramatic events leave stronger impression than others). The described effects underline the importance to explicitly address human factors in decision making. Decision are not only rational. In his book *Systems Thinking – Managing Chaos and Complexity*, Gharajedaghi describes three dimensions of decisions: rational (interest of the decision maker, reflecting the perceived interest of the decision maker at the time), emotional (domain of beauty and excitement) and cultural (norms of the collective, of which the decision maker is a member) [Gha06]. Also in negotiation it is well known that human factors need to be considered well. Fisher and Ury describe effective negotiation techniques, where people-related aspects and interests are addressed separately [FU91].

5.3 Summary

Software engineering is done by people and therefore is heavily constrained by how people think and behave. Social psychology helps us to understand these effects and how to deal with them. Not only ‘hard facts’ like cost, time or quality, but also ‘soft facts’ such as relationship and human behavior are measurable, as it was demonstrated in the Robbers Cave experiment or the shipping business game. Relationship between subteams at different locations is an important aspect which cannot be neglected in GSE projects. We conclude that to improve GSE projects, it is necessary to focus on human factors and team relations, not only on formal organization or formal communication.

Chapter 6

Communication Metrics

In Chapter 4 we have seen that communication is a key challenge in GSE. In this chapter we take a closer look at the role of communication in software engineering as well as communication theory, and we introduce the concepts of software measurement and communication metrics.

6.1 The Role of Communication in Software Engineering

Bruegge and Dutoit define communication as "an activity during which developers exchange information, either synchronously or asynchronously, and either spontaneously or according to a schedule." [BD10, p.760] Software engineering is a collaborative activity that brings together people from different backgrounds, e.g. domain experts, analysts, designers, programmers, managers, and users for the purpose of the software development project [BD10, p.111]. In fact, Bruegge and Dutoit see communication as central activity in software development with high impact on project success: "Communication is the most critical and time-consuming activity in software engineering. Misunderstandings and omissions often lead to faults and delays that are expensive to correct later in the development." [BD10, p.55]. It is also confirmed by other studies that miscommunication is a major reason for cost overruns, quality problems and project delays [HLRP99]. Communication in software engineering includes *planned communication* such as status meetings, client and project reviews or project retrospectives, and *unplanned communication*, e.g. a request for clarification or change, or communication for issue resolution. Communication mechanisms are either *synchronous* (e.g. phone call) or *asynchronous* (e.g. e-mail), depending on whether the communication mechanism requires the sender and receiver to be available at the same time [BD10, p.136]. Communication also involves unplanned, informal exchange of information, for example hallway conversations [BD10, p.138].

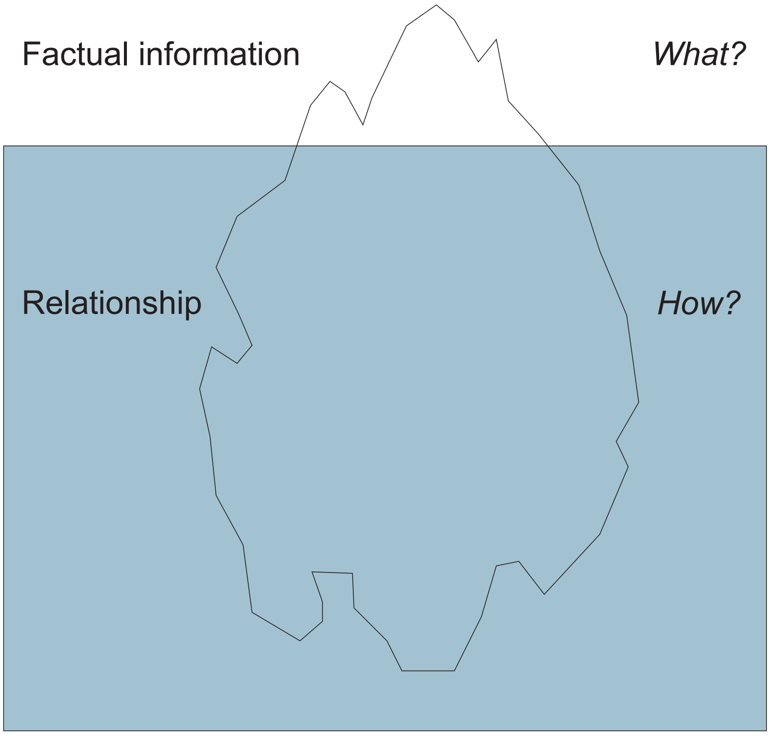


Figure 6.1: The iceberg model of communication

The number of possible communication links in a team with n persons is $n * (n - 1)/2$, i.e. it grows geometrically [Car99, p.55]. For example, a team of 8 members has 28 possible communication links, in a team of 30 members there are already 435 possible communication links. While in production *economy of scale* is an important principle which means the higher the quantities the cheaper the product gets and the better the overall economic result, this is not true with respect to the size of software engineering teams. Boehm calls this effect *diseconomy of scale*: "The more members that are added to a team, the more time is consumed in communication with other team members. This decrease in programmer productivity is called a diseconomy of scale in economic terms." [Boe81]

6.2 Models from Communication Theory

In communication, not only the 'what', but also the 'how' is important. This is often illustrated by an iceberg representation, see Figure 6.1. While the visible tip of the iceberg represents the factual information which is communicated (the 'what'), a large portion of communication – the 'how' – at the same time is 'below the surface', determining the relationship between communicants.

Watzlawick, a pioneer in communication theory, postulated five axioms of communication[1]:

1. One cannot *not* communicate
2. Every communication has a content and a relationship aspect
3. The nature of a relationship is contingent upon the communication patterns
4. Human beings communicate both verbally and non-verbally
5. Depending on the role of communicants, communication interchanges are either symmetrical or complementary

One cannot not communicate. According to Watzlawick, any kind of behavior is communication. As soon as two persons meet, they communicate with each other not only with words, but through their whole behavior. According to Watzlawick, it is impossible *not* to communicate, because activity or inactivity, words or silence all have message value [WJB67, p.49]. Behavior has no opposite. If we think of a woman sitting in a cabin of a train and listening to music from her iPod via ear-phones and with her eyes closed, she is communicating that she doesn't want to be disturbed, which is usually respected by others.

Every communication has a content and a relationship aspect. A communication message comprises both a factual information (content level) and information about the relationship of the communicants (relationship level). The latter classifies the former and is therefore a *metacommunication*, i.e. information about how to interpret the factual information. The two sentences "It is important to release the clutch smoothly" and "Just let the clutch go, it'll ruin the transmission in no time." have a very similar meaning on content level, but the relationship implied is very different: the first statement is factual, the second statement is dismissive with respect to the target person.

The nature of a relationship is contingent upon the communication patterns. This axiom states that the nature of the relationship between communicants corresponds to the communication patterns between them. Punctuation[2] organizes interactional sequences [WJB67, p.56] and is culturally influenced, since people have conventions of how to behave and communicate in specific situations. Analyzing communication patterns is particularly useful when investigating communication difficulties. Watzlawick gives an example of a couple which

[1]abbreviated, the full text can be found in chapter 2 of Watzlawick's *Pragmatics of Human Communication* [pp.48–71][WJB67]

[2]Instead of communication patterns, Watzlawick uses the term 'punctuation of the communication sequence'.

has a marital problem, where "the problem lies primarily in [...] their inability to metacommunicate about their respective patterning of their interaction." [WJB67, p.58]. To see the communication patterns is the prerequisite to improve the situation.

Human beings communicate both verbally and non-verbally. As human beings, we communicate both verbally – using language – and non-verbally, e.g. with gestures or facial expressions. Watzlawick uses the terms 'digital' and 'analogic' to characterize these two communication types. Verbal communication as 'digital message' has a much higher degree of complexity, versatility and abstraction compared to non-verbal communication, but at the same time it is less intuitive. Verbal communication is often used to express factual information (content level), while non-verbal communication usually refers to the relationship level.

Depending on the role of communicants, communication interchanges are either symmetrical or complementary. In his fifth axiom Watzlawick postulates that communication between two communicants is either symmetrical or complementary, depending on their roles in the communication interchange. Symmetrical means that communicants have an equal status position. Complementary refers to a constellation where one person is in the role of the superior, while the other takes the role of the inferior position.

Schulz von Thun extended the communication model of Watzlawick and introduced the so-called *4-ear model*. According to the 4-ear model, a message can be understood from four perspectives (the different 'ears'): 1) the factual information, 2) self statement, 3) relationship indicator, and 4) appeal. For example when the superior tells his developer: "Your code is very buggy.", the factual information (1) is: "This code contains many bugs.". It also gives a self statement (2) of the superior: "I care about quality.". Furthermore, the message can be understood as relationship indicator (3): "I am the boss." Finally, one can also comprehend it as an appeal (4): "Work more carefully and avoid bugs!"

6.3 Software Measurement

Introducing objective *metrics* is a well established management technique to monitor the current status of a project. We described basic measurement concepts in an earlier publication [LP07]. Software measurement helps to understand more about software development as a basis for making operational decisions as well as estimation and prediction. It is the basis for specifying and achieving objectives [ED07, p.2], in accordance with the saying 'You can't control what you can't measure' [DM82]. Measurement must be goal-oriented [ED07, p.21]. A well-

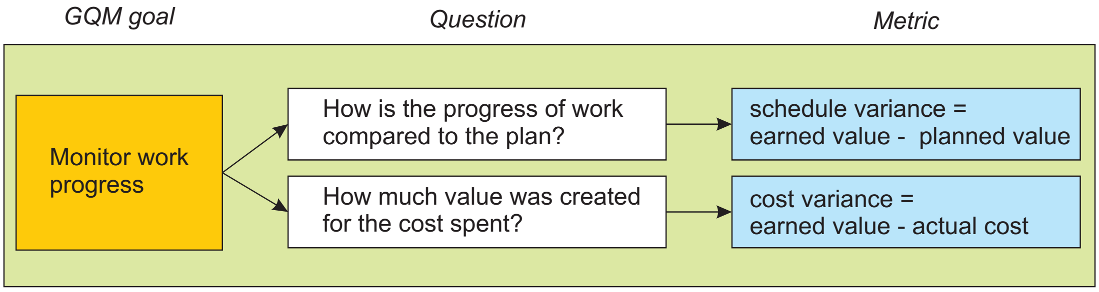


Figure 6.2: Earned value analysis: Goal, question and metric

established approach for definition of goal-oriented metrics is the Goal Question Metric (GQM) paradigm introduced by Basili and his colleagues [BLR^{+}10], which consists of three steps:

1. **Goal:** Which goal shall be achieved by the measurement?
2. **Question:** Which question(s) shall be answered by the measurement?
3. **Metric:** Which metric(s) are suitable to answer the question?

We use the term *measurement system* to refer to a defined set of metrics associated with a process for execution of the measurement.

In project management, metrics can provide essential decision support: Project managers use metrics to observe the status of their project, including cost (e.g. actual and planned cost), time (e.g. milestone trend analysis) and quality aspects (e.g. open defects).

One disadvantage of monitoring cost trend only is that it doesn't relate to the progress of the project with respect to value and schedule, i.e. one knows how much money was spent, but it is not clear how much output was created for the expenditures. A more advanced measurement technique that copes with this challenge is the so-called *earned value analysis*. In this approach, cost is set into relation with the progress of the work. Depending on the percentage of completion of individual work packages of the project, the 'earned value' of the respective work packages are taken into account. For instance, if a work package is 50% complete, it accounts for 50% of its planned cost. It is a common practice only to consider four different values for the percentage of completion of work packages: 0% for work that not started yet, 10% for work that started, 50% for work in progress and 100% for work completed [ED07, p.213]. Two metrics are defined to measure progress:

$$schedule\ variance = earned\ value - planned\ value$$
$$cost\ variance = earned\ value - actual\ cost$$

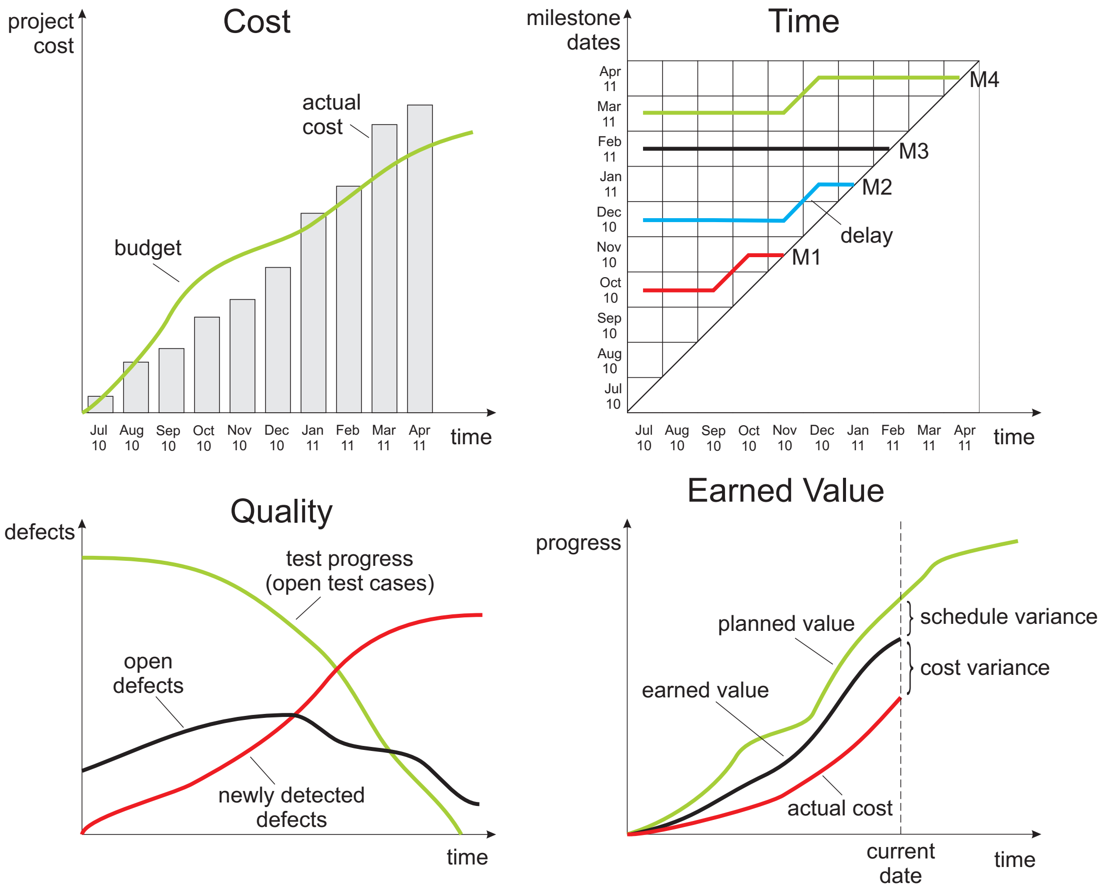


Figure 6.3: A simple measurement cockpit, adapted from Ebert and Dumke [ED07, p.208] and Lescher and Paulisch [LP07]

Figure 6.2 shows the goal, question and metric of earned value analysis. The *schedule variance* compares the progress of completed work to the progress which was originally planned. The *cost variance* looks at the deviation between the earned value achieved and the actual cost spent. In this way, it is possible to monitor the progress of work based on cost and schedule at the same time.

Like the instruments of a pilot in an aircraft, a *measurement cockpit* (also known as *measurement dashboard*) gives an overview of the current situation of the project. Measurement cockpits support project managers in their decisions and help them to perform corrective actions. Figure 6.3 shows a simple measurement cockpit.

6.4 Communication Metrics

Metrics-based approaches, when used in the right way, are a powerful tool for project management. However, traditional measurement indicators are more related to the symptoms instead of the root-causes: Typically measurement indicators consider *cost* (e.g. cost trend analysis), *time* (e.g. schedule compliance) and *quality* (e.g. defect occurrences) aspects. They provide valuable information about the current status of a project, but they cannot predict e.g. integration problems in a later stage of the project. Key questions are: How can collaboration problems be detected early on *before* they have a serious impact? What are suitable metrics and visualizations that can serve as early warning indicators? In this context, communication metrics – measuring communication inside development teams – are more useful than traditional measurement indicators. For example, analyzing communication data turned out to be a helpful approach for understanding behavior of globally distributed development teams. Bruegge and Dutoit showed that metrics on communication artifacts – message counts of inter-team and intra-team communication – can be used to gain significant insight into the development process that produced them [DB98].

A common approach for analysis of communication data is the so-called *social network analysis*[3]: Based on communication data – either collected via a survey (questionnaire filled out by team members) or derived from other sources such as e-mail, or the change management system – the communication relationships between team members are analyzed (e.g. who is communicating more than 3 times per week over e-mail or phone). The same method can be also used for awareness relationships (e.g. does person X know what person Y is working on). The resulting network can be displayed as a graph, and analyzed applying methods of graph theory. Ehrlich and Chang [EC06] provide an overview of the social network analysis methodology in Global Software Engineering. Damian and her colleagues conducted social network analysis in various contexts, e.g. analyzing communication relationships in so-called requirement-dependency social networks (i.e. social networks of teams, which work on requirements that are dependent on each other) [MDSS08]. Even prediction of build failures based on social network analysis was shown to be possible under certain conditions [WSDN09].

A prominent example for the effectiveness of analyzing communication is the American energy company Enron, based in Houston, Texas. Enron was awarded 'America's Most Innovative Company' by the Fortune magazine for six consecutive years. However, the real 'innovative' power seemed to be in the company's accounting practices, which turned out to be illegal, making financial misstatements, which became known as the Enron scandal. As a consequence, Enron went bankrupt in December 2001. In the context of the public investigations that followed, the company's e-mail data were made public, which consisted of (after duplicates were removed) half a million e-mail messages from about 150

[3]The roots of social network analysis go back to social psychology. Jacob Moreno developed a network analysis technique called sociometry in the 1930s, introducing the so-called sociogram (a pictoral social network) and sociomatrix (a social network data table) [Mor34]. Moreno used sociometry for exploration of social structures via measurement of sympathies and antipathies of group members.

accounts, including top executives of Enron. These e-mails were extensively analyzed by researchers using social network analysis. They found evidence that the e-mail traffic patterns tracked major events at Enron, e.g. the manipulation of California energy prices and in fact these events could have been detected earlier based on communication metrics [Kol05].

The Enron example as well as existing studies in Global Software Engineering are post mortem views of projects, i.e. projects were analyzed after their completion. In this way, the projects themselves had no use of it. However, we believe that these concepts can also be adapted for usage during project lifetime: The new aspect of our work is to use communication metrics during project lifetime to support project managers in their decisions.

Chapter 7

Model of GSE Metrics

In this chapter we introduce a model of metrics for Global Software Engineering projects and define an set of collaboration-based metrics to be used as early warning indicators by project managers.

7.1 Overall Approach

Our research question is: *How can collaboration problems in GSE projects be detected early on before they have serious impact?*. Our research goal is to introduce a communication-based measurement model to support project management of GSE projects (see Chapter 2). In the previous chapters we dealt with aspects related to our research question: We discussed challenges in GSE, the psychology of distributed teams, the importance of measurement in project management of software development projects and the concept of communication metrics. Putting these aspects together, we now define a model of GSE metrics for project management. In the following we briefly summarize the overall approach in pattern format, which is well known as an effective form of knowledge representation and as an approach for structured thinking [Ale79].

Problem: How can one detect problems in GSE projects early on?

Forces: *[brackets refer to main chapters where the forces were discussed]*

- Global distribution makes collaboration difficult [Ch. 4]
- Cost, time, quality metrics reveal problems late (symptoms, not root-causes) [Ch. 6]
- Team members are often aware of problems, but don't dare to speak openly [Ch. 1]

- Human factors such as trust or team cohesion have strong impact on collaboration effectiveness [Ch. 5]

Solution: Use a measurement system to monitor collaboration and take corrective actions, where necessary

Consequences: *(parantheses indicate positive or negative impact)*

- Problems can be detected and solved before they have serious impact *(positive)*
- Data collection and analysis can cause high effort *(negative)*
- Measurement influences behavior *(can be both a positive and a negative consequence)*

For the design of the GSE metrics and the related measurement questions, we considered the information from the previous chapters and three specific sources: 1) the social network analysis methodology as described by Ehrlich and Chang [EC06], 2) a GSE survey conducted by Herbsleb and Mockus [HM03], and 3) characteristics of dysfunctional cultures as described by Kerth for project retrospectives [Ker01].

Ehrlich and Chang [EC06] introduce six categories for social network analysis in GSE projects: *communication* (how often have you communicated with this person?), *availability* (how easy is it for you to reach this person?), *general awareness* (how aware are you of this person's professional background?), *current awareness* (how aware are you of the current set of tasks that this person is working on?), *familiarity* (how closely did you work with this person on your last project together?), and *importance* (how important is it for you to interact with this person?).

Herbsleb and Mockus [HM03] conducted studies on globally distributed development projects at Lucent. They used a survey questionnaire with 54 questions, structured in eight categories: patterns of communication, working relationships, work atmosphere, communication and coordination, information exchange, best practices, technology, and demographic information.

Kerth [Ker01, p.42] described 13 characteristics for observation of elements of dysfunctional and functional cultures in software development organizations.

In the definition of our model, we take a three-step approach: 1) abstraction of measurement levels, 2) introduction of measurement classes and 3) definition of metrics. For the definition of metrics, we adapted the GQM^{+}Strategies approach, an extension of the Goal Question Metric (GQM) approach [BGR94].

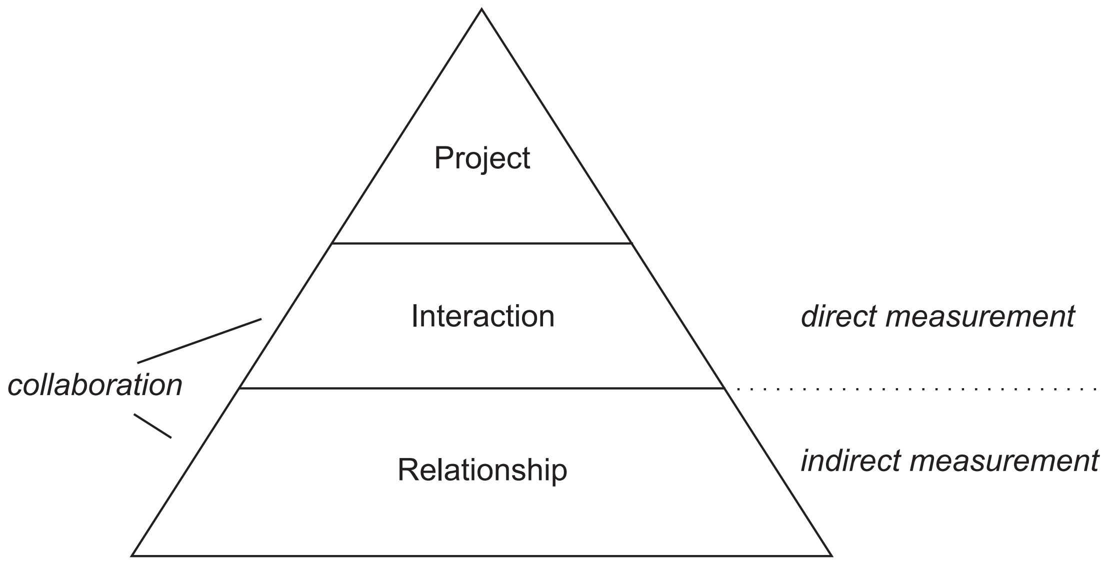


Figure 7.1: Model of GSE Metrics: The GSE metrics pyramid

7.2 Abstraction of Measurement Levels: The GSE Metrics Pyramid

The GSE metrics model consists of three levels: Project level, Interaction level, Relationship level. The first level, the *Project level* represents traditional measurement of cost, time and quality status as it is state of the art today. In addition to the Project level we introduce two more levels in our model to measure collaboration, the *Interaction level* and the *Relationship level*, which are not state of the art today. They refer to both interaction (e.g. communication) and relationship (e.g. team cohesion) between team members of a globally distributed development project and correspond to the 'togetherness' described by Colin Matthews, Chief Executive of BAA in the Heathrow Terminal 5 example in Chapter 1. The three measurement levels of our model are illustrated in Figure 7.1.

Project level. This level refers to project management metrics for progress monitoring, which are state of the art in today's software development practice and literature and used by project managers in industry today. Measurement on the Project level usually includes cost, time and quality metrics for monitoring the project staying within the budget (budget compliance), keeping milestone dates (schedule compliance) and achieving quality goals (defect rates). Also more sophisticated metrics like *earned value* are contained in this level.

Interaction level. The Interaction level aims at measuring visible interactions between project team members such as (formal) communication, traveling and delay of work items due to missing information. For metrics of the Interaction level a *direct measurement* is possible, e.g. by analyzing e-mail data, traveling records or dates in a change management system. Because of the *direct measurement*, data required for Interaction metrics can be collected either manually or automatically.

Relationship level. This level refers to measuring interpersonal relationship between team members. It includes aspects such as team cohesion, informal communication, awareness, culture and language and team satisfaction. It is not possible to measure this data directly or derive it from existing data. Instead only an *indirect measurement* is possible by asking or observing the team members, e.g. by surveys or interviews, which implies a manual data collection.

The three levels are presented in the form of a pyramid (see Figure 7.1). It reflects the assumption that the explanatory power for detecting problems in a project early on is assumed to be higher for collaboration metrics than for traditional metrics.

7.3 Classes of GSE Metrics

Figure 7.2 depicts the classes of metrics included in our model and gives an overview of the GSE metrics, which we define in detail in the following sections.

The class Project Metrics represents the traditional metrics. It includes: Cost, Time, Quality and Earned Value, as typical state of the art project management metrics.

In the class Interaction Metrics we introduce collaboration metrics for which a direct measurement is possible. This class includes Directly Measurable Communication such as formal communication in e-mail or meetings, Communication Media, i.e. how much communication is conducted via a specific communication medium, Delay Due To Missing Information e.g. based on information from a change management system, and Traveling for example referring to frequency of traveling between development sites based on traveling records.

Finally the class Relationship Metrics subsumes collaboration metrics for which only an indirect measurement is possible, e.g. by surveys or interviews. It includes Team Cohesion, i.e. the extent to which team members feel as one team, communicate openly, assist each other in case of difficult situation, etc., and Informal Communication such as discussion of non-work related matters or exchange of valuable project information in an informal setting. Furthermore, the class Relationship Metrics includes Awareness, e.g. do team members know what others are currently working on, Culture & Language referring to language skills and influence of cultural differences, and Team Satisfaction, i.e. the extent to which team members are satisfied with the project's achievements and how well they feel in the team.

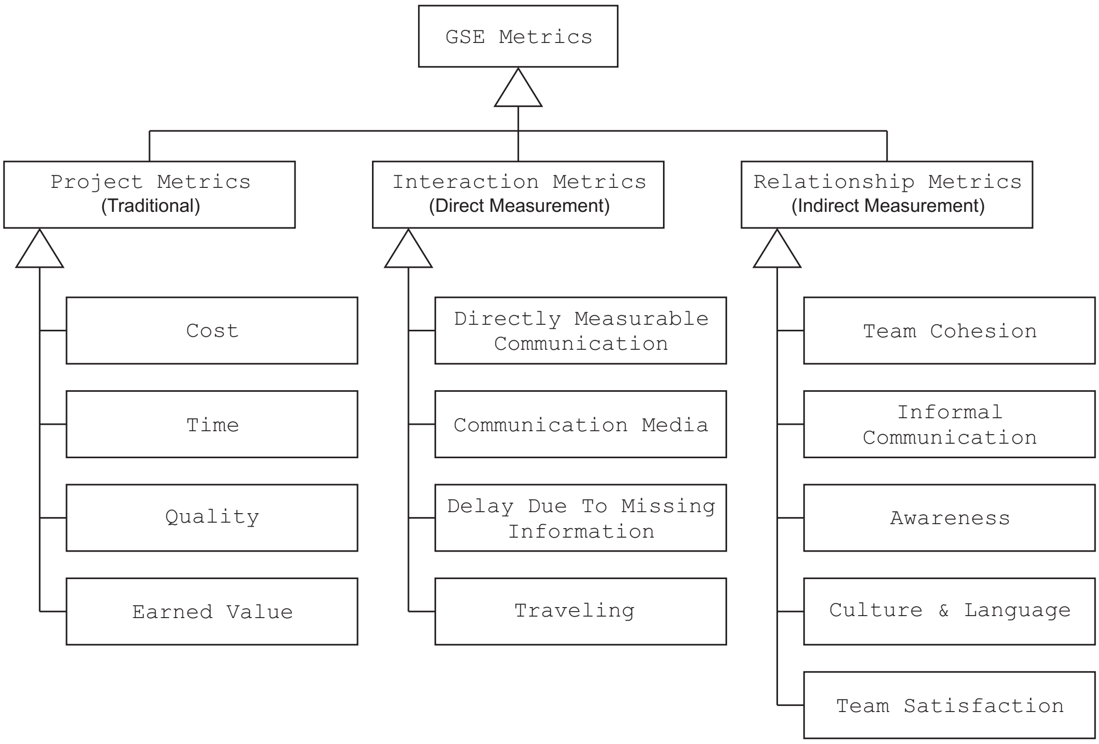


Figure 7.2: Class diagram: GSE Metrics

Figure 7.3 shows goals and strategies on the three levels Project, Interaction and Relationship as well as context factors and assumptions. We adapted the GQM$^+$Strategies approach introduced by Basili and his colleagues [BLR$^+$10] for definition of our GSE Metrics and also for the graphical illustrations. On the Project level, the goal is to conduct a successful development project, where according to the context the software portion is an important part of the product being developed (C1).

Based on the assumption that the effectiveness of project management can be improved by using a measurement system (A1), the strategy on the Project level is to use a measurement system to support project management. Global Software Engineering projects are characterized by collaboration difficulties caused by geographic separation, different time zones, cultural differences and different native languages (C2).

The goal on the Interaction level is to ensure effective team collaboration across sites. Influenced by the fact that the later problems are found the more costly it is to fix them (C3) and the assumption that monitoring team collaboration (e.g. communication) can reveal latent problems early on (A2), the strategy on the Interaction level is to identify collaboration problems early on by monitoring team interactions.

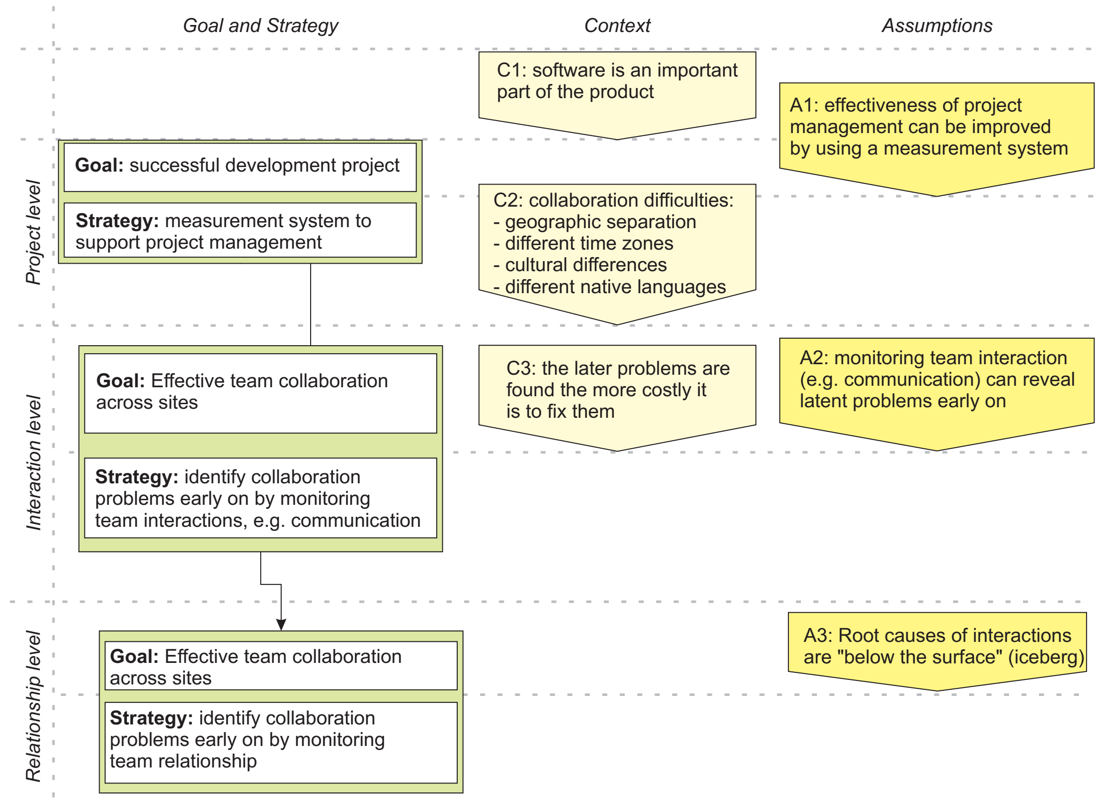


Figure 7.3: Level overview

On the Relationship level, the goal also is to ensure effective team collaboration across sites, but here the focus is on relationship instead of visible interactions. As according to the iceberg model of communication (see Chapter 6), the root-causes of interactions are 'below the surface' caused by relationship (A3), the strategy on the Relationship level is to identify collaboration problems early on by monitoring team relationship.

7.4 Project Level

Figure 7.4 details the metrics on the Project level. The goal of these metrics is to ensure a successful development project. This is achieved by the strategy to use a measurement system of traditional metrics to support project management. The measurement goal on the Project level is to analyze cost, time, and quality progress of the project (G0), which is broken down to the following measurement questions:

- Q0.1: Is development cost on track?
- Q0.2: Is development time on track?

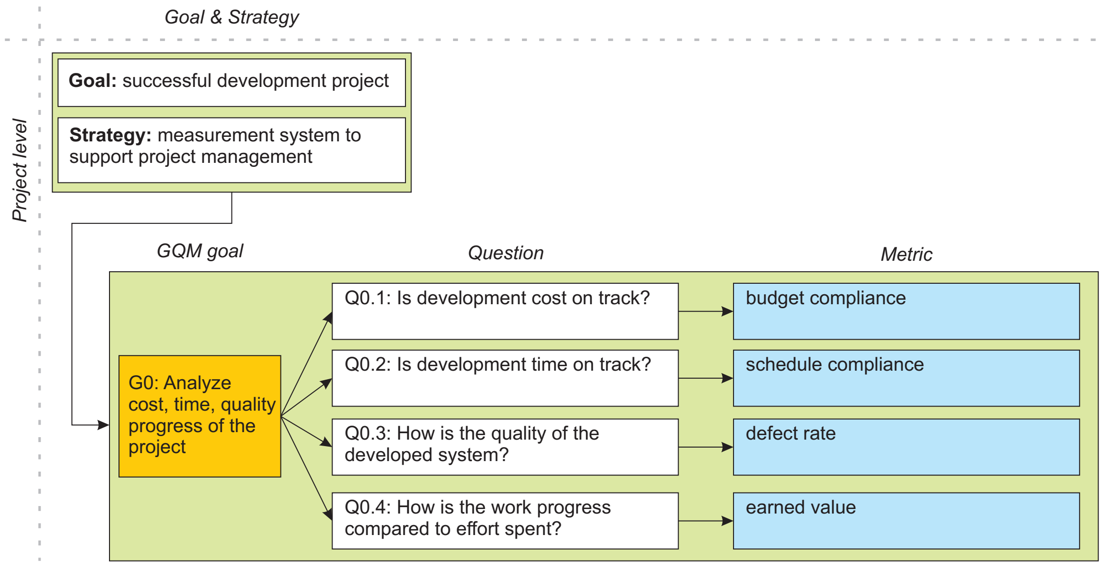


Figure 7.4: Project level

- Q0.3: How is the quality of the developed system?
- Q0.4: How is the work progress compared to the effort spent?

The corresponding metrics are budget compliance, schedule compliance, defect rate, and earned value. As these metrics are not in the main scope of this dissertation, we refer to Ebert and Dumke [ED07] or Lescher and Paulisch [LP07] for detailed definitions.

7.5 Interaction Level

The overall goal of Interaction level metrics is to achieve effective team collaboration across sites (see Figure 7.5). This is to be achieved through the strategy to identify collaboration problems early on by monitoring team interactions, which are directly measurable. This is further detailed in the following measurement goals:

- G1: Analyze communication frequency
- G2: Analyze reachability
- G3: Analyze communication media
- G4: Analyze delay due to missing information
- G5: Analyze traveling records

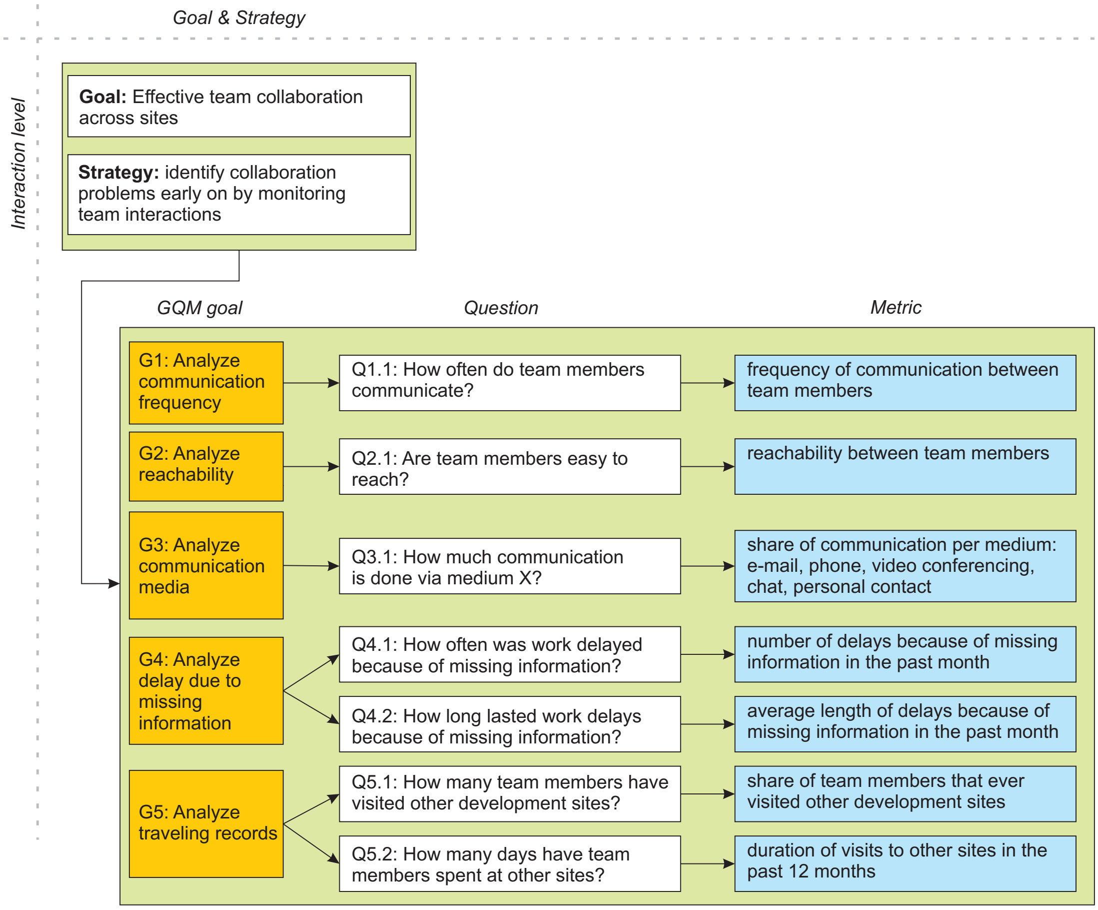


Figure 7.5: Interaction level

G1 and G2 correspond to social network analysis introduced in Chapter 6 and the social network analysis categories communication and availability by Ehrlich and Chang [EC06], asking *How often do team members communicate?* (Q1.1), and *Are team members easy to reach?* (Q2.1). For this, the frequency of communication between roles in the team resp. the reachability between roles in the team is measured. We chose to measure communication based on roles in order to preserve privacy and make results more comparable (cf. approach of Coplien and Harrison [CH04]). G3 aims at analyzing *How much communication is done via medium X?* (Q3.1), measuring the share of communication per medium: e-mail, phone, video conferencing, chat and personal contact. The underlying assumption is that the communication media profile can reveal latent problems, for instance a high share of e-mail communication can be an indicator for ineffective communication due to missing personal contact or misunderstandings. Q3.1 refers to question 38 by Herbsleb and Mockus [HM03]. As delay is a key issue of distributed collaboration [HM03], the goal G4 is to analyze delay due to missing information: *How often was work delayed because of missing* *information?* (Q4.1), measured by number of delays because of missing information in the past month, and *How long lasted work delays because of missing information?* (Q4.2), measured in average length of delays because of missing information in the past month. Q4.1 and Q4.2 were adapted from question 29 and 30 by Herbsleb and Mockus [HM03]. G5 aims at analyzing traveling records: *How many team members have visited other development sites?* (Q5.1) with the metric share of team members that ever visited other development sites, and *How many days have team members spent at other sites?* (Q5.2), measured by duration of visits to other sites in the past 12 months. The underlying assumption is that rare meetings can indicate insufficient personal contact, or increased traveling can indicate coordination effort due to problems. Q5.1 and Q5.2 correspond to question 52 of Herbsleb and Mockus [HM03].

7.6 Relationship Level

Figure 7.6 introduces our metrics on the Relationship level. The overall goal on the Relationship level is – as on the Interaction level – to achieve effective team collaboration across sites, however, here the strategy is to identify collaboration problems early on by monitoring team relationship. As opposed to the Interaction level, on the Relationship level only indirect measurements are possible, e.g. by interviews or surveys. The Relationship level comprises the following measurement goals:

- G6: Analyze team cohesion
- G7: Analyze informal communication
- G8: Analyze awareness
- G9: Analyze culture & language
- G10: Analyze team satisfaction

G6 aims at analyzing team cohesion and is further detailed in the following questions:

- Q6.1: Is communication inside the project team honest and open?
- Q6.2: Can team members rely on their colleagues?
- Q6.3: Do team members assist each other with heavy workloads?
- Q6.4: Are new ideas from colleagues seen as valuable?
- Q6.5: Are meetings/phone conferences constructive?

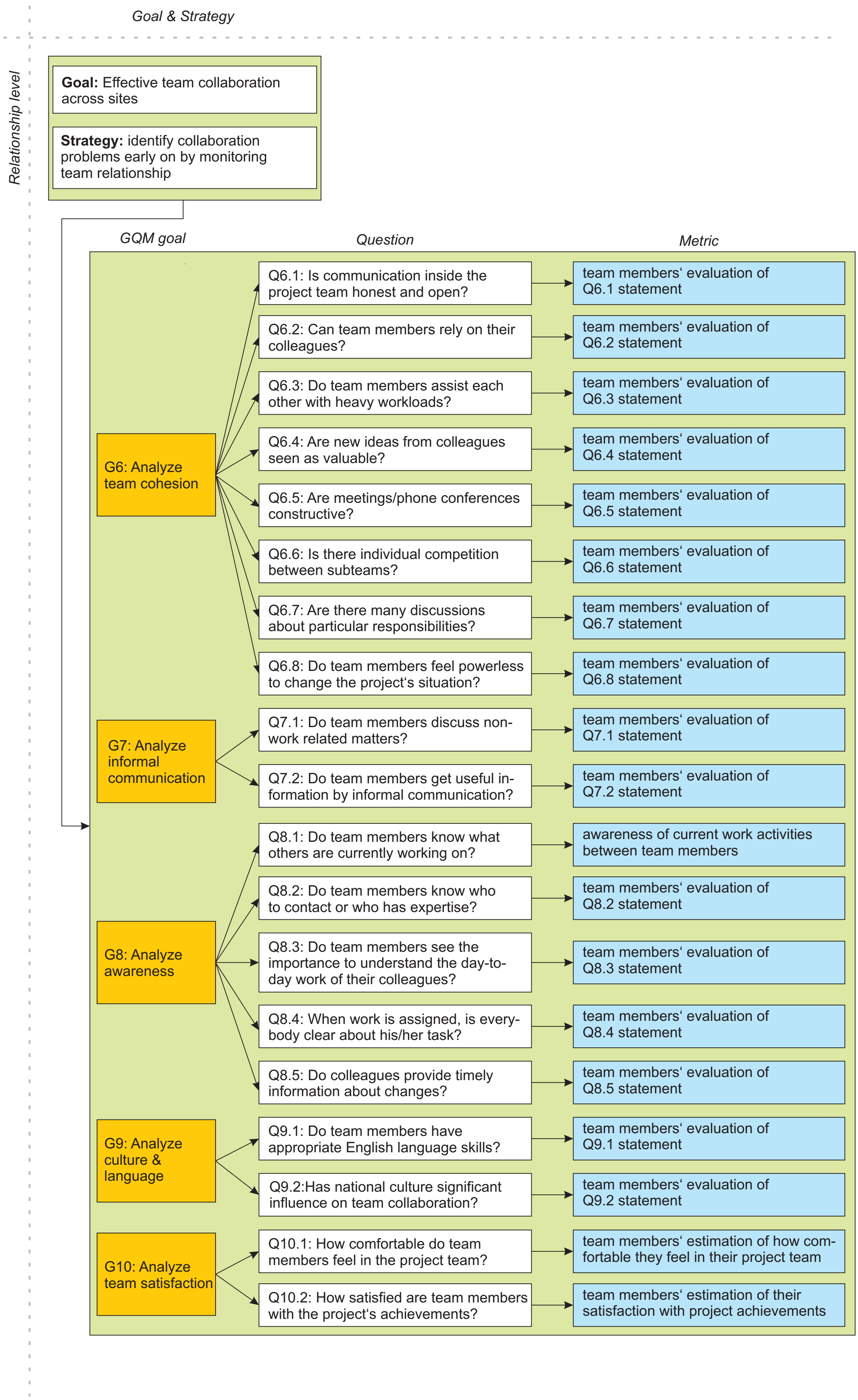


Figure 7.6: Relationship level

- Q6.6: Is there individual competition between subteams?
- Q6.7: Are there many discussions about particular responsibilities?
- Q6.8: Do team members feel powerless to change the project's situation?

Q6.2 and Q6.3 are based on questions 12 and 16 in the survey of Herbsleb and Mockus [HM03]; Q6.1 as well as Q6.4-Q6.8 are adapted from Kerth's characteristics of dysfunctional cultures [Ker01, p.42]. As no direct measurement is possible on the Relationship level, measurements are based on team members' evaluation. Please note that even though the measurement questions are fomulated as closed questions, the team members are supposed to answer according to their degree of confirmation, e.g. "Please evaluate the statement: *I feel powerless to change the project's situation* – select from the range of values: *fully agree, tend to agree, tend to disagree, fully disagree*".

Goal G7 is to analyze informal communication. It comprises the questions *Do team members discuss non-work related matters* (Q7.1) and *Do team members get useful information by informal communication?* (Q7.2). Q7.1 and Q7.2 are related to question 13 and 24 of Herbsleb and Mockus [HM03].

G8 aims at analyzing awareness of team members. It refers to the following questions:

- Q8.1: Do team members know what others are currently working on?
- Q8.2: Do team members know who to contact or who has expertise?
- Q8.3: Do team members see the importance to understand the day-to-day work of their colleagues?
- Q8.4: When work is assigned, is everybody clear about his/her task?
- Q8.5: Do colleagues provide timely information about changes?

Out of the social network analysis categories defined by Ehrlich and Chang [EC06] related to awareness – general awareness, current awareness and familiarity – we chose to adapt current awareness for a detailed role-based social network analysis in Q8.1, as it is supposed to be the most relevant aspect for monitoring team collaboration. General awareness was considered in Q8.2, but only with a simple statement instead of a complete social network analysis. Familiarity and also importance were omitted in order to keep the number of questions and GSE metrics at a reasonable size. Q8.2 is also related to question 37 of Herbsleb and Mockus [HM03]. Q8.3 is adapted from question 23, Q8.4 from question 22, and Q8.5 from question 19 of Herbsleb and Mockus [HM03].

G9 refers to culture and language. While in particular culture is a broad field, as opposed to the other aspects on the Relationship level culture and language are rather static and don't change as much and as quickly during project lifetime as team cohesion, informal communication, awareness and team satisfaction. Typically, a change in the team constellation would also change the situation with respect to culture and language. Despite the complexity of culture and language, therefore we decided to include them with just two questions (Q9.1 is related to question 50 and 51 of Herbsleb and Mockus [HM03]):

- Q9.1: Do team members have appropriate English language skills?
- Q9.2: Has national culture significant influence on team collaboration?

G10 aims at analyzing team satisfaction. Questions include *How comfortable do team members feel in the project team?* (Q10.1), and *How satisfied are team members with the project's achievements?* (Q10.2).

Detailed definitions of the indicators can be found in Appendix A.

7.7 Measurement Cockpit View

Figure 7.7 and 7.9 show a generic measurement cockpit for a current project status on the Interaction level and the Relationship level. Furthermore, Figure 7.8 and 7.10 depict changes over time. In this generic example, a project with three development sites is shown: site 1, 2 and 3. The portions of the system developed at the three sites are all interdependent, i.e. all the cross-site communication links are relevant. On the Interaction level (Figure 7.7) the following can be observed: From the Communication Frequency diagram it can be seen that between site 1 and site 2 as well as between site 2 and site 3 there is frequent cross-site communication. However, the communication between site 1 and site 3 is striking: on average, they are communicating only less than every two weeks. Also the Reachability between site 1 and 3 is much worse than between the other sites. The Communication Media Profile site 3 is different from the other sites: they use less personal contact in their communication. Delays Due to Missing Information occur more frequently across-sites than within sites. The highest number of delays and also the longest delays are experienced at site 1 and site 3. Regarding Traveling, almost half of the team members from site 1 and about 30% of the team members from site 2 have ever visited other development sites, while people from site 3 have never visited other sites. The duration of visits in case of site 1 was about 25 days within the past 12 months, for site 2 it was about 13 days. The changes over time on the Interaction level of the past four measurement cycles (Figure 7.8) indicate that delays had increased significantly since the last measurement. Also the duration of visits went up in case of site 1.

On the Relationship level (Figure 7.9) the following is visible: Team Cohesion within the own site (local) is at a high level and well-balanced. With respect to the cooperation with other sites (distant) the Team Cohesion is lower. In particular the aspects open communication, reliability of colleagues and clear responsibilities show deficiencies. Also the Informal Communication across sites is at a low level. In Culture & Language no problems are visible. The Awareness ratings indicate a problem with information about changes as well as the knowledge of contact persons and about the task assignments. The Current Awareness graph shows a high awareness level between site 1 and 2 as well as site 2 and site 3, but again there is a problem between site 1 and site 3 visible. Finally, the team satisfaction shows positive ratings for site 2, but negative values in case of site 1 and site 3. The changes over time on the Relationship level (Figure 7.10) indicate the values have been going down since the last measurement, which affects most of all Team Cohesion and Awareness with respect to the distant site. Also in the Team Cohesion a significant decrease can be observed at both site 1 and site 3. Site 2 is not impacted in their Team Cohesion.

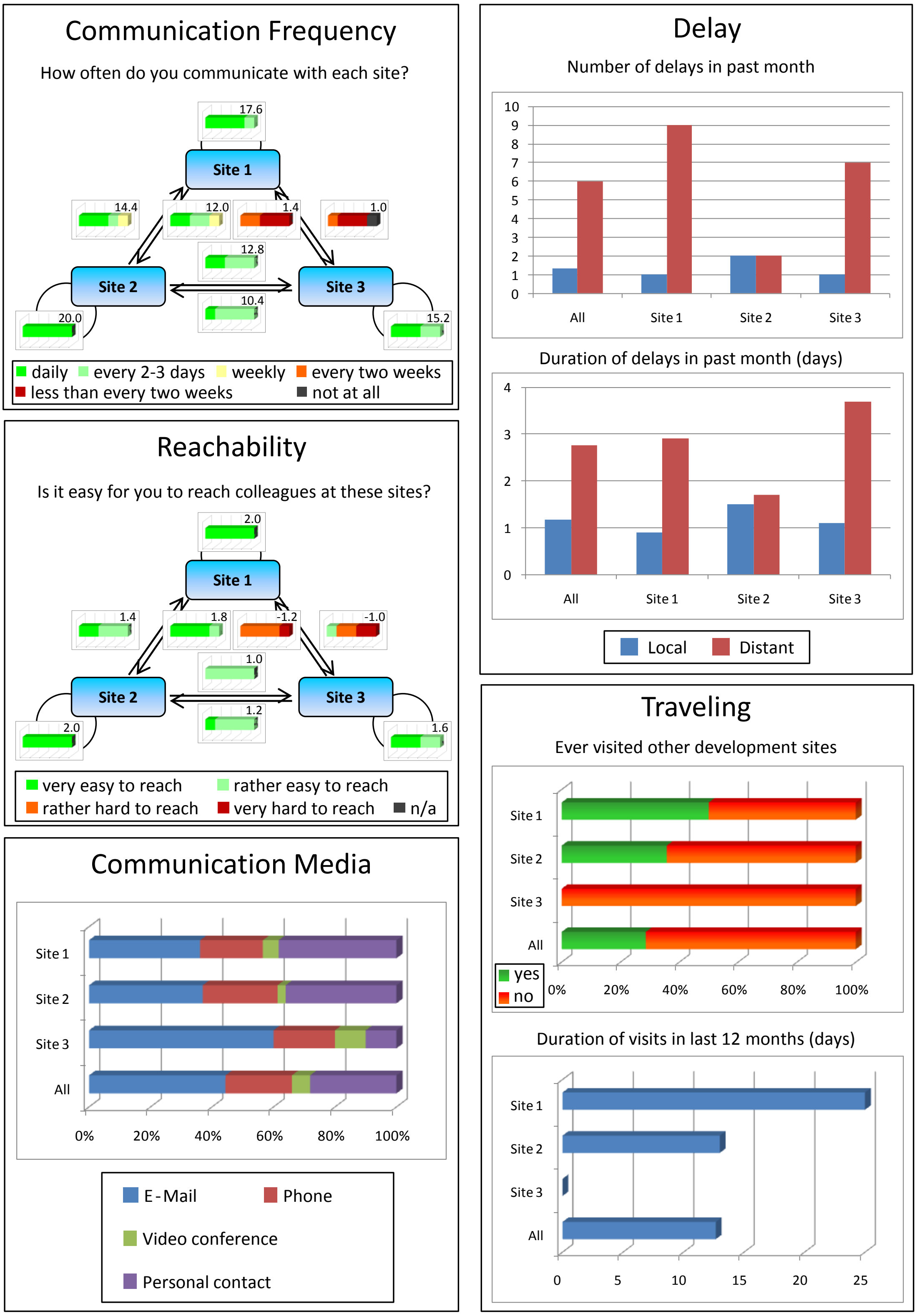
From this sample data it results that there is a collaboration problem between site 1 and site 3 in almost all aspects of the analysis that should be urgently clarified and mitigated.

7.8 Summary

In this chapter, we defined a model of collaboration-based metrics for project management support in GSE. We introduced an abstraction of measurement levels (the GSE metrics pyramid), defined measurement classes and derived corresponding GSE metrics. We gave an example for a cockpit view which provides a summary of the current project status to a project manager.

The metrics we defined above were also discussed with a psychologist from TU München, specialized on social psychology and also with some software engineering background, and confirmed to be a good selection.

In the next chapter, we will describe how the metrics were applied for the first time.



communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Figure 7.7: A generic measurement cockpit: Interaction level

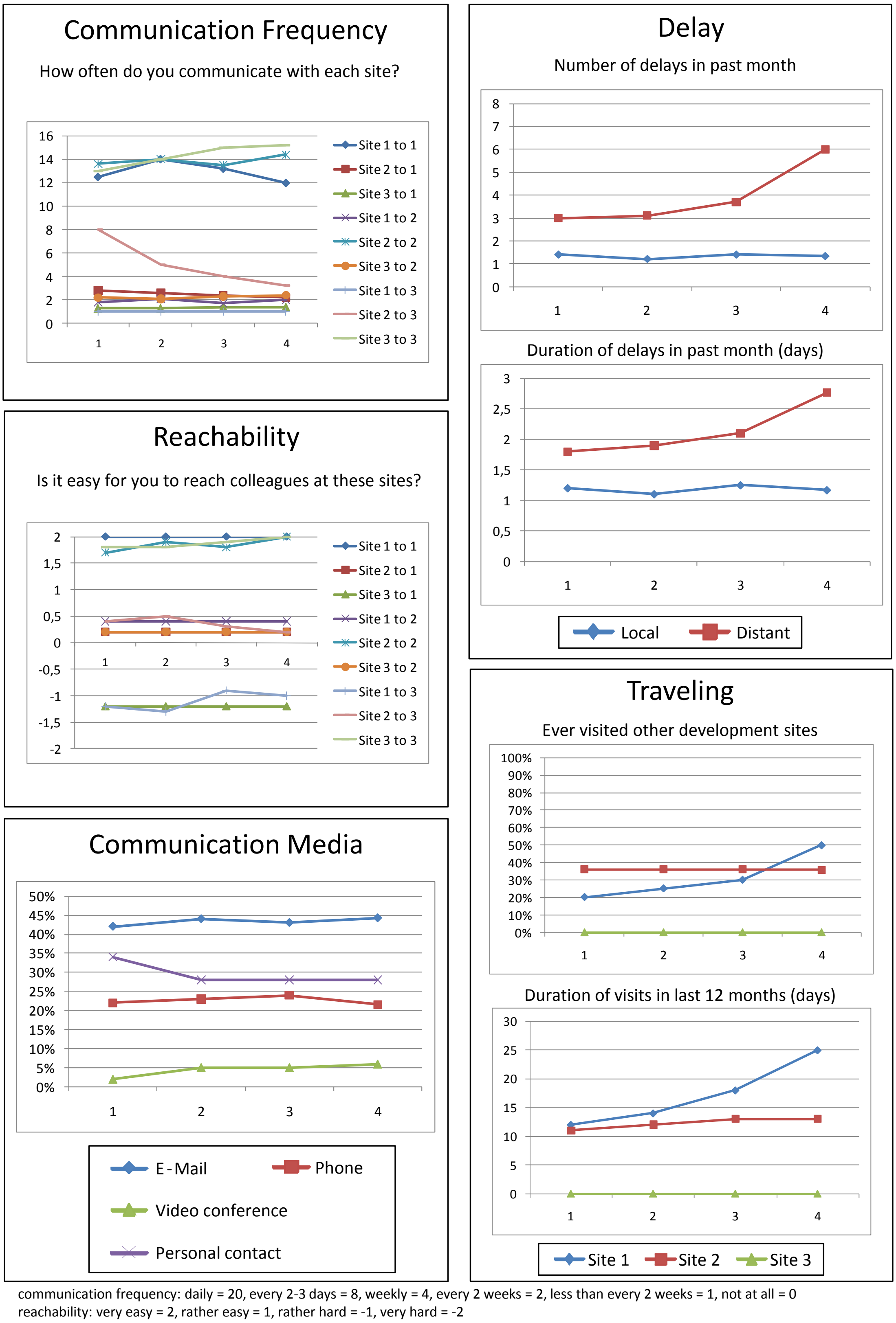


Figure 7.8: A generic measurement cockpit: Interaction level, changes over time

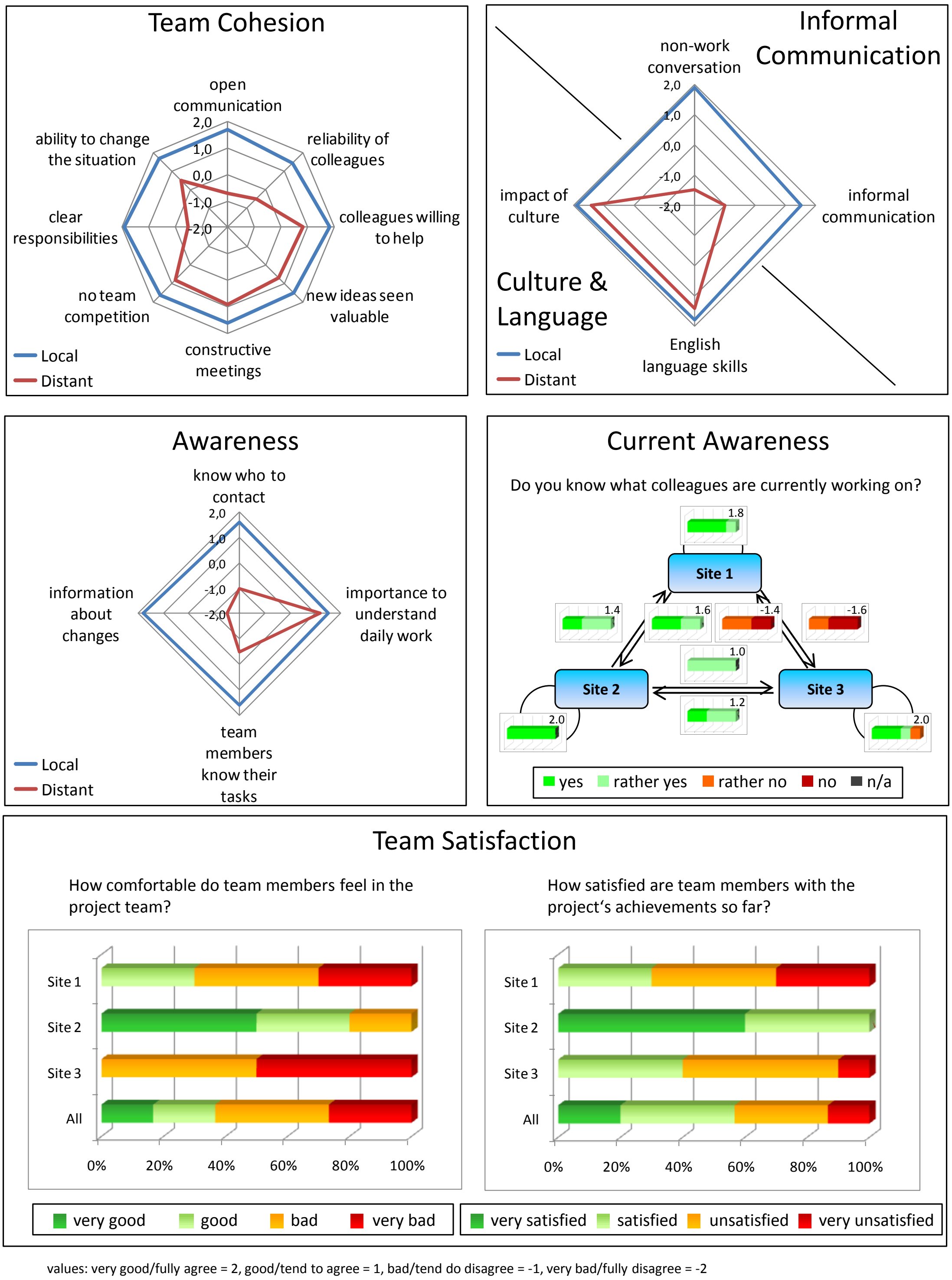


Figure 7.9: A generic measurement cockpit: Relationship level

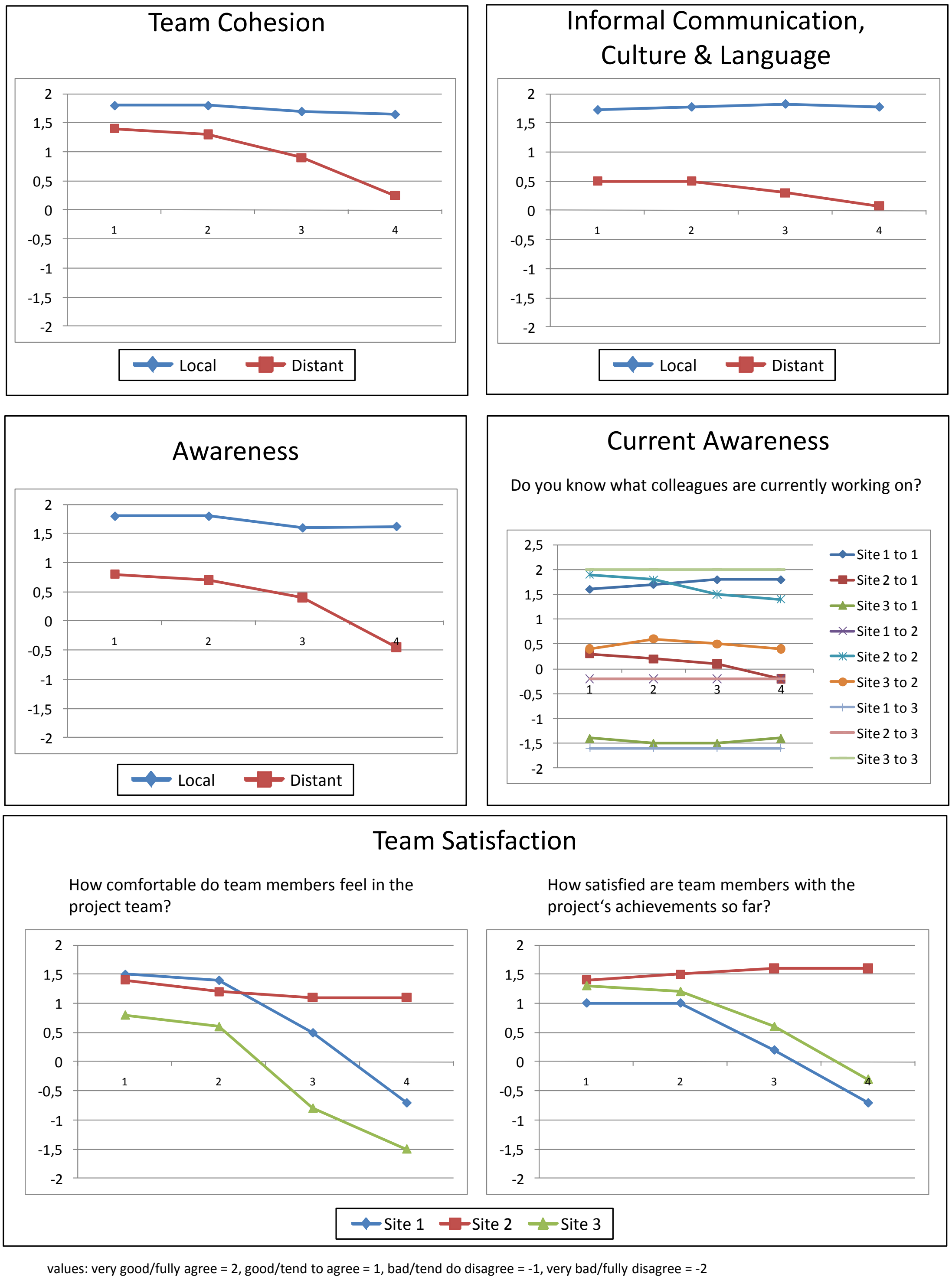


Figure 7.10: A generic measurement cockpit: Relationship level, changes over time

Chapter 8

Industrial Case Study

Following our research process, we defined and applied our GSE measurement model in realistic context. For this purpose, we selected a real-world industry project according to the criteria which we described in Chapter 3 (see Section 3.3.1): *Industrial project, globally distributed project, medium size project, risk project, willingness to support research.*

We were able to find a suitable project at the Energy Sector of Siemens AG and started a cooperation on the application and refinement of our GSE metrics. Siemens is a globally operating, integrated technology company with more than 400,000 employees and business activities in around 190 countries [Sie11a, p.116]. The Siemens company is active in the Sectors Industry, Energy and Healthcare and is – with a consolidated revenue of 75.978 billion Euro [Sie11a, p.116] and a Total Sectors profit of 7.789 billion Euro in fiscal year 2010 [Sie11a, p.145] – highly successful. Although Siemens is not perceived as a software development company, its products and solutions contain a significant software share. The company employs approx. 20,000 software engineers worldwide [Sie11b] and occupies leading market positions in the majority of businesses [Sie11a, p.116]. The project identified for our study fulfilled all the criteria mentioned above, as shown in Table 8.1. In the following, we call it *Project X*. **Since we deliberately selected a crisis project, Project X is not representative for the performance of other Siemens projects.**

8.1 The Approach

Figure 8.1 depicts the approach of our project study. It consists of six steps: analysis of the project context, conception, execution and evaluation of a survey with our GSE metrics, then improvement, and a final feedback. Figure 8.1 shows the detailed steps, along with the objective, task, result and responsible of each step.

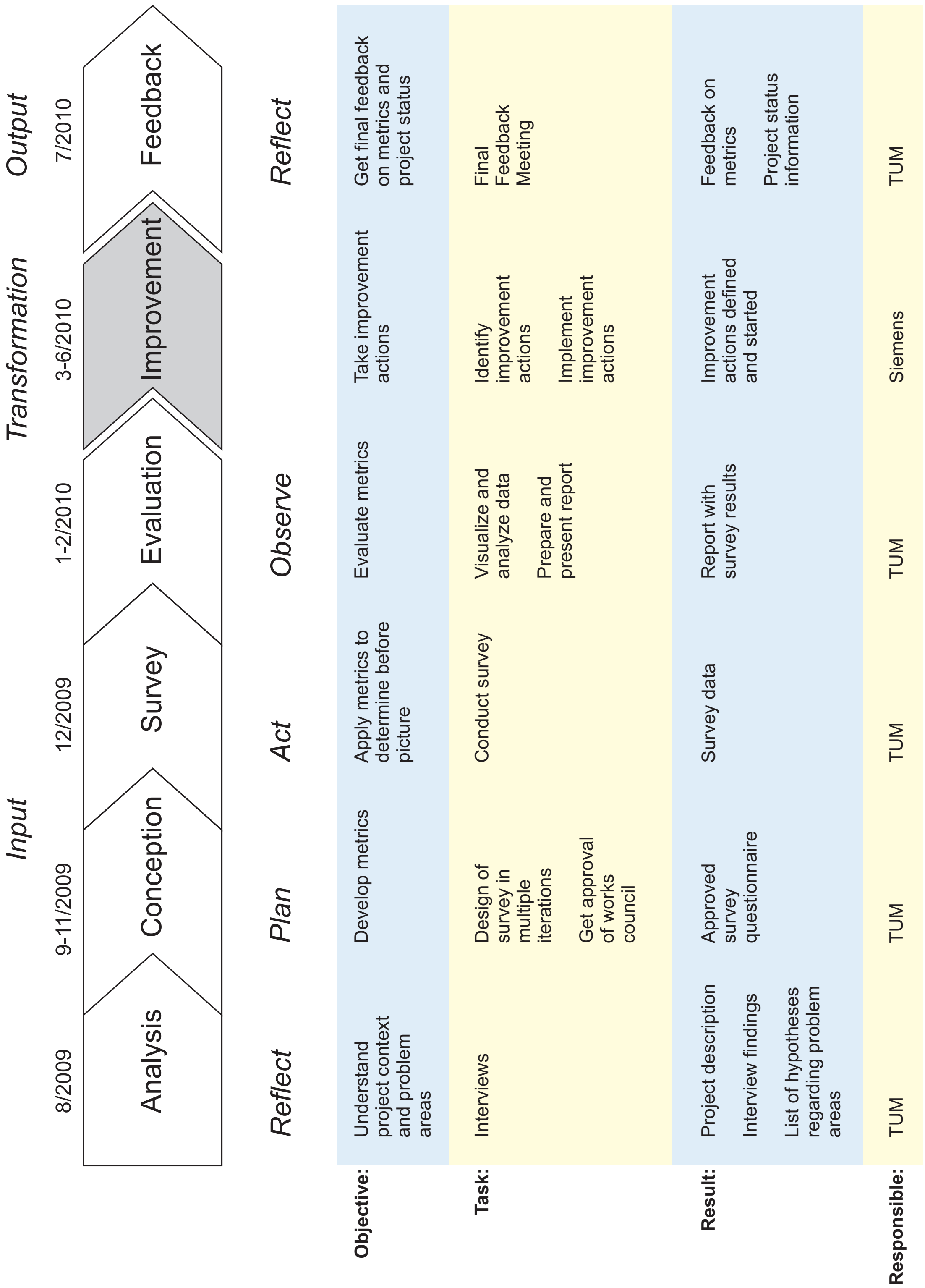


Figure 8.1: The approach of the project study (process view)

Criteria	Situation at Project X
Industrial project	Software Engineering project at Siemens AG, Energy Sector
Globally distributed project	Three sites in three different countries: Germany, USA (high cost) and one country in Central Eastern Europe (low-cost)
Medium size project	Approx. 50 project team members: 28 in Germany, 20 in Central Eastern Europe and 2 in USA
Risk project	At the starting point of our study, Project X was in a very critical state with a significant delay and no usable results visible
Willingness to support research	For the project selected, the tradeoff was to invest a certain amount of time and knowledge to contribute to our research, while gaining relevant insights from the project study to improve their situation

Table 8.1: Criteria for project selection and their fulfillment by Project X

- **Analysis.** Objective of the first step was to understand the project context and the problem areas. This was done by interviews, which delivered a project description, a documentation of interview findings, and first hypotheses regarding the problem areas. The interviews took place in August 2009.

- **Conception.** The purpose of the second step was to develop a suitable set of metrics to be applied in Project X. We used a survey questionnaire for data acquisition. This process step involved the design of a survey which was done in multiple iterations in cooperation with the project organization, and getting approval of works council for conducting the survey. The delivered result was an approved survey questionnaire.

- **Survey.** In order to determine the status of Project X, we applied the defined metrics and used them to identify major problem areas. This was done by conducting a survey in December 2009, which delivered data on communication and collaboration in Project X.

- **Evaluation.** Finally, the survey results were evaluated. Tasks included visualization and analysis of data, and preparation and presentation of the report with survey results, which took place in February 2010.

- **Improvement.** The objective of this step was to take improvement actions, based on the survey findings. This steps was driven by the project organization. Improvement actions were defined and implemented.

- **Feedback.** The objective of this final step was to gather the final feedback on our GSE metrics and the project status after the improvement. For this purpose, a follow up meeting was organized. This step took place in July 2010, about half a year after application of the survey.

8.2 Project Description

In this section, we give an overview of the characteristics of Project X. This information was gathered in the Analysis process step. Information sources were interviews with the project manager of Project X at that time, the Integration sub-project and line management in August 2009.

8.2.1 Project Setup: Locations and Organization Structure

Project X was conducted by a GSE project team, located at three sites: one site in Germany, one site in Central Eastern Europe, and one site in the US. The site in Germany was the headquarters location. The site in Central Eastern Europe was operated by an external company working as supplier to the German site (offshore outsourcing). In the following, we call the company in Central Eastern Europe *Supplier S*. Project X had approx. 50 project team members, thereof 28 in Germany, 20 in Central Eastern Europe and 2 in USA. Project X was started in spring 2007 and was intended to be completed in March 2010 (our study started in August 2009). The system under development was a collection of 14 applications in an application suite. Five of the applications were in the responsibility of the German site, partly with internal supply from Central Eastern Europe for development efforts. For the other nine of the applications, requirements were defined by the headquarters site in Germany, while Supplier S was in charge of the entire development, including design, code and user documentation. Finally, system test was done at the German site. While in principle the applications were independent from each other, one of the applications developed in Central Eastern Europe used a component of an application developed in Germany, i.e. Supplier S had to follow changes in the component developed in Germany.

At the time of the study, Project X was structured into the following sub-projects:

- Application 1–5
- Documentation
- Test
- Systems Engineering & Deployment, Configuration Management
- Integration

Each of the sub-projects *Application 1–5* were in charge of a set of applications. Furthermore, the *Documentation* sub-project was taking care of user documentation, *Test* was conducting the system test. Additionally there was a sub-project *Systems Engineering & Deployment, Configuration Management* responsible for handling the engineering of the system as well as configuration management of code and documents. A specific sub-project was the so-called *Integration* sub-project, a dedicated team which was introduced in March 2009 and was responsible for improving collaboration inside Project X in order to achieve a better integration of work results. The Integration sub-project was our main contact for the project study. Project X had very low attrition both in Germany and Central Eastern Europe. Except for one person in Germany who changed jobs for a different reason, at the point of time when we started the study the team was still the same as in the early phases of Project X. The organization was implemented in the way of a matrix organization, i.e. besides the aforementioned project structure with the project and sub-project managers as technical superiors, there was a line organization with a team lead acting as disciplinary superior.

Project X had a predecessor project implementing similar functionality, which was done about 10 years earlier, at that time developed as part of another system. However, the applications developed at that time were not usable in practice, in particular due to insufficient performance, usability and maintainability. Therefore Project X was started as new approach and as a new and independent product, based on old experience, but featuring a new team. Right after the start of Project X, product management intervened because the project as it was started was foreseen to get way to expensive and too late, and consequently product management looked for alternative ways to improve the situation. As result – after about one year of contract negotiations – a contract with Supplier S in Central Eastern Europe was concluded to take over major portions of the project. The new constellation with Supplier S resulted in the situation that at the German site there were now only managers, i.e. people who were developers before now were promoted to be sub-project managers. Project management and sub-project management was done by the German site, while most of the development was handled by Supplier S. There used to be also one sub-project manager located in the US, however, due to capacity reasons (USA was handling the pilot customer and therefore needed to spend much time there) this responsibility was transferred to Germany also. Project X used a waterfall process, enhanced with so-called *synchronization points* or *delivery points* – a specific set of milestones to ensure early testability. While the concept of synchronization points worked well in other projects, it was not successful in Project X, because the goal of early testability was not reached. The overall organization was assessed to have a high process maturity (CMMI[1]) level (Level 4), and a sophisticated organization-wide measurement system was in place, including regular cost, time, quality and earned value analysis.

8.2.2 Role of the Supplier Company

To engage the external supplier for Project X in Central Eastern Europe was a management decision, driven by product management. The contract was concluded in spring 2008. The supplier company had about 400–600 employees. Project X was not the first project done in cooperation with Supplier S. However, it was the first with this volume and with own responsibility for software development (before that, Supplier S had done service and commissioning projects). What made it particularly challenging was the fact that Supplier S was in the area of service and commissioning business a competitor of Siemens and also "snatched offers away from Siemens". In the attitude of the project manager at that time, it was not clear what the position of the own management was with respect to the cooperation with Supplier S: "Do they want the cooperation to be successful or not?" He mentioned there was no consistent attitude of persons in the management role. It was observed that managers used the problems with the Supplier S as argument for demonstrating that outsourcing doesn't work, or that a good collaboration with the development department is not possible. According to the interviewees, the challenges in working with the external supplier were underestimated. It was expected that the process would work smoothly: The German team would hand over the specification to the supplier, test the quality of the deliverables received, and integrate them. However, it turned out that the know-how at the supplier was often insufficient. The German site was dependent on the supplier as the German team did not have sufficient resources to do the development tasks on their own. Therefore the team tried not to attack the members of the supplier team, but to motivate them. According to the project manager at that time there was a lack of transparency. The team in Germany had no capacity to supervise what the team at Supplier S is doing. As a consequence, problems were detected very late. Overall Supplier S was recognized as cooperative partner. Several persons working for Supplier S in Central Eastern Europe were known for years, there were well established contacts. For example,

[1]Capability Maturity Model Integration, see Chrissis et al. [CKS06]

one of the project managers of Supplier S used to be for two years in Germany. Employees of the supplier company were eager to travel to foreign countries (it was seen as incentive). A large majority (about 90%) of the employees at Supplier S spoke English well; some of them were also able to speak German. In the cooperation with Supplier S, both fixed price and hourly based contracts were in place. Project X made the experience that fixed price contracts were bad for communication, as people tended to communicate less in this case. The project had about 500 requirements in about 300 pages of development specifications.

8.2.3 Communication and Development Infrastructure

With respect to the communication infrastructure and development environment, the German site and the supplier site had different prerequisites. As external company, Supplier S had no access to the Siemens network. Access to the source code was possible via a mirrored server (ClearCase MultiSite), which was synchronized regularly (approx. every 30–60 minutes). However, due to the missing access to the Siemens network, simple logging in and debugging was not possible. Furthermore, it was impractical to exchange large chunks of data in a secure way. As communication media, e-mail, telephone, and Microsoft Live Meeting/Communicator were used. Additionally, there were frequent personal meetings and traveling. However, because travel expenses had to be paid by the German company, the German team watched that no unnecessary traveling occurred.

8.2.4 Past Challenges and Measures

According to our interviewees, at the time when our project study started, there was a high frustration level within the project team. Appreciation for work results and feelings of success were completely missing, "The project is no fun". From employee's point of view, it was not worth fighting the problems or doing overtime, when they would get trouble anyway. Quality was a main issue for the de-motivation. With respect to the aforementioned delivery points, in Project X only few delivery points were achieved without drawbacks, e.g. only 1 of 10 delivery points. Team members, sub-project managers and the project manager felt powerless to change the situation. The overall atmosphere was negative. People were working against each other, which was seen different from any other projects before. This could be recognized by symptoms like missing readiness to help, forsaking others, and not accepting ideas from others. According to the interviewees, Supplier S had overextended itself on doing the development, but didn't want to show that they didn't have sufficient domain knowledge. There were significant quality problems with deliveries from Supplier S. In the interviewees' opinion, the majority of problems was caused by interpersonal relationship, not political or strategic reasons (estimation: 80% interpersonal relationship, 20% political). The technical people on the team were interested in getting the system to work, not in politics. Errors were made unwittingly, not by intention. However, there were also some reservations in the German team because Supplier S was a competitor to Siemens. Measures taken so far to improve the situation included sending one person from the German team for two weeks to Supplier S in Central Eastern Europe. This person acted as a bridge, as he was also able to speak the local language, which helped to improve the communication. Also, focus was put on that he would act as a colleague and not as a manager or supervisor, which created a trust basis for collaboration. Help was readily accepted by Supplier S.

Although Project X applied the metrics of the organization-wide measurement system, problems were detected very late. For example, the earned value analysis which is a recommended approach in literature to monitor the real progress of a project (cf. Section 6.3), didn't reveal the problems for a long time. According to the opinion of the interviewees a major reason was the missing openness in the team; work packages were declared to be 'done', even though they were not completely finished. Being asked about what the project study should focus on, the interviewees replied that focus should be on the collaboration between Supplier S and the German site, in particular investigating project-internal conflicts in the distributed team and to understand the frame conditions and root causes.

8.3 Project Retrospective

In order to identify problems in Project X and improve team collaboration, a project retrospective themed "Future Workshop" ("Zukunftswerkstatt") was organized. The workshop took place in August 2009, shortly before our interviews. Seven key people from the project participated: the project manager and six sub-project managers of the central site. Line management was deliberately excluded in order to allow for open discussion. The workshop was led by an external facilitator and lasted two days. As it was hard to convince the management to dedicate time for the workshop while the project was late anyway, the team found the compromise to have the workshop on Friday and Saturday (i.e. one working day and one day of the weekend, while preserving one day of the weekend as free time), so that the cost for the project was only one day. In the opinion of the project team, the optimal duration would have been three days, as the time to derive and agree on measures was perceived as too short (note: three days is also the duration recommended by Kerth for project retrospectives [Ker01]). The project retrospective took place in the office facilities due to cost reasons. If affordable, a location outside the usual work place would be preferable, as there is the high risk for people to be distracted from concentrating on the project retrospective by their regular work environment.

Figure 8.2: Traffic signs characterizing the project's situation drawn by the participants of the project retrospective

8.3.1 Activities in the Project Retrospective

The project restropective comprised the following activities:

Round of introduction. In a classical round of introduction, participants introduced each other. Although the team members were known to each other before, the round of introduction was helpful as warm up exercise and revealed new information about team colleagues (e.g. hobbies).

Reflection on the project's situation. All workshop participants were asked to reflect on the current situation of the project. This was done in the way that every participant made a drawing of a traffic sign which stands for the project's situation, which then were put on a pin board and discussed. Results were traffic signs for a building site (everything still under construction), slip danger (risk to go astray), no standing at any time (high time pressure) and falling rocks (dangerous way, risk that one's road will be blocked), see Figure 8.2.

Appreciation of project achievements. Positive aspects and achievements of the project team were collected and appreciated. The instruction of the facilitator to the participants was to collect snapshots like in a photo album. The metaphor was that a photo album usually contains nice pictures only. Each 'snapshot' of a positive project achievement was written down by the workshop participants on a moderation card and put on a pin board as 'project photo album'. Snapshots included statements such as

- Chance/possibility to realize something new
- Augmentation of one's horizon: technical and personal
- Long discussions, finally reasonable solution
- High engagement – will to succeed
- High willingness to acquire new knowledge at the supplier site

Analysis of relationships via sociograms. Each participant was asked to draw a sociogram from his/her perspective to illustrate relationships, conflicts as well as informal relationships. The results were discussed by all participants. This was a key element on the first day on the project retrospective workshop, and discussion of the results continued on the second day of the workshop. Figure 8.3 shows two examples out of the seven sociograms (to protect the privacy of the participants, names are replaced by letters in this figure; each letter represents the same person in both figures). Lines illustrate communication relationships, the line width indicates the intensity of communication. Dotted lines were used to represent informal relationships; the flash symbol indicates conflicts or communication problems. In the two examples in Figure 8.3 had in common a conflict between A and C as well as intensive communication links between A and B/D/E/F, while the role of G was understood differently. The sociograms exercise was well received by the participants of the project retrospective workshop, because the sociograms helped to visualize and discuss their views of team relationships.

Team exercise on requirements communication. Participants were asked to group in teams of two persons. They sat on chairs back-to-back so that they could hear each other but didn't see what the other was doing. In this setting, the task was for the one team member to verbally specify a drawing which the other team member had to produce on a sheet of paper. The learning experience was that communication can be ambiguous and that good communication is necessary to achieve a good result.

Agree on improvement measures. At the end of the workshop, measures were derived and agreed on. This was done in the form of a written agreement – kind of a contract – signed by all participants. Unfortunately there was no

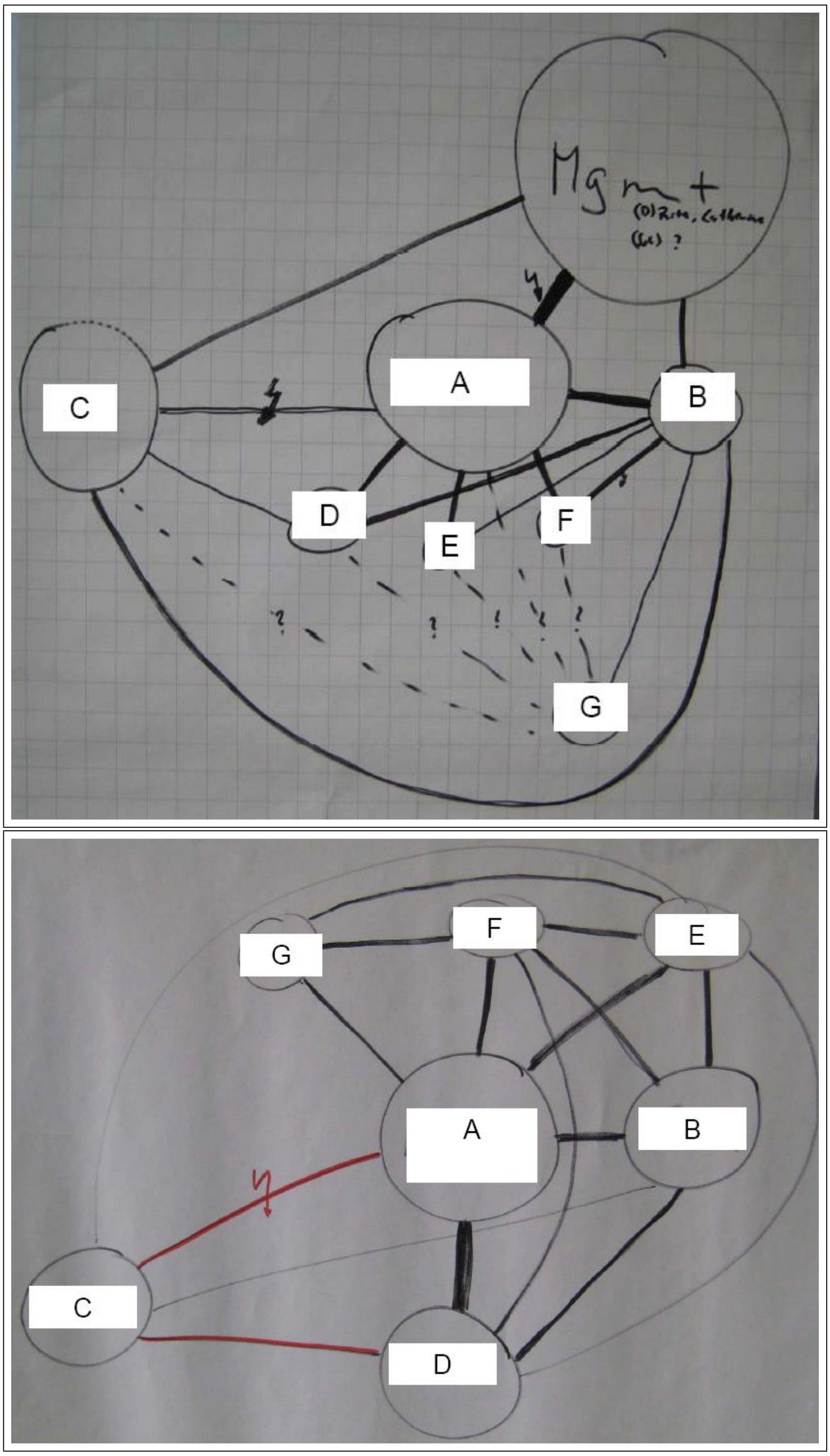


Figure 8.3: Two examples of the sociograms of the project team drawn by the participants of the project retrospective

sufficient time to complete this step.

Additionally, basic knowledge like the iceberg model of communication and the 4-ears-model of Schulz von Thun [SvT81] were explained (cf. Section 6.2), which was new for most of the team members and lead to an 'aha effect' of the participants when they noticed that this was exactly the case in certain situations of the project. To focus on team collaboration was seen as valuable by all participants. It was decided to conduct monthly meetings of the project manager and the sub-project managers with duration of approx. 1 hour to regularly address collaboration issues.

8.4 Design of the Survey

The goals of the application of our GSE metrics was to identify collaboration problems in the project team (project goal) and at the same time the construction of our GSE metrics as well as to get feedback on them (research goal). The metrics to be applied in Project X were designed based on the model of GSE metrics as described in Chapter 7. The metrics were discussed and refined over several iterations with our project partner. In order to keep the effort for data collection low, we decided to use a survey questionnaire for data acquisition to be filled out by the team members of Project X. This also had the advantage of avoiding privacy issues, as people provide data voluntarily as opposed to analyzing e.g. their communication data such as personal e-mail.

We adapted the metrics of our model defined in Chapter 7 for the application in Project X. We added questions to validate our first hypotheses on problem areas from statements we heard in our interviews. Problem areas, which were not covered by the original set of GSE metrics and therefore were additionally introduced were related to management and strategy, requirements and quality problems:

- Management has a clear strategy towards the cooperation with the supplier company
- The challenges of distributed collaboration were underestimated when the project was initiated
- Project successes are adequately appreciated by management
- Requirements are clearly defined and communicated
- Quality problems are one of the major root-causes for project delay

The two questions *Are there many discussions about particular responsibilities?* and *Do team members feel powerless to change the project's situation?* which in our GSE model belong to team cohesion were grouped together with the five Management&Strategy-related questions listed above because of their close relationship. Furthermore, we added questions on two additional statements to get information about the degree of reachability and current awareness of people, as the survey just included a yes/no question:

- People I need to communicate with are difficult to reach
- I have sufficient insight into what my colleagues are currently working on

Major aspects in designing the survey were the target group and privacy. While originally the metrics were planned to be collected from all team members at all three sites (about 50 data points), the scope was narrowed down in several steps. First, instead of collecting identifiable personalized data, it was decided to ask for the roles instead of names, to protect the privacy of the team members. Eight roles were identified: Developer, Tester, Configuration Manager, Project Lead, Sub-Project Lead, Architect, Quality Manager, and System Integrator (up to 8 data points per site). After that, it was decided by management that only the German site should be asked to fill in the survey, in particular because Supplier S was a separate legal entity and external to Siemens (8 data points, only for the German site). Finally, the survey questionnaire had to be approved by the works council in Germany. The approval was successful, but under the restriction to reduce the number of roles to 3 groups to ensure that there is a sufficient number of people in every group so that individual respondents are not identifiable (3 data points). Further restrictions requested by the works council were that the data collection had to be done via a paper questionnaire instead of an electronic data collection, as it was originally planned. The paper questionnaire had to be sent directly to the author of this dissertation rather than to a member resp. manager of Project X. Participation in the survey had to be fully voluntary.

The full survey questionnaire can be found in Appendix B.

8.5 Results of the Survey

The survey was distributed as paper questionnaire to 28 team members of the German site. In total, 24 persons answered the survey which equates to a response rate of 86%. The survey therefore is considered representative for the team at the German site. The distribution of roles of respondents is as follows: 9 answers (37.5%) were from the roles *developer and architect*, 10 answers (41.7%) from the group *project lead, sub-project lead, and quality manager* and 3 answers (12.5%) from the roles *tester, system integrator, and configuration manager*. In 2 of the returned questionnaires (8.3%) the role was not specified. For ease of reading, in the following we abbreviate the 3 groups of roles with *Developer*, *Project Lead* and *Tester*. In the following we focus only on key results. The detailed results can be found in Appendix C.

8.5.1 Interaction Level

Figure 8.4 shows the measurement cockpit view on Interaction Level. Due to the constraints explained above with respect to collecting communication data, the social network analysis contained in the questionnaire only distinguished 3 groups of roles instead of individual team members. Therefore the possibilities to analyze the data are also very limited. However, it turns out that even this highly aggregated data is still very valuable. The measurement cockpit view in Figure 8.4 contains the communication frequency and reachability of the 3 groups. The arrows in the figure illustrate the direction of communication, e.g. the arrow from Developer towards Project Lead represents what developers stated about their communication with the role Project Lead. The circles at the roles refer to communication with colleagues of the same group of roles. The diagram on communication media depicts the share of communication media used in Project X. Furthermore, the cockpit view includes the frequency and duration of delay due to missing information, as well as frequency and duration of traveling (average values of all answers). Based on this aggregated view, the following observations can be made:

Communication Frequency. As we can see from the figure on communication frequency, most respondents stated that they communicate *daily* or *at least once a week*. A striking result is that some people of the group Project Lead answered that they communicate *at least once a month* only, which seems insufficient. This is valid for all directions in which Project Lead are communicating: towards Developer, Tester as well as to colleagues of the same role.

Reachability. According to the social network diagram on how easy it is to reach team members of the given roles, reachability in total is very positive. The only deviance is visible with the internal reachability of the role Project Lead, where 30% state that it is not easy to reach colleagues of the same role, while for all other relationships in the same diagram the value is never above 10%. In addition, we asked to evaluate the sentence *People I need to communicate with are difficult to reach*. The results of this question conform with the social network analysis on reachability: A vast majority of respondents state that reachability is not a problem, where the local site appears better than the distant site, which represents the usual effects of distribution. There were also hints at hidden communication problems in the comments from the survey: “Between sub-project

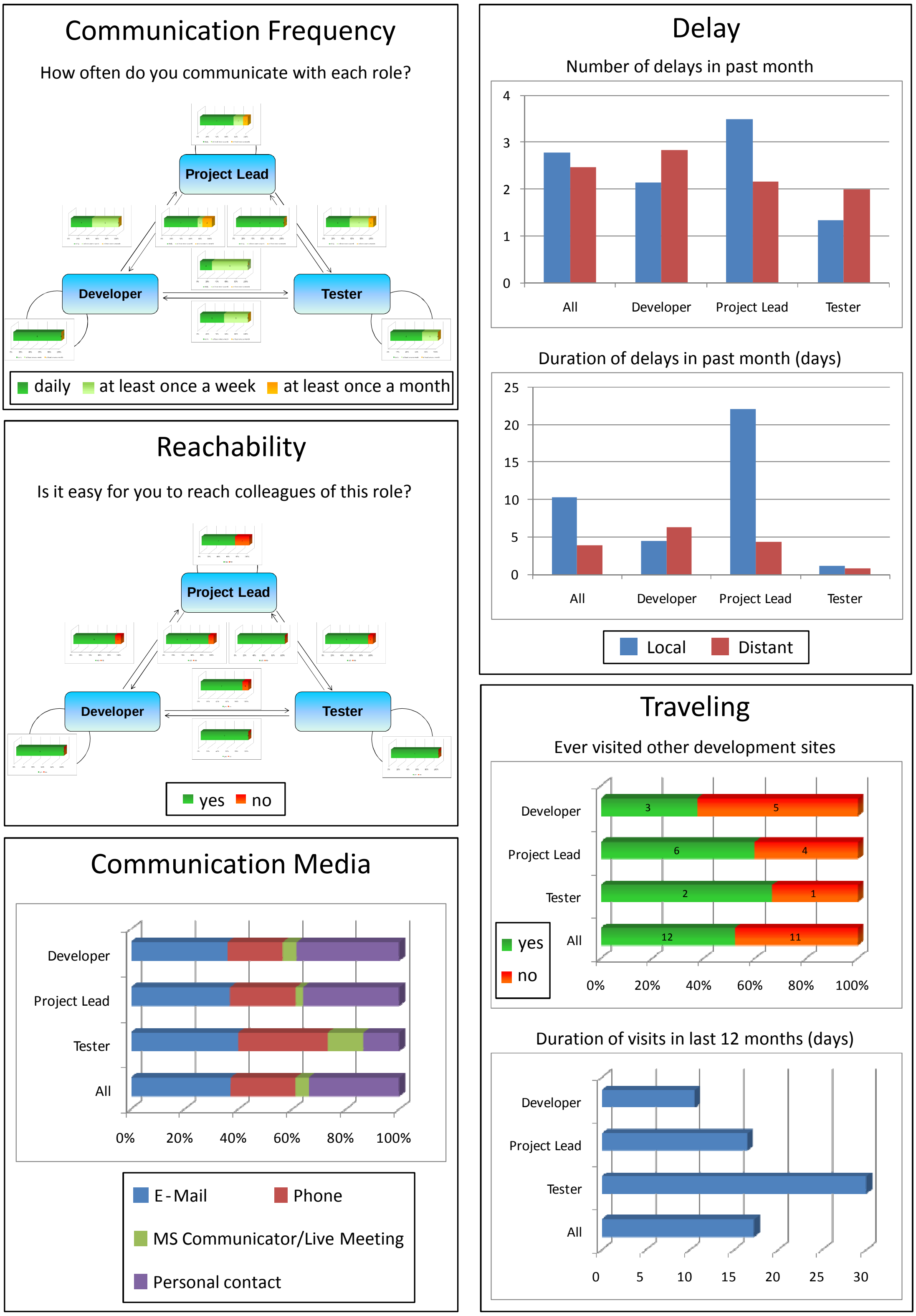


Figure 8.4: Measurement Cockpit for Project X: Interaction Level

leads there are tensions (possibly interpersonal issues); because everybody is under high pressure and all are delayed, there is very low tolerance." While overall reachability is very positive, the limited internal reachability of the role Project Lead was conspicuous. Also the comments of survey respondents hint at tensions between sub-project leads.

Communication Media. The project team at the German site on average uses 37% e-mail communication, 24% phone communication, 5% Microsoft Communicator/Live Meeting and 35% personal contact. There are no noticable problems in usage of communication media.

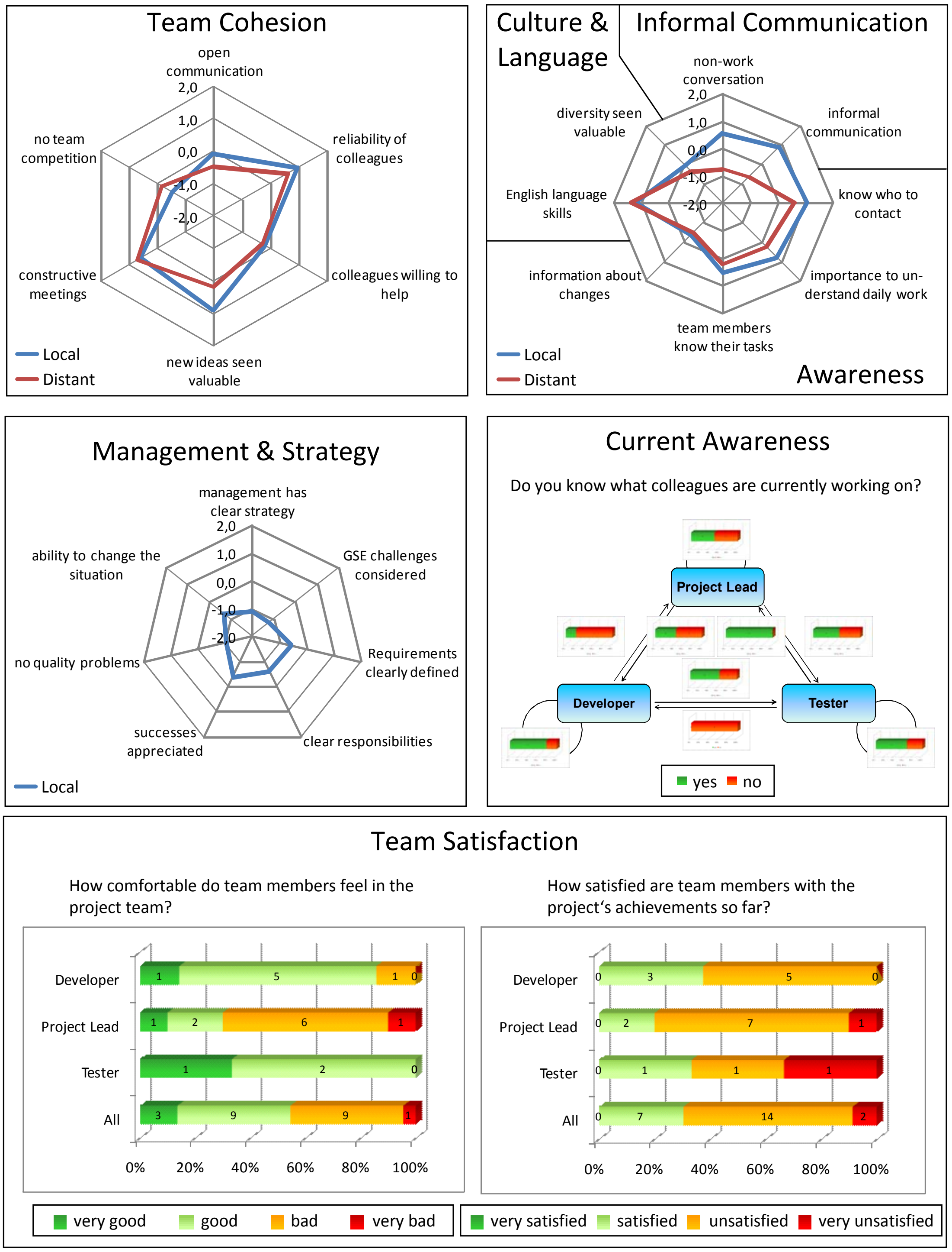
Delay Due to Missing Information. On average, the role Developer had 2.1 delays in the local case and 2.8 in the distant case. In the group Project Lead it were 3.5 for local and 2.2 for distant. For the role Tester the number of delays was on average 1.3 local and 2.0 distant. The expected result would have been that there are more delays with respect to the distant site than within the local site. This is confirmed in case of Tester (50% more distant delays than local delays) as well as Developer (32% more distant delays than local delays). However, in case of the role Project Lead, the opposite was noticed: There were 62% more local delays than distant delays. This result hints at local problems that cause significant delays in the Project Lead's work. Also with respect to the duration of delays the group Project Lead shows a striking result: here the average local delay is 22.1 days, while for the distant site it is only 4.4 days. Although it is a well-know effect that delay occurs in a distributed setting, the local delays in case of the role Project Lead outweigh that. The conclusion is that there must be local problems that cause significant delays in the Project Lead's work.

Traveling. A high portion of the team in Germany has already visited the other development site of the project: 38% of the role Developer have already visited other development sites, 60% of Project Lead and 67% of Tester. The average duration (number of days) spent at other sites within the last 12 months amounts to 10.5 days for Developer, 16.5 days for Project Lead, and 30 days for Tester. In total, the results show an intensive exchange across sites.

In summary, on Interaction level our early warning indicators hint at local problems with respect to the group Project Lead.

8.5.2 Relationship Level

The measurement cockpit view on Relationship Level is shown in Figure 8.5. Kiviat diagrams illustrate the results regarding team cohesion, informal communication, awareness, culture & language as well as management & strategy. The survey answers (*fully agree*, *tend to agree*, *tend to disagree* and *fully disagree*)



values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Figure 8.5: Measurement Cockpit for Project X: Relationship Level

were translated to values from +2 to –2 (*fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2) to calculate the average answers for the Kiviat diagrams. In case of negative statements, the additive inverse value was taken so that positive values always represent a positive state. In the following, values in parantheses depict the average results regarding local and distant site, e.g. '(+1.0;+0.6)' refers to a results of +1.0 for the local site and +0.6 for the distant site. Additionally, the social network analysis diagram of current awareness and the ratings related to team satisfaction are shown in the measurement cockpit. The following can be observed from the data:

Team Cohesion. The most positive ratings were obtained for *reliability of colleagues* (+1.0;+0.6), *new ideas seen valuable* (+0.9;+0.2) and *constructive meetings* (+0.6;+0.7). Comparing the results of local and distant site, colleagues from the distant site are seen as less reliable compared to local colleagues. Similarly, new ideas from the distant site are seen less valuable compared to the local site. Both are explainable as the usual effect of distribution and impact on trust; the effect might be emphasized in case of the project study through the fact that the supplier is a competitor company. With respect to the statement *Meetings/phone conferences are constructive*, the distant site received a better result than the local site, which is against the effect of ingroup favoritism and might hint at problems with meetings/phone conferences being constructive at the German site. The answers regarding *open communication* (–0.1;–0.5), *colleagues willing to help* (–0.2;–0.3) and *no team competition* (–0.5;–0.2) revealed serious problems in team cohesion. A clear problem with communication not being honest and open is indicated. Furthermore, the statement *There is individual competition between sub-project teams* is strongly confirmed. It is particularly striking that there seems to be much more competition between sub-project teams at the local site compared to the distant site. A breakdown of team cohesion results per role is shown in Figure 8.6. From this diagram it is obvious that the role Project Lead provided the most negative ratings in almost all aspects of team cohesion. The strongest effects are observed with respect to *no team competition* and *constructive meetings.* We conclude that there is individual competition in particular between team members of the role Project Lead (i.e. between sub-project leads and project lead).

Comments from the survey confirm this view:

> "The problem is that even on local side the team is split. Colleagues which don't have knowledge about technologies used for [Project X] try to prove that they are anyway useful while pointing out what the others are doing wrong. Small problems are presented as big, very big issues!!! They are playing politics in the hope that the managers will not realize what is really going on. HOW SHOULD BE THE COMMUNICATION HONEST?"

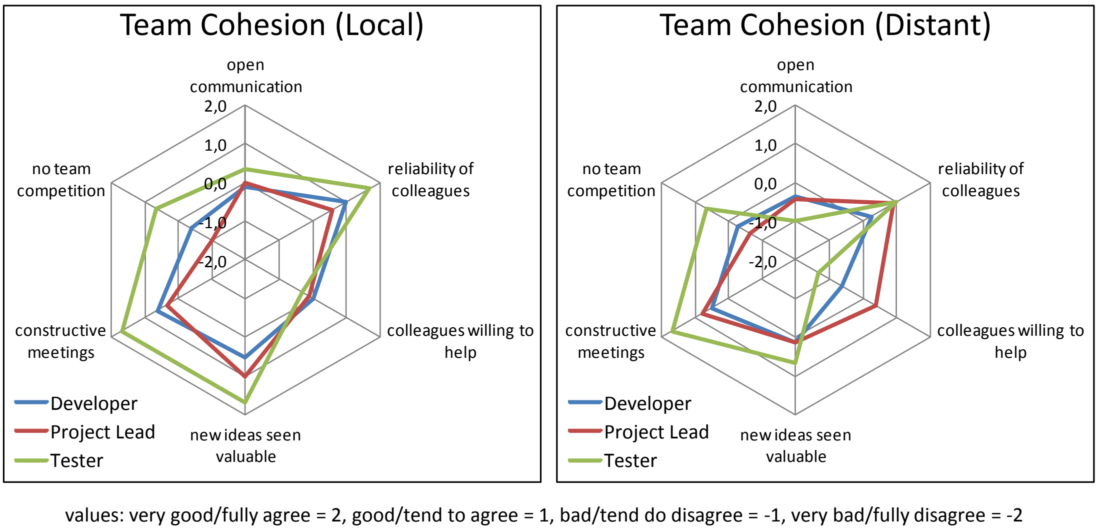


Figure 8.6: Measurement Cockpit for Project X: team cohesion per role

Also problems with the supplier company, similar to those mentioned in the interviews (see Section 8.2) were mentioned in the comments:

> "The problem with the distant site is that the company we are working with is not serious and not professional, there is a big lack of knowledge and experience, and again politics is used to hide it. They are presenting themselves as experts, but expect us to teach them ON PHONE basics about data processing and power systems, but managers stick to the way how they are presenting themselves."

And another comment related to team cohesion:

> "In my own team I can rely 100% on my colleagues (with everything), my team means all colleagues below one sub-PM. Between other teams there is a competition. New ideas are not appreciated at all, they are understood as an attack on old establishment. Meetings are time consuming and not intended to solve any problems (it looks like that), they are used as presentation platform for the participants."

Informal Communication. To probe informal communication, the two statements *I discuss non-work related matters with other project team members* (+0.5; –0.8) and *I often get useful work-related information through informal communication* (+0.9;–0.7) were evaluated. The results constitute that useful work-related information through informal communication is gathered much more locally than distant. This conforms to the expected behavior in GSE.

Awareness. The questions on awareness consisted of the aspects *know who to contact* (+1.1;+0.6), *importance to understand daily work* (+0.8;+0.2), *team members know their tasks* (+0.5;+0.2) and *information about changes* (–0.3;–0.5). Additionally, we asked for social network analysis data on current awareness as well as the evaluation of the statement *I have sufficient insight into what my colleagues are currently working on* (see Figure 8.7 and 8.8). The negative ratings of the statement *My colleagues provide timely information about changes in current plans* suggest that there are problems in the project team with being informed on time about changes. According to the social network analysis results in Figure 8.5, the overall situation with respect to current awareness is not as good as communication frequency or reachability investigated with social network analysis above. We found it noticeable that 50% of the role Project Lead don't know what others of this role are currently working on. Furthermore, the majority (56%) of the role Project Lead stated that they don't know what colleagues of the role Developer are currently working on. On the other hand, even 78% of the group Developer don't know what the Project Lead is working on. Also 44% of Project Lead don't know what the Testers are currently concerned with. Figure 8.7 and 8.8 show that awareness with respect to current work activities of colleagues is not optimal. In particular in case of the distant site, there is overall an insufficient insight into what colleagues are currently working on, which is symptomatic for a distributed team. In case of the local site, the values of the role Project Lead are noticeable: All of the negative values originate from this role, which indicates a communication problem with respect to the role Project Lead in collaboration at the local site. Referring to the distant site, more effects can be observed: half (50%) of the role Developer state that they fully disagree to have sufficient insight into what their colleagues at the distant site are currently working on. The other 50% tended to agree with the statement. The role Project Lead showed 22% agreement (0% fully agree, 22% tend to agree) and 78% disagreement (44% tend to disagree and 33% fully disagree). This is a critical situation as the Project Lead persons are in charge of coordination with the distant team members. While part of the results related to awareness can be explained by the effects of distribution, the survey results indicate communication problems with respect to the role Project Lead in collaboration at the local site.

Culture & Language. Regarding culture and language, survey respondents rated the statements *Project team members have appropriate English language skills* (+1.2;+1.4) and *Diversity in the project team (e.g. national cultures) is valuable for team collaboration* (0.0;–0.4). There were no significant problems noticeable with respect to culture and language. English language skills are no problem.

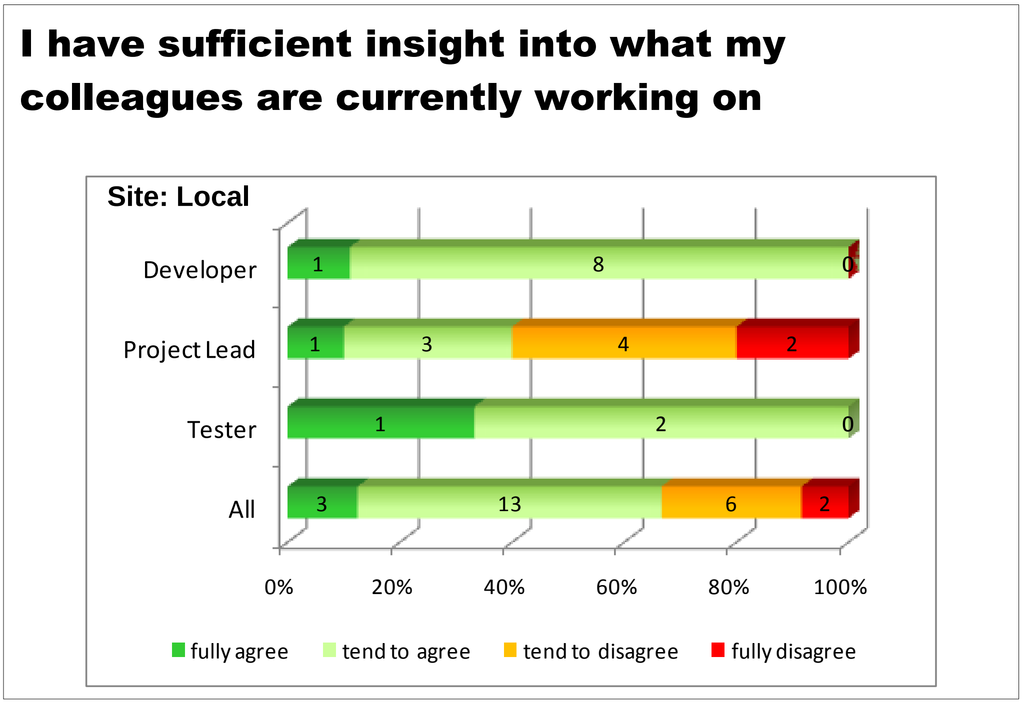


Figure 8.7: Answers for *I have sufficient insight into what my colleagues are currently working on* per role, local

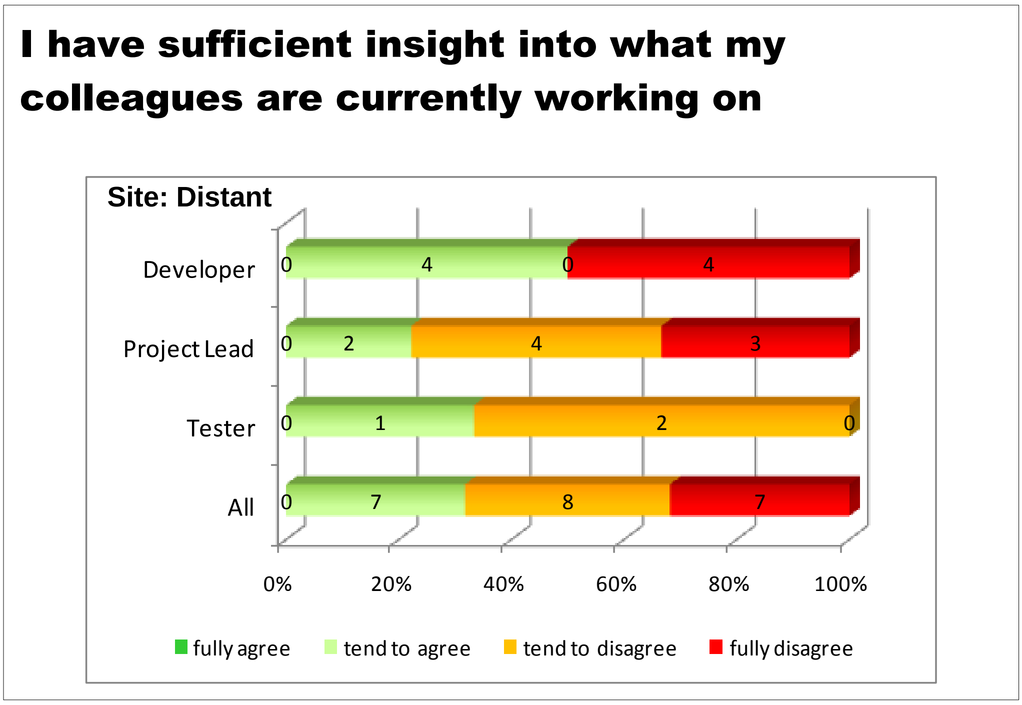


Figure 8.8: Answers for *I have sufficient insight into what my colleagues are currently working on* per role, distant

Management & Strategy. The team members' opinion related to management and strategy were investigated with seven questions (please note that team members were asked only for a rating of the local site, therefore in the following parantheses include only one value): *management has clear strategy* (–1.1), *GSE challenges considered* (–1.2), *requirements clearly defined* (–0.5), *clear responsibilities* (–0.6), *successes appreciated* (–0.4), *no quality problems* (–1.0), *ability to change the situation* (–0.7). As all seven questions yielded negative ratings, the situation seems critical with respect to multiple aspects:

- From the team's point of view, there is no clear strategy towards the cooperation with the supplier company
- Challenges of distributed collaboration were underestimated
- Project successes are not adequately appreciated by management
- Requirements are not clearly defined
- Quality is a major root cause for project delay (interrelated to unclear requirements)
- Project team members feel powerless to change the situation

Team Satisfaction. Concerning team satisfaction, we asked the project team members how comfortable they feel in the project team and how satisfied they are with the project's achievements so far (see Figure 8.5). Regarding *How comfortable do team members feel in the project team?* almost all of the negative answers originate from Project Lead. Obviously people of the group Project Lead don't feel well in their situation. With respect to *How satisfied are team members with the project's achievements so far?*, only few respondents stated that they are satisfied, and again the role Project Lead gave the most negative answers. The text comments reveal further insights:

> "The Sub-PMs are not able to detect gaps and issues inside the subproject. Partially the unqualified Sub-PMs are responsible for the sub-projects. It brings huge problems in the understanding and communication. Normally they try to transmit all technical problems to developers and do a pure administrating work."

> "Individual goals are given more importance as team goals. The skills of many persons (developers) are not properly utilized. Task planning is weak. Despite all this, we are moving forward."

> "With better internal collaboration and a better contact of all team members to the remote site the project would be in a better shape."

In total, nobody is really satisfied with the current situation of the project. In particular team members of the role Project Lead don't feel well.

8.6 Summary of Findings

In summary, the following aspects turned out to be positive:

- Overall team situation not too bad
- Good English skills
- Exchange, traveling

We identified the following major problem areas:

- Local problems with respect to the role Project Lead
- Competition between teams
- Management and Strategy
- Handling of requirements

One of the respondents aptly summarized the situation:

> "The way how is communicated is important. Respect for everyone from everyone would be helpful. Only together as one team the challenges could be managed."

8.7 Recommendations

Based on the findings from our early warning indicators, three major recommendations can be derived:

Improve the local Project Lead and Sub-Project Lead organization. From our study, it results that there are significant collaboration problems with the role Project Lead, including communication frequency, reachability, individual competition, and possibly qualification. Therefore special attention to the Project Lead and Sub-Project Lead organization is necessary. We recommend to discuss the situation with the team members of the role Project Lead and jointly decide about improvement actions.

Clearly communicate the management strategy. The strategy of management is not seen from the team's point of view. We recommend to further the communication between management and project team and clearly communicate the management strategy.

Improve Requirements Engineering. The survey indicated major problems related to requirements definition and management. Focus shall be put on ensuring that requirements are understood and requirements changes are communicated well within the project team.

8.8 Final Feedback

In July 2010, about half a year after the application of our metrics, a follow up meeting with the project partner took place to get their feedback and learn about the current status of Project X.

Based on the results of our survey, the management of Project X had taken fundamental measures:

- **Sub-Project Management Level.** The sub-project management level was completely eliminated. The individual situation of the former sub-project managers was discussed in one-on-one interviews between them and line management. Five of the former sub-project managers moved to a different organization, eight of them remained in the team and now work as 'senior developers'.

- **Line Organization.** Differing from the rest of the organization, Project X was transferred into a line organization with technical and disciplinary responsibility, i.e. an own department. The purpose of this step was to strengthen the team identity and the team cohesion.

- **Domain Knowledge Training.** A domain knowledge training was conducted inside the team, and coaching was introduced.

- **Management Strategy.** The management strategy and the rationale behind was explained to the team.

At the beginning, the team members showed some reservation with respect to the changes. However, after a short period the measures were well received. The line management emphasized the importance of the role of the senior developers. It was observed that due to these measures, the team was now more coherent. At the point of time of the follow up meeting, the product developed by Project X was tested and approved by system test, and thus only a few days were missing to the final release. The same team was about to start the development of the next version (V1.10) of the application suite, which was considered to be comparable in complexity and function volume to the first version.

With respect to the metrics applied, we received a very positive feedback from the follow up meeting. In particular the communication network analysis, the analysis of the team cohesion and the analysis of delays in the project were seen as very valuable. The measurements confirmed the gut-feeling which the team members and line management had before. However, only after seeing the 'facts' of the objective measurements, a change process in Project X started. According to the final feedback, the GSE metrics were very helpful to improve the situation of Project X.

Chapter 9

Validation

In the validation phase, we introduced our GSE metrics to multiple projects and observed their effectiveness in project management. In a first case study, we applied the metrics to four GSE student projects. In a second case study, we dealt with the application of the GSE metrics to a large co-located student project.

In Chapter 3 we established four validation criteria to evaluate if our approach was successful: accuracy, usefulness, effort and innovation grade. We now utilize the following information sources to evaluate these criteria:

- **GSE metrics data.** The metrics data itself reveals information about accuracy and usefulness. Because multiple projects are analyzed, a comparison between projects is possible, in particular also between a globally distributed and a co-located setting.

- **Semi-structured interviews.** In semi-structured interviews with team members, experiences in the application of the metrics as well as feedback regarding accuracy, usefulness and effort are investigated.

- **Feedback survey.** Feedback surveys filled in by team members resp. the team coaches and supervisors offer data regarding the accuracy, effort and usefulness of the GSE metrics.

- **Project reports.** In the first case study, part of the students' work was to author a report about the project, which provides an additional information source.

- **Own observations.** Finally also the own observations of the author of this dissertation reveal insights into the validity of the GSE metrics.

In the feedback surveys we used four-point evaluation scales, for example: *accurate, rather accurate, rather not accurate, not accurate*. To be able to interpret the feedback results we defined threshold values. For the criteria accuracy, usefulness and effort we demanded a clear majority vote, therefore we established a threshold value of 67% (two-thirds majority), i.e. at least two thirds of the feedback ratings had to be positive ratings – e.g. either *accurate* or *rather accurate*). Please note that a close-to 100% accuracy is not required because the metrics are used as early warning indicators which trigger a further investigation of the root causes. Regarding the effort besides the threshold value on the feedback results we also looked into the response rates from the GSE metrics surveys, i.e. how many people invested their time to fill in the surveys and postulated a target value of at least 50% response rate. Regarding the innovation grade, we demanded at least 90% positive feedback to accept our approach as innovative.

9.1 NEREID

In Section 4.3 we studied GSE student projects of the so-called NEREID course of winter term 2009/2010. In the following year, in winter term 2010/2011, we had the opportunity to accompany the new NEREID projects and apply the GSE metrics in four GSE student projects within this course.

Altogether, seven student teams were formed in the winter term 2010/2011 NEREID course[1]. For our study, we selected four out of the seven projects. The selection was done in such a way that we could cover a wider range with respect to the degree of distribution, i.e. 2 to 4 sites.

9.1.1 Project Description

Similar to the projects described in Section 4.3, the four NEREID projects were conducted by GSE project teams consisting of students from multiple globally distributed universities:

- Institut National des Sciences Appliquées de Lyon, France
- Tecnológico de Monterrey campus Puebla, Mexico
- Technische Universität München, Germany
- Universidad Tecnica Federico Santa Maria, Chile

The four teams had four different tasks to solve: The team *Tricia Google Translate* had the task to extend Tricia – an open source knowledge management and

[1] Originally eight GSE project teams were planned, including two teams working on the same topic of Carpooling and Ride Sharing. However, at the beginning of the projects it was decided by the supervising professor in Mexico to have one large Carpooling and Ride Sharing team instead of two small teams.

Project	Number of Sites	Team Size (All Sites)	Persons per Site
Tricia Google Translate	2	4	2
Touch Screen Tablet	3	6	1–3
Electronic Billing	3	8	2–3
Carpooling and Ride Sharing	4	9	2–3

Table 9.1: Characteristics of NEREID Projects

collaboration tool, which provides wikis, blogging, social networking, file sharing, etc. but only with English user interfaces – with multiple language support by means of the Google Translate API [UP11]. The task of the *Touch Screen Tablet* team was to develop a touch screen menu system for restaurants, which replaces traditional menu cards and allows for automatization of the ordering process [Sch11]. The team *Electronic Billing* worked on an electronic billing system for companies in Mexico, where a new law required companies to provide electronic invoices effective January 2011 [JKG11]. Finally, the team *Carpooling and Ride Sharing* had to develop an Internet platform as market place for ride sharing (i.e. offer and search for rides) [NS11]. The distribution involved different time zones and three different native languages (none of them English). It should be noted that also the local teams were often culturally mixed, because the NEREID course was offered in English language and therefore preferred by international students. For example, at the Munich site of the Electronic Billing team there were three students, one from China, one from Pakistan and one from Germany. The team members of a specific site never met in person with their team colleagues from the other sites. Table 9.1 shows the characteristics of the projects: number of sites, team size and persons per site. Differently from the year before was that the students in Germany were offered an intercultural training of one day, conducted by the company Computacenter, to create awareness for cultural differences and how to deal with them.

Due to the global distribution of the four universities of NEREID our communication for the purpose of this dissertation with the students and the supervisors in France, Mexico and Chile was also only virtual (mostly e-mail), while there were also in-person meetings with the German students and supervisors.

9.1.2 Design of the Study

The four NEREID projects started in October 2010 and ended in December 2010. Table 9.2 gives an overview of the time line regarding the NEREID case study.

The metrics for the NEREID projects were designed based on the model of GSE metrics as described in Chapter 7. The metrics were tailored to the NEREID

Date	Description
Oct., 2010	Project kick-off
Nov. 4, 2010	Initial project presentation
Dec. 22, 2010	Final project presentation
Jan. 16, 2011	Final project reports due
Nov. 4-12, 2010	Survey 1
Dec. 1-7, 2010	Survey 2
Jan. 19, 2011	Semi-structured interviews with student teams in Munich
Jan. 24-Feb. 28, 2011	Supervising professors feedback survey
Mar. 21-28, 2011	Student feedback survey

Table 9.2: NEREID time line

projects in the following aspects:

- **Aggregation.** We decided to use the sites as the level of aggregation, i.e. Lyon (France), Puebla (Mexico), Munich (Germany), Valparaiso (Chile). (The members of the student teams didn't have specific roles such as tester, developer or project lead, therefore an aggregation according to roles was not even possible.)
- **Traveling.** We left out the Traveling indicator, because it was clear from the beginning that the students will not be able to travel to the other sites which they are collaborating with.
- **Communication Media Profile.** The list of communication media was adapted to: *e-mail, phone, video conference (e.g. skype with webcam), voice over IP (e.g. skype, voice only), chat (e.g. skype chat), personal contact, other*
- **Communication Frequency.** As options for evaluating the communication frequency, we established: *daily*, *every 2–3 days*, *weekly*, *every two weeks*, *less than every two weeks*, *not at all*
- **Delay Due to Missing Information.** We chose 2 weeks as the reference period in the question about delay due to missing information: *How many times in the past two weeks was your own work delayed because you needed information, discussion, or a decision from someone at your site or another site?*

Jointly with the course owners, we decided to apply the GSE metrics twice: beginning of November 2010, shortly after the first interim presentation by the students and beginning of December 2010, about two weeks before end of the projects. The students had about one week time to fill in the survey. Shortly after that (e.g. 2–3 days later) the teams received the results in form of a so-called *collaboration analysis report*. It consisted of an executive summary with the interpretations of the metrics data, recommendations to the team and the detailed data. A sample report can be found in Appendix D. They were offered to use it for improvement of their project. The data collection was done via an online survey platform (LimeSurvey) [lim]. We refer to this survey as the *GSE metrics survey*. The answering of the survey was anonymous, however, during a survey period, we used tokens to monitor who has not yet filled in the survey and to send reminder e-mails.

At the end of the projects we conducted semi-structured interviews with the student teams in Munich and used an online questionnaire to collect detailed feedback from the students at all four sites involved in the projects to assess the accuracy, usefulness and the effort involved in our approach. Additionally, we asked the supervisors to fill in a feedback questionnaire. We refer to the feedback survey of the students as *student feedback survey* and to the feedback survey of the supervising professors as the *supervising professors feedback survey*. We organized the interviews and feedback surveys after the students had submitted their work and final reports and received their grades, in order to prevent that they would have reservations to speak openly. Additional sources of feedback were the final reports compiled by the student teams.

9.1.3 Results

In the following, we refer only to main results out of the GSE metrics survey in the four NEREID projects. The detailed results can be found in Appendix E.

Validation of General Hypotheses

To assess the reliability of the data collected, we validated the following hypotheses:

Hypothesis 1 *Communication Frequency, Reachability and Current Awareness decrease with the degree of distribution (number of sites, team size).*

Hypothesis 2 *There are more frequent and longer Delays Due to Missing Information across sites than within a site.*

Hypothesis 3 *Team Cohesion, Informal Communication and Awareness are better within the local site than across sites.*

Figure 9.1 depicts Communication Frequency, Reachability and Current Awareness of the four GSE student projects, sorted by degree of distribution (from

2 sites to 4 sites). On the left hand side (a) the status of the first GSE metrics survey is shown (Nov. 2010), on the right hand side (b) the status after the second GSE metrics survey (Dec. 2010). In the status of the first survey a clear trend is visible: The higher the degree of distribution, the less frequent is the average communication. Also Reachability and Current Awareness decrease with the degree of distribution. Hypothesis 1 is confirmed. The second survey changed the picture. This is explained by different performance of the projects which until then had either improved or worsened their situation. In particular, project 1 (Tricia Google Translate) has worsened and project 4 (Carpooling and Ride Sharing) has improved.

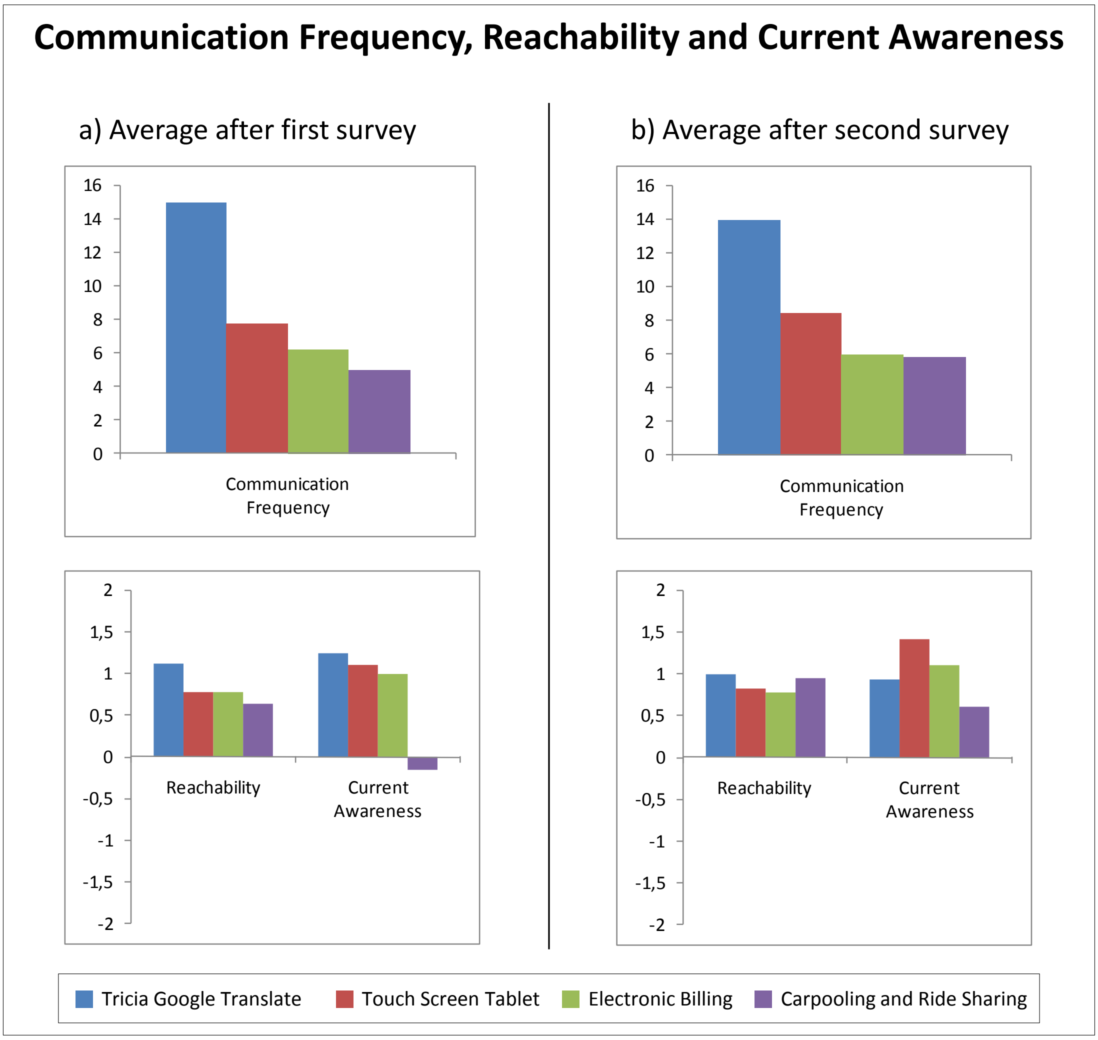
In Figure 9.2 the average Delay Due to Missing Information of the NEREID projects is shown. In all cases there are more frequent and longer Delays Due to Missing Information across sites than within a site. Thus Hypothesis 2 is confirmed. The average ratio of the duration of distant delays and local delays is 2.4, which even corresponds well to the factor 2.5 identified by Herbsleb [HM03].

The results for Team Cohesion, Informal Communication, Culture & Language and Awareness of the four GSE student projects are illustrated in Figure 9.3. The ratings of Awareness and Informal Communication are better for the local site compared to the distant sites, which is in conformance with Hypothesis 3. With respect to Team Cohesion, 6 of the 8 subitems show better results for the local than for the distant site. The two exceptions are: no team competition and clear responsibilities. The former is explainable by the fact that there were multiple NEREID student projects run in parallel under the same supervisor which was perceived as kind of competition between the teams. The latter can be justified by the fact that teams had established a clear worksplit between sites, but not for individual team members at one site, which lead to increased discussion over who has a particular responsibility within one site. The total value of Team Cohesion is 1.0 with respect to the local site and 0.7 for the distant sites (recall that the range of values is from +2.0 to –2.0, where +2.0 is the best and –2.0 is the worst value). We therefore consider Hypothesis 3 as well as confirmed.

Even though the data collected is based on subjective measurements, Hypothesis 1, 2 and 3 were confirmed which provides confidence regarding the reliability of the GSE metrics survey data (i.e. no arbitrary data).

Project Results

In the NEREID projects of winter term 2010/2011 the students experienced very similar collaboration challenges compared to the NEREID projects of the year before (see Section 4.3). Collaboration challenges explicitly mentioned in the final project reports [UP11, Sch11, JKG11, NS11] and the semi-structured interviews included the following areas:



Communication Frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
Reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2
Current Awareness: very good = 2, good = 1, bad = -1, very bad = -2

Figure 9.1: NEREID: Communication Frequency, Reachability and Current Awareness per project, sorted by degree of distribution

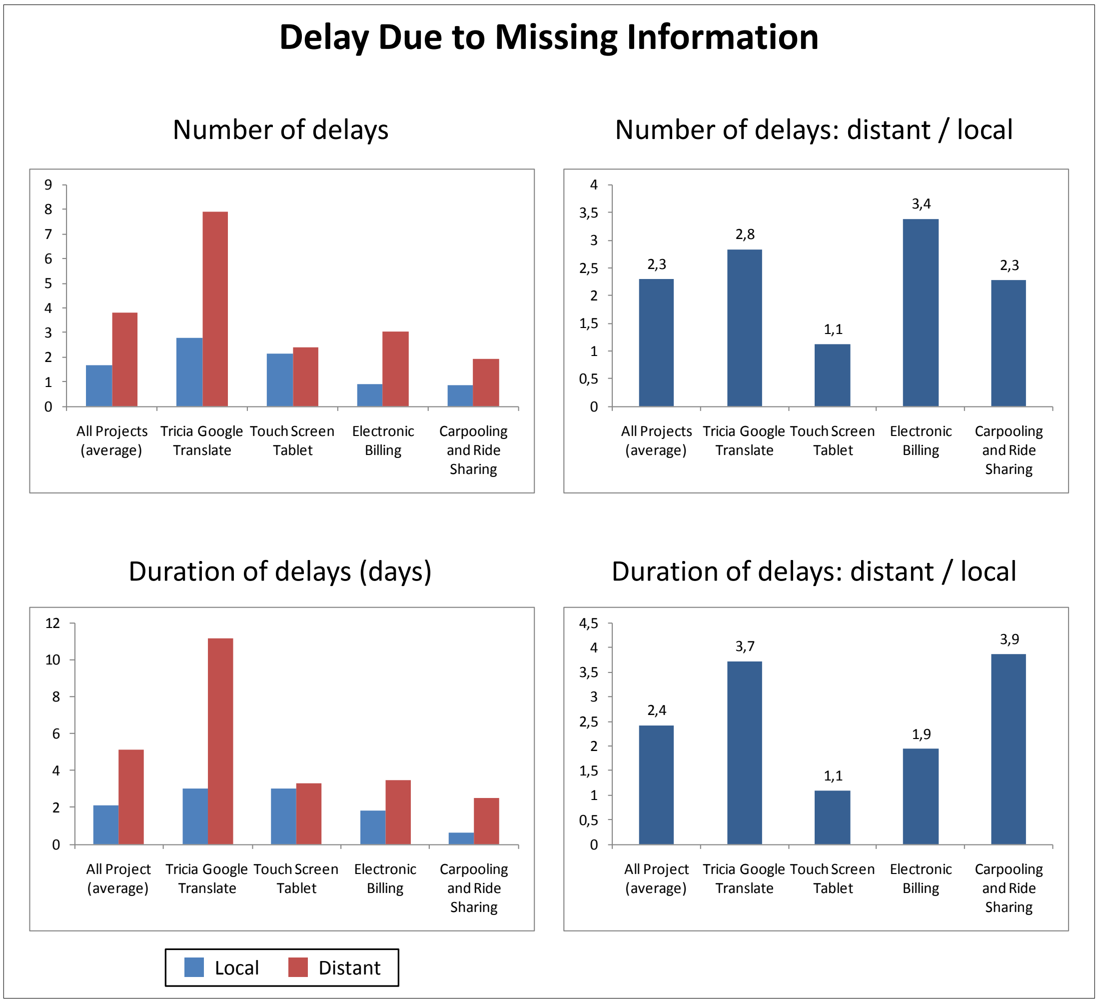


Figure 9.2: NEREID: Delay Due to Missing Information, average of all project and per project

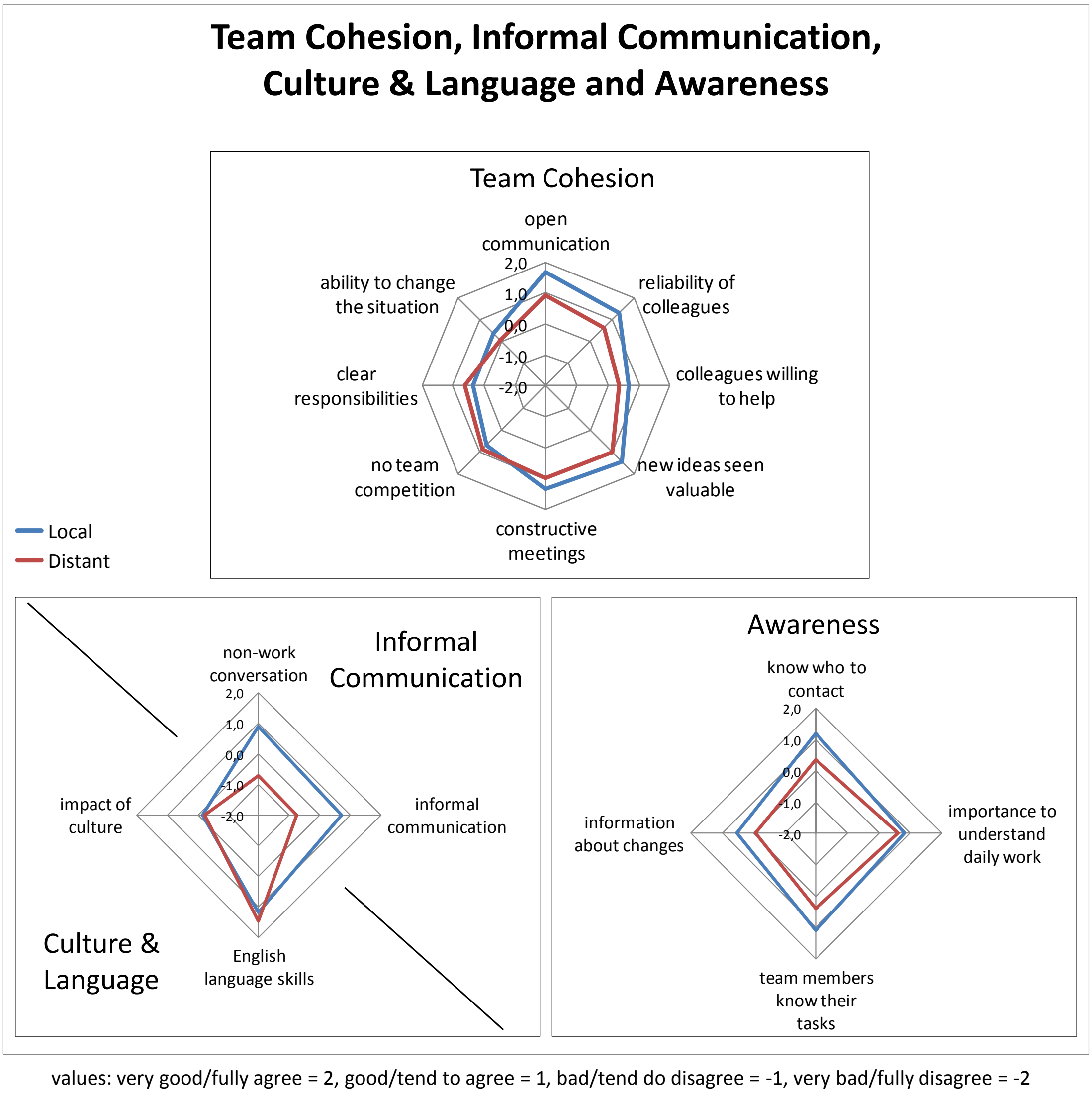


Figure 9.3: NEREID: Team Cohesion, Informal Communication, Culture & Language and Awareness, average of all projects

- Time coordination and availability
- Communication delays
- Cultural differences
- Language challenges
- Different software engineering backgrounds
- Different motivations per site

Because the challenges in NEREID were realistic also compared to industrial projects, the projects provided a suitable context for validation of our GSE metrics.

Right after each round of data collection via the GSE metrics survey, we prepared a collaboration analysis report, which was then distributed to the students of each project team of the four NEREID projects. To each collaboration analysis report, we added our interpretation of the GSE metrics survey data in the form of an executive summary. Figure 9.4 lists the main findings which we derived from the survey data. Please note that the interpretations were prepared from an 'outsider' perspective, without being involved in what was going on in the project teams, and before the semi-structured interviews took place.

Feedback

As mentioned above, we used both semi-structured interviews and a feedback questionnaire (the student feedback survey) to collect the feedback of the students. Furthermore, we collected also the feedback of the supervising professors using a feedback questionnaire (the supervising professors feedback survey). The student feedback survey was answered by 17 of the 26 students[2] which equates to a response rate of 65%.

In the first part of the student feedback survey, we collected the general feedback of the students: We asked them to evaluate the following statements related to accuracy, effort and usefulness:

- The collaboration analysis provided insights into the team dynamics of our project team

[2] Of the 27 students at the beginning of the projects, one person of the Carpooling and Ride Sharing team dropped out early out of the team due to collaboration issues as therefore was not participating in the student feedback survey. Additionally, two team members of the team Touch Screen Tablet left the project because of a late technology switch from Android/Java to PHP according to customer request. However, these two students still answered the feedback questionnaire, therefore we use 26 as denominator in calculating the response rate.

Tricia Google Translate

- S1.1 From Survey 1 to Survey 2 the collaboration has worsened
- S1.2 At the time of Survey 2, the majority of the team was unsatisfied with the project's achievements
- S1.3 Survey 2: There was honest and open communication within the local team, but not with the distant colleagues
- S1.4 Survey 2: Distant colleagues didn't provide timely information about changes in current plans
- S1.5 There was almost no informal communication across sites
- S1.6 English language skills were no major problem

Touch Screen Tablet

- S2.1 From Survey 1 to Survey 2 the overall status of collaboration was more or less constant
- S2.2 Communication inside the team was honest and open
- S2.3 There was almost no informal communication across sites
- S2.4 The majority of the team felt powerless to change the project's situation
- S2.5 At the time of Survey 2, the collaboration between Valparaiso and the other sites was particularly challenging
- S2.6 There were some issues related to English language skills

Electronic Billing

- S3.1 There were issues with communication and reachability across sites
- S3.2 There was almost no informal communication across sites
- S3.3 At the time of Survey 2, there were issues with timely information about changes by the distant colleagues
- S3.4 The team felt rather powerless to change the project's situation
- S3.5 English language skills were no major problem

Carpooling and Ride Sharing

- S4.1 From Survey 1 to Survey 2 the collaboration has improved
- S4.2 At the time of Survey 2, most of the team knew what other colleagues are currently working on
- S4.3 Survey 2: The majority of the team was satisfied with the project's achievements
- S4.4 There were local problems inside the Munich team
- S4.5 English language skills were no major problem

Figure 9.4: NEREID: Main findings based on GSE metrics data

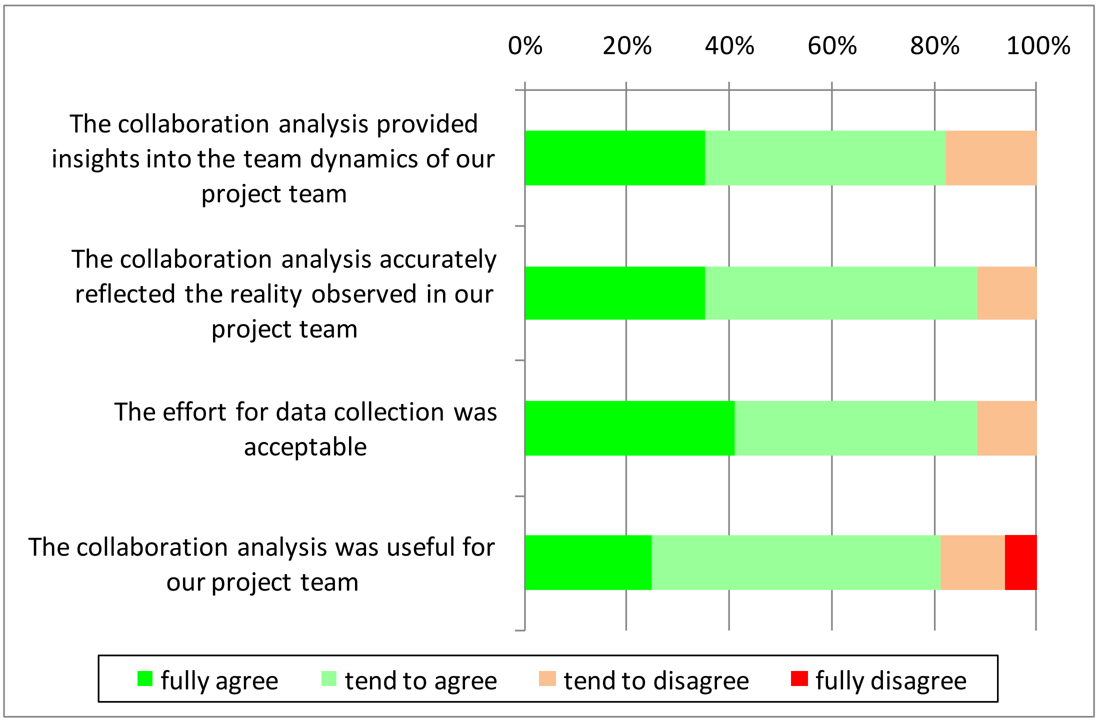


Figure 9.5: NEREID: General feedback from student feedback survey

- The collaboration analysis accurately reflected the reality observed in our project team
- The effort for data collection was acceptable
- The collaboration analysis was useful for our project team

Additionally, we asked them how much time they needed to fill in the GSE metrics survey. Figure 9.5 contains the feedback results. The overall feedback was very positive: all four statements were confirmed at an agreement level (*fully agree* and *tend to agree*) of 81% to 88%. The answers regarding the time needed to fill in the GSE metrics survey were in between 8 and 25 minutes, where 14.1 minutes was the average. This was seen acceptable according to the feedback of the students, see Figure 9.5.

Furthermore, we asked the team members to evaluate the main findings from our interpretation (see above). The results are shown in Figure 9.6.

The majority of our interpretations were confirmed to be correct: the degree of confirmation (share of answers that *fully agree* or *tend to agree*) ranged between 68% for the Electronic Billing team to 100% for the team Carpooling and Ride Sharing. It is to be noted that some findings were controversial to the team members: Statement S2.4 (*The majority of the team felt powerless to change the project's situation*) of the team Touch Screen Tablet was confirmed by half of the

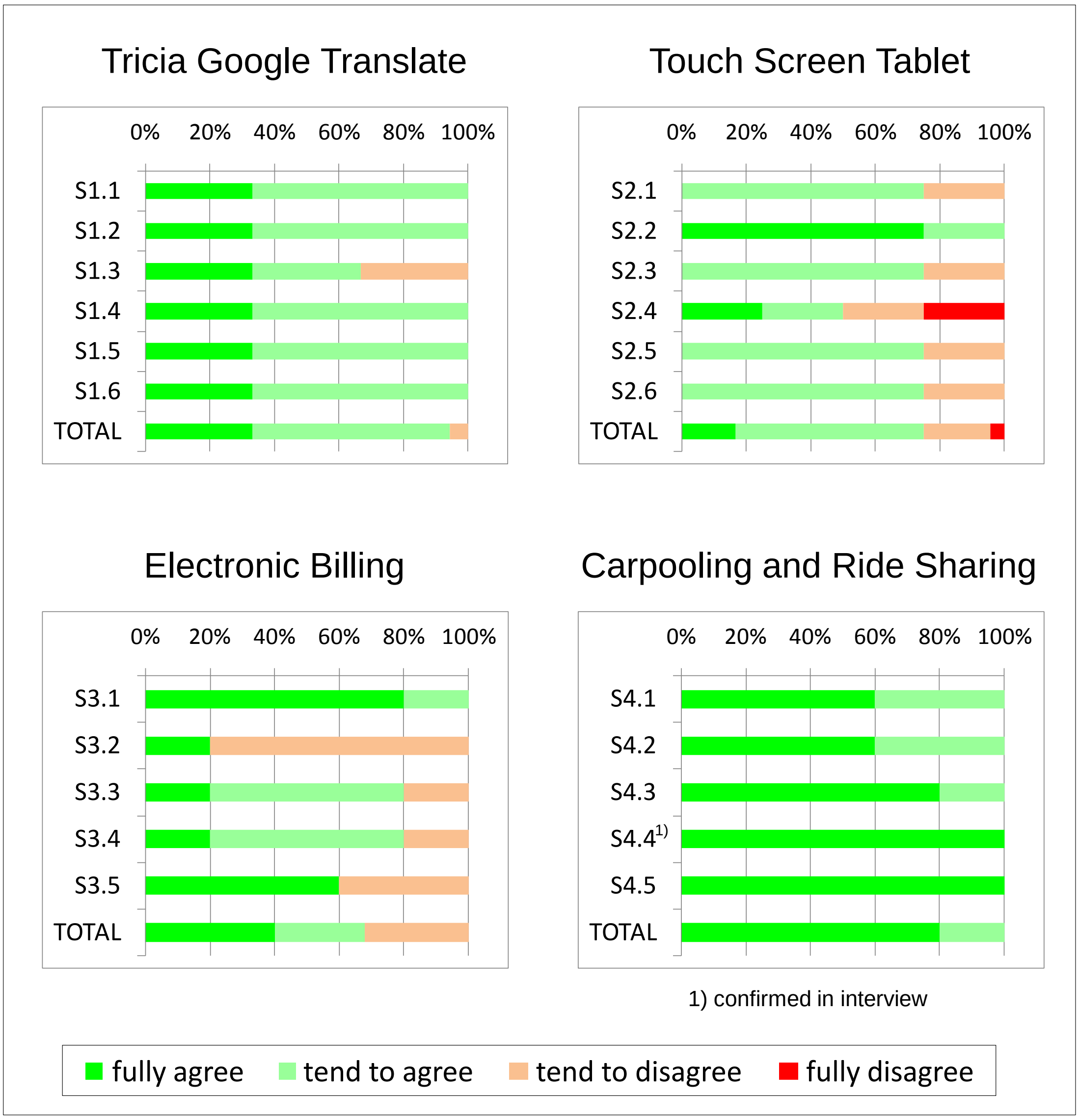


Figure 9.6: NEREID: Evaluation of main findings from student feedback survey

team and rejected by the other half. The explanation we found from the interviews was that the project sponsor which was an external company in Mexico was seen by some team members as a higher authority which is setting targets and frame conditions that can't be changed by the project team, while others didn't see a problem. Other irregularities were noticed with the statements S3.2, S3.5 and S1.3. The statement S3.2 (*There was almost no informal communication across sites*) is striking since it is the only finding which team members more disagreed than agreed with. Here our explanation is that we formulated our finding in an too extreme way the team members could not agree with. In fact, in the two GSE metrics surveys, the majority answered the questions regarding *I discuss non-work related matters with other project team members* and *I often get useful work-related information through informal communication (e.g. hallway conversations)* negatively for the distant site. However, even though informal communication was impeded, there was no complete lack of informal communication across sites. Statement S3.5 (*English language skills were no major problem*) was fully agreed with by 60%, but 40% answered they tend to disagree. The explanation we found from the interviews was that the English language skills were seen differently depending on who was communicating with whom. The Munich students mentioned that it was challenging for them to communicate with the Mexicans because they found the South American English hard to understand. They also mentioned that there was no need for them to communicate with the colleagues in Chile, because Mexico was the central team. On the other hand, English language skills seemed not to have been an issue within South America: When drilling down into the feedback results, we found that all team members from Chile fully agreed with S3.5.

From the GSE metrics data – without knowing what was going on in detail in the projects –, we had observed local problems inside the Munich team of the Carpooling and Ride Sharing project (see S4.4). It occurred in the interview that one person of the two at the Munich site had left the team due to collaboration problems and without having significantly contributed to the project, which explained the observation from the data. Besides, in the interview we heard that the Carpooling and Ride Sharing team had been too large which caused collaboration difficulties. In the team Touch Screen Tablet, we had noticed that at the time of survey 2 the collaboration between Valparaiso (Chile) and the other sites was impeded, we interpreted it as 'particularly challenging' (See S2.5). As it turned out from the interviews, two students from Chile had left the project team because of a late technology switch from Android/Java to PHP according to customer request. However, these two students had contributed to the project before and there were no major collaboration issues. The reason why the metrics indicated an abnormal behavior was the absence of the Chilean team after this event.

Besides the evaluation of the main findings, we also asked the team members to rate the accuracy (*How accurately did the indicators reflect the reality observed*

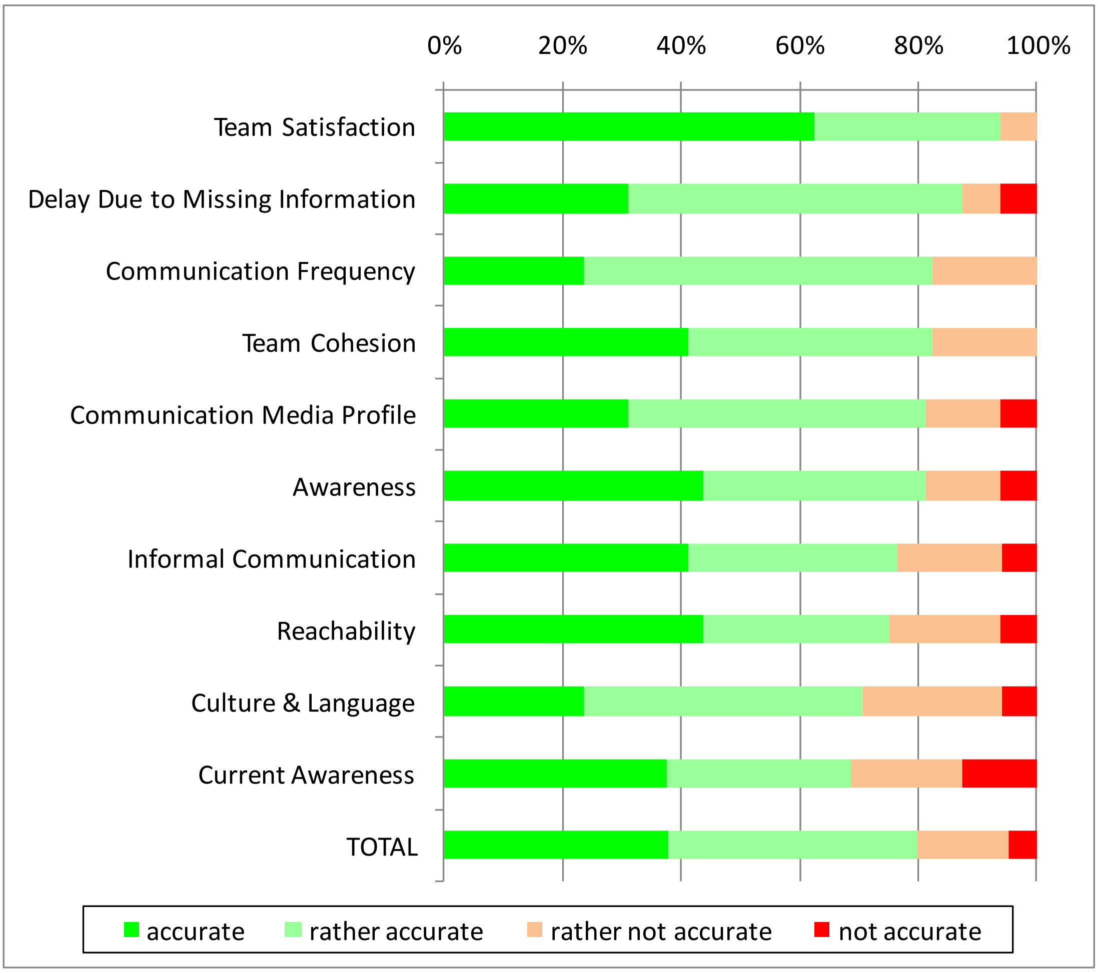


Figure 9.7: NEREID: Accuracy of indicators according to student feedback survey

in your team?) and usefulness (*How useful are the indicators?*) of the GSE metrics. Figure 9.7 shows the evaluation of accuracy. Altogether it was stated in the feedback that the indicators reflected the team reality well: Positive accuracy ratings of the indicators (*accurate* and *rather accurate*) range from 94% to 69%, the average of all indicators is 80%. In Figure 9.8 the evaluation of usefulness is depicted. Positive ratings of usefulness (*useful* and *rather useful*) are between 100% and 60%, the average of all indicators is 86%. While seven of the ten indicators received very high ratings (4 times 100%, twice 93% and once 86%), three of the indicators were seen less useful: Communication Media Profile (69%), Culture & Language (64%) and Informal Communication (60%).

Another indication that the GSE metrics were useful was the fact that the Touch Screen Tablet team had used diagrams out of the collaboration analysis within their final report to illustrate the situation inside their team. In fact, the Munich student of the Touch Screen Tablet team (note that it was the only site

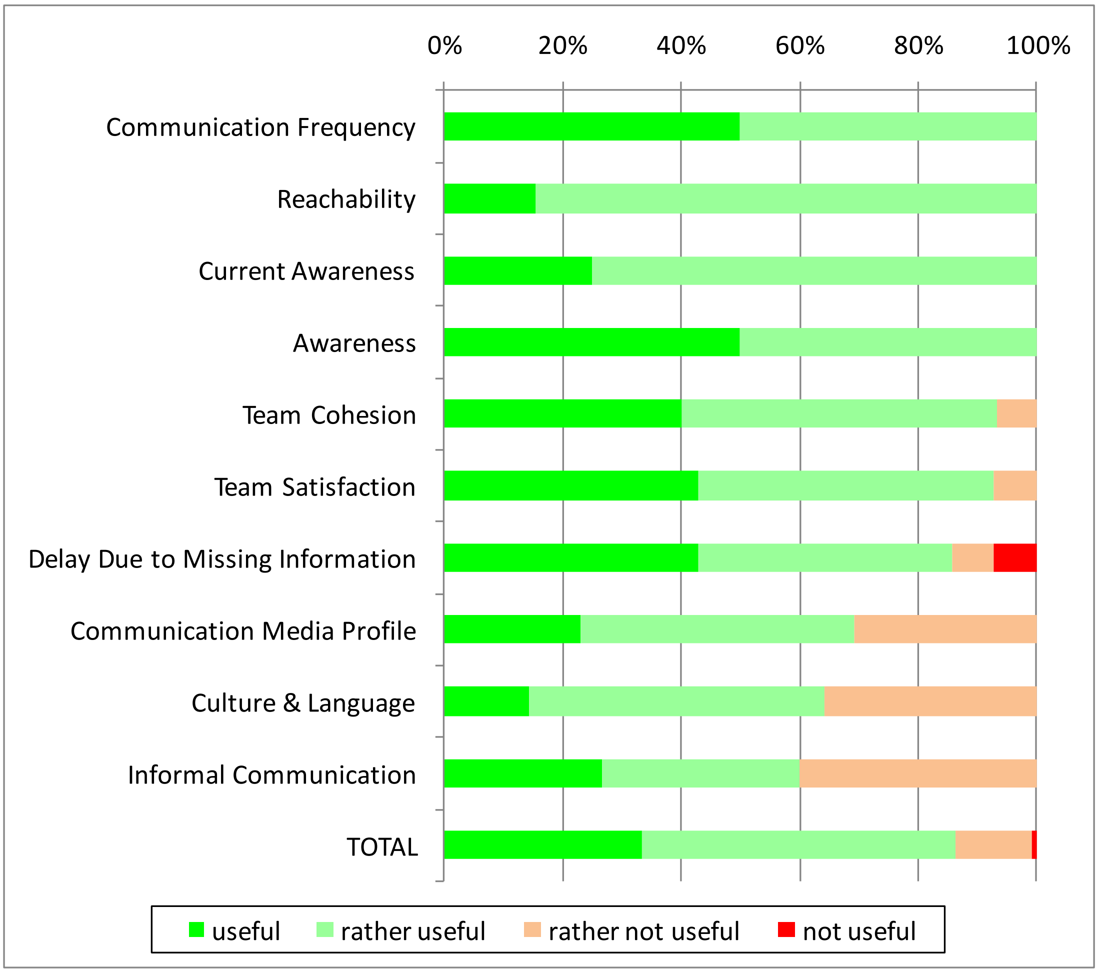


Figure 9.8: NEREID: Usefulness of indicators according to student feedback survey

with a 1-person team) explicitly spoke highly of the collaboration analysis diagrams: "The graphs were cool!" He explained to us that they were very helpful to create awareness for collaboration-related issues inside the team and "bring the problems to the point". They lead to intensive discussions inside the team and provided thought-provoking impulses. According to the Munich student of the Touch Screen Tablet team, the interpretations in the executive summary which had been included in the collaboration analysis were sometimes fuzzy, because they were formulated just based on the data but without knowing in detail what was going on in the project team. Furthermore, he criticized that the GSE metrics survey questionnaire was not tailored for a 1-person-team at one site.

In an open question we asked the students to describe in text how they used the collaboration analysis reports: *Please describe shortly what your team did with* *the two collaboration analysis reports (e.g. discuss them in the team, take action, etc.).* 12 of the respondents answered this question, the other 5 left the text field empty. 10 of the 12 respondents mentioned that they had discussed the collaboration analysis reports within their teams. In 8 of the answers it was mentioned that the analysis was useful to the team resp. that the team had used it to take corresponding action based on the results.

Among the answers were the following:

> "As the project manager, I was already taking action by the time the analysis report came in, but the report completely agreed with what I had perceived and helped a lot to focus on the solutions necessary to get things going. The second report indicated that the actions had worked. But while the reports and the development of the project itself agreed this time, had it not been the case the reports would've been invaluable in getting the team back on track, or at the very least a fantastic post-mortem tool to learn from."

> "Through the analysis report, we have found common problems which trouble our teamwork. The efficiency has improved a lot after that. Moreover, we communicate more often with other team members, which is a great help for improving our working quality."

> "With the results of collaborative analysis the team was becoming aware of shortcomings in equipment and also of virtues or benefits we had as a team. No doubt the polls provide information valuable in making decisions."

2 of 12 had mentioned they didn't really use the collaboration analysis reports:

> "In my team we did nothing with the analysis reports"

> "To be honest, we didn't use the collaboration analysis reports as we should have used. I mean, we should have used your work as an important way to describe our team attitudes and efficiency, but in reality, I think it was only in the last month of the project when we finally realize the importance of your work, and how useful it would had been to our development. Anyway, I think the reports described well the reality of the team, so our communication worked better and we were more focused on the relevant tasks of the software."

The supervising professors feedback survey contained questions similar to the general section of the students feedback survey. Additionally, we asked the supervising professors also to rate the innovation grade of our approach:

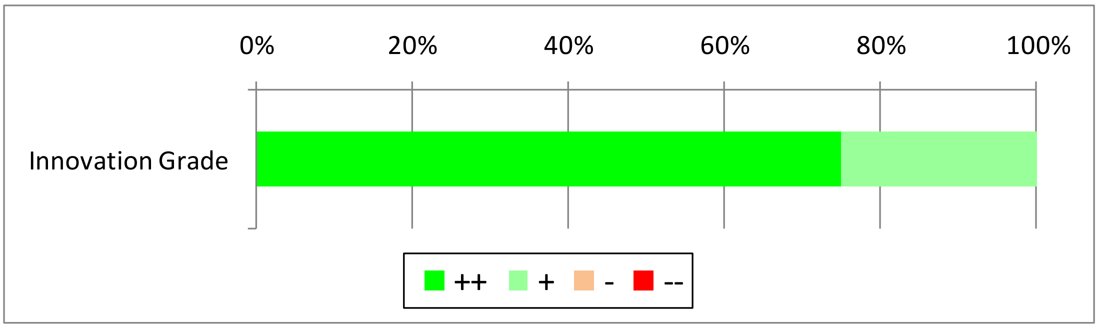


Figure 9.9: NEREID: Innovation grade of the collaboration analysis approach according to supervising professors feedback survey

> *Please estimate the innovation grade: How new is the collaboration analysis approach in project management of Global Software Engineering projects?*
>
> *++ = innovative, not used in Global Software Engineering project management today*
>
> *– – = not innovative, already state of the practice in Global Software Engineering today*

It turned out that the supervising professors of the four involved universities were not sufficiently involved in the students' work and therefore we disqualified their answers for the scientific validation (see also comments below). However, the question regarding the innovation grade was still meaningful as it required a general understanding and GSE background, but no detailed insight into the student projects. As it can be seen from Figure 9.9, all four supervising professors saw the collaboration analysis approach as new in GSE (75% ++ and 25% +) and therefore as innovative.

Comments by the supervising professors included:

> "I don't know of other 'collaboration analysis approaches' (CAP) in PM of GSE. Nevertheless, I found the carried CAP very complete and valuable. It can be part of an evaluation process in GSE, for example to asses Quality of PM, or to deal and/or manage team performance."
>
> "I was not really involved in the evaluation of the student teams and do not really think that I can contribute scientifically relevant data to this [feedback] survey."
>
> "The surveys were very useful for giving feedback to my students. I saw only the results, but I didn't know the questions. The variables to estimate the degree of collaboration, the cultural differences and technical differences will be very useful."

Summary

In summary, the GSE challenges in the 2010/2011 NEREID project were comparable to the year before and provided a realistic context for our study. The GSE metrics helped to identify collaboration problems and address them. Although in some cases there was a lack of precision in our interpretations derived from the GSE metrics due to missing context information, the GSE metrics were seen as very valuable by the team members. According to the feedback results we collected and the interviews we conducted, the GSE metrics were accurate, useful and the effort involved was reasonable. The overall approach was rated as very innovative and new to GSE project management.

9.2 DOLLI4

9.2.1 Project Description

Our second case study in the validation phase was the DOLLI4 project of TU München, department of informatics, chair for applied software engineering. The project was the fourth in a series of practical courses, which were conducted in cooperation with Flughafen München GmbH on a yearly basis. Flughafen München GmbH is the operator of Munich airport and has an own IT department responsible for ground handling (e.g. baggage handling from check-in to the aircraft), information services (e.g. display of arrival and departure of flights) and IT field services (e.g. maintenance of IT devices) [NB10b]. The project name DOLLI is derived from the airport terminology: 'dolly' (same pronunciation, different spelling) is the technical term for a transport cart used at airports e.g. in baggage handling. While student projects are often organized in such a way that there is one problem to solve which is then handled by multiple teams of 2–3 students which deliver functionally identical solutions, this was not the approach of this project. In case of DOLLI4, there were complex problems that had to be solved by individual subteams inside a large team (30 students) and only one solution per problem was to be delivered. The distinctiveness of the DOLLI4 project is characterized by the so-called '6 Rs' [NB10a]:

- **Real customer.** The project has a real customer: Munich airport (Flughafen München GmbH)
- **Real data.** The DOLLI4 project deals with real data, e.g. from the airport's building management systems, geographic data, flight data
- **Real problems.** The project has to solve real problems, i.e. support of passengers, security and building management

- **Real team experience.** The DOLLI4 project consists of four teams of 6–9 persons each, the collaboration inside the teams offers a real team experience
- **Real project experience.** DOLLI4 provides real project experience, including meetings, reviews and acceptance tests
- **Real deadline.** The results have to be presented at the customer acceptance test with presence of the customer and the press in mid of March 2011

DOLLI4 consisted of the four teams: Building Management, Security, MUC App and Landside Server. These teams had to deal with different tasks as follows [NB10b]: The Building Management team had to develop a system to enable the employees in the airport offices to control the lighting and air conditioning inside their office rooms. The task of the Security team was to develop a system to notify the security staff which is patroling at the airport about alarm events and to instruct them for inspection of the situation. The objective of the MUC App team was to develop a mobile application for the Apple iPhone to provide airport-related information to passengers or other airport visitors, e.g. flight status information or a shopping guide. Finally, the Landside Server team was responsible for the interfacing with the real data sources in the airport IT systems, i.e. to provide the server part for the other three teams. For each of the four teams, a *team coach* was established who was responsible for guiding the team and was acting as kind of project manager. The team coaches regularly had internal meetings to discuss the situation in their teams and take improvement actions. Therefore the team coaches were the ideal target group for our GSE metrics.

Table 9.3 shows the project characteristics. Even though the DOLLI4 project was not globally distributed, we expected the team members to have similar collaboration challenges because of the dependencies between the sub-project teams (in particular Landside Server and the other teams) and because of the communication with a real customer. Furthermore, we wanted to have a possibility to compare the usage of our GSE metrics in a co-located setting with globally distributed projects.

9.2.2 Design of the Study

Along the lines of the NEREID projects, the metrics for the DOLLI4 project were designed based on the model of GSE metrics as described in Chapter 7 and then tailored to the DOLLI4 project. The following adaptations were made:

- **Aggregation.** We decided to use the subteams as the level of aggregation, i.e. Building Management, Security, MUC App and Landside Server. (Sites

Project Subteams	Number of Sites	Team Size
Building Management	1	7
Security	1	8
MUC App	1	6
Landside Server	1	9
DOLLI4 Total	1	30

Table 9.3: Characteristics of the DOLLI4 project teams

were not relevant to the co-located project, and again specific roles were practically not existent.)

- **Customer communication.** Since the DOLLI4 project had a real customer, located at a different site, we included the customer communication from the team's perspective in the survey (Communication Frequency, Reachability, and Delay Due to Missing Information with respect to the customer).

- **Constructive Meetings.** We slightly adapted the wording and asked to evaluate the statement *Meetings are constructive* instead of *(Virtual) meetings are constructive*

- **Team Competition.** We used the formulation: *There is a competition between individual in the own team resp. competition with other teams*

- **Traveling.** Instead of Traveling, we asked: *How many meetings did you have during the past two weeks with the own team, the other teams, or with the customer?*

- **Communication Media Profile.** The list of communication media was adapted to the specific media of the DOLLI4 project: *phone, chat, forum, video conference (e.g. skype with webcam), voice over IP (e.g. skype, voice only), personal contact, other*

- **Culture & Language.** The aspect of Culture & Language was excluded from the DOLLI4 case study because of its co-located setting.

- **Language of surveys.** Questions were translated into German, because the business language of DOLLI4 was supposed to be German. All results and quotations from DOLLI4 given below were translated from German to English for this dissertation.

DOLLI4 had a series of interim milestones and associated presentations, they are recorded in Table 9.4. The project had two major distinct phases: A prototyping

Date	Description
Oct. 25, 2010	Project Kick-off
Nov. 25, 2010	Analysis Status Review (internal)
Dec. 2, 2010	Analysis Status Review (with customer)
Dec. 16, 2010	System Design Status Review (with customer)
Jan. 20, 2011	Object Design Status Review (internal)
Feb. 3, 2011	System Integration Test (internal)
February	break for examinations
Mar. 7-18, 2011	Development Phase at Airport Munich (Scrum)
Mar. 23, 2011	Customer Acceptance Test (with customer and press)
Nov. 11-19, 2010	Survey 1
Dec. 15-22, 2010	Survey 2
Feb. 1-8, 2011	Survey 3
Mar. 7-18, 2011	Daily Measurements
Mar. 17-24, 2011	Survey 4
Mar. 28-Apr. 16, 2011	Team coaches feedback survey

Table 9.4: DOLLI4 time line

phase which started with the project kick-off (October 25, 2010) and ended with the system integration test (February 3, 2011). In this phase the Rational Unified Process was followed with the core process steps business modeling, requirements, analysis & design, implementation, test, and deployment [IBM07]. The second phase was conducted as agile development using scrum [scr] and required full time on-site attendance at the airport (Mar. 7-18, 2011).

The GSE metrics were applied on a monthly basis during the prototyping phase (survey 1–3). In addition, there was a control survey with the same questions at the end of the scrum phase at the airport (survey 4). Furthermore, we experimented with daily measurements of a small set of metrics during the scrum phase in conjunction with the daily scrum meetings. The project ended with the customer acceptance test (March 23, 2011), where the students officially presented their results to the customer and offered exhibition stands to demonstrate and discuss their final products. To gather data regarding our validation criteria, we collected the feedback of the team coaches in an online feedback survey, after the whole project had been completed and grades had been established.

9.2.3 Results

Comparison with GSE Student Projects

Our expectation was that the DOLLI4 team members would have similar collaboration challenges as globally distributed teams due to the structuring in four subteams with dependencies (at least between Landside Server and the other teams) and the required collaboration with a real customer. We formulated the following hypothesis:

Hypothesis 4 *In a large co-located project there are similar collaboration challenges between subteams as in a GSE project team, i.e. less communication, lower awareness, reduced team cohesion and longer delays due to missing information.*

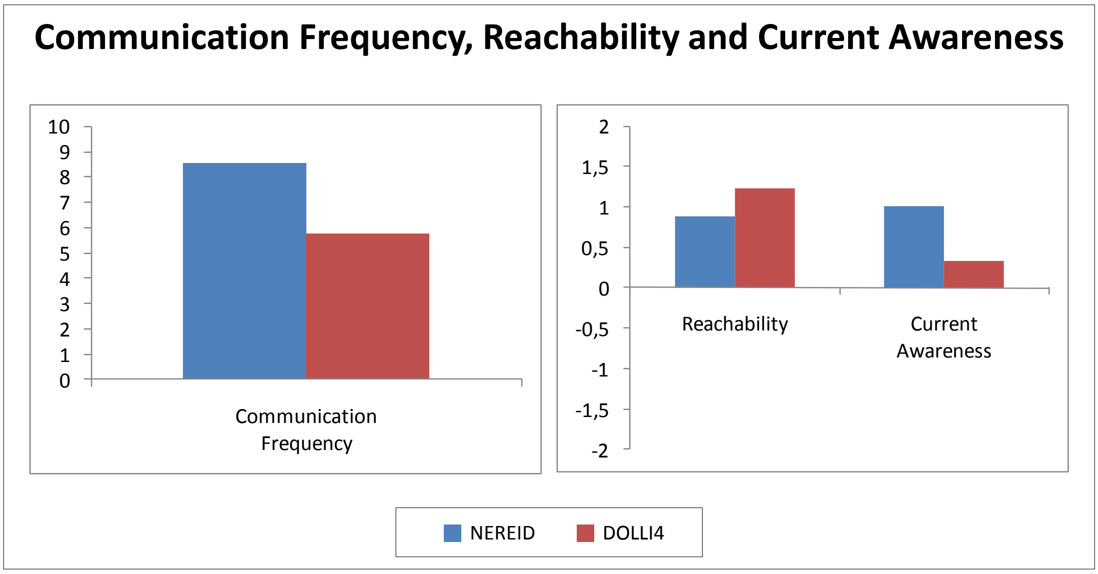
To validate Hypothesis 4, we compared the survey results from NEREID with the survey results from DOLLI4. In Figure 9.10 the Communication Frequency, Reachability and Current Awareness of the GSE student projects (NEREID) and DOLLI4 are shown. While the Reachability is better for the DOLLI4 project than NEREID, which is reflecting the fact that the DOLLI4 team members are working at the same site, the Communication Frequency and Current Awareness is lower in DOLLI4 compared to NEREID. This is explainable by the fact that the DOLLI4 project is significantly larger than the four NEREID projects. Delay Due to Missing Information of DOLLI4 is depicted in Figure 9.11. The duration of delays shows a similar tendency as observed in the GSE student projects: The average duration of delays increases from waiting for information from the own team to other teams and customer. In Figure 9.12 Team Cohesion, Informal Communication, and Awareness are shown[3]. The values represent better ratings of the own team compared to the other teams for all subitems of the metrics. Hypothesis 4 is confirmed.

Project Results

Similarly to the NEREID projects, also in DOLLI4 we prepared a collaboration analysis report right after each round of data collection via the GSE metrics survey. This analysis was prepared within a short time frame (e.g. 2–3 days) and then sent to the team coaches. Each time five different views were provided: one out of the perspective of each team and a fifth one with an overview and comparison of all four teams for the team coaches. The team coaches were asked to discuss the results of the collaboration analysis and take appropriate action where necessary. If and how they discussed the collaboration analysis reports with their teams was up to the team coaches.

Altogether, the collaboration went smoothly in the DOLLI4 project. However, one major issue with the cross-team collaboration was detected in the analysis: There were collaboration problems between the Landside Server team and the other teams, in particular low awareness of the work of the others and low communication frequency, which was critical because the server part had to integrate well with the client part and vice versa. As a consequence, the team coaches

[3]Please note that Culture & Language is not applicable in the case of DOLLI4.



Communication Frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
Reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2
Current Awareness: very good = 2, good = 1, bad = -1, very bad = -2

Figure 9.10: NEREID and DOLLI4: Average Communication Frequency, Reachability and Current Awareness

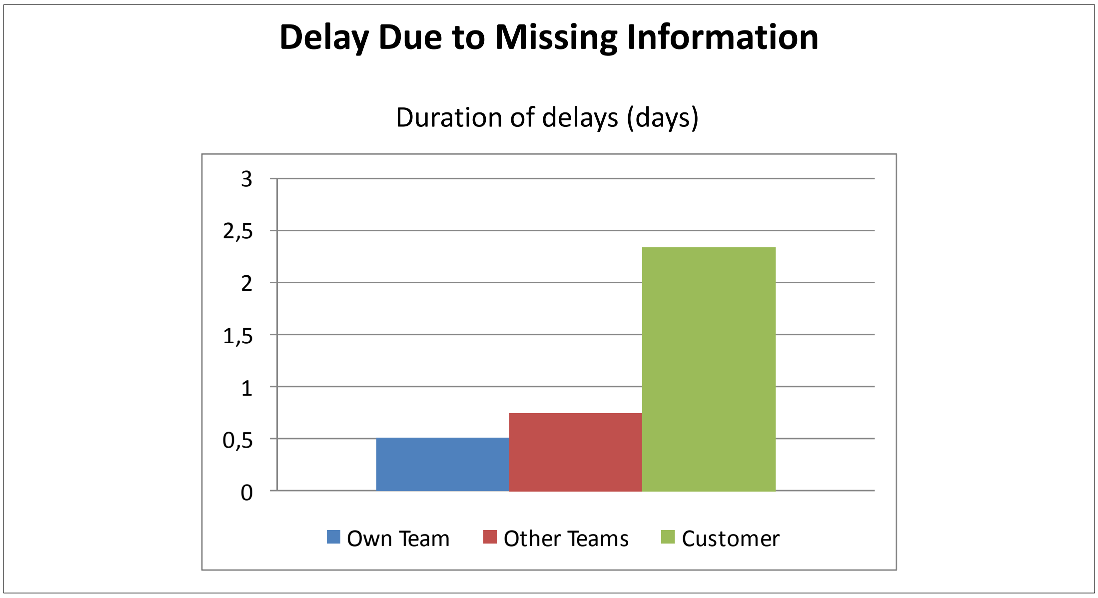


Figure 9.11: DOLLI4: Average Delay Due to Missing Information

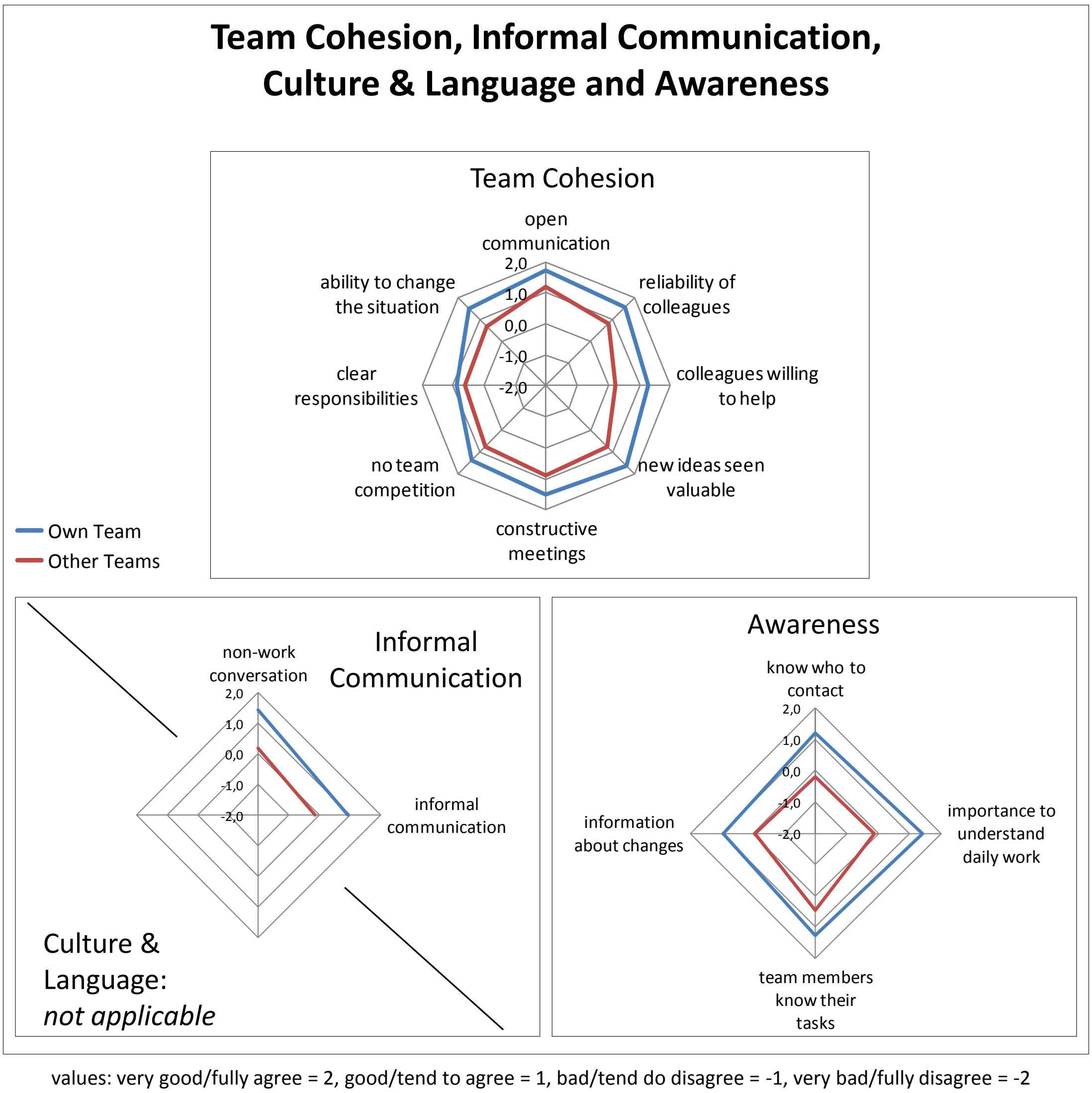


Figure 9.12: DOLLI4: Team Cohesion, Informal Communication, Culture & Language and Awareness, average

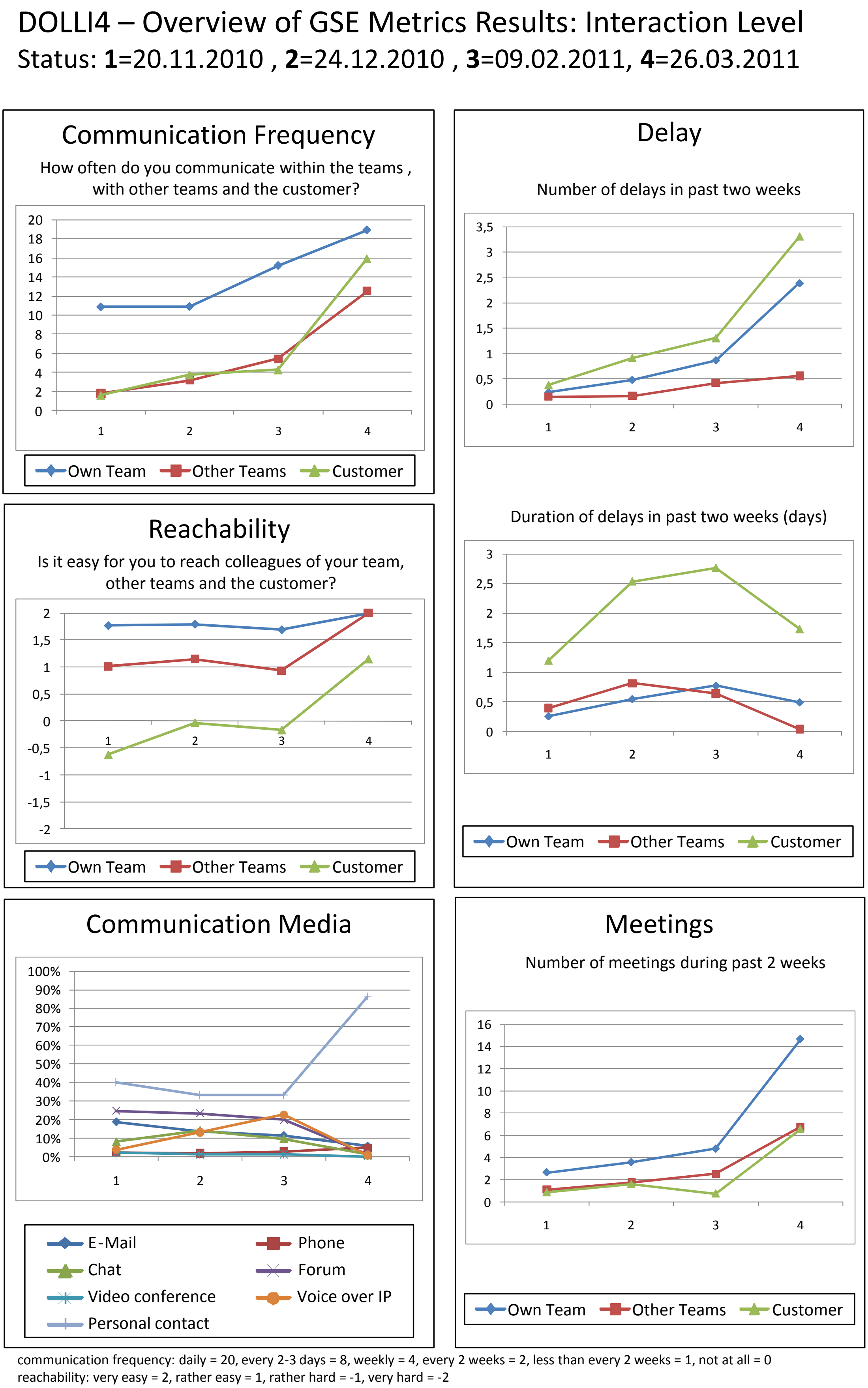


Figure 9.13: DOLLI4: Overview of GSE metrics results of survey 1–4, Interaction level

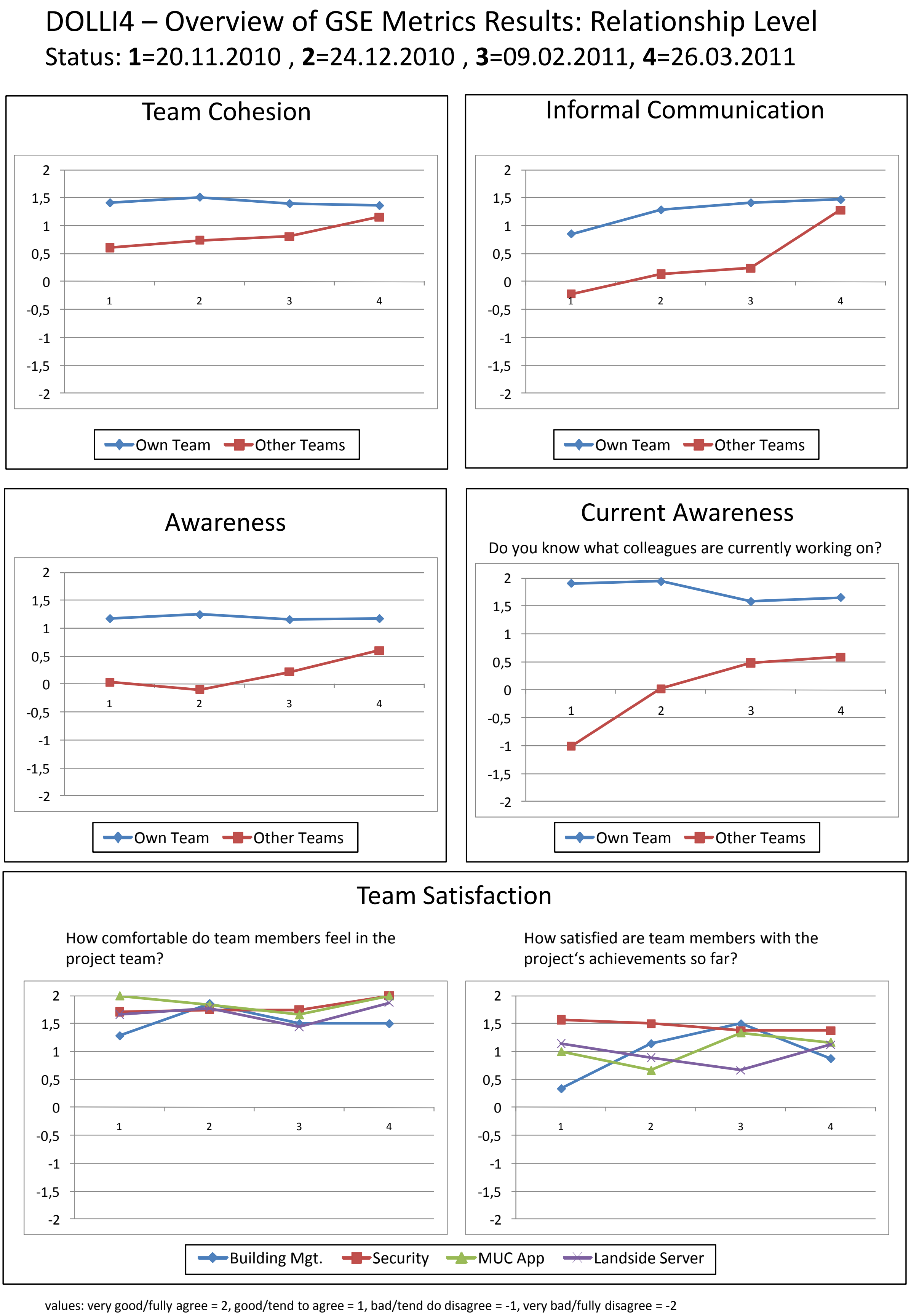


Figure 9.14: DOLLI4: Overview of GSE metrics results of survey 1–4, Relationship level

and the course owners finally decided to introduce a major reorganization: Between survey 2 and 3, the Landside Server team was dissolved. The former team members of the Landside Server team were integrated into the other three teams. This is in line with Conway's law according to which "a design effort should be organized according to the need for communication" [Con68]. However, this organizational requirement was not recognized prior to the collaboration analysis. To ensure comparability of the GSE metrics surveys, we asked the DOLLI4 project members to fill in also survey 3 and 4 according to the old team structures (they should consider the team members which came from the Landside Server team to them but are still working on the Landside Server part for their team still as 'other team'), even though the Landside Server team was no longer existent as an own team. Another problem area highlighted was the communication with the customer. In particular Reachability and Delay Due to Missing Information were obstacles. The issues in communication with the customer were addressed by escalation via the course owners.

Figure 9.13 and Figure 9.14 provide an overview of the results of the GSE metrics surveys in DOLLI4. From these figures, the development of the team situation over time is visible. Communication Frequency and the number of Delays Due to Missing Information have increased steadily over time. Also the Team Cohesion, Informal Communication, Awareness and Current Awareness with respect to the other teams grew steadily. These effects are explainable by the increased intensity of collaboration and also by the reorganization of the Landside Server team between survey 2 and 3, which improved cross-team collaboration. It is also obvious that the situation at the time of survey 4 was much different from the situation during the previous three surveys. The reason is the difference between the prototyping phase and the on-site scrum phase. During the scrum phase, Communication Frequency increased heavily, Reachability was optimal within and between the teams, and also the Reachability of the customer was much better (even though not as good as within and between the student teams), and as it can be seen from the Communication Media Profile almost only personal communication was used. Due to the intensive work, the number of Delays Due to Missing Information increased, however, the duration of Delays Due to Missing Information dropped notably. The number of meetings increased as well. On the Relationship level, the values related to other teams improved significantly according to survey 4, i.e. Team Cohesion, Informal Communication, Awareness, and Current Awareness. Noticeable is also that the Team Cohesion inside the own team went down slightly. The intensive work in the scrum phase and the prevailing time pressure increased tensions between the team members, e.g. if somebody made changes to the code which caused problems to another team member's work or broke the build.

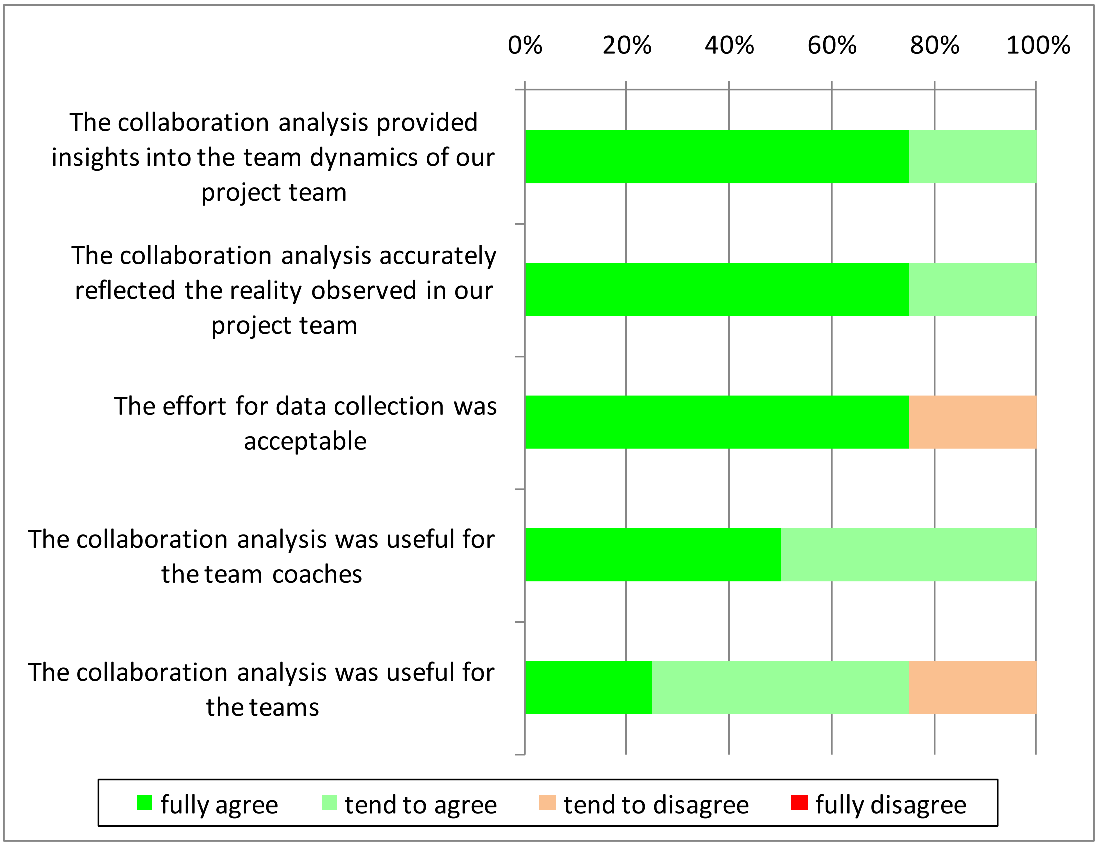


Figure 9.15: DOLLI4: General feedback from coaches feedback survey

Feedback

After the DOLLI4 project was completed and the grades had been established, we asked the four team coaches for their feedback, using an online feedback questionnaire. In the following we refer to this survey as the *coaches feedback survey*. We received a 100% response rate. The feedback questionnaire followed a similar structure as the student feedback survey of the NEREID project mentioned above and consisted of an overall feedback, the evaluation of the individual indicators, and questions related to the usage of the collaboration analysis results. Figure 9.15 presents the overall feedback of the team coaches. All five statements were confirmed at a high agreement level. One of the four respondents criticized the effort for data collection (tend to disagree). Regarding the usefulness of the collaboration analysis, it was seen by the team coaches as more useful for themselves than for the individual team members. Another indication that the effort for data collection was acceptable and in favor of the usefulness of the metrics are the high response rates in the GSE metrics surveys, which were always close to 100% (response rate from survey 1 to 4: 97%, 100%, 97%, 100%).

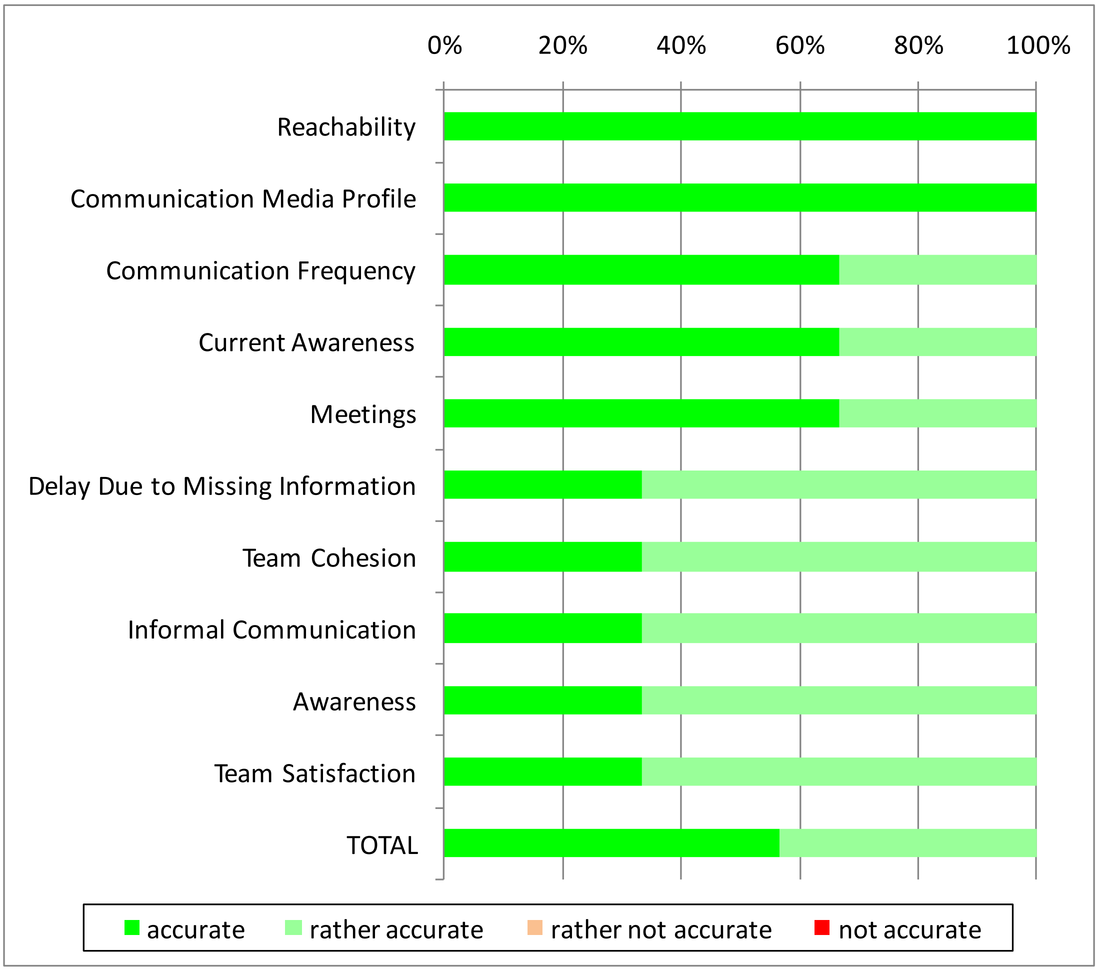


Figure 9.16: DOLLI4: Accuracy of indicators according to coaches feedback survey

In Figure 9.16 the accuracy ratings for the indicators are shown, Figure 9.17 illustrates the feedback on usefulness of the indicators. The accuracy of all indicators was seen very positive. On the other hand, there were some drawbacks in the usefulness. It should be noted that the indicators Team Cohesion and Team Satisfaction received very high ratings on the aspect of usefulness, even though they were less accurate compared to other indicators, which is in line with their nature as 'subjective' indicators. In fact, in the comments of the coaches feedback survey it was explicitly stated that the relationship-oriented indicators were the most useful for the team coaches. The Meetings indicator was criticized in two ways: 1) One of the team coaches stated in his feedback, that this indicator was not useful at all because he knew anyway how many meetings they had. 2) Another team coach commented that it was underspecified what a meeting is referring to – is a meeting with half of the team also a meeting? What if a person

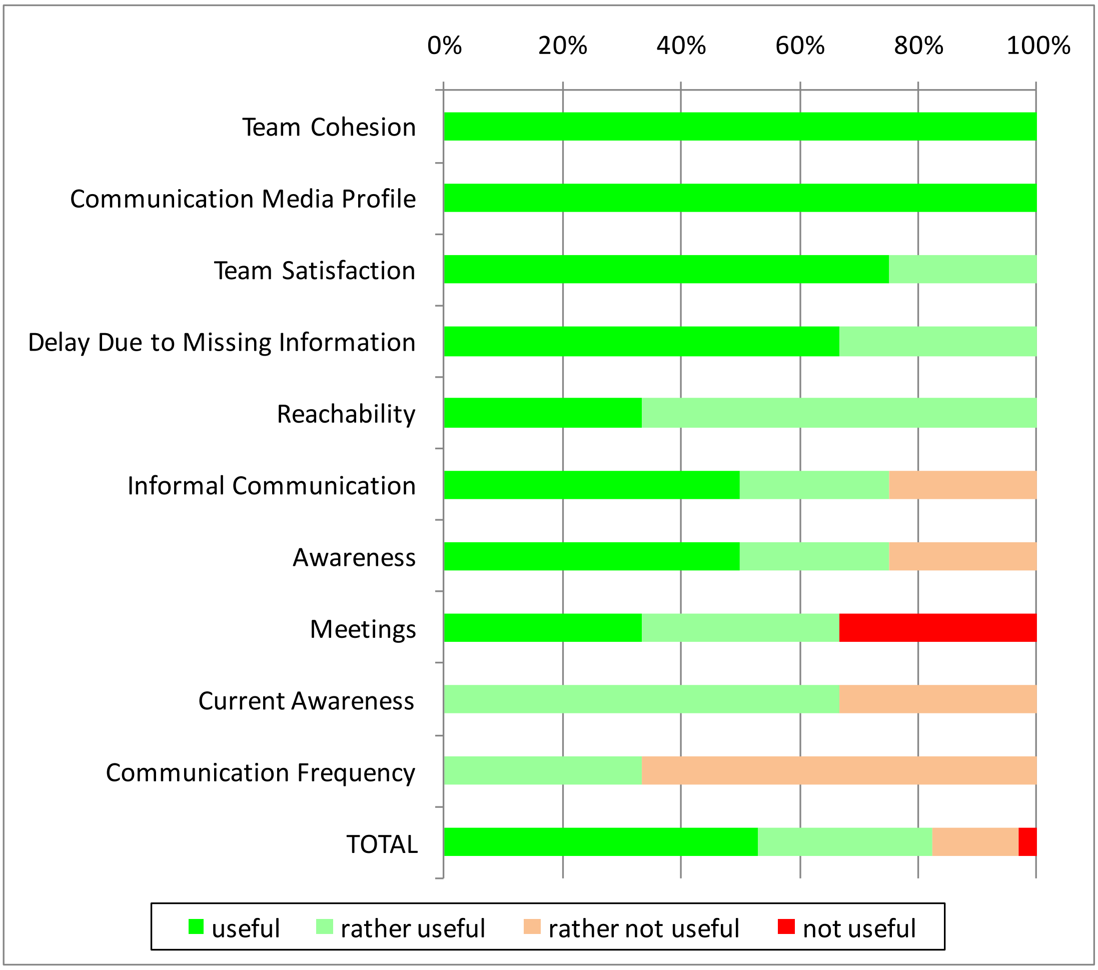


Figure 9.17: DOLLI4: Usefulness of indicators according to coaches feedback survey

spoke to just another person of the team? Another point of criticism was the usefulness of the social network analysis indicators: Communication Frequency, Reachability and Current Awareness. While the indicator Reachability still was seen as very useful, in particular with respect to the reachability of the customer, the Communication Frequency and Current Awareness were seen less useful. The explanation from the feedback comments was the reorganization of the Landside Server team: On the one hand, the responsible Landside Server colleagues were now part of the own team and therefore the social network across teams was not really needed anymore. On the other hand, people found it confusing that the survey answers still had to be given according to the old organization structure. Regarding the Informal Communication indicator it was mentioned by one of the feedback respondents that it was not clear to him what the expected behavior would have been.

In their feedback, the team coaches mentioned concrete actions which they had taken based on the collaboration analysis. Besides the reorganization of the Landside Server team which was the most obvious action, they for example escalated the issues with Delay Due to Missing Information because of the customer collaboration. A bad result related to the aspect *there are many discussion about who has a particular responsibility* was used by the coach to discuss the issue in the team and to clearly define the assignment of tasks.

Two of the respondents explicitly stated that the collaboration analysis approach has high potential for larger projects with more conflicts:

> "Because the communication in our team went well anyway, the usefulness of the analysis was limited, however in case of problems it would have been definitely valuable."
>
> "The usefulness was limited because of [...] few irregularities. A project at a larger scale and in the real world would certainly have a big benefit."

Daily Measurements During the Scrum Phase

During the on-site scrum phase we experimented with daily application of a small set of metrics in order to investigate the accuracy, usefulness and effort of collaboration-based indicators when increasing the measurement frequency to a daily basis. We defined the following three questions to be answered by all team members:

1. How comfortable do I feel in my team?
2. How satisfied am I with the project's achievements so far?
3. How good is the communication with the customer for technical clarifications?

Question 1 and 2 refer to the indicator Team Satisfaction in our GSE measurement model. Question 3 refers to the communication with the customer with respect to technical clarifications, which is related to the indicators Reachability as well as Delay Due to Missing Information. Instead of using an online survey, we prepared flipcharts for the teams and equipped the team coaches with glue dots. The team coaches were instructed to ask their team members to place one dot per question on the flipcharts every day before or right after the daily scrum meeting. Figure 9.18 gives an example. The results were used on-site by the teams to discuss and resolve issues. After the completion of the DOLLI4 project, we analyzed the data from the flipcharts and prepared a graphical summary, see Figure 9.19. Based on these results, we asked the team coaches for background information for the effects observed.

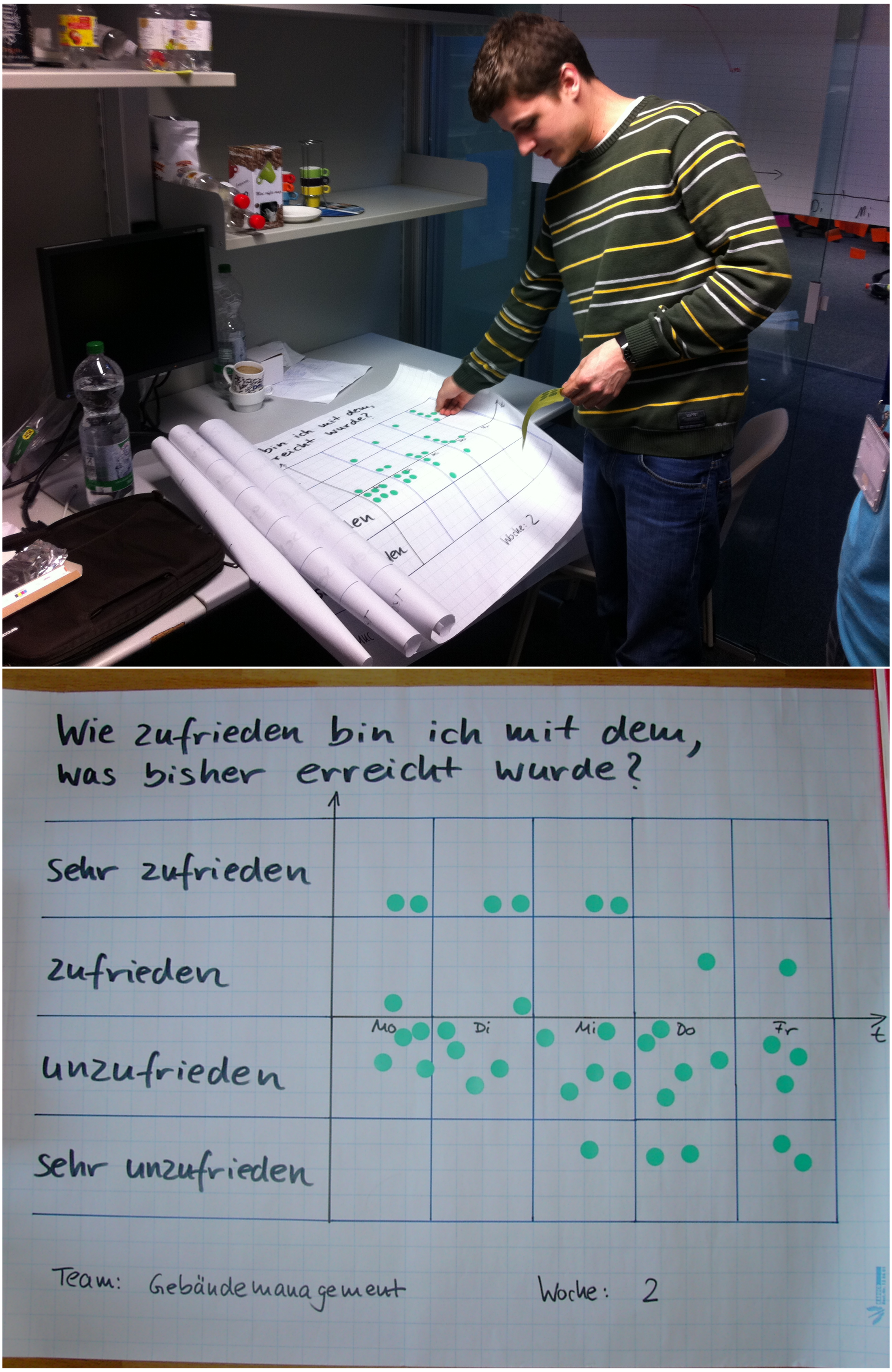


Figure 9.18: DOLLI4: Participant of the team MUC App during dot placement and sample flipchart of the team Building Management regarding the question *How satisfied am I with the project's achievements so far?*

Figure 9.19: DOLLI4: Results of daily measurements during the scrum phase

In the results of the team Building Management is it striking that team satisfaction and customer communication for technical clarifications went down in particular after day 7. The explanation which we received from the team coach was the following: During the prototyping phase and also the first week of the scrum phase, the team thought they were well informed because they had two contact persons at the customer and received presentations during the prototyping phase at the airport about the system. But at the end of week 1 at the airport, it turned out that there were conflicting requirements since the the two contact persons had different login systems in mind, which were entirely incompatible. Finally login functionality could not be offered at all and a lot of development effort was wasted. The Building Management team was frustrated because they were of the opinion that it was not their fault but that deficiencies in the customer-internal communication lead to their problem. Another issue occurred with the CORBA system which they needed for data exchange with the building management systems at the airport: The team based their work on a publicly available CORBA implementation, however, there were interfacing problems because the airport uses a specific CORBA implementation. The expert at the customer, who could potentially help, did only have very limited time for the DOLLI4 project. At the end of week 1, when a higher ranking person came to see the team, the situation suddenly changed: The higher ranking person immediately took action and sent the expert who didn't have time before to provide support to the Building Management team, which finally helped to solve the problem. However, the team had spent much effort before to implement a workaround, which was of no use anymore and had to be discarded. These events explain why both the team satisfaction and the rating of the customer communication went down towards the end of the project.

In the Security team both the customer communication for technical clarification as well as the team satisfaction increased significantly after day 5. The team coach of the Security team explained these effects in the following way: At the beginning of the scrum phase, the team had been behind schedule. In particular, an architecture change by the customer on day 2 caused an additional delay, which had impact on the satisfaction with the project's achievements so far. From day 4, the main contact person at the customer was back from holiday and provided very prompt support. From that day the communication with the customer increased continuously. Besides, on day 4 the architecture changes were completed successfully and the team could now focus on the implementing of the functionality related to the user stories.

In the team MUC App the team satisfaction correlated with the degree of completion of the system.

In the coaches feedback survey we explicitly asked for the feedback regarding the three flipchart questions. The results can be seen in Figure 9.20 and Figure 9.21. In total, the feedback again was very positive. It is to be noted that the

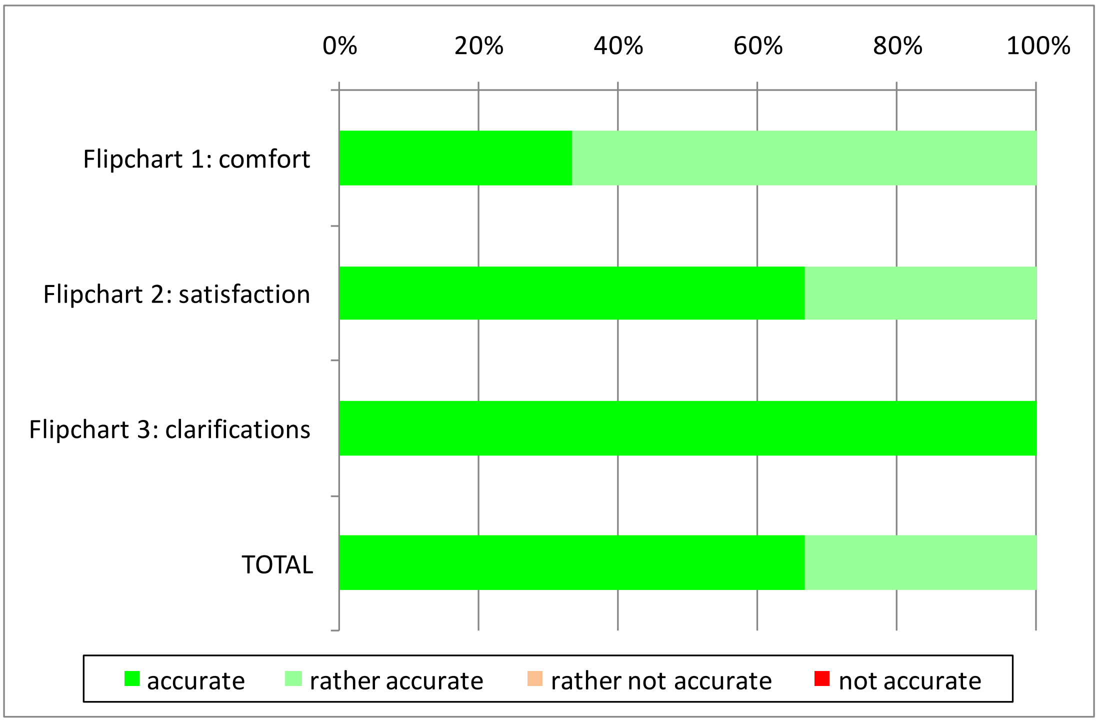


Figure 9.20: DOLLI4: Accuracy of daily feedback indicators according to coaches feedback survey

first question – How comfortable do I feel in my team? – was seen less accurate than the other two and was rated by one of the team coaches as not useful. He explained it in the feedback comments by the fact that in his team nobody dared to admit not feeling comfortable inside the team, because the placement of the dots was not anonymous and people were afraid of negative consequences, e.g. to be asked in front of the other team members to explain their answer, if their rating differed from the others'. A second feedback comment mentioned the same tendency towards good ratings on flipchart 1: Although there were problems, the scale of answers was contorted. Therefore, answers under 'satisfied' instead of 'very satisfied' were used as inducement for a crisis talk. The question of Flipchart 2 received one *rather not useful* rating, which was explained by the fact that the status correlated with the overall work progress which was obvious for the team. Regarding the third question it was stated in the feedback that they had paid attention to hide this flipchart from the eyes of the customer.

Complementary to the coaches feedback survey, we had conversations with team members at the customer acceptance test, which revealed that they found the daily measurements valuable. The team members stated that the flipcharts helped to address issues by discussing them openly, and that they found it helpful to see that others had the same problems as they had.

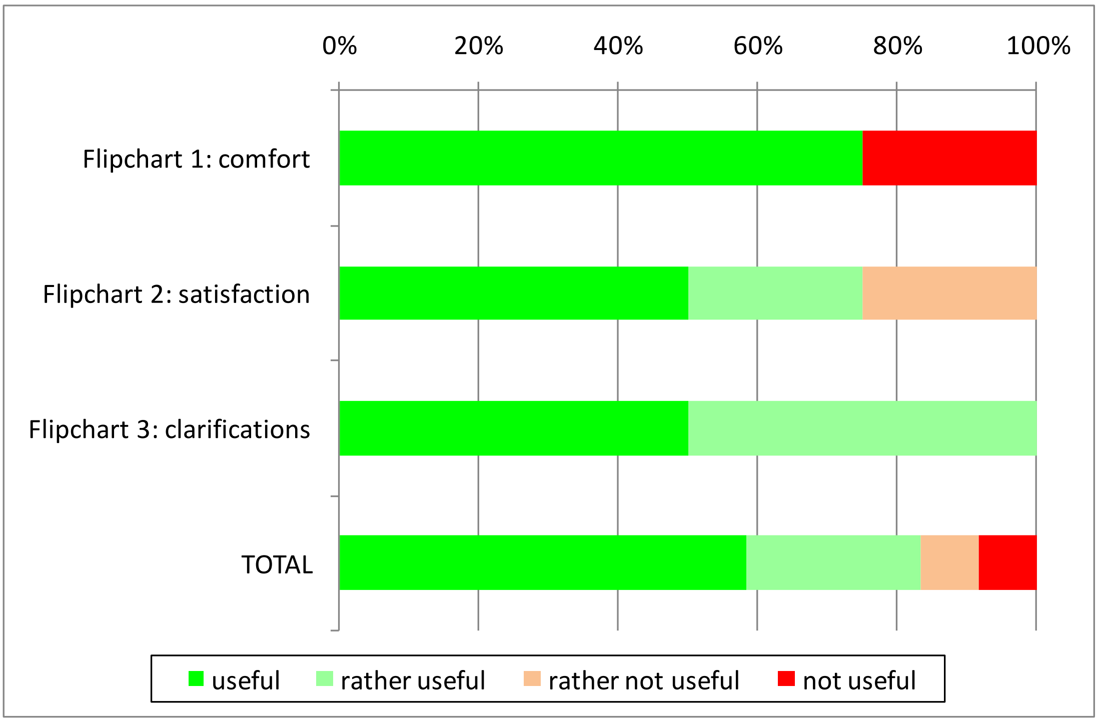


Figure 9.21: DOLLI4: Usefulness of daily feedback indicators according to coaches feedback survey

Summary

Although DOLLI4 was not a GSE project, similar collaboration challenges were observed. In particular the communication between subteams that had dependencies on each other and the communication with the real customer were challenging. The GSE metrics data showed plausible results over time in comparison with the NEREID projects which supports the reliability of the data collected. A major issue identified with the help of our GSE metrics was the impeded collaboration between the Landside Server team and the other teams, which was resolved by a reorganization of teams. We experimented with the daily measurement of three questions during the on-site scrum phase of the DOLLI4 project and found evidence that the measurements reflected events well that impeded collaboration. It was noted that when the answers of the team members are not anonymous their ratings are influenced by the fear of negative consequences, e.g. being asked in front of the audience to explain a negative rating.

In the feedback of the team coaches on the overall approach, it was expressed that the GSE metrics were accurate, useful also in case of this large co-located project, even though people mentioned it would have even more potential in case of larger projects with more collaboration problems. The effort involved was seen as acceptable.

9.3 Threats to Validity

In this section we discuss threats to the validity of our results.

9.3.1 Construct Validity

A central threat to validity of our study is the *subjectiveness of data*. The data collection for our GSE metrics was done based on survey questionnaires that were filled in by the team member of the projects under observation. Consequently, the data reflects the subjective view of the respondents. There is even the risk that project team members just check arbitrary answers in the survey questionnaire. However, although there might be a certain fuzziness in data collected by surveys, the project team members and team coaches clearly confirmed the accuracy and usefulness of the metrics in their feedback as well as in the interviews, as we described above. To probe the reliability of the data, we established general hypotheses of expected behavior, which were confirmed by the data (see Section 9.1.3). This would not have been the case, if the data was arbitrary or meaningless. Also the findings from the interviews were consistent with the data, e.g. the statements from the industrial case study in Chapter 8 or the internal conflicts at the Munich site in the Carpooling and Ride Sharing project of NEREID.

9.3.2 Internal Validity

A core problem in the interpretation of the survey results was the lack of context information: Because the author of this dissertation was not part of the team, there is the risk that interpretations may be weak. One example we have seen above was the anticipated collaboration problem between Valparaiso and the other sites in the Touch Screen Tablet project of NEREID, where the real cause was that the team members from Valparaiso had left the team due to technical reasons. It is to be noted that the GSE metrics are intended for project managers, insofar the situation in this dissertation was artificial. If the approach is used by project managers, they also have the required context information at hand.

9.3.3 External Validity

In this dissertation, a small set of projects was analyzed, which provides anecdotal evidence only. However, although the set was small, we tried to cover multiple project types: a real-world industrial project, GSE student projects as well as a large co-located student project. Although there is the risk that the student projects might be not representative for the reality in industry, we found that these projects were facing quite realistic challenges. As described in Section 4.3, the collaboration challenges inside the NEREID projects were comparable in many aspects to what is known about collaboration challenges in industrial GSE

Activity	Effort (hours)
Initial setup of the online survey (create survey questionnaire, enter list of team members)	20
Initial setup of the tooling for automatic generation of the collaboration analysis reports (linked Excel and Powerpoint)	40
Administration of each survey (send out invitations, send out reminders)	1
Preparation of the collaboration analysis report (generate report, write executive summary, send to teams)	2–4

Table 9.5: Effort for preparation, execution and analysis of the GSE metrics surveys

projects. Also the DOLLI4 project had a quite realistic setting, in particular because of the 6 Rs (real customer, real deadline, etc.) described above.

9.4 Own Observations and Lessons Learned

In the following we describe experiences and observations during the validation phase from our own perspective.

We prepared an overview of the effort data for handling the GSE metrics surveys. It refers to the basic efforts required for the preparation, execution and analysis of the surveys. Note that the development of the GSE model and the industrial case study was already done when we started to prepare the surveys for the validation projects. Furthermore, we were already familiar with conducting and analyzing surveys in general. Also the preparation of graphical summaries for this dissertation and activities such as the coordination with the course owners, the feedback surveys and semi-structured interviews, writing this dissertation, etc. are *not* included in these figures. Table 9.5 lists the activities and the required number of working hours. Please note that the effort data in this table refers to handling a whole set of projects (4 NEREID projects or 4 DOLLI4 teams) with the same survey.

The initial efforts to prepare the online survey tool and the tooling for automatic generation of the collaboration analysis reports from the survey data took approx. 60 hours. These were one-time efforts for the whole set of NEREID projects resp. all DOLLI4 teams. The recurring efforts for the execution of a single survey were comparably low: it took 1 hour to send out the invitations to fill in the online survey to the project team members and to send reminders after a few days to those people who had not answered yet. Also the effort for preparation of the collaboration analysis reports was with 2–4 hours rather low: Since the collaboration analysis report documents were generated automatically from linked Excel and Powerpoint tooling, the main effort was in examining the results and writing the executive summary for the collaboration analysis reports. From our experience, automating the report generation was crucial to be able to evaluate a large number of surveys within short time (in total we handled 24 surveys in the validation phase: NEREID: $2 * 4 = 8$ surveys, DOLLI4: $4 * 4 = 16$ surveys). The data collection via a survey actually turned out to be a lightweight, but very useful approach. In the future, it could be combined well with other data sources (e.g. e-mail data or traveling records).

Because of the usage of our metrics during project lifetime instead of a post mortem analysis, three major aspects were important from our point of view:

Just enough measurement. As data collection and analysis also can cause high effort during project lifetime, it is important to define a suitable measurement strategy: How much measurement is necessary in order to achieve reasonable improvements? In our experience it is more appropriate to have selective measurements and random samples instead of an all-embracing approach. The set of metrics in our GSE model and the effort associated with collecting this data was acceptable according to the feedback which we received from the validation projects. However, there were some comments that the number of metrics might be reduced to lower the effort. In particular, if the frequency of measurements should be further increased from monthly e.g. to bi-weekly or weekly, it would be crucial to further reduce the effort per survey.

Privacy and open feedback. It is essential to preserve a relationship of mutual trust with the team members and use the collaboration analysis results only for the good of the team. If the team members have to fear negative consequences, they will not answer openly. For example the daily feedback in the on-site scrum phase of DOLLI4 were done with flipcharts and thus it was visible to all team members who was voting in which way. Team coaches reported in the feedback that their team members didn't answer the question *How comfortable do I feel in my team?* openly because they didn't want to be conspicuous and found it embarrassing to be asked in front of the other team members to explain their rating, if they voted differently from the others. However, open feedback is essential to get the real picture, therefore the data should be collected anonymously. Furthermore, one should be careful with setting target values, because the survey respondents may easily 'fake' the data just to achieve the target.
Depending on the legal situation of a certain country, data collection of personal data is restricted or prohibited. In this case, the data needs to be anonymized and kept confidential, or the approval of the team members is a necessary prerequisite for data collection. Our original motivation to collect social network analysis data on a higher aggregation level – on team level instead of individuals – was to protect the privacy of the team members. However, it turned out that this reduction in granularity actually was beneficial also by increasing the lucidity

(simple graphs) and reducing the effort for analysis, while the results were still meaningful.

Interpretation of results. Although we were in an 'outsider' position to the projects, we got lots of insights into their team dynamics and what was currently going on in the projects. However, our interpretations derived from the GSE metrics data were sometimes imprecise due to lack of context information. The GSE metrics should be actively used by the project manager instead of an outside research person. In our opinion, the team coaches concept of DOLLI4 worked well and would be recommended. To increase the quality of interpretations, the GSE metrics should be used in conjunction with other, traditional project metrics as well as the observations of the project manager.

We experienced that the introduction of our GSE metrics in case of the globally distributed projects was more challenging than in the co-located project because of the discussed effects of distribution. In contrast to the DOLLI4 project, we didn't have the chance to meet the NEREID project team members in person and needed to communicate mostly via e-mail, and there was a risk that project team members would not participate in the surveys. Therefore one should not underestimate the impact of distribution also when introducing such a measurement system. In case of the NEREID projects, we argued via the benefits for the team members (they could use it to improve their projects and finally to deliver a better result), but did not exert pressure on them.

9.5 Summary

In summary, our approach of the collaboration-based project management metrics was successful in both the GSE student projects (NEREID) as well as the co-located student project (DOLLI4) which we studied during the validation phase. In total the metrics reflected the reality in the project teams well. Furthermore, they were seen as useful by the teams and the effort involved was acceptable. The finding that the GSE metrics were useful was also expressed through the improvement actions taken by the teams, e.g. the reorganization of the DOLLI4 teams after the collaboration analysis revealed collaboration deficiencies. Our approach was rated to be very innovative and new to GSE project management. We presented our own observations from the validation phase and highlighted the importance of focusing on the right metrics, as data collection and analysis might cause high effort during project lifetime, and anonymous feedback to preserve privacy and increase the openness of the respondents.

Chapter 10

Conclusion and Future Work

In this dissertation, we dealt with the research question *How can collaboration problems in Global Software Engineering (GSE) projects be detected early on before they have serious impact?* The fundamental hypothesis of our research was that collaboration-based measurement indicators can help to detect and address collaboration problems in GSE projects. To validate our hypothesis we constructed a set of collaboration-based project management metrics for GSE and evaluated them. We structured our research in three major steps: First, we performed a literature research and conducted case studies and controlled experiments to get a profound understanding of the challenges in GSE projects. We took an interdisciplinary view on the social psychology of distributed teams and studied the basic principles of communication. Furthermore, we presented the concepts of software measurement and communication metrics. Second, we introduced a model of GSE metrics and defined a set of ten collaboration-based project management metrics for GSE. Third, we applied our GSE metrics to several projects in order to gain experience with them and evaluate their accuracy, usefulness, the effort involved as well as the innovation grade of our approach. The objects of our research were an industrial case study of a real-world crisis project, four GSE student projects as well as one large co-located student project.

The application of our GSE metrics in the six projects gave anecdotal evidence that our collaboration-based GSE metrics are helpful to detect and address collaboration problems in GSE projects. In particular, we found that the GSE metrics data provided accurate and useful information to the project teams, while the effort for data collection was acceptable. Our approach was evaluated to be very innovative and new to GSE project management.

10.1 Contributions

With this dissertation we made the following contributions to enhance the existing body of knowledge on Global Software Engineering:

1. **Structure-Behavior-Function Meta Model for GSE.** We systematized existing work on GSE by introduction of a Structure-Behavior-Function Meta Model of GSE Models representing different views on GSE. For example, the Location Model and the Organization Model deal with predefined structures, while the Collaboration Model refers to dynamic team behavior and actual communication structures.

2. **Experiments in GSE education.** We presented group exercises for teaching GSE which at the same time were experiments in GSE research and provided insights on fundamental effects in GSE, e.g. delay.

3. **Model of collaboration-based project management metrics for GSE.** We defined a model of collaboration-based metrics to support GSE project management. The metrics were constructed using an interdisciplinary approach, utilizing knowledge from social psychology as well as software measurement expertise from computer science.

4. **Industrial case study.** We constructed and applied the GSE measurement model in an industrial case study of a real-world project with a significant size (50 team members). By studying a risk project, we presented also pitfalls in GSE, where industrial publications often tend to describe success stories.

5. **Validation in two case studies.** We applied the defined collaboration-based project management metrics to a set of GSE projects as well as a co-located project and evaluated their accuracy and usefulness as well as the effort involved. Furthermore, also the innovation grade of our approach was evaluated.

10.2 Future Directions

With our research we opened up the new field of collaboration-based project management metrics for GSE. However, there are several aspects that we could not deal with in this dissertation, but will be important next steps from our perspective.

Quantifying the benefit of improved collaboration. Our approach was based on the assumption that improved collaboration also implies improved project performance. This assumption can be challenged and analyzed. Future research should explore the relationship between collaboration-based metrics and project performance and quantify the benefit of improved collaboration.

Summative evaluation. The six projects we explored gave anecdotal evidence for the effectiveness of our approach. However, more case studies in application of the GSE metrics are necessary to get a broad and representative picture of their accuracy and usefulness. On this basis, also a statistical analysis of results will be possible. Furthermore, the data collected can be used to build an experience database with reference values for future measurements.

Refinement of metrics. Also the selection of metrics should be challenged. Specific studies should be designed to find the optimal set of metrics for a given context to cope with the tradeoff between effort and effectiveness. In our research, all social network analysis data referred to the aggregation of sites respectively roles. The original motivation was to preserve the privacy of the individual team members. However, it turned out that this aggregated view still provided actionable information and at the same time reduced the effort for preparing the analysis and increased the lucidity of the results. However, it should be investigated how valuable a social network analysis based on individuals would be as a project management metric in GSE.

Alternative ways of data collection. Our model of GSE metrics distinguishes between the Interaction level – metrics related to visible interaction between team members – and the Relationship level referring to the interpersonal relationship between team members. Per definition, for the metrics on Interaction level a direct measurement is possible. In this dissertation, the whole data collection relied on survey questionnaires filled in by the team members. Alternative ways of data collection should be explored in future research, i.e. the metrics on Interaction level could be calculated by analyzing e-mail data, traveling records, etc. Also on Relationship level, alternative ways of data collection could be explored, i.e. observation by team coaches either instead or in addition to the survey data provided by the team members themselves.

Variation of measurement frequency. Furthermore, the influence of the frequency of data collection should be studied in future work. We tapped different possibilities in this dissertation, ranging from a one time application in the industrial case study over a monthly application in the GSE student projects and the co-located student project up to an experiment with daily feedback in the scrum phase of the co-located student project. However, a systematic analysis of the measurement frequency and its positive and negative consequences is still to be done.

The interdisciplinary view into social psychology provided valuable insights to us. We see a great potential of conducting more interdisciplinary work between computer scientist and other disciplines in the future, e.g. incorporating more knowledge from communication theory.

Appendix A

Definitions of GSE Metrics

This appendix contains the detailed definitions of our GSE metrics which we introduced in Chapter 7. It is structured into *Base Measure* definitions which represent the primary data collected and *Indicator* definitions which are derived from the base measures and used in measurement cockpits.

A.1 Templates

We structured the Indicator and Base Measure definitions according to the following templates; the template concept was adapted from Card et al. [MCJ+01, p.160] and Lescher and Paulisch [LP07].

A.1.1 Indicator Template

Indicator.
Name of the indicator

Level.
Interaction Level or Relationship Level

Business Goal.
Business goal which the indicator is related to

Strategy.
Strategy that is followed to achieve the business goal

Measurement Goal.
Goal that shall be achieved by the measurement

Questions.
Questions which are answered by the indicator

Definitions.
Definitions related to the indicator

Base Measures.
List of base measures that are used by the indicator

Interpretation and Decision Criteria.
Guidance for the interpretation of the indicator, e.g. typical behavior, thresholds or target values

Diagram.
A graphical representation of the indicator

References.
References to external sources

A.1.2 Base Measure Template

Base Measure.
Name of the base measure

Identifier.
Short identifier used to refer to the base measure in formulas (e.g. NWD)

Reference Question.
Reference to a question number in the model of GSE metrics (see Chapter 7), e.g. Q6.1

Description.
Description of the base measure

Value Range.
The range of values which the base measure can take

Data Source.
The data sources from which the base measure can be obtained

A.2 Indicators: Interaction Level

A.2.1 Communication Frequency

Indicator.
Communication Frequency

Level.
Interaction Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team interactions

Measurement Goal.
Analyze communication frequency

Questions.
Q1.1 How often do team members communicate?

Definitions.
$communication\ frequency(a, b) = avg(CF(a, b))$

where $CF(a, b)$ depicts the set of answers for communication frequency between two individuals or two subgroups of the team, i.e. roles or sites

Base Measures.
Communication Frequency (CF)

Interpretation and Decision Criteria.
Frequent communication is the prerequisite for a good collaboration. Therefore watch the communication frequency for the local sites as well as across sites. The expected behavior is that the local communication is more frequent than the communication across sites, because the ability to communicate of co-located team members is higher. However, sufficient cross-site communication is crucial for a good collaboration and to avoid misunderstandings and rework, thus particular attention to the cross-site communication is required. The communication frequency should match the system dependencies (cf. Conway's Law [Con68]). Sudden increases in communication frequency can indicate problems, such as issues with system integration.

Diagram.
See Figure A.1

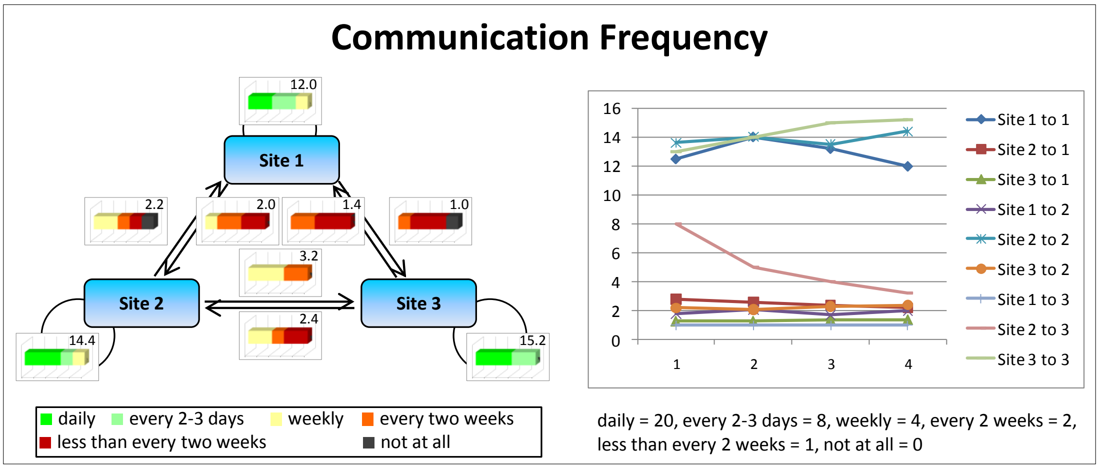


Figure A.1: Indicator: Communication Frequency

References.
Moreno [Mor34], Bruegge and Dutoit [DB98], Damian et al. [MDSS08], Ehrlich and Chang [EC06], Wolf et al. [WSDN09], Coplien and Harrison [CH04]

A.2.2 Reachability

Indicator.
Reachability

Level.
Interaction Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team interactions

Measurement Goal.
Analyze reachability

Questions.
Q2.1 Are team members easy to reach?

Definitions.

$reachability(a, b) = avg(TR(a, b))$

where $TR(a, b)$ depicts the set of answers for team reachability between two individuals or two subgroups of the team, i.e. roles or sites

Base Measures.

Team Reachability (TR)

Interpretation and Decision Criteria.

When team members can't reach their colleagues they need to communicate with, it impedes collaboration and causes friction losses. Therefore watch the reachability of team members. The expected behavior is that the local reachability is higher than reachability of distant colleagues. Poor reachability can be neglected, if there is no need for communication, e.g. because there are no system dependencies between two particular sites. Besides poor reachability, also changes over time (e.g. three consecutive declines) shall be investigated.

Diagram.

See Figure A.2

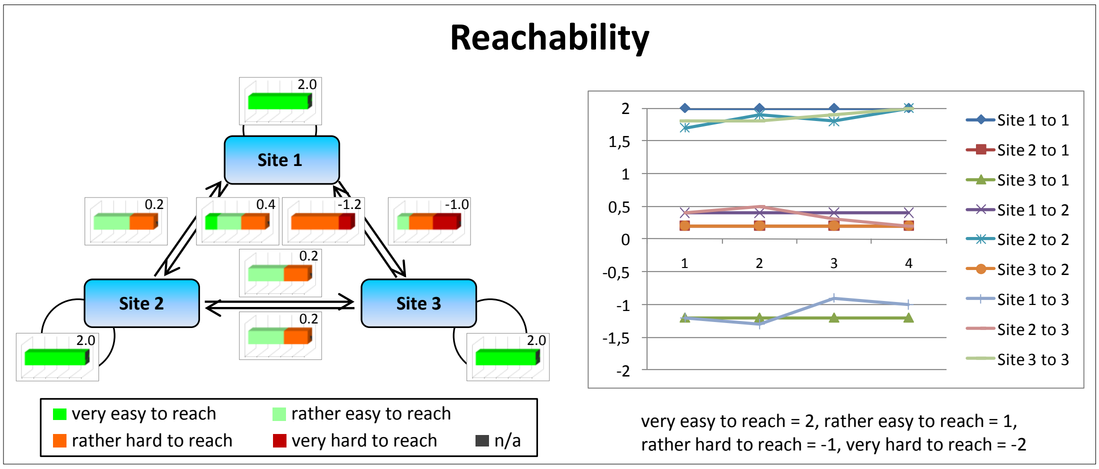


Figure A.2: Indicator: Reachability

References.

Moreno [Mor34], Bruegge and Dutoit [DB98], Damian et al. [MDSS08], Ehrlich and Chang [EC06], Wolf et al. [WSDN09], Coplien and Harrison [CH04]

A.2.3 Communication Media Profile

Indicator.
Communication Media Profile

Level.
Interaction Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team interactions

Measurement Goal.
Analyze communication media

Questions.
Q3.1 How much communication is done via medium X?

Definitions.
$communication\ media\ profile = \{(m, avg(CMU(m)))|m \in M\}$

where M is the set of communication media and $CMU(m)$ depicts the set of answers for communication media utilization of medium m

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Communication Media Utilization (CMU)

Interpretation and Decision Criteria.
The Communication Media Profile, i.e. the share of communication media which team members use, can indicate communication problems. Watch for unusual profiles, e.g. high share of e-mail communication and missing personal contact, which might lead to misunderstandings and ineffective communication. In addition to the current Communication Media Profile, watch also for changes over time.

Diagram.
See Figure A.3

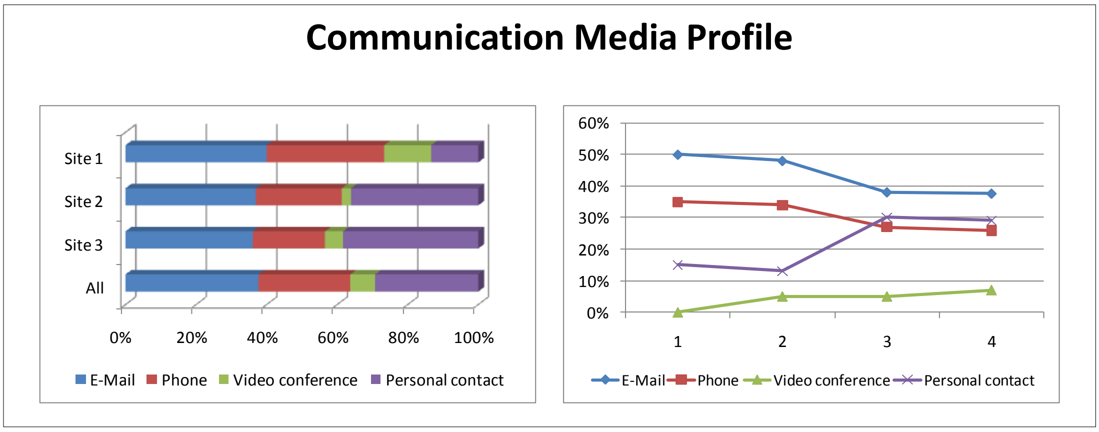


Figure A.3: Indicator: Communication Media Profile

References.
Herbsleb and Mockus, question 38 [HM03]

A.2.4 Delay Due to Missing Information

Indicator.
Delay Due to Missing Information

Level.
Interaction Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team interactions

Measurement Goal.
Analyze delay due to missing information

Questions.
Q4.1 How often was work delayed because of missing information?
Q4.2 How long lasted work delays because of missing information?

Definitions.
$number\ of\ delays = avg(NWD)$
$duration\ of\ delays = avg(DWD)$

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Number of Work Delays (NWD), Duration of Work Delays (DWD)

Interpretation and Decision Criteria.
Delay is a key phenomenon in GSE and can impact the project performance significantly (e.g. drift of the project end date, time pressure during integration and testing, etc.). Therefore watch the number and duration of delays due to missing information. The expected behavior is that there are more frequent and longer delays with respect to the distant site than locally. According to Herbsleb and Mockus, distributed work items can take about 2.5 times longer to complete than similar items where all the work is co-located [HM03]. Investigate the root causes for frequent and/or long delays (e.g. more than 5 calendar days). Furthermore, watch for sudden escalation of delays and changes over time (e.g. three consecutive increases).

Diagram.
See Figure A.4

References.
Herbsleb and Mockus, question 29 and 30 [HM03]

A.2.5 Traveling

Indicator.
Traveling

Level.
Interaction Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team interactions

Measurement Goal.
Analyze traveling records

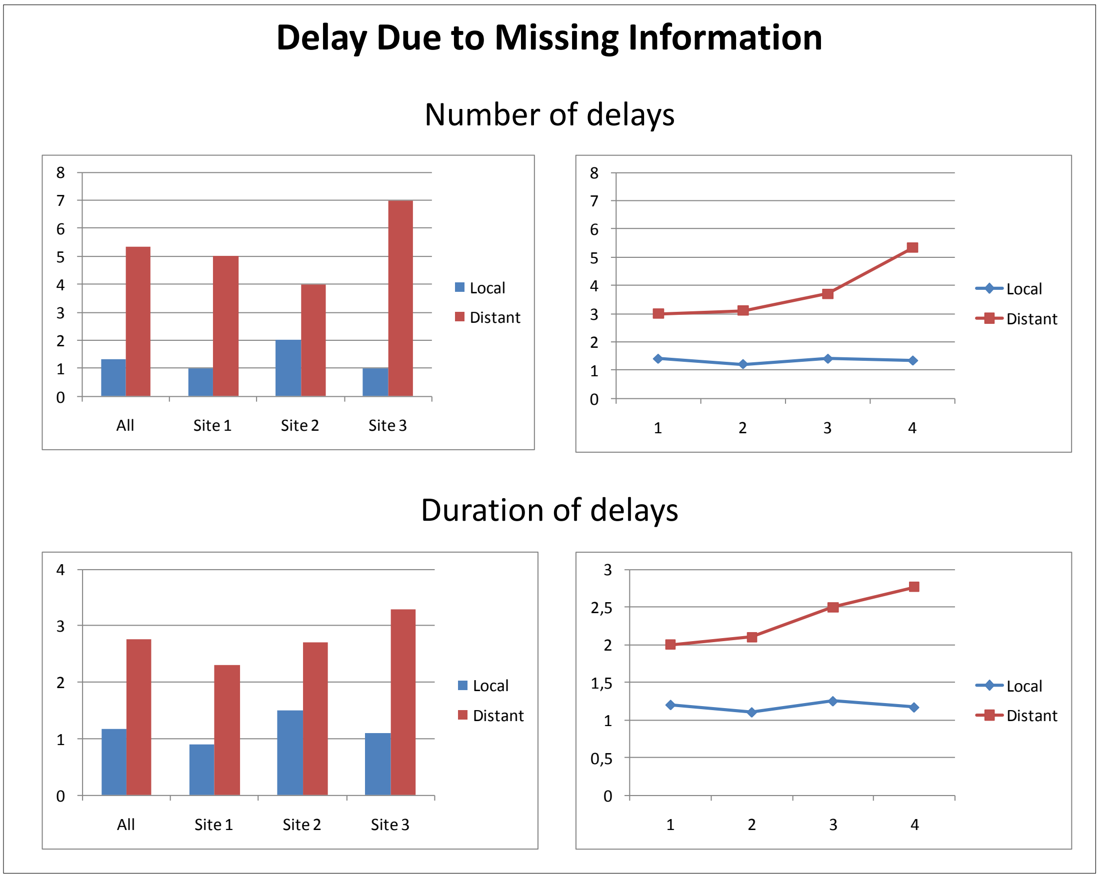


Figure A.4: Indicator: Delay Due to Missing Information

Questions.
Q5.1 How many team members have visited other development sites?
Q5.2 How many days have team members spent at other sites?

Definitions.
$traveling\ rate = \frac{\sum CST}{N}$
$duration\ of\ visits = avg(TD)$

where N represents the total number of team members.

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Cross-Site Traveling (CST), Traveling Duration (TD)

Interpretation and Decision Criteria.

Traveling of team members provides implicit information about team communication and personal contact. Meeting in person can help to clarify issues that are difficult to solve over virtual media. Therefore watch traveling behavior. A low traveling rate can reveal insufficient personal contact. Sudden increases in traveling can indicate coordination efforts due to problems. The duration of visits substantiates the intensity of the contact. Watch also for changes over time, e.g. three consecutive declines.

Diagram.

See Figure A.5

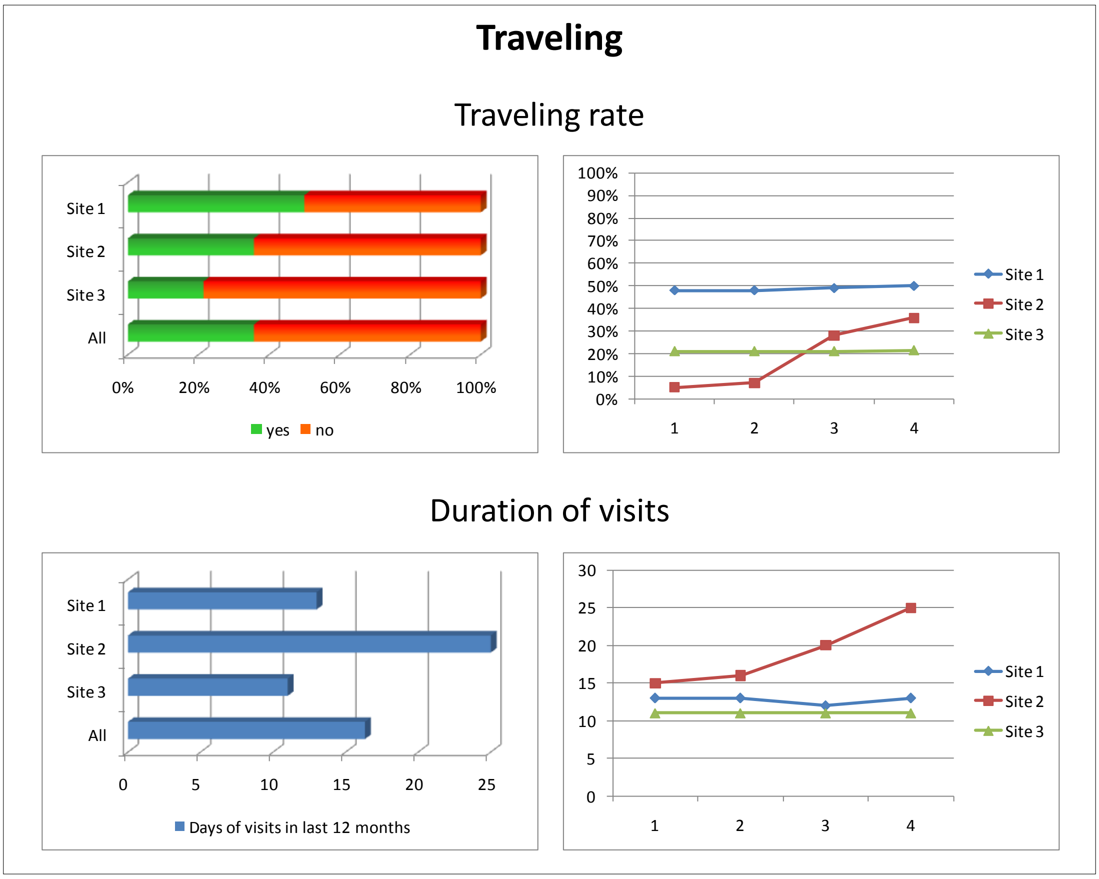


Figure A.5: Indicator: Traveling

References.

Herbsleb and Mockus, question 52 [HM03]

A.3 Indicators: Relationship Level

A.3.1 Team Cohesion

Indicator.
Team Cohesion

Level.
Relationship Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team relationship

Measurement Goal.
Analyze team cohesion

Questions.

- Q6.1: Is communication inside the project team honest and open?
- Q6.2: Can team members rely on their colleagues?
- Q6.3: Do team members assist each other with heavy workloads?
- Q6.4: Are new ideas from colleagues seen as valuable?
- Q6.5: Are meetings/phone conferences constructive?
- Q6.6: Is there individual competition between subteams?
- Q6.7: Are there many discussions about particular responsibilities?
- Q6.8: Do team members feel powerless to change the project's situation?

Definitions.
$team\ cohesion = (avg(OC), avg(RC), avg(WH), avg(VNI), avg(CM),$
$avg(TC), avg(CR), avg(ACS))$

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Open Communication (OC), Reliability of Colleagues (RC), Willingness to Help

(WH), Valuation of New Ideas (VNI), Constructive Meetings (CM), Team Competition (TC), Clear Responsibilities (CR), Ability to Change the Situation (ACS)

Interpretation and Decision Criteria.
The Team Cohesion indicator can reveal problems which are 'below the surface' of visible interaction. The values of the 8 dimensions of Team Cohesion should be between 1 and 2. The expected behavior of this indicator is that the local site is seen slightly better than the distant sites (e.g. difference of 0.5). Watch for low scores in individual values, discrepancies between sites as well as changes over time (e.g. three consecutive declines).

Diagram.
See Figure A.6

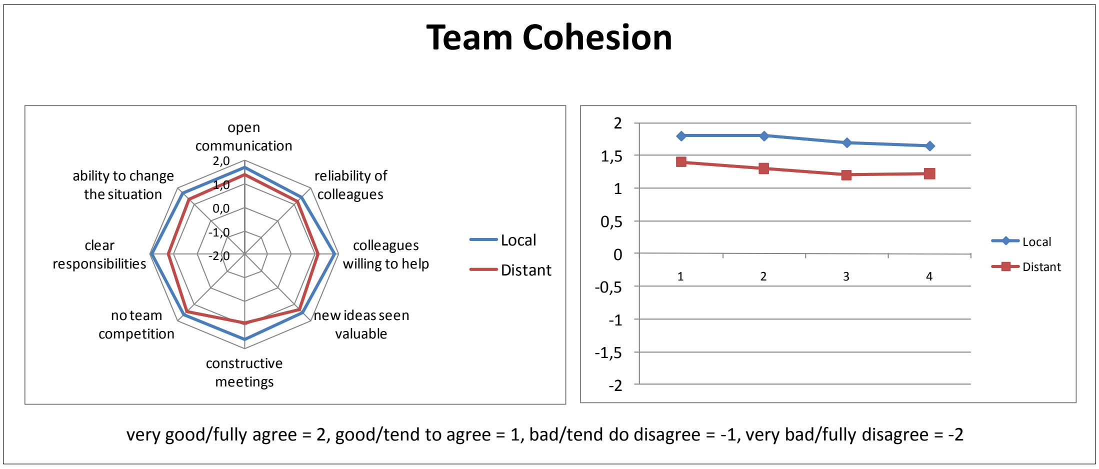


Figure A.6: Indicator: Team Cohesion

References.
Herbsleb and Mockus, question 12 and 16 [HM03], Kerth [Ker01, p.42]

A.3.2 Informal Communication

Indicator.
Informal Communication

Level.
Relationship Level

Business Goal.
Effective team collaboration across sites

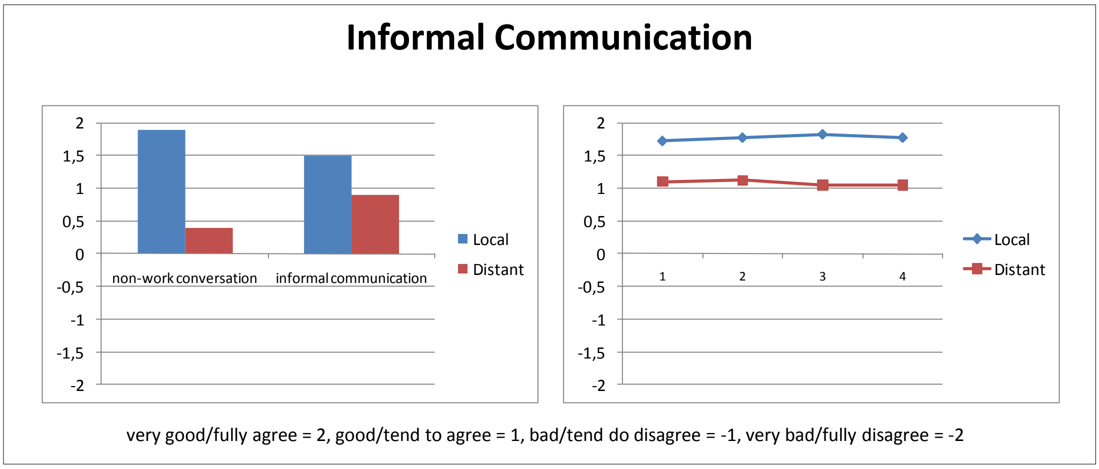


Figure A.7: Indicator: Informal Communication

Strategy.
Identify collaboration problems early on by monitoring team relationship

Measurement Goal.
Analyze informal communication

Questions.

- Q7.1: Do team members discuss non-work related matters?
- Q7.2: Do team member get useful information by informal communication?

Definitions.
$informal\ communication = (avg(NWC), avg(IC))$

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Non-Work Conversation (NWC), Informal Communication (IC)

Interpretation and Decision Criteria.
Informal communication is a sign of a good relationship between team members. Therefore watch the level of informal communication within sites as well as across sites. Low values can indicate limited trust and limited openness in communication. Due to the geographic separation of teams at different sites, the level of informal communication is lower across sites than locally. Watch also for changes over time (e.g. three consecutive declines).

Diagram.
See Figure A.7

References.
Herbsleb and Mockus, question 13 and 24 [HM03]

A.3.3 Awareness

Indicator.
Awareness

Level.
Relationship Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team relationship

Measurement Goal.
Analyze awareness

Questions.

- Q8.1: Do team members know what others are currently working on?
- Q8.2: Do team members know who to contact or who has expertise?
- Q8.3: Do team members see the importance to understand the day-to-day work of their colleagues?
- Q8.4: When work is assigned, is everybody clear about his/her task?
- Q8.5: Do colleagues provide timely information about changes?

Definitions.
$awareness = (avg(GA), avg(IUDW), avg(CT), avg(IAC))$
$current\ awareness(a, b) = avg(CA(a, b))$

where $CA(a, b)$ depicts the set of answers for current awareness between two individuals or two subgroups of the team, i.e. roles or sites

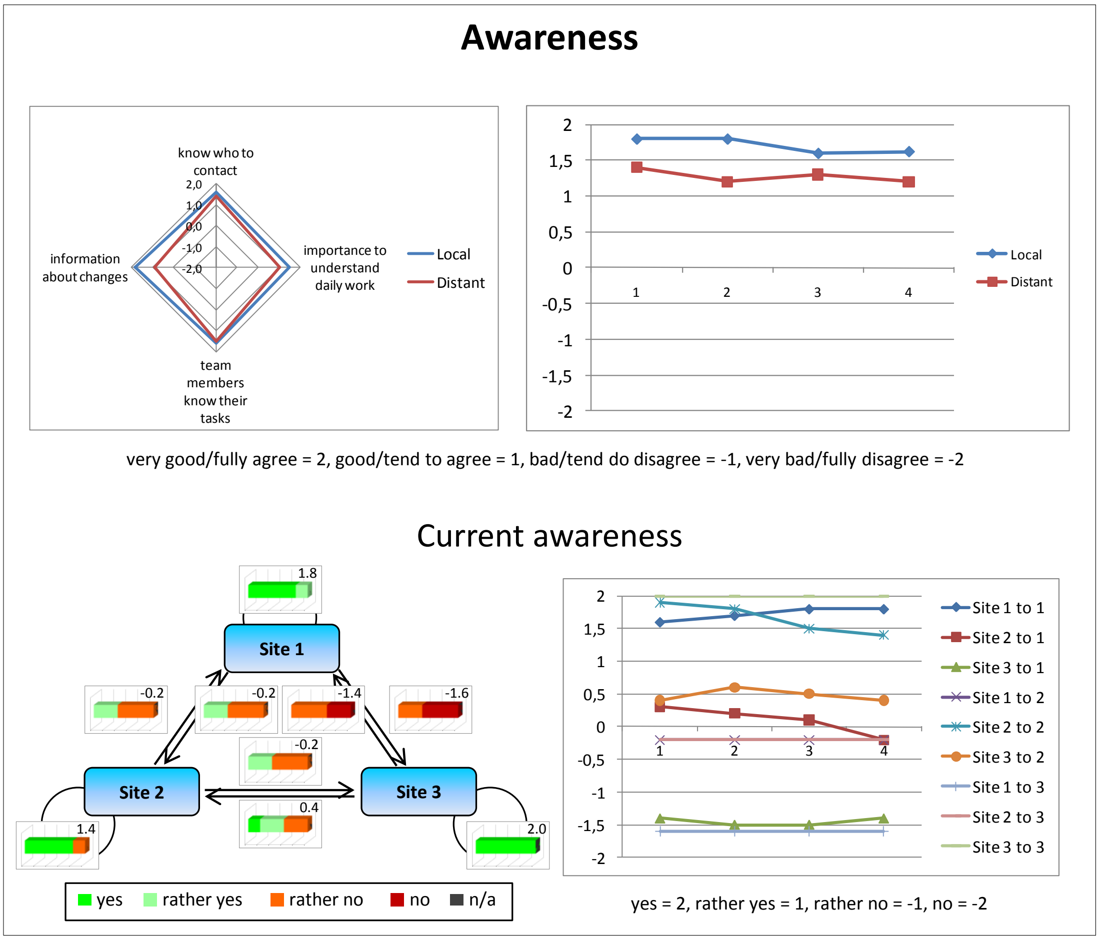


Figure A.8: Indicator: Awareness

Base Measures.

Current Awareness (CA), General Awareness (GA), Importance of Understanding Daily Work (IUDW), Clear Tasks (CT), Information About Changes (IAC)

Interpretation and Decision Criteria.

To know who to contact or who has expertise, the awareness of what colleagues are currently working on, and timely information about changes are important factors in a smooth collaboration. Therefore watch the Awareness level in the GSE teams. Because of the distribution it is often difficult for team members to maintain awareness about the work of colleagues at the distant sites. Therefore particular attention to the awareness level across sites is necessary. Watch for low values and clarify root causes. Besides poor awareness, also changes over time (e.g. three consecutive declines) shall be investigated.

Diagram.
See Figure A.8

References.
Herbsleb and Mockus, question 19, 22, 23 and 37 [HM03], Moreno [Mor34], Bruegge and Dutoit [DB98], Damian et al. [MDSS08], Ehrlich and Chang [EC06], Wolf et al. [WSDN09], Coplien and Harrison [CH04]

A.3.4 Culture & Language

Indicator.
Culture & Language

Level.
Relationship Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team relationship

Measurement Goal.
Analyze culture & language

Questions.

- Q9.1: Do team members have appropriate English language skills?
- Q9.2: Has national culture significant influence on team collaboration?

Definitions.
$culture\&language = (avg(ELS), avg(NCI))$

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
English Language Skills (ELS), National Culture Impact (NCI)

Interpretation and Decision Criteria.
Language skills and cultural differences can have significant impact on team collaboration. Poor language skills as well as cultural differences can lead to inefficient communication and the risk of misunderstandings. Therefore watch the level of Culture & Language. Low values can indicate latent problems. Usually the value of this Indicator is stable, if the team is stable (e.g. no new team members).

Diagram.
See Figure A.9

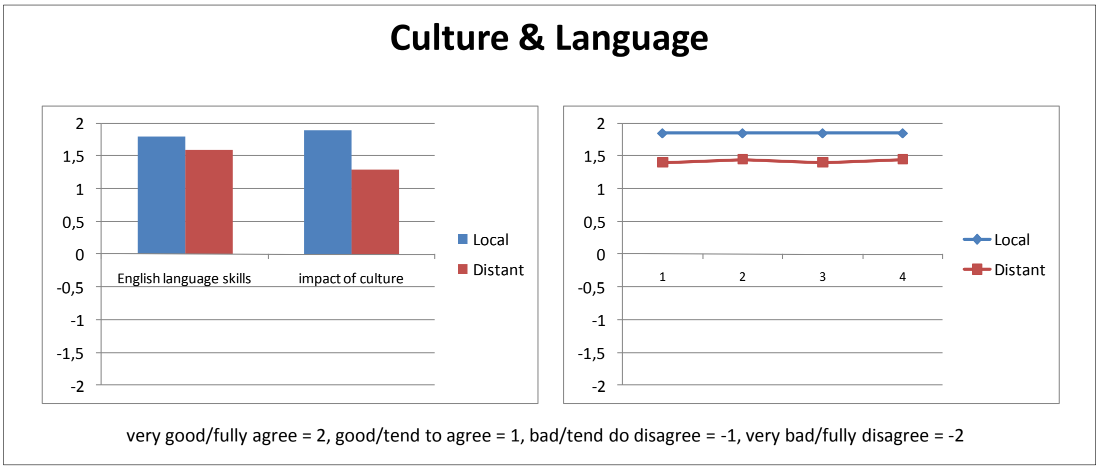


Figure A.9: Indicator: Culture & Language

References.
Herbsleb and Mockus, question 50 and 51 [HM03]

A.3.5 Team Satisfaction

Indicator.
Team Satisfaction

Level.
Relationship Level

Business Goal.
Effective team collaboration across sites

Strategy.
Identify collaboration problems early on by monitoring team relationship

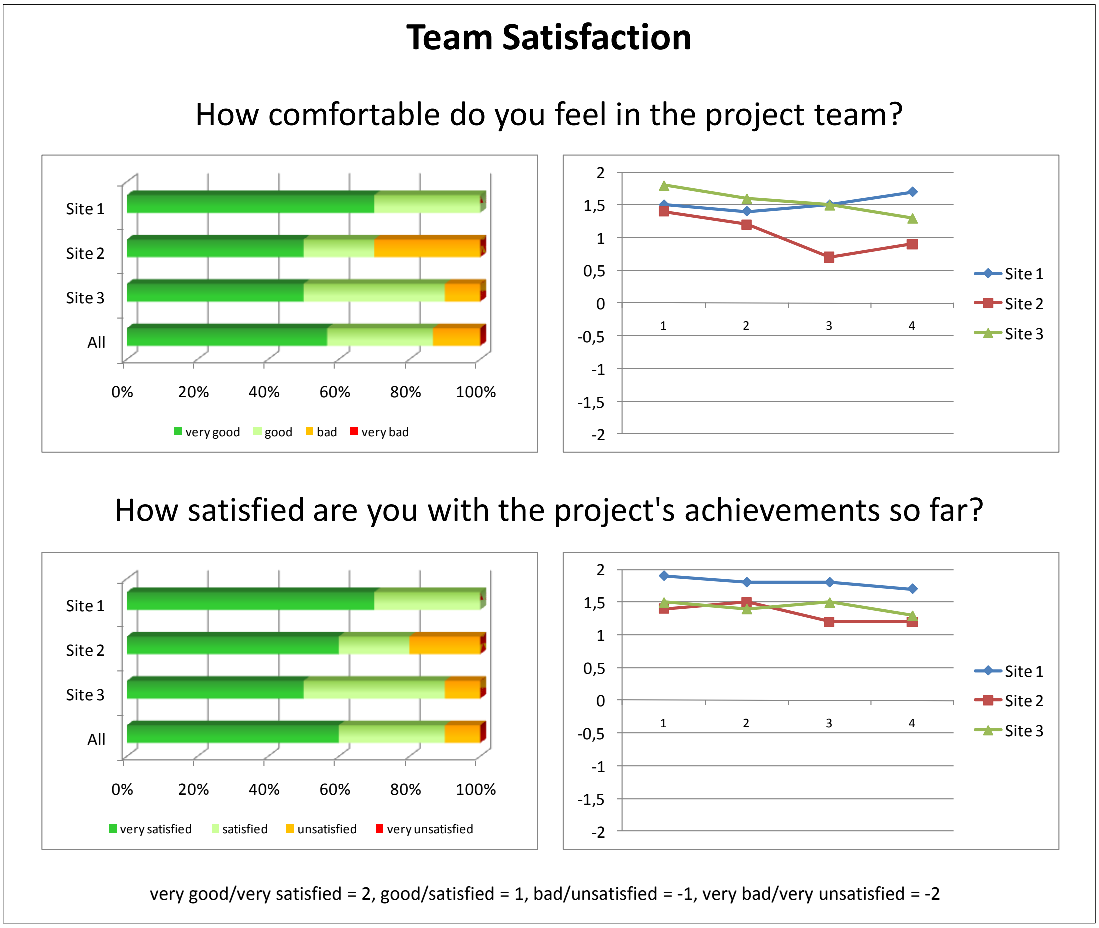


Figure A.10: Indicator: Team Satisfaction

Measurement Goal.
Analyze team satisfaction

Questions.

- Q10.1: How comfortable do team members feel in the project team?
- Q10.2: How satisfied are team members with the project's achievements?

Definitions.
$team\ satisfaction = (avg(CPT), avg(SPA))$

This indicator can also be clustered according to subgroups of the team, i.e. roles or sites

Base Measures.
Comfort in the Project Team (CPT), Satisfaction with Project Achievements (SPA)

Interpretation and Decision Criteria.
The Team Satisfaction indicator can reveal problems in the project team and in work progress. If team members don't feel comfortable in their team, there might be problems in the team that need to be clarified in order to avoid a bad impact on collaboration. Furthermore, the degree of team members' satisfaction with the project's achievements is an indicator for the perceived progress of the project. Therefore watch the level of Team Satisfaction. Low values as well as changes over time (e.g. three consecutive declines) shall be investigated.

Diagram.
See Figure A.10

References.
Kerth [Ker01]

A.4 Base Measures

A.4.1 Communication Frequency (CF)

Base Measure.
Communication Frequency

Identifier.
CF

Reference Question.
Q1.1

Description.
Determines the frequency of communication interactions between groups of team members of the development team in a given time period. The groups of team members can be defined according to roles (e.g. developer, tester, project lead) and/or location of the team members (e.g. development site 1, development site 2).

Value Range.
Positive integer: $0 - n$
Instead of the absolute number of communication interactions, also predefined

clusters can be used, e.g. *daily, every 2–3 days, weekly, every two weeks, less than every two weeks, not at all.* For value aggregation, the clusters are translated to Integer values (e.g. *daily*=20, *every 2–3 days*=8, *weekly*=4, *every two weeks*=2, *less than every two weeks*=1, *not at all*=0)

Data Source.
Questionnaire, observation or existing communication data repositories (e.g. e-mail communication data)

A.4.2 Team Reachability (TR)

Base Measure.
Team Reachability

Identifier.
TR

Reference Question.
Q2.1

Description.
Depicts the reachability between groups of team members of the development team in a given time period. The groups of team members can be defined according to roles (e.g. developer, tester, project lead) and/or location of the team members (e.g. development site 1, development site 2).

Value Range.
Degree of reachability, e.g. *very easy to reach, easy to reach, hard to reach, very hard to reach.* For value aggregation, the degree of reachability is translated to Integer values (e.g. *very easy to reach*=+2, *easy to reach*=+1, *hard to reach*=–1, *very hard to reach*=–2)

Data Source.
Questionnaire, observation or existing communication data repositories (e.g. telephone connection data)

A.4.3 Communication Media Utilization (CMU)

Base Measure.
Communication Media Utilization

Identifier.
CMU

Reference Question.
Q3.1

Description.
Represents the share of utilization of specific communication media by the project team members in a given time period.

Value Range.
Percentage values for specified communication media.

Data Source.
Questionnaire, observation or existing communication data repositories (e.g. e-mail and telephone connection data)

A.4.4 Number of Work Delays (NWD)

Base Measure.
Number of Work Delays

Identifier.
NWD

Reference Question.
Q4.1

Description.
Refers to the number of work delays due to missing information experienced by the project team members in a given time period. This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Positive integer: $0 - n$

Data Source.
Questionnaire, observation or existing communication data repositories (e.g. change management system)

A.4.5 Duration of Work Delays (DWD)

Base Measure.
Duration of Work Delays

Identifier.
DWD

Reference Question.
Q4.2

Description.
Refers to the average length of work delays due to missing information experienced by the project team members in a given time period. This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Positive integer: $0 - n$ days

Data Source.
Questionnaire, observation or existing communication data repositories (e.g. change management system)

A.4.6 Cross-Site Traveling (CST)

Base Measure.
Cross-Site Traveling

Identifier.
CST

Reference Question.
Q5.1

Description.
Depicts if team members have ever visited other development sites.

Value Range.
1 (yes) and 0 (no)

Data Source.
Questionnaire, observation or existing data repositories (e.g. traveling records)

A.4.7 Traveling Duration (TD)

Base Measure.
Traveling Duration

Identifier.
TD

Reference Question.
Q5.2

Description.
Refers to the duration of visits to other sites in the past 12 months.

Value Range.
Positive integer: $0 - n$ days
0, if no visits

Data Source.
Questionnaire, observation or existing data repositories (e.g. traveling records)

A.4.8 Open Communication (OC)

Base Measure.
Open Communication

Identifier.
OC

Reference Question.
Q6.1

Description.
Refers to the team members' evaluation of the statement:
Communication inside the project team is honest and open
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.9 Reliability of Colleagues (RC)

Base Measure.
Reliability of Colleagues

Identifier.
RC

Reference Question.
Q6.2

Description.
Refers to the team members' evaluation of the statement:
I can rely on my colleagues
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.10 Willingness to Help (WH)

Base Measure.
Willingness to Help

Identifier.
WH

Reference Question.
Q6.3

Description.
Refers to the team members' evaluation of the statement:
My colleagues assist me with heavy workloads, beyond what they are required to do
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.11 Valuation of New Ideas (VNI)

Base Measure.
Valuation of New Ideas

Identifier.
VNI

Reference Question.
Q6.4

Description.
Refers to the team members' evaluation of the statement:
New ideas of colleagues from the local/distant site are seen as valuable
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.12 Constructive Meetings (CM)

Base Measure.
Constructive Meetings

Identifier.
CM

Reference Question.
Q6.5

Description.
Refers to the team members' evaluation of the statement:
(Virtual) meetings are constructive
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.13 Team Competition (TC)

Base Measure.
Team Competition

Identifier.
TC

Reference Question.
Q6.6

Description.
Refers to the team members' evaluation of the statement:
There is individual competition between sub-project teams
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values (note that the question is using a negative statement): *fully agree*=–2, *tend to agree*=–1, *tend to disagree*=+1, *fully disagree*=+2

Data Source.
Questionnaire or observation

A.4.14 Clear Responsibilities (CR)

Base Measure.
Clear Responsibilities

Identifier.
CR

Reference Question.
Q6.7

Description.
Refers to the team members' evaluation of the statement:
There are many discussion over who has a particular responsibility
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values (note that the question is using a negative statement): *fully agree*=–2, *tend to agree*=–1, *tend to disagree*=+1, *fully disagree*=+2
disagree=+1, *fully disagree*=+2)

Data Source.
Questionnaire or observation

A.4.15 Ability to Change the Situation (ACS)

Base Measure.
Ability to Change the Situation

Identifier.
ACS

Reference Question.
Q6.8

Description.
Refers to the team members' evaluation of the statement:
I feel powerless to change the project's situation
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values (note that the question is using a negative statement): *fully agree*=–2, *tend to agree*=–1, *tend to disagree*=+1, *fully disagree*=+2

Data Source.
Questionnaire or observation

A.4.16 Non-Work Conversation (NWC)

Base Measure.
Non-Work Conversation

Identifier.
NWC

Reference Question.
Q7.1

Description.
Refers to the team members' evaluation of the statement:
I discuss non-work related matters with other project team members
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.17 Informal Communication (IC)

Base Measure.
Informal Communication

Identifier.
IC

Reference Question.
Q7.2

Description.
Refers to the team members' evaluation of the statement:
I often get useful work-related information through informal communication (e.g. hallway conversations)
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.18 Current Awareness (CA)

Base Measure.
Current Awareness

Identifier.
CA

Reference Question.
Q8.1

Description.
Depicts the awareness of current work activities between groups of team members of the development team in a given time period. The groups of team members can be defined according to roles (e.g. developer, tester, project lead) and/or location of the team members (e.g. development site 1, development site 2).

Value Range.
Degree of awareness, e.g. *very aware, aware, unaware, very unaware.* For value aggregation, the degree of awareness is translated to Integer values (e.g. *very aware*=+2, *aware*=+1, *unaware*=–1, *very unaware*=–2)

Data Source.
Questionnaire or observation

A.4.19 General Awareness (GA)

Base Measure.
General Awareness

Identifier.
GA

Reference Question.
Q8.2

Description.
Refers to the team members' evaluation of the statement:
I nearly always know who to contact or who has a specific kind of expertise
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values:
fully agree=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.20 Importance of Understanding Daily Work (IUDW)

Base Measure.
Importance of Understanding Daily Work

Identifier.
IUDW

Reference Question.
Q8.3

Description.
Refers to the team members' evaluation of the statement:
It is important for me to understand the day-to-day work of my colleagues
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values:
fully agree=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.21 Clear Tasks (CT)

Base Measure.
Clear Tasks

Identifier.
CT

Reference Question.
Q8.4

Description.
Refers to the team members' evaluation of the statement:
When work is assigned, everyone is clear about his or her task
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.22 Information About Changes (IAC)

Base Measure.
Information About Changes

Identifier.
IAC

Reference Question.
Q8.5

Description.
Refers to the team members' evaluation of the statement:
My colleagues provide timely information about changes in current plans
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values:

fully agree=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.23 English Language Skills (ELS)

Base Measure.
English Language Skills

Identifier.
ELS

Reference Question.
Q9.1

Description.
Refers to the team members' evaluation of the statement:
Project team members have appropriate English language skills
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.24 National Culture Impact (NCI)

Base Measure.
National Culture Impact

Identifier.
NCI

Reference Question.
Q9.2

Description.
Refers to the team members' evaluation of the statement:
National culture has significant influence on team collaboration
This base measure can be elicited along a specific partitioning of the project team, for example with respect to roles (e.g. developer, tester, project lead) and/or locations (e.g. development site 1, development site 2).

Value Range.
Degree of confirmation: *fully agree, tend to agree, tend to disagree, fully disagree*
For value aggregation, the degree of confirmation is translated to Integer values: *fully agree*=+2, *tend to agree*=+1, *tend to disagree*=–1, *fully disagree*=–2)

Data Source.
Questionnaire or observation

A.4.25 Comfort in the Project Team (CPT)

Base Measure.
Comfort in the Project Team

Identifier.
CPT

Reference Question.
Q10.1

Description.
Refers to the team members' answers to the question:
How comfortable do you feel in the project team?

Value Range.
Degree of comfort: *very good, good, bad, very bad*
For value aggregation, the degree of comfort is translated to Integer values: *very good*=+2, *good*=+1, *bad*=–1, *very bad*=–2)

Data Source.
Questionnaire or observation

A.4.26 Satisfaction with Project Achievements (SPA)

Base Measure.
Satisfaction with Project Achievements

Identifier.
SPA

Reference Question.
Q10.2

Description.
Refers to the team members' answers to the question:
How satisfied are you with the project's achievements so far?

Value Range.
Degree of satisfaction: *very satisfied, satisfied, unsatisfied, very unsatisfied*
For value aggregation, the degree of satisfaction is translated to Integer values:
very satisfied=+2, *satisfied*=+1, *unsatisfied*=–1, *very unsatisfied*=–2)

Data Source.
Questionnaire or observation

Appendix B

Industrial Case Study: Questionnaire

The following pages show the questionnaire as it was used in the industrial case study (see Chapter 8).

INSTITUT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Survey on Distributed Collaboration

This questionnaire is part of a project study on distributed collaboration, conducted in cooperation with Technische Universität München.

Motivation

Many companies today conduct development projects in globally distributed teams. Geographic separation, different time zones and cultural differences have brought new challenges to development projects. Key challenges are collaboration and communication over distance. Miscommunication is one major reason for issues such as cost overruns, quality problems and project delays.

Research Objectives

Purpose of this research is to learn more about communication and collaboration in distributed projects, analyzing hypotheses such as:
- There is less frequent communication across sites than within a site
- Distributed collaboration introduces delays
- People at different sites are less aware of the tasks other team members are currently working on
- Distributed social networks are significantly smaller than same-site social networks

Results will be generalized to identify early warning indicators and metrics to improve project management in distributed projects. While classical measurement systems typically focus on cost, time and quality aspects, which are rather the “symptoms” that the “root-causes”, communication metrics and social network data are promising to help the project team detecting problems early on *before* they have serious impact.

Confidentiality of Survey Data

- Data will be used for research purposes and to provide feedback to the project team.
- Participation in this survey is voluntary. You may skip questions that you feel uncomfortable with.
- Confidentiality of your responses will be strictly protected. No identifiable individual data will be used in reports or publications. Your responses will be used to learn more about communication in distributed projects, but NOT to evaluate your personal performance.

Results of the Study

You will receive a report with the results of the study.

Contact

In case of questions, please contact:
- Christian Lescher, lescher@in.tum.de or christian.lescher@siemens.com
- [redacted]

Survey on Distributed Collaboration

This questionnaire is part of a project study on distributed collaboration, conducted in cooperation with Christian Lescher, who is a Siemens employee and currently working on his Dissertation at TU München. Data will be used for research purposes and to provide feedback to the project team.

- You will receive a report with the results of the study.
- Participation in this survey is voluntary. You may skip questions that you feel uncomfortable with.
- Confidentiality of your responses will be strictly protected. No identifiable individual data will be used in reports or publications. Your responses will be used to learn more about communication in distributed projects, but NOT to evaluate your personal performance.

Please use the return envelope to send the survey answers to: Christian Lescher, [redacted]

My Role:

- ○ Developer / Architect
- ○ Project Lead / Sub-Project Lead / Quality Assurance
- ○ Tester / System Integration / Configuration Manager

Please evaluate for your **local site ([redacted])** and the **distant sites ([redacted])** you are collaborating with:

Note: If you are not collaborating with other sites, you can leave the column "distant site" empty.

Communication and Awareness	**Local Site** fully agree	tend to agree	tend to disagree	fully disagree	**Distant Site** fully agree	tend to agree	tend to disagree	fully disagree
1. People I need to communicate with are difficult to reach	○	○	○	○	○	○	○	○
2. I discuss non-work related matters with other project team members	○	○	○	○	○	○	○	○
3. I nearly always know who to contact or who has a specific kind of expertise	○	○	○	○	○	○	○	○
4. It is important for me to understand the day-to-day work of my colleagues	○	○	○	○	○	○	○	○
5. I have sufficient insight into what my colleagues are currently working on	○	○	○	○	○	○	○	○
6. I often get useful work-related information through informal communication (e.g. hallway conversations)	○	○	○	○	○	○	○	○
7. Communication inside the project team is honest and open	○	○	○	○	○	○	○	○
8. Project team members have appropriate English language skills	○	○	○	○	○	○	○	○

Comments:

Team Collaboration	Local Site: fully agree	Local Site: tend to agree	Local Site: tend to disagree	Local Site: fully disagree	Distant Site: fully agree	Distant Site: tend to agree	Distant Site: tend to disagree	Distant Site: fully disagree
9. I can rely on my colleagues	○	○	○	○	○	○	○	○
10. My colleagues assist me with heavy workloads, beyond what they are required to do	○	○	○	○	○	○	○	○
11. New ideas of colleagues from the local/distant site are seen as valuable	○	○	○	○	○	○	○	○
12. Meetings/phone conferences are constructive	○	○	○	○	○	○	○	○
13. When work is assigned, everyone is clear about his or her task	○	○	○	○	○	○	○	○
14. My colleagues provide timely information about changes in current plans	○	○	○	○	○	○	○	○
15. There is individual competition between sub-project teams	○	○	○	○	○	○	○	○
16. Diversity in the project team (e.g. national cultures) is valuable for team collaboration	○	○	○	○	○	○	○	○

Management and Strategy	fully agree	tend to agree	tend to disagree	fully disagree
17. Management has a clear strategy towards the cooperation with the supplier company	○	○	○	○
18. The challenges of distributed collaboration were underestimated when the project was initiated	○	○	○	○
19. Requirements are clearly defined and communicated	○	○	○	○
20. There are many discussions over who has a particular responsibility	○	○	○	○
21. Project successes are adequately appreciated by management	○	○	○	○
22. Quality problems is one of the major root-causes for project delay	○	○	○	○
23. I feel powerless to change the project's situation	○	○	○	○

Comments:

Work Delays	Local Site	Distant Site
24. How many times in the past month was your own work delayed because you needed information, discussion, or a decision, from someone at your site or another site?	times	times
25. What was the average length of the delays you experienced before acquiring the needed information, having the discussion, or being informed of the decision by the person from your site or the other site?		

Comments:

Communication between Roles

26. Please indicate for roles you are communicating with to get your work done for the project:
- How often do you communicate with each role?
- Is it easy for you to reach colleagues of this role?
- Are you aware of the current set of tasks that colleagues of this role are working on?

Role	daily	at least once a week	at least once a month	Easy to reach?	I know what they are currently working on
Developer / Architect	○	○	○	○ yes ○ no	○ yes ○ no
Project Lead / Sub-Project Lead / Quality Assurance	○	○	○	○ yes ○ no	○ yes ○ no
Tester / System Integration / Configuration Manager	○	○	○	○ yes ○ no	○ yes ○ no

27. Where/between which roles do you see communication problems?

28. Please indicate how many percent of your communication is done via the following media:

Communication media	How many percent?
E-Mail	%
Phone	%
MS Communicator/Live Meeting	%
Personal contact	%
Other, please specify:	%

29. Have you ever visited other development sites?
In the last 12 months, how many days have you spent at other sites?

Ever visited other sites?	Days of visits in last 12 months:
○ yes ○ no	

Comments:

30. How well do I feel in the project team?

○ very bad ○ bad ○ good ○ very good

31. How satisfied am I with the project's achievements so far?

○ very unsatisfied ○ unsatisfied ○ satisfied ○ very satisfied

32. Additional Comments:

Appendix C

Industrial Case Study: Detailed Results

In the following the detailed results of the industrial case study (see Chapter 8) are described. Please note that the question numbers (e.g. Q25) refer to the numbering in the survey questionnaire (see Appendix B).

C.1 Detailed Results of the Survey

The survey was distributed as paper questionnaire to 28 team members of the German site. In total, 24 persons answered the survey which equates to a response rate of 86%. The survey therefore is considered representative for the team at the German site. Figure C.1 shows the distribution of roles of the respondents: 9 answers (37.5%) were from the roles *developer and architect*, 10 answers (41.7%) from the group *project lead, sub-project lead, and quality manager* and 3 answers (12.5%) from the roles *tester, system integrator, and configuration manager*. In 2 of the returned questionnaires (8.3%) the role was not specified. For ease of reading, in the following we abbreviate the 3 groups of roles with *Developer*, *Project Lead* and *Tester*.

C.1.1 Team Cohesion

Multiple questions of the survey were concerned with team cohesion. In the following we describe the detailed survey results, followed by a summary of conclusions.

Survey Results

Figure C.2,C.3 and C.4 show the results of the questions *I can rely on my colleagues* (Q09), *New ideas of colleagues from the local/distant site are seen as valuable* (Q11) and *Meetings/phone conferences are constructive* (Q12). 29% of

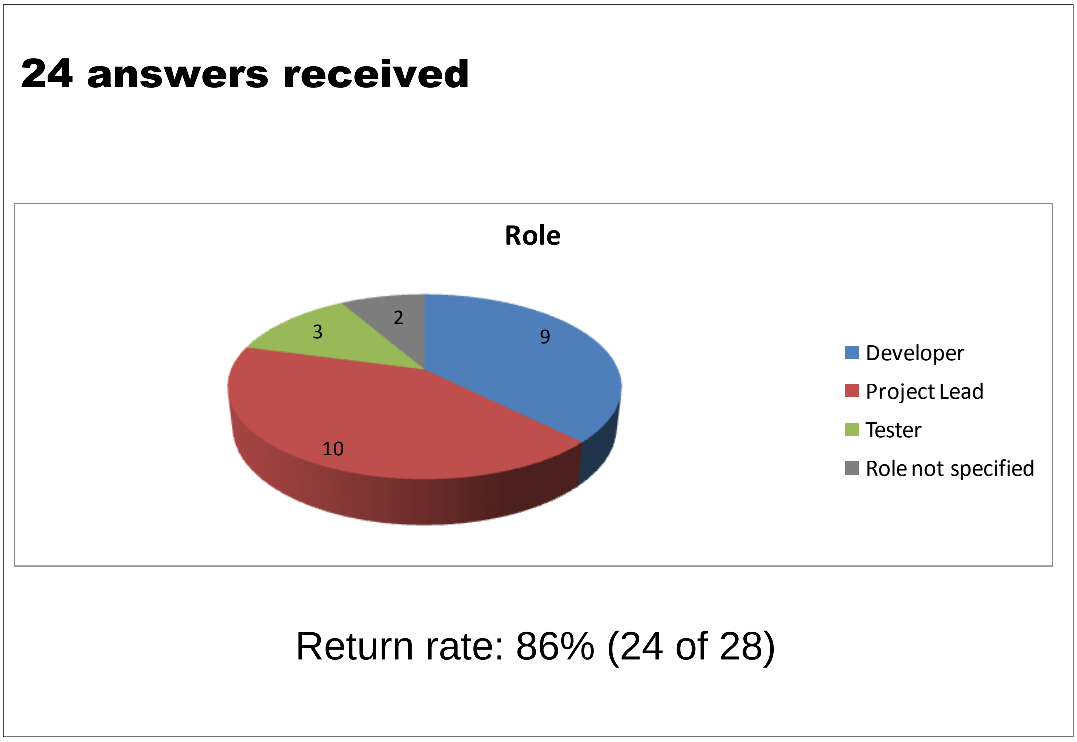


Figure C.1: Roles of respondents and return rate

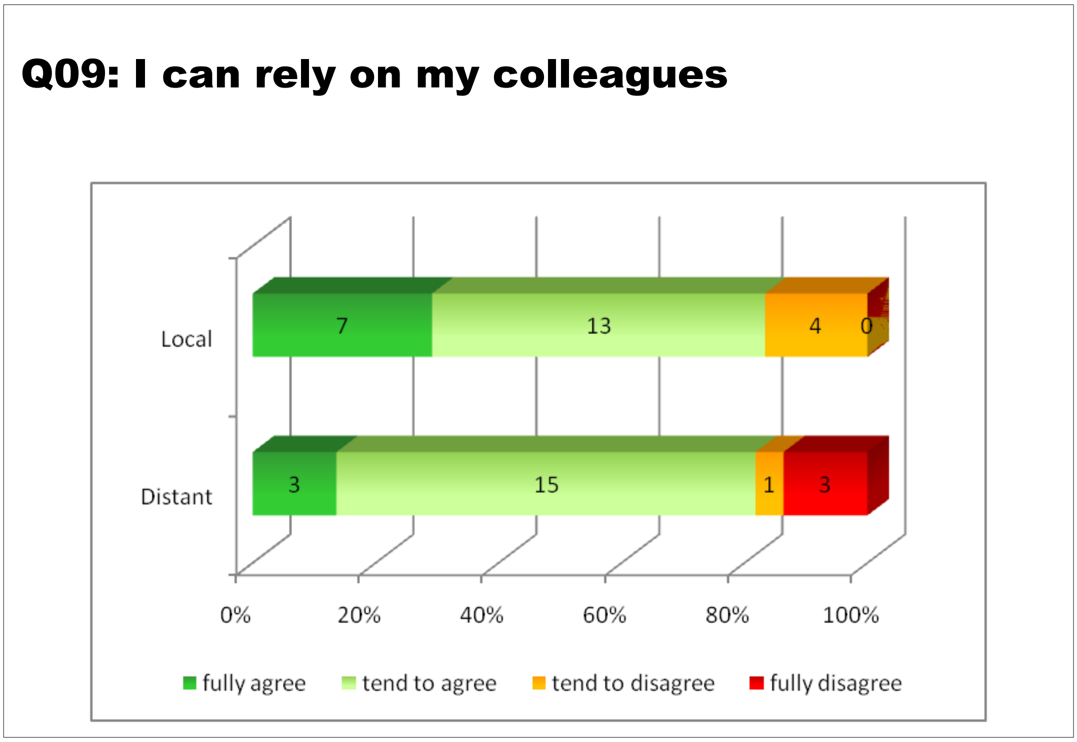


Figure C.2: Question Q09

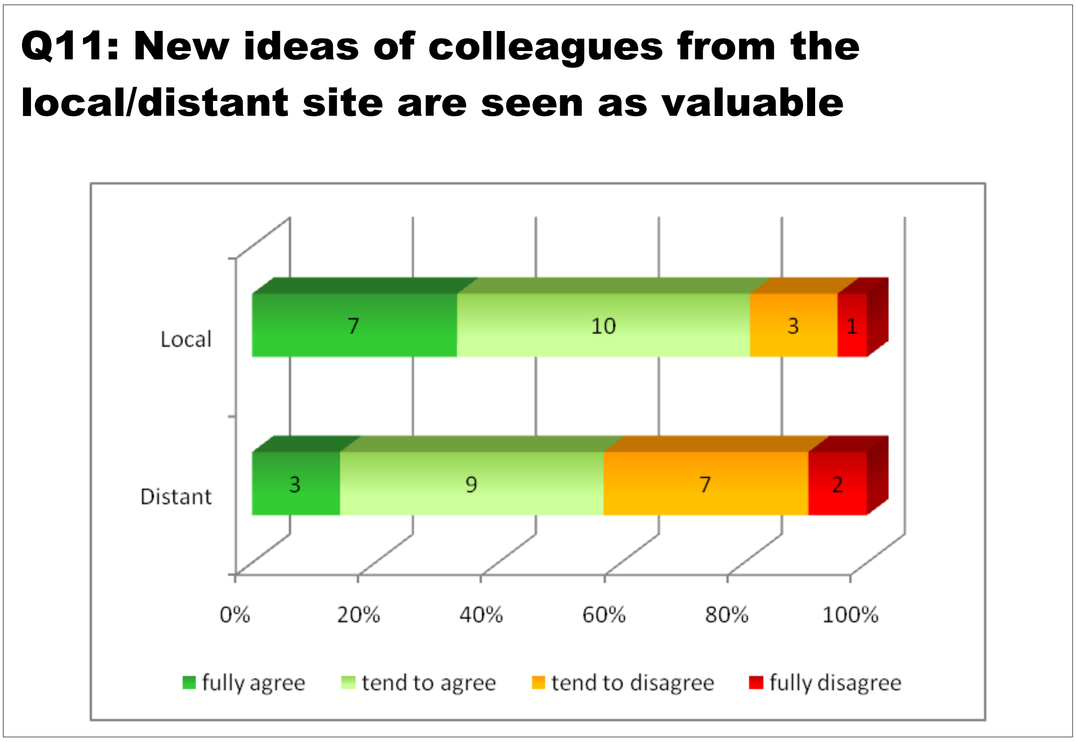


Figure C.3: Question Q11

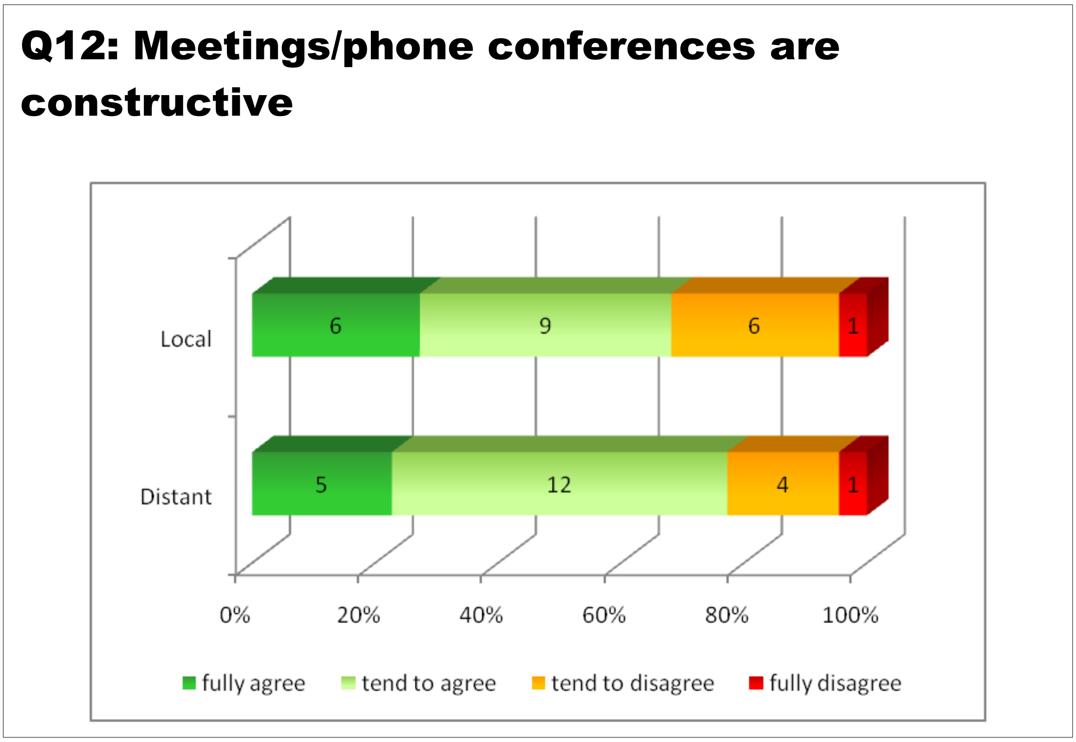


Figure C.4: Question Q12

respondents stated with respect to the local site (i.e. the German site) they fully agree that they can rely on their colleagues, 54% answered that they tend to agree; together this amounts to 83% of respondents. The remaining 17% responded that they tend to disagree with the statement. With respect to the distant site (i.e. the supplier site in Central Eastern Europe), 14% answered that they fully agree, 68% that they tend to agree with the statement *I can rely on my colleagues*; together 82% of respondents. The remaining answers are distributed as follows: 5% replied 'tend to disagree', while 14% answered 'fully disagree'. It results from the answers to question Q09 that colleagues from the distant site are seen as less reliable compared to colleagues from the local site, which is explainable as the usual effect of distribution and impact on trust; the effect might be emphasized in case of the project study through the fact that the supplier is a competitor company. However, in summary reliability of colleagues seems not to be an issue. The answers to question Q11 give a similar picture: While 33% of respondents fully agree to the statement that *new ideas of colleagues from the* <u>*local*</u> *site are seen as valuable*, and 48% tend to agree to the same statement, which together amounts to 81%, only 14% fully agree that new ideas of colleagues from the <u>distant</u> site are seen as valuable and 43% tend to agree, which sums up to 57%. Again the overall result is positive, with the clear tendency that the local site is seen better than the distant site. Also meetings/phone conferences are overall seen as being constructive (Q12): With respect to the local site, 68% answered 'fully agree' or 'tend to agree' (27% fully agree, 41% tend to agree). With relation to the distant site, even 77% stated that they fully agree or tend to agree with the statement *Meetings/phone conferences are constructive*. Against the effect of ingroup favoritism of the local site, the distant site received a better evaluation, which might hint at problems with meetings/phone conferences being constructive at the German site. Again, the overall situation depicted in question Q12 is rather positive.

Questions Q07, Q10, Q15 and Q20 revealed more serious problems with team cohesion; Figures C.5 and C.6 illustrate the detailed results of Q07 and Q10. Being asked if *communication inside the project team is honest and open* (Q07), with respect to the local site only 43% answered with fully agree or tend to agree (17% fully agree, 26% tend to agree) and 57% with tend to disagree or fully disagree (43% tend to disagree, 13% fully disagree). With respect to the distant site, 36% expressed their agreement (5% fully agree, 32% tend to agree) and 64% showed disagreement (41% tend to disagree, 23% fully disagree). While the effect in the local and distant figures is again explainable by the effects of distribution, the overall figures indicate a clear problem with communication not being honest and open. In question Q10 we asked for the evaluation of the sentence *my colleagues assist me with heavy workloads, beyond what they are required to do*. With respect to the local (German) site, 43% of the answers were positive (4% fully agree, 39% tend to agree), while 57% were negative (48% tend to disagree, 9% fully disagree). For the distant (Central Eastern Europe) site,

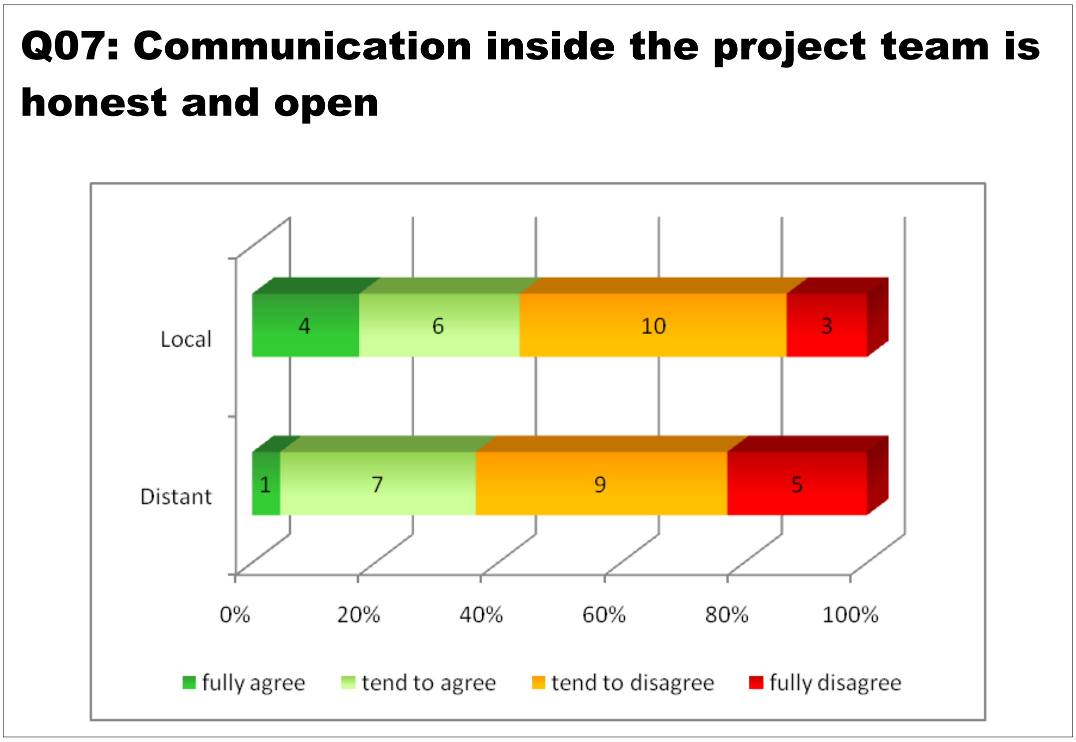


Figure C.5: Question Q07

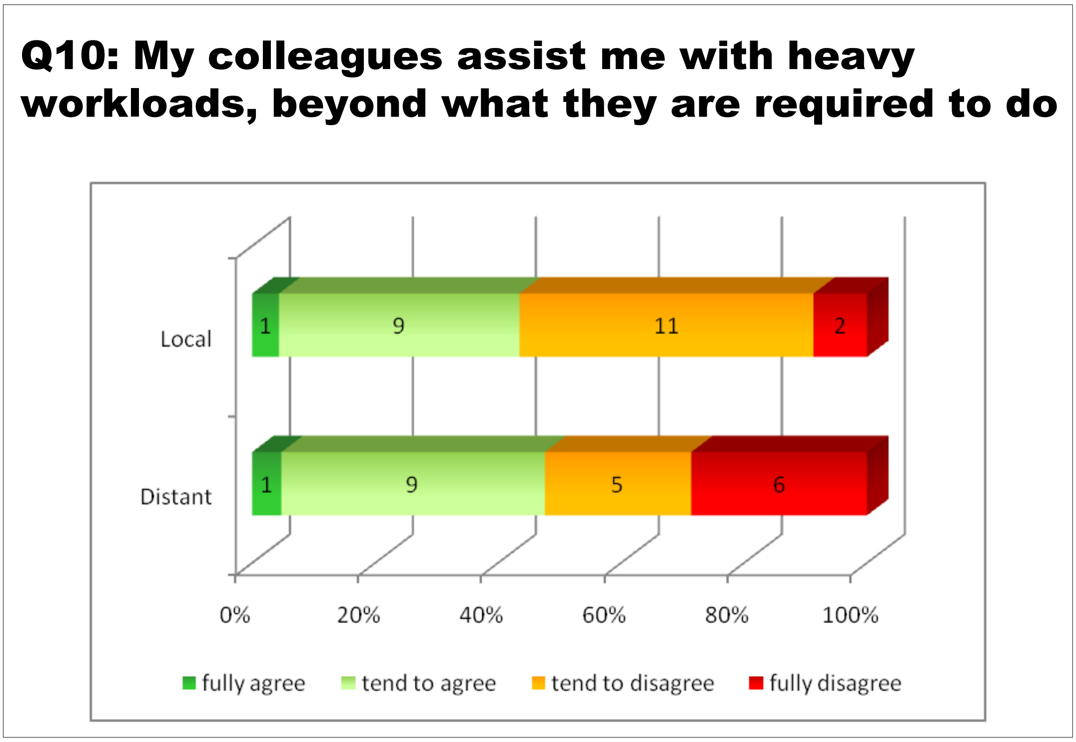


Figure C.6: Question Q10

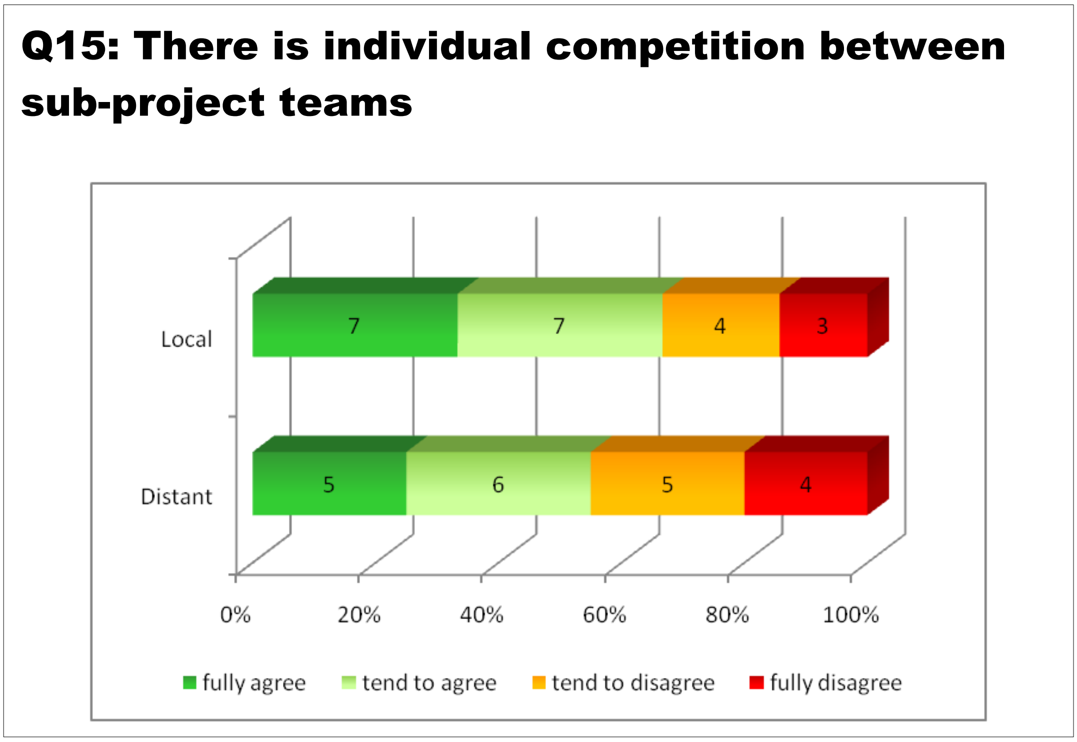


Figure C.7: Question Q15

48% of the answers were positive (5% fully agree, 43% tend to agree) and 52% negative (24% tend to disagree, 29% fully disagree). In summary, only few people fully agreed with the statement that colleagues assist them with heavy workloads beyond what they are required to do, while there is a clear disagreement with the statement (strongest for the distant site). The low values of readiness to help out colleagues indicate a serious problem in team cohesion.

The purpose of Q15 was to identify competition between sub-project teams of Project X. Figure C.7 shows the results. According to 67% of the respondents, there is individual competition between sub-project teams with respect to the local site (33% fully agree, 33% tend to agree), only 33% disagreed with the statement (19% tend to disagree, 14% fully disagree). Related to the distant site, 55% of respondents answered with fully agree or tend to agree (25% fully agree, 30% tend to agree), while 45% didn't agree (25% tend to disagree, 20% fully disagree). The confirmation of individual competition is very high (clear majority), which hints at a serious problem inside the project team. In addition, it is striking that there seems to be much more competition between sub-project teams at the local site compared to the distant site. Figure C.8 and C.9 show a breakdown of results by roles who gave the answers. As we can see from these figures, the group Project Lead shows significantly higher agreement than the other two groups: With respect to the local site, 75% of the group Project Lead

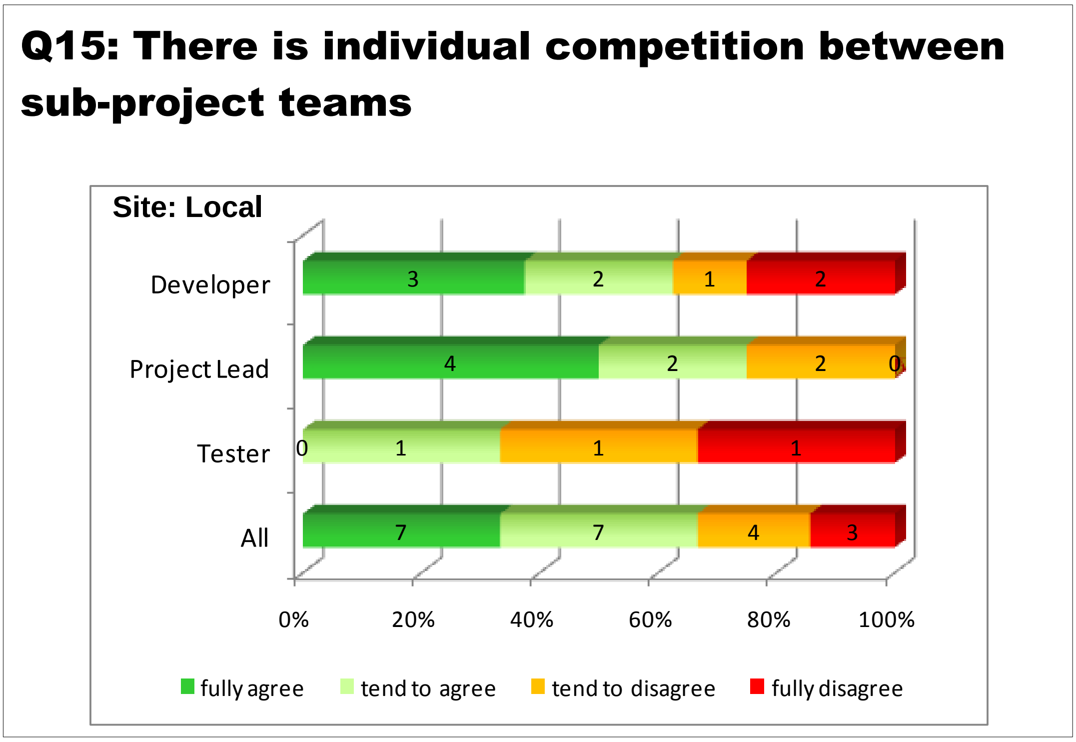


Figure C.8: Question Q15 per role, local

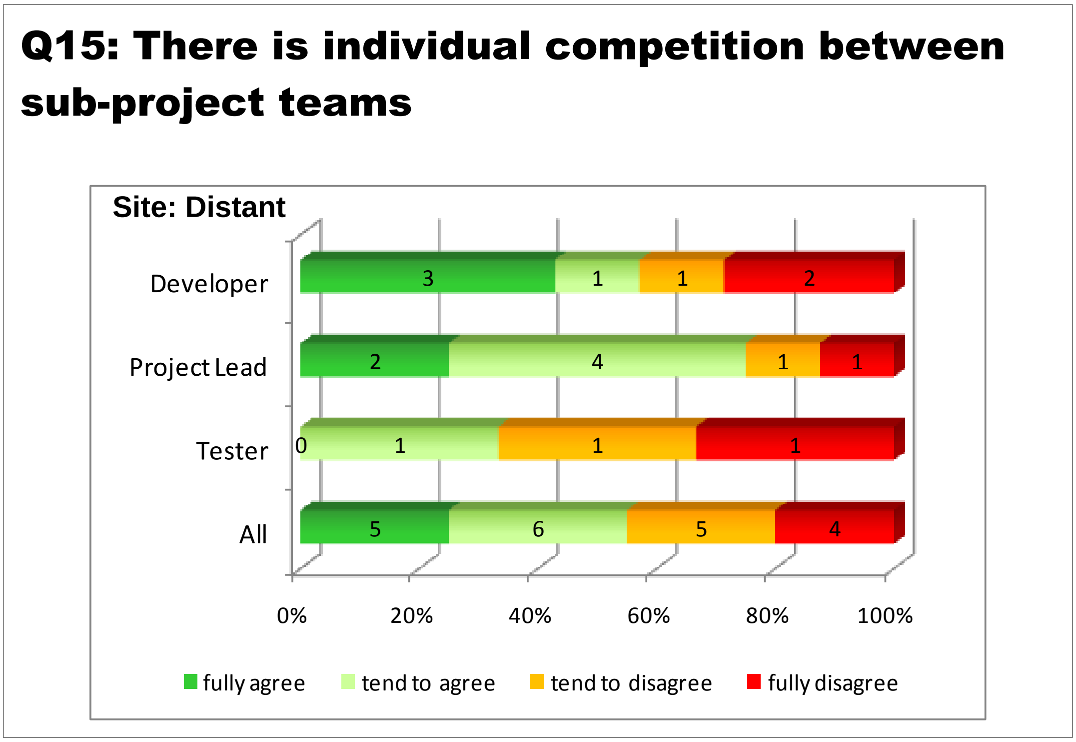


Figure C.9: Question Q15 per role, distant

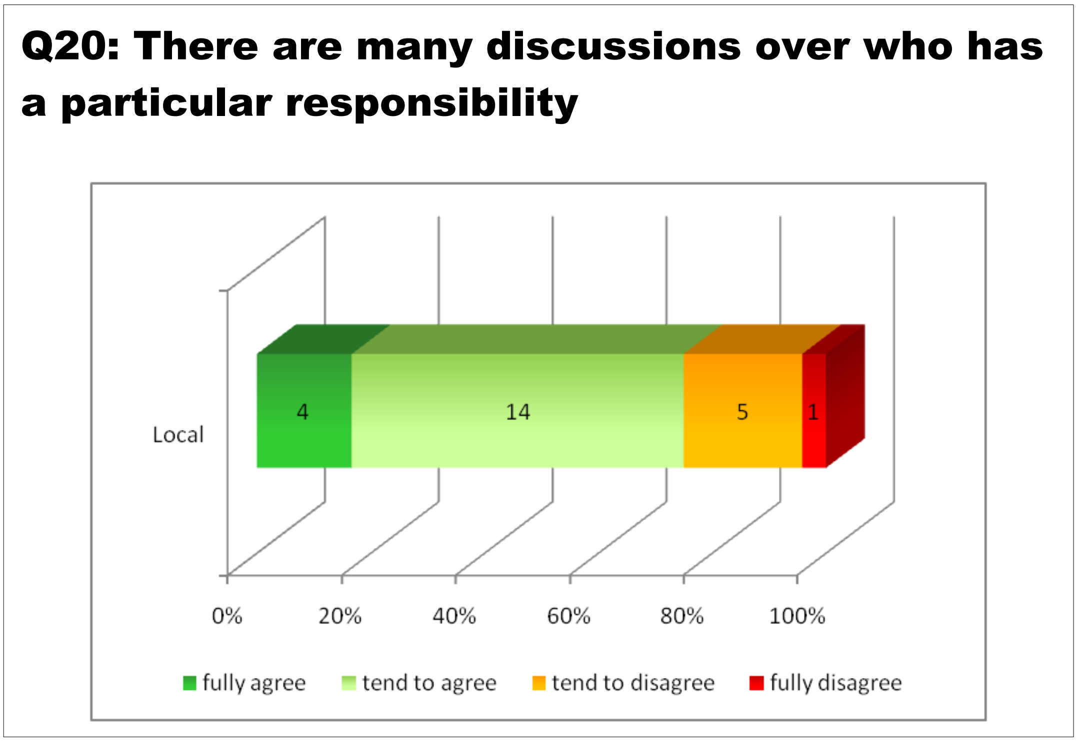


Figure C.10: Question Q20

answered fully agree or tend to agree (50% fully agree, 25% tend to agree), but only 25% tend to disagree and 0% fully disagree. For the distant site, still the group Project Lead expressed an agreement of 75% (25% fully agree, 50% tend to agree), i.e. the effect is strongest in case of the local site. Based on the aforementioned answers to question Q15, we conclude that there is individual competition in particular between project lead/sub-project leads.

A similar picture is given by the questions *There are many discussions over who has a particular responsibility* (Q20). Respondents showed a strong agreement of 75% with this statement (17% fully agree, 58% tend to agree), while only 25% disagreed (21% tend to disagree, 4% fully disagree), as shown in Figure C.10. Analyzing the breakdown of roles who gave the answers, again the group Project Lead has the highest values of agreement: 90% agreement (30% fully agree, 60% tend to agree), 10% disagreement (10% tend to disagree, 0% fully disagree) (see Figure C.11).

Comments from the survey confirm this view:

> "The problem is that even on local side the team is split. Colleagues which don't have knowledge about technologies used for [Project X] try to prove that they are anyway useful while pointing out what the others are doing wrong. Small problems are presented as big, very big issues!!! They are playing politics in the hope that the managers

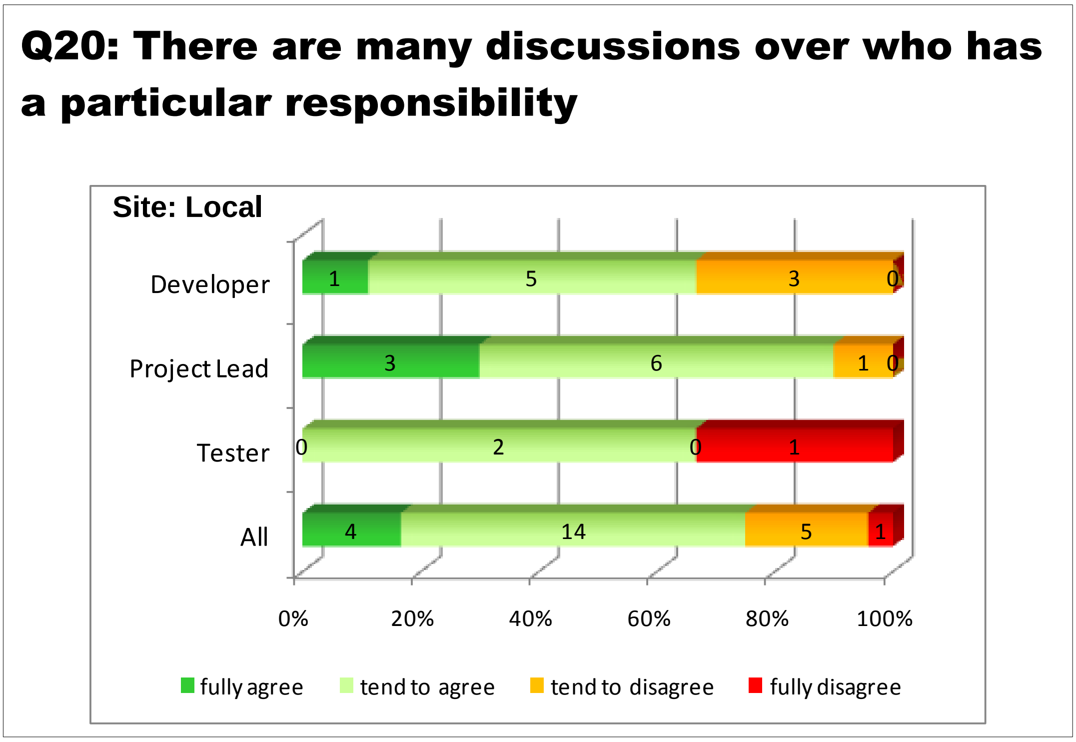


Figure C.11: Question Q20 per role, local

> will not realize what is really going on. HOW SHOULD BE THE COMMUNICATION HONEST?"

Also problems with the supplier company, similar to those mentioned in the interviews (see Section 8.2) were mentioned in the comments:

> "The problem with the distant site is that the company we are working with is not serious and not professional, there is a big lack of knowledge and experience, and again politics is used to hide it. They are presenting themselves as experts, but expect us to teach them ON PHONE basics about data processing and power systems, but managers stick to the way how they are presenting themselves."

And another comment related to team cohesion:

> "I my own team I can rely 100% on my colleagues (with everything), my team means all colleagues below one sub-PM. Between other teams there is a competition. New ideas are not appreciated at all, they are understood as an attack on old establishment. Meetings are time consuming and not intended to solve any problems (it looks like that), they are used as presentation platform for the participants."

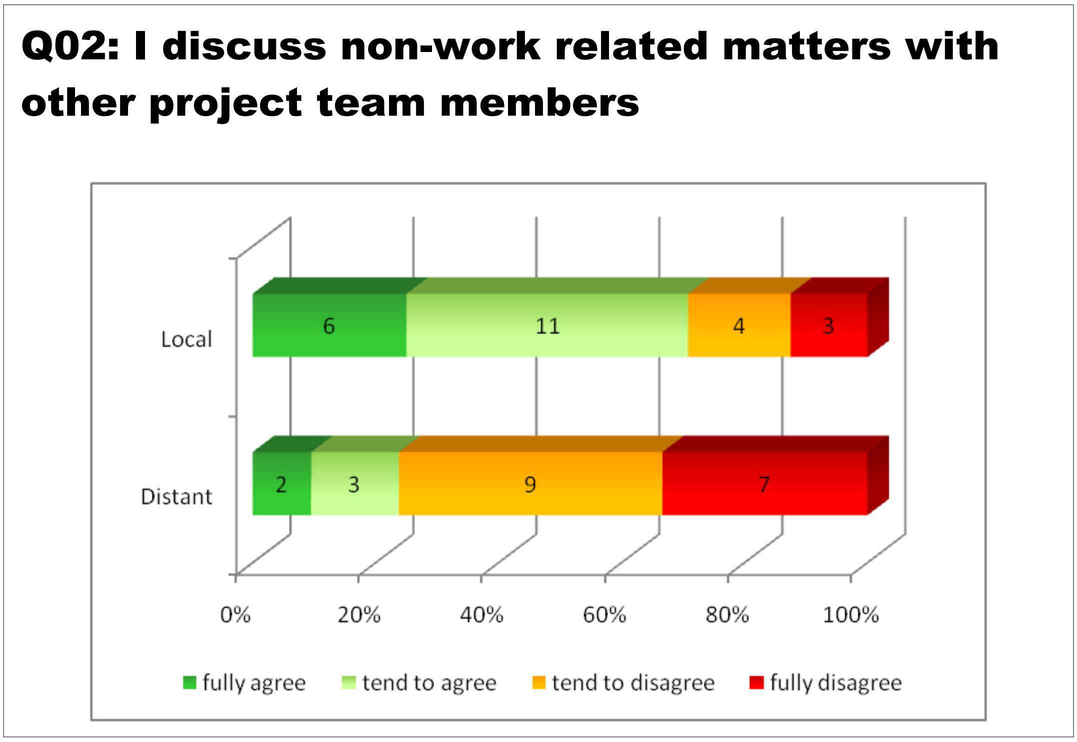


Figure C.12: Question Q02

Conclusions

With respect to the aforementioned questions on team cohesion (Q07, Q09, Q10, Q11, Q12, Q15 and Q20), we draw the following conclusions:

- The overall team situation is not too bad (*I can rely on my colleagues, new ideas are seen as valuable, meetings are constructive*)
- Major problems are:
  - No open and honest communication
  - Individual competition between sub-project teams
  - In particular, the local site seems to be affected

C.1.2 Informal Communication

Informal communication was investigated in question Q02 and Q06.

Survey Results

Figure C.12 and C.13 show the results of question Q02 and Q06 related to informal communication. Q02 was *I discuss non-work related matters with other*

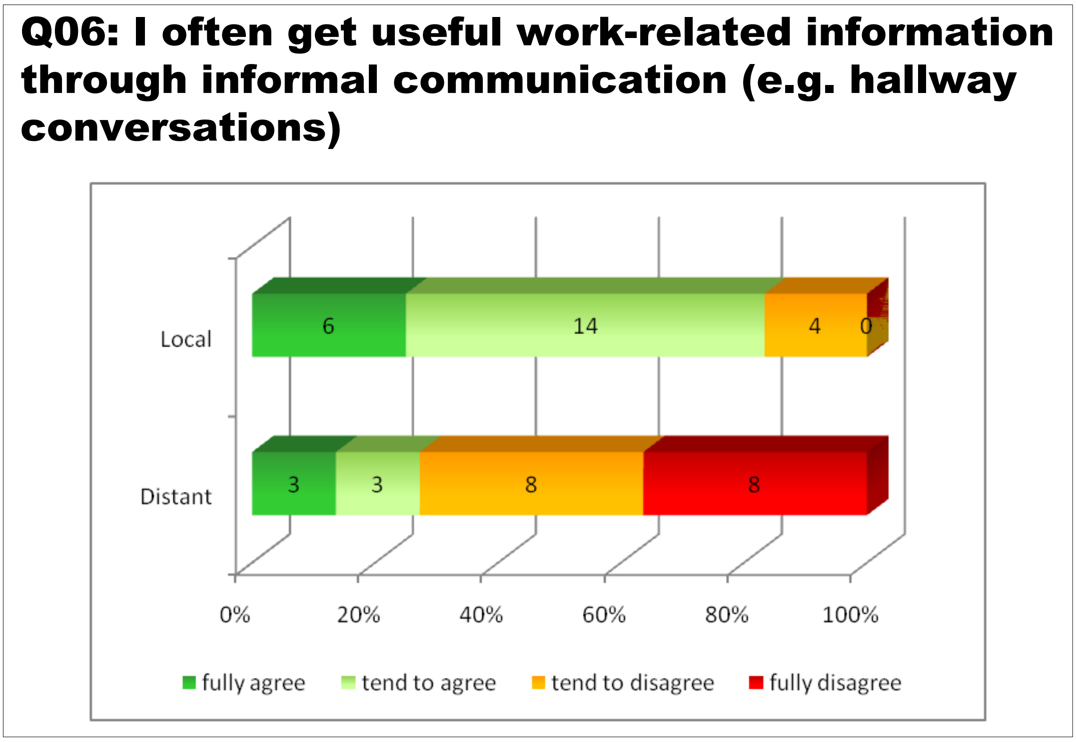


Figure C.13: Question Q06

project team members. The answers conform with the expected behavior: Being asked about the local site, 71% of respondents show agreement with the statement (25% fully agree, 46% tend to agree), 29% show disagreement (17% tend to disagree, 13% fully disagree). For the distant site, the result is reversed: only 24% expressed agreement (10% fully agree, 14% tend to agree) and 76% disagreement (43% tend to disagree, 33% fully disagree). The results of question Q06 were comparable. The statement *I often get useful work-related information through informal communication (e.g. hallway conversations)* was evaluated with respect to the local site as follows: 83% answered fully agree or tend to agree (25% fully agree, 58% tend to agree), while 17% answered tend to disagree, and 0% fully disagree. With respect to the distant site, the results were: 28% agreement (14% fully agree, 14% tend to agree) and 73% disagreement (36% tend to disagree, 36% fully disagree).

Conclusions

There was a strong statement that useful work-related information through informal communication is gathered much more locally than distant. This conforms to the expected behavior in GSE.

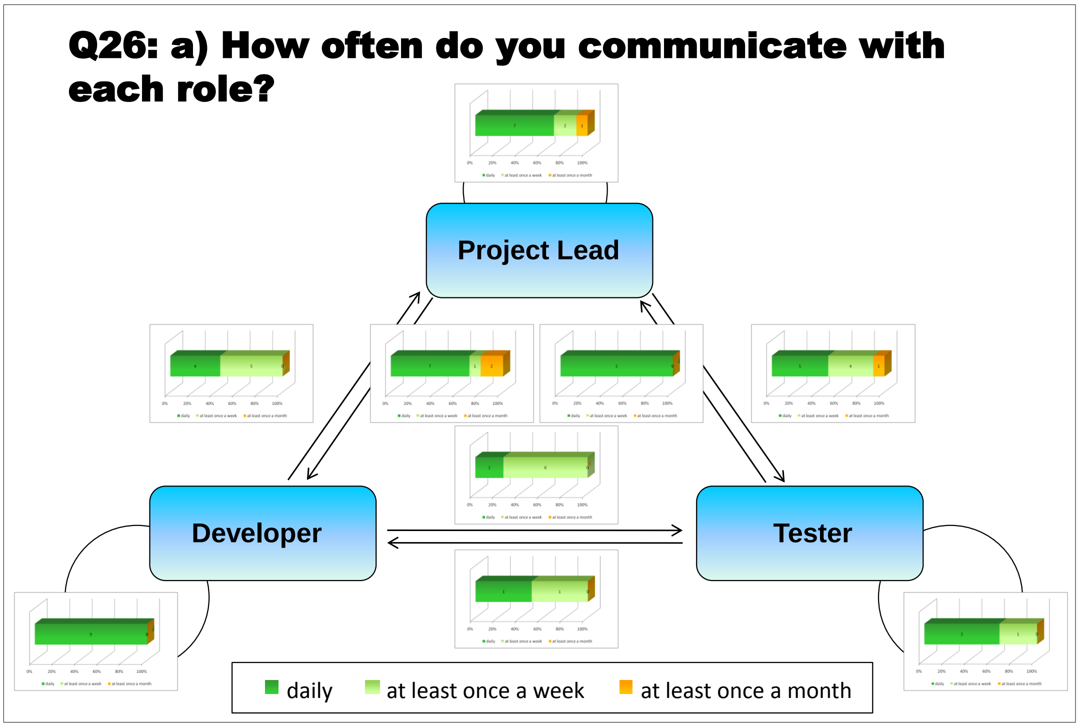


Figure C.14: Question Q26a

C.1.3 Communication Frequency

Due to the constraints explained above with respect to collecting communication data, the social network analysis contained in the questionnaire only distinguished 3 groups of roles instead of individual team members. Therefore the possibilities to analyze the data are also very limited. However, it turned out that even this highly aggregated data is still very valuable.

Survey Results

Figure C.14 shows the results of question *How often do you communicate with each role?* (Q26a) The arrows in the figure illustrate the direction of communication, e.g. the arrow from Developer towards Project Lead represents what the Developers stated about their communication with the role Project Lead. The circles at the roles refer to communication with colleagues of the same group of roles. As we can see from the figure, most respondents stated that they communicate *daily* or *at least once a week*.

Conclusions

A striking result is that some people of the group Project Lead answered that they communicate *at least once a month* only, which seems insufficient. This is valid

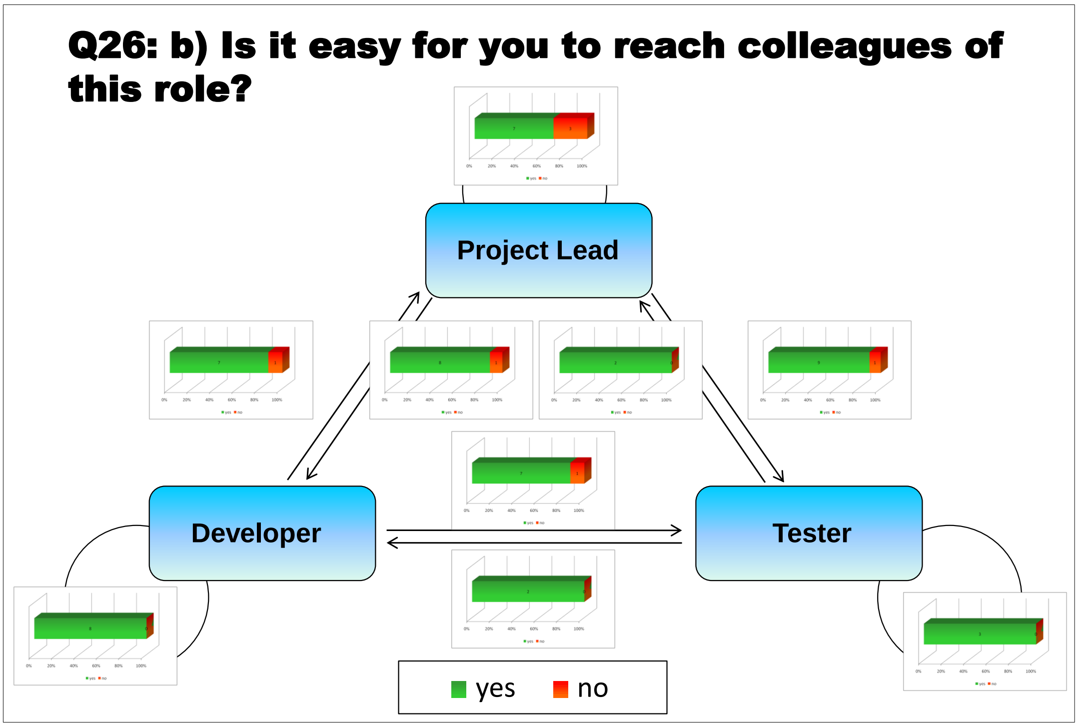


Figure C.15: Question Q26b

for all directions in which people of the group Project Lead are communicating: towards Developer, Tester as well as to colleagues of the same role.

C.1.4 Reachability

Social network analysis data collected in the survey included the aspect of reachability of team members at the local and the distant site.

Survey Results

In question Q26b we asked participants of the survey *Is it easy for you to reach colleagues of this role*? In addition, Q01 asked to evaluate the sentence *People I need to communicate with are difficult to reach*. Results of Q01 and Q26b are shown in Figure C.16 and C.15. According to the social network diagram in Figure C.15, reachability overall is very positive. The only deviance is visible with the internal reachability of the role Project Lead, where 30% state that it is not easy to reach colleagues of the same role, while for all other relationships in the same diagram the value is never above 10%. Q01 gives a more fine-granular answer to reachability. The results conform with the results of Q26b. A vast majority of respondents state that reachability is not a problem: Related to the local site, only 21% agreed with the statement *People I need to communicate*

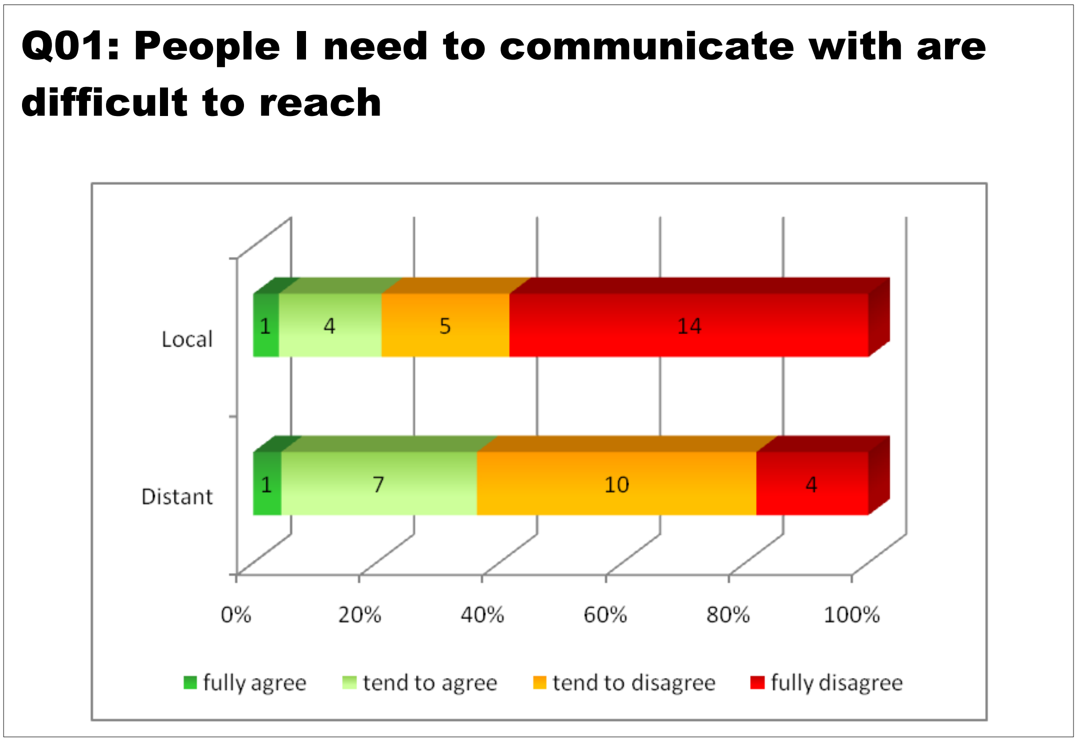


Figure C.16: Question Q01

with are difficult to reach (4% fully agree, 17% tend to agree) and 79% disagree (21% tend to disagree, 58% fully disagree). With respect to the distant site, 36% agreed with the statement (5% fully agree, 32% tend to agree) and 64% disagreed (45% tend to disagree, 18% fully disagree), which represents the usual effects of distribution.

In Q27 we asked the open question *Where/between which roles do you see communication problems?* The answers were manifold. Communication problems between almost any role were reported: developer—sub-project lead, project lead—sub-project lead, developer—developer, developer—system integration, developer—project lead, tester—developer. In addition, there were also hints at hidden communication problems:

- “Between teams, especially those on distant site there clearly exist problems.”

- “Between sub-project leads there are tensions (possibly interpersonal issues); because everybody is under high pressure and all are delayed, there is very low tolerance.”

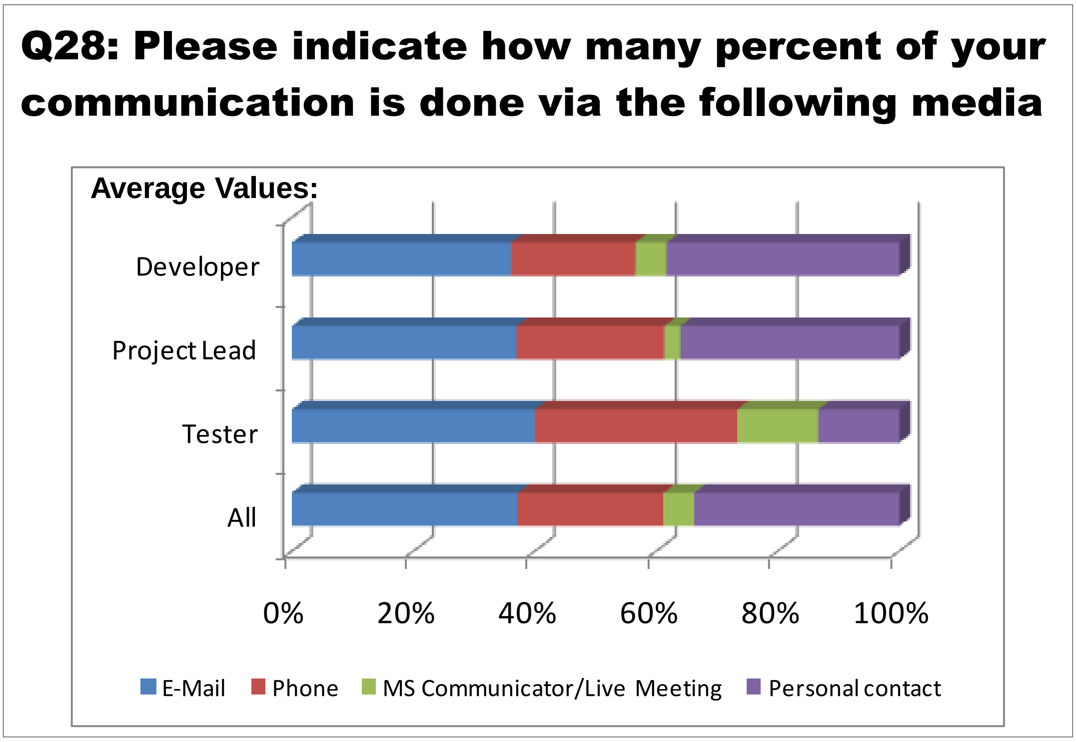


Figure C.17: Question Q28

Conclusions

While overall reachability is very positive, the limited internal reachability of the role Project Lead was conspicuous. Also the comments of survey respondents hint at tensions between sub-project leads.

C.1.5 Communication Media

In question Q28 we analyzed the usage of communication media.

Survey Results

Figure C.17 illustrates the average share of communication media based on the question *Please indicate how many percent of your communication is done via the following media* (Q28). The figure shows average values for the 3 role groups. The project team at the German site on average uses 37% e-mail communication, 24% phone communication, 5% Microsoft Communicator/Live Meeting and 35% personal contact. While project members of the role Developer and Project Lead have a high share of personal communication (38% for Developer resp. 36% for Project Lead), the role Tester has a significantly lower share (13%) of personal communication. This is compensated by an above average usage of MS

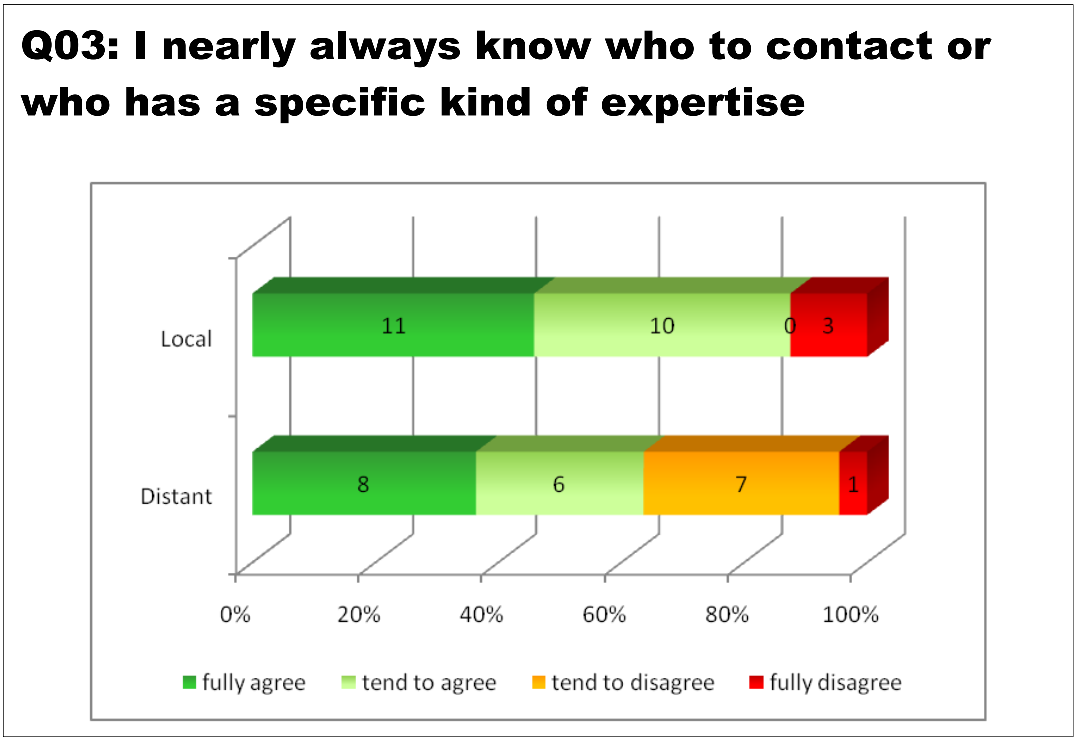


Figure C.18: Question Q03

Communicator/Live Meeting (13% compared to 5% for Developer and 3% for Project Lead) and also phone and e-mail communication.

Conclusions

There were no noticeable problems in usage of communication media.

C.1.6 Awareness

Multiple questions in our survey were related to awareness.

Survey Results

Q03 asked to evaluate the sentence *I nearly always know who to contact or who has a specific kind of expertise*, while Q04 asked if *It is important for me to understand the day-to-day work of my colleagues.* The results of Q03 and Q04 are illustrated in Figure C.18 and C.19. Finding the right contact person seems not to be a major problem in Project X: With respect to the local site, 88% fully agree or tend to agree that they nearly always know *who to contact or who has a specific kind of expertise* (46% fully agree, 42% tend to agree), while 0% tend to disagree and 13% disagree. In case of the distant site, 64% of respondents

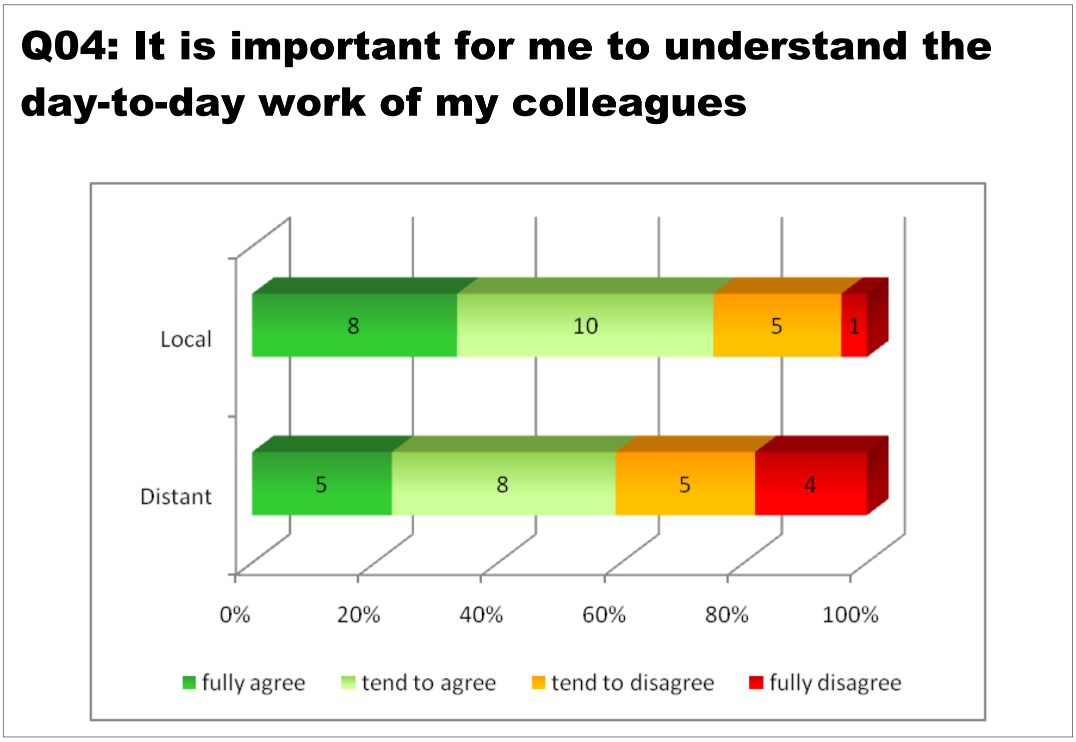


Figure C.19: Question Q04

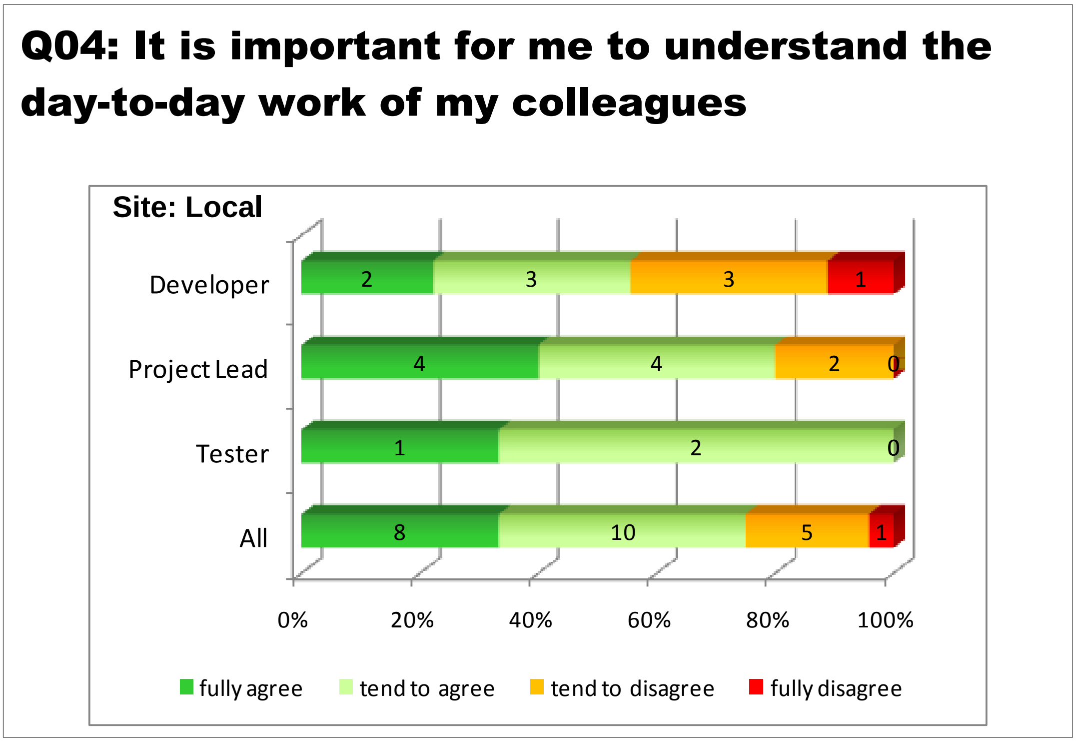


Figure C.20: Question Q04 per role, local

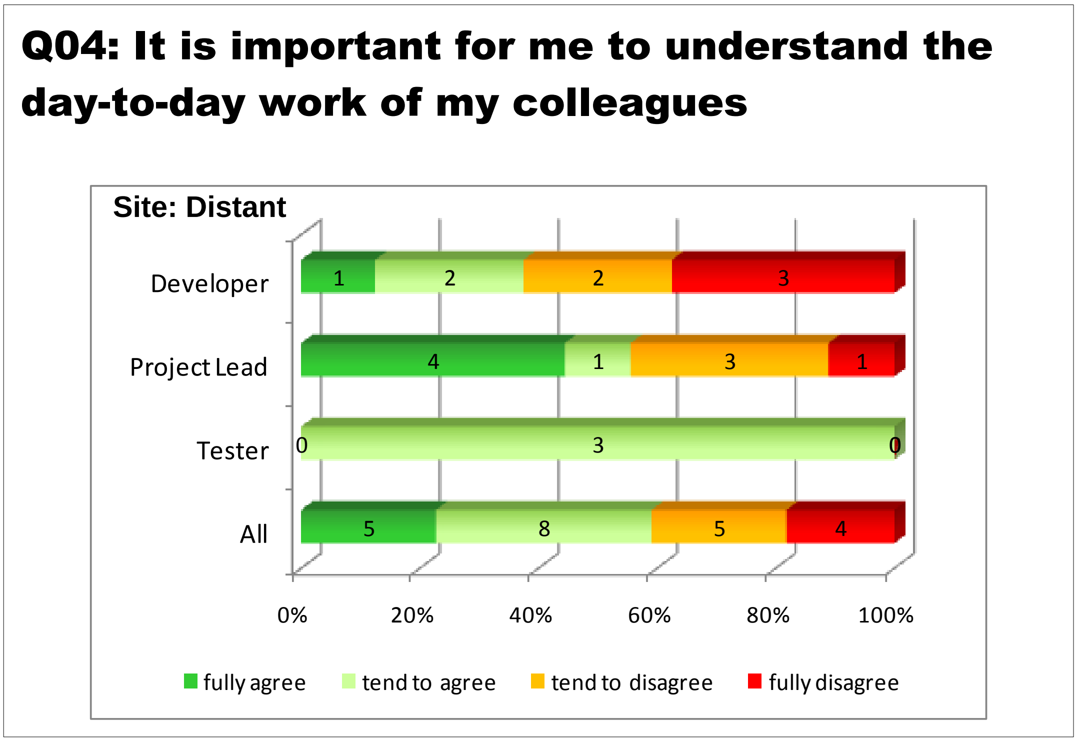


Figure C.21: Question Q04 per role, distant

express agreement (36% fully agree, 27% tend to agree) and 36% disagreement (32% tend to disagree, 5% fully disagree). The situation of both local and distant site show a high awareness of who is the right contact person within the project team. The difference between local and distant site again can be explained by the usual effects of distribution (lower awareness in case of the distant site). In question Q04, a high portion of respondents stated that it is important for them to understand the day-to-day work of their colleagues: For the local site, 75% agreed (33% fully agree, 42% tend to agree) and 25% disagreed with the statement (21% tend to disagree, 4% fully disagree). With respect to the distant site, 59% expressed agreement (23% fully agree, 36% tend to agree) and 41% showed disagreement (23% tend to disagree, 18% fully disagree). The lower (but still high) agreement for the distant site is not surprising, as the work between the German site and the supplier site in Central Eastern Europe is more decoupled than the local activities. If we look at the role-specific results with respect to the local site (see Figure C.20), we can notice that the need for understanding the day-to-day work of colleagues increases from Developer (56% agreement) over Project Lead (80% agreement) to Tester (100% agreement). The same holds for the distant site (see Figure C.21: Developer: 38%, Project Lead: 56%, Tester: 100% agreement). This effect can be explained by the extent of how much a person of this role depends on other colleagues: While a Developer has lower

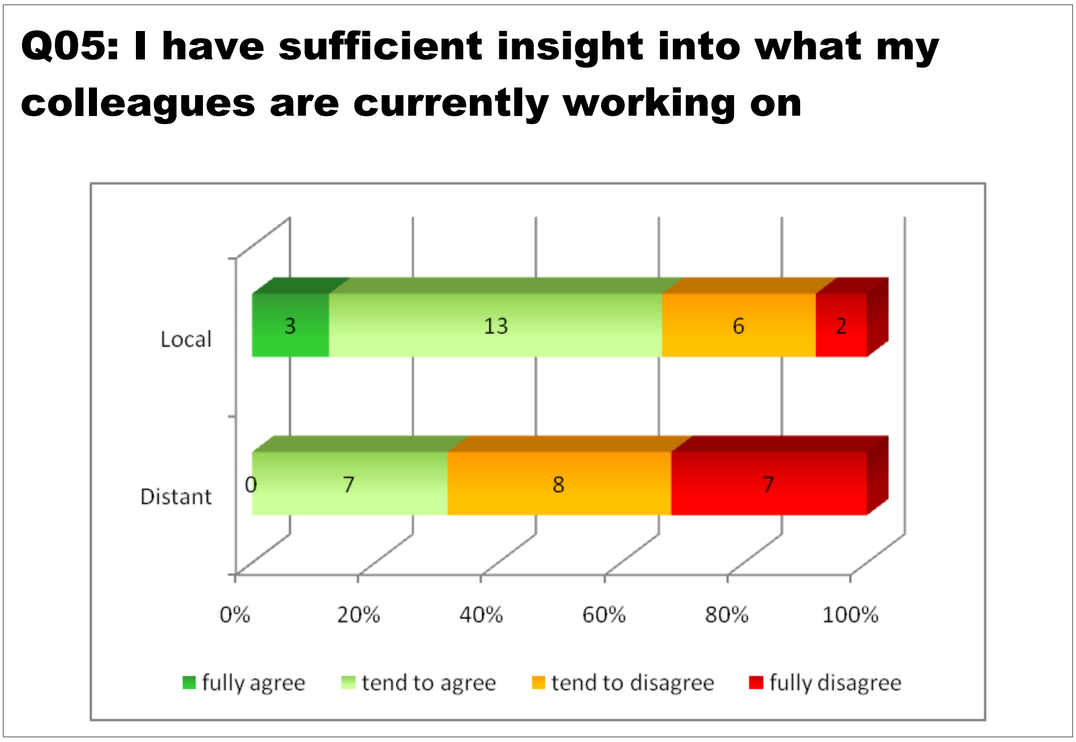


Figure C.22: Question Q05

dependencies as (s)he is working on their respective work packages, Project Lead need to coordinate with multiple persons from the team. Finally, a Tester has to integrate the whole system and therefore has maximum dependencies on the work of colleagues.

In question Q05 we asked for an evaluation of the statement *I have sufficient insight into what my colleagues are currently working on*. Figure C.22 shows the result. With respect to the local site, there is 67% of agreement, where only 13% fully agree and 54% tend to agree. 25% expressed disagreement, with 25% tend to disagree and 8% fully disagree. For the distant site, there was 32% agreement with the statement *I have sufficient insight into what my colleagues are currently working on*, where no one fully agreed, but 32% tended to agree. Disagreement was 68% (36% tend to disagree, 32% fully disagree). The results of both local and distant site show that awareness with respect to current work activities of colleagues are not optimal. In particular in case of the distant site, there is overall an insufficient insight into what colleagues are currently working on, which is symptomatic for a distributed team. Figure C.23 and C.24 provide a role-breakdown of Q05 with respect to the local and the distant site. In case of the local site, the values of the role group Project Lead are noticeable: All of the negative values originate from this role group; 60% of this role expressed disagreement with the statement (40% tend to disagree, 60% fully disagree). This result indicates a

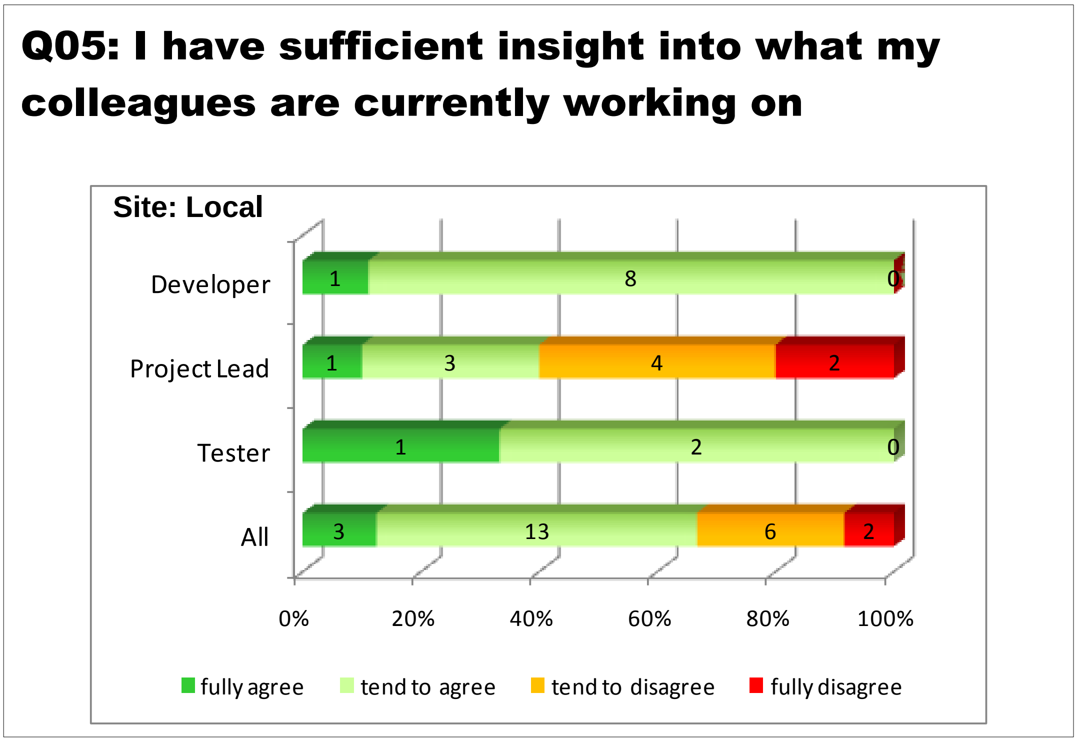


Figure C.23: Question Q05 per role, local

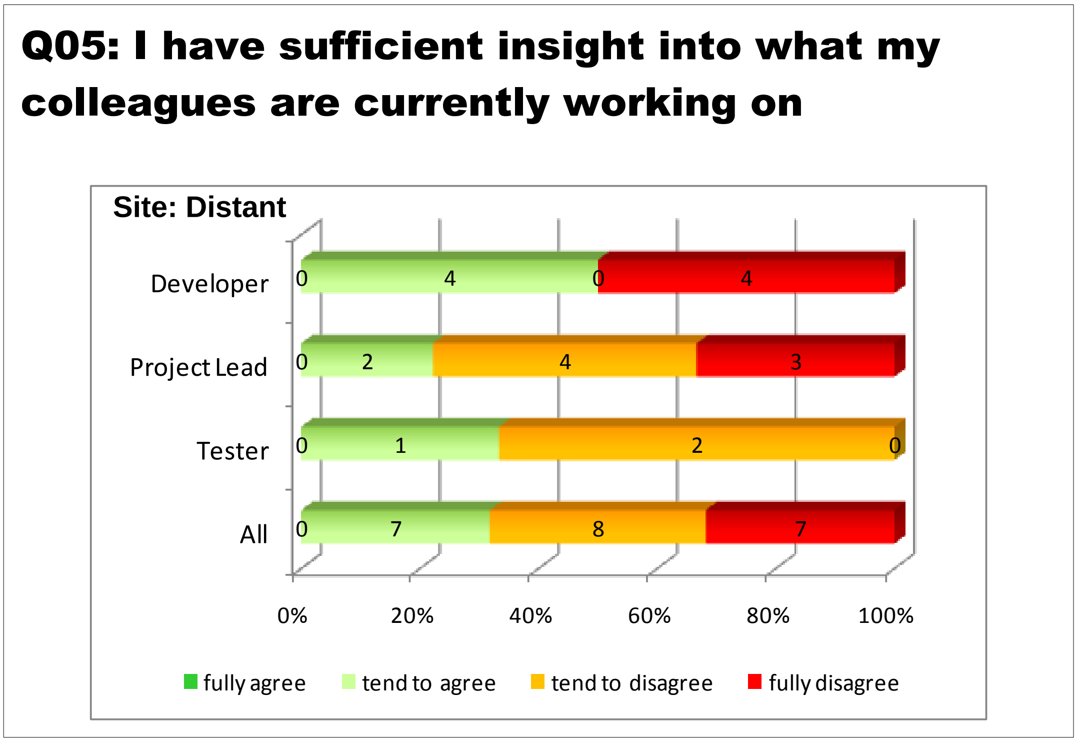


Figure C.24: Question Q05 per role, distant

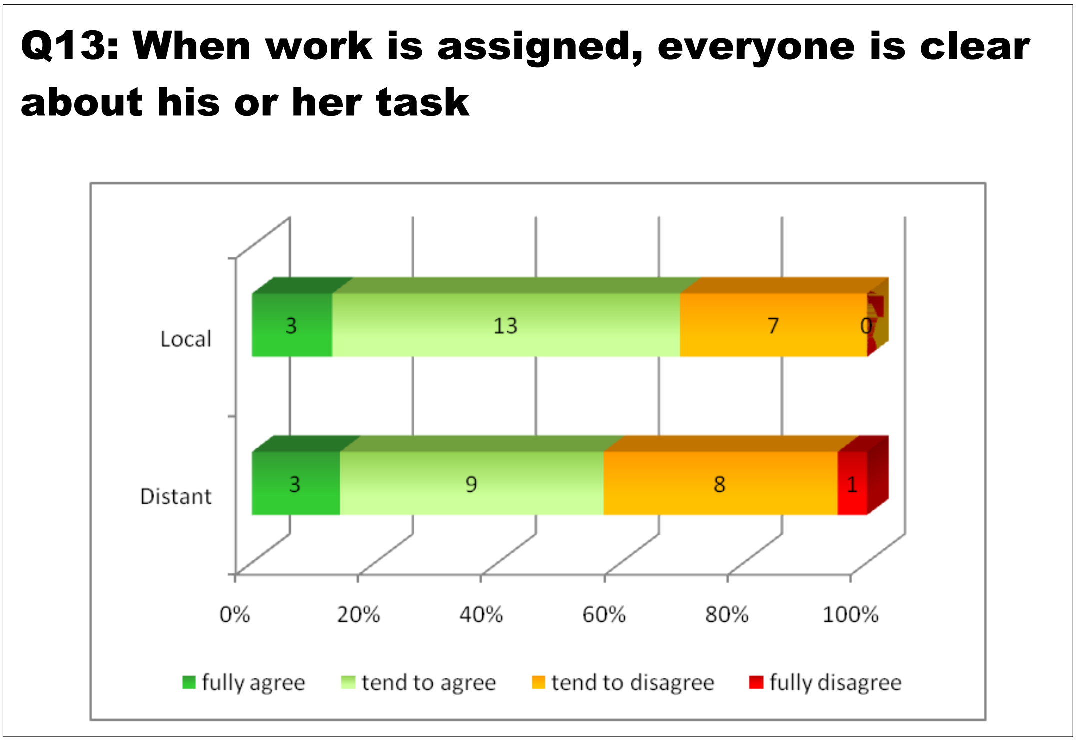


Figure C.25: Question Q13

communication problem with respect to the role Project Lead in collaboration at the local site. Referring to the distant site, more effects can be observed: half (50%) of the role Developer state that they fully disagree to have sufficient insight into what their colleagues at the distant site are currently working on. The other 50% tended to agree with the statement. The role Project Lead showed 22% agreement (0% fully agree, 22% tend to agree) and 78% disagreement (44% tend to disagree and 33% fully disagree). This is a critical situation as the Project Lead persons are in charge of coordination with the distant team members. The role Tester expressed 33% agreement (0% fully agree, 33% tend to agree) and 67% disagreement (67% tend to disagree).

Figure C.25 and C.26 contain the results of the questions *When work is assigned, everyone is clear about his or her task* (Q13) and *My colleagues provide timely information about changes in current plans* (Q14). For the local site, 70% agreed to the statement, where only 13% fully agreed and 57% tended to agree. 30% expressed disagreement (30% tend to disagree, 0% fully disagree). With respect to the distant site, 57% agreed with 14% as 'fully agree' and 43% 'tend to agree'. Disagreement was 43% (38% tend to disagree, 5% fully disagree). Although the overall result of Q13 is not too bad, the low portion of 'fully agree' answers is alarming, as most of the team members only "more or less" know what they are supposed to do when work is assigned. There seems to be also a prob-

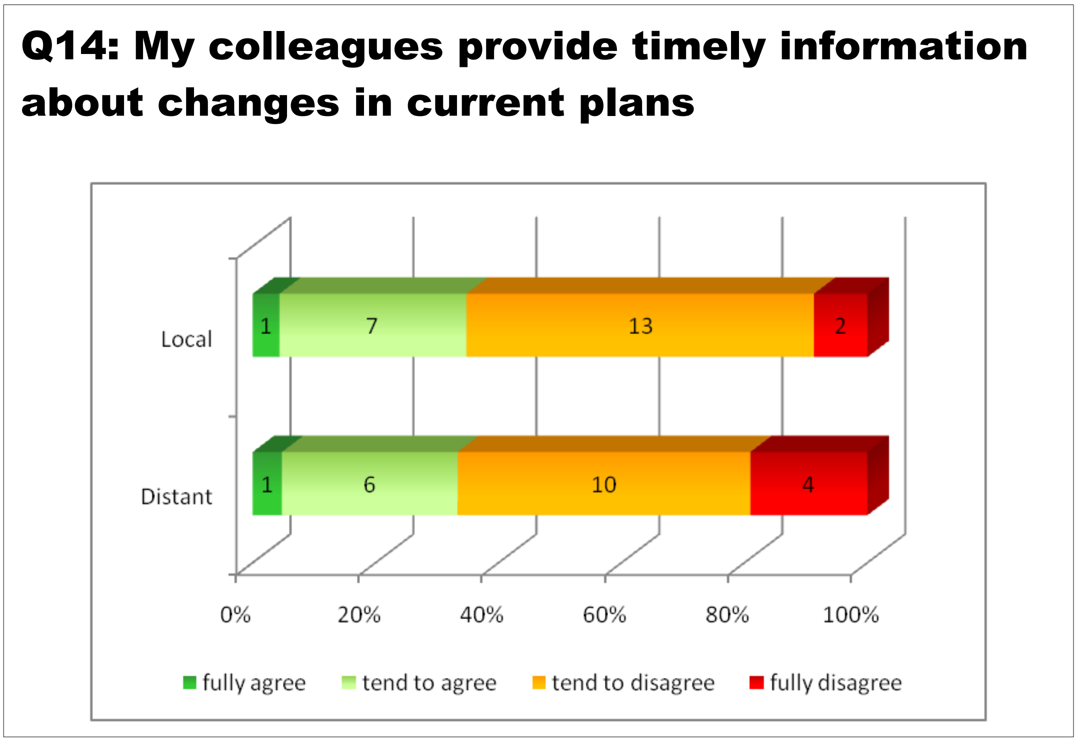


Figure C.26: Question Q14

lem with timely information about changes in current plans, as investigated with question Q14: In case of the local site, 35% agreed to the statement *My colleagues provide timely information about changes in current plans* (4% fully agree, 30% tend to agree), while 65% expressed disagreement (57% tend to disagree, 9% fully disagree). This means the vast majority has a problem with being informed on time about changes (recall, only 4% answered fully agree). With respect to the distant site 33% expressed agreement (5% fully agree, 29% tend to agree) and 67% disagreement (48% tend to disagree, 19% fully disagree). This means the situation is comparable, but slightly worse than the local case (19% fully disagree as opposed to 9% fully disagree in case of the local site).

Figure C.27 shows the social network analysis with respect to the question *Are you aware of the current set of tasks that colleagues of this role are working on?* (Q26c) The overall situation with respect to awareness is not as good as communication frequency or reachability investigated with social network analysis above. We found it noticeable that 50% of the role Project Lead don't know what others of this role are currently working on. Furthermore, the majority (56%) of the role group Project Lead stated that they don't know what colleagues of the role Developer are currently working on. On the other hand, even 78% of the group Developer don't know what the Project Lead is working on. Also 44% of Project Lead don't know what the Tester is currently concerned with.

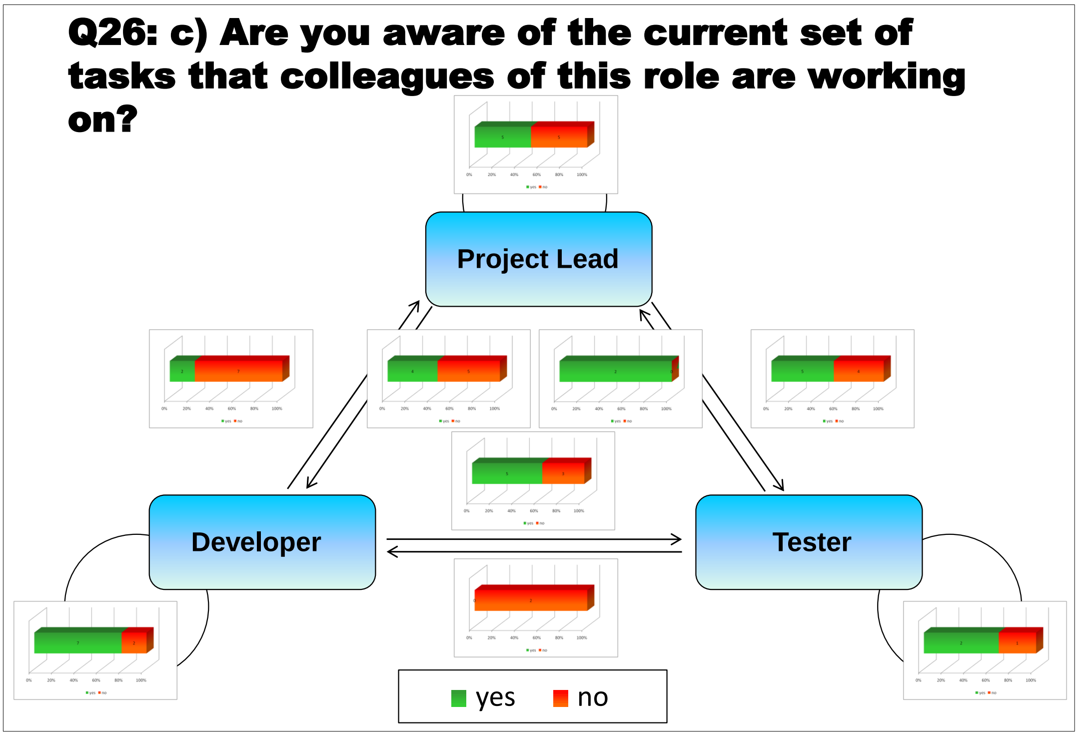


Figure C.27: Question Q26c

Conclusions

While part of the results can be explained by the effects of distribution, the survey results indicate communication problems with respect to the role Project Lead in collaboration at the local site. Furthermore, awareness with respect to the current work activities is not optimal, and tasks are not fully clear to the team when they are assigned.

C.1.7 Traveling

In question Q29, we investigated the frequency and duration of traveling across sites.

Survey Results

In Q29a we asked *Have you ever visited other development sites?* Additionally, we wanted to know in Q29b: *In the last 12 months, how many days have you spent at other sites?* The answers are summarized in Figure C.28 and C.29. A high portion of the team in Germany has already visited the other development site of the project: 38% of the role Developer have already visited other development sites, 60% of Project Lead and 67% of Tester. The average duration (number

Q29: a) Have you ever visited other development sites?

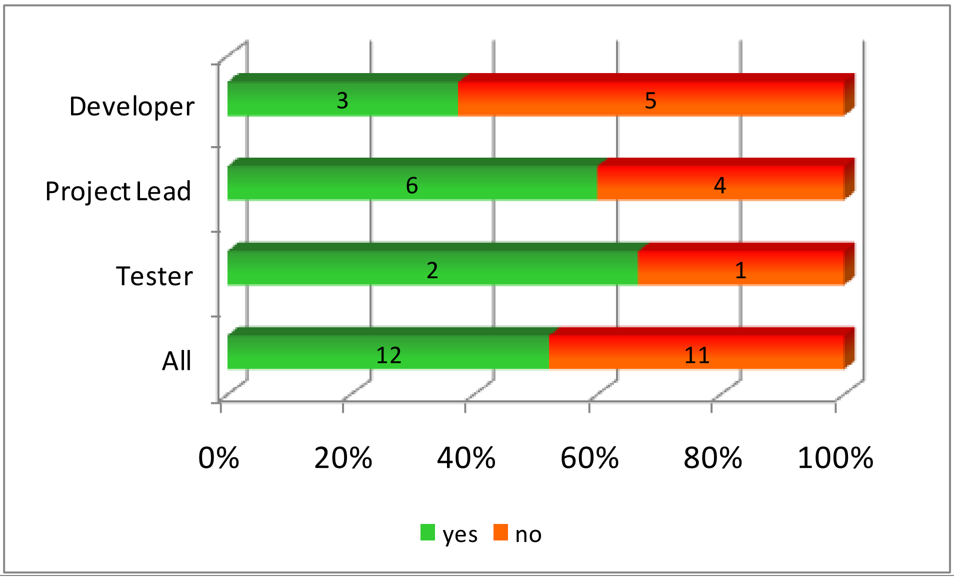


Figure C.28: Question Q29a

Q29: b) In the last 12 months, how many days have you spent at other sites?

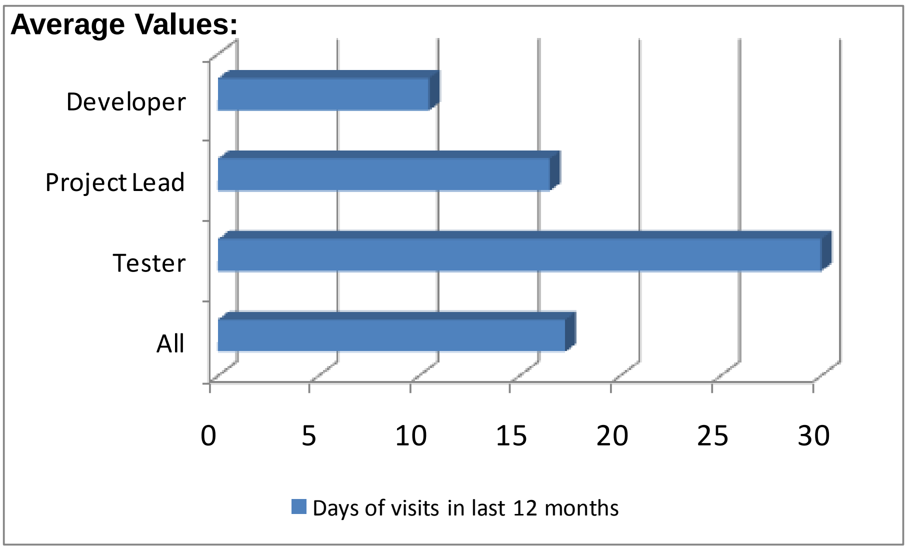


Figure C.29: Question Q29b

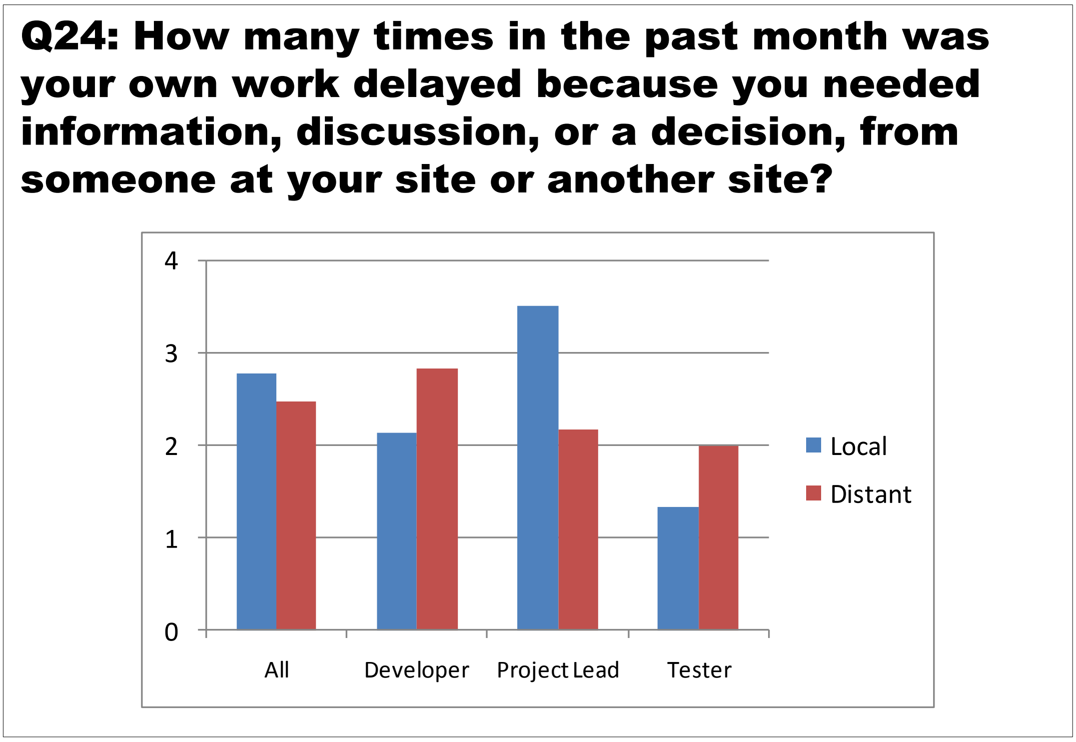


Figure C.30: Question Q24

of days) spent at other sites within the last 12 months amounts to 10.5 days for Developer, 16.5 days for Project Lead, and 30 days for Tester.

Conclusions

In summary, Q29a and Q29b show an intensive exchange across sites.

C.1.8 Delay

As we have seen in Chapter 4, delay is a key phenomenon in GSE. Questions Q24 and Q25 were concerned with analyzing delays in Project X.

Survey Results

Figure C.30 and C.31 show the results of questions Q24 and Q25. In Q24 we asked *How many times in the past month was your own work delayed because you needed information, discussion, or a decision, from someone at your site or another site?* On average, the role group Developer had 2.1 delays in the local case and 2.8 in the distant case. In the group Project Lead it were 3.5 for local and 2.2 for distant. For the role Tester the number of delays was on average 1.3 local and 2.0 distant. The expected result would have been that there are

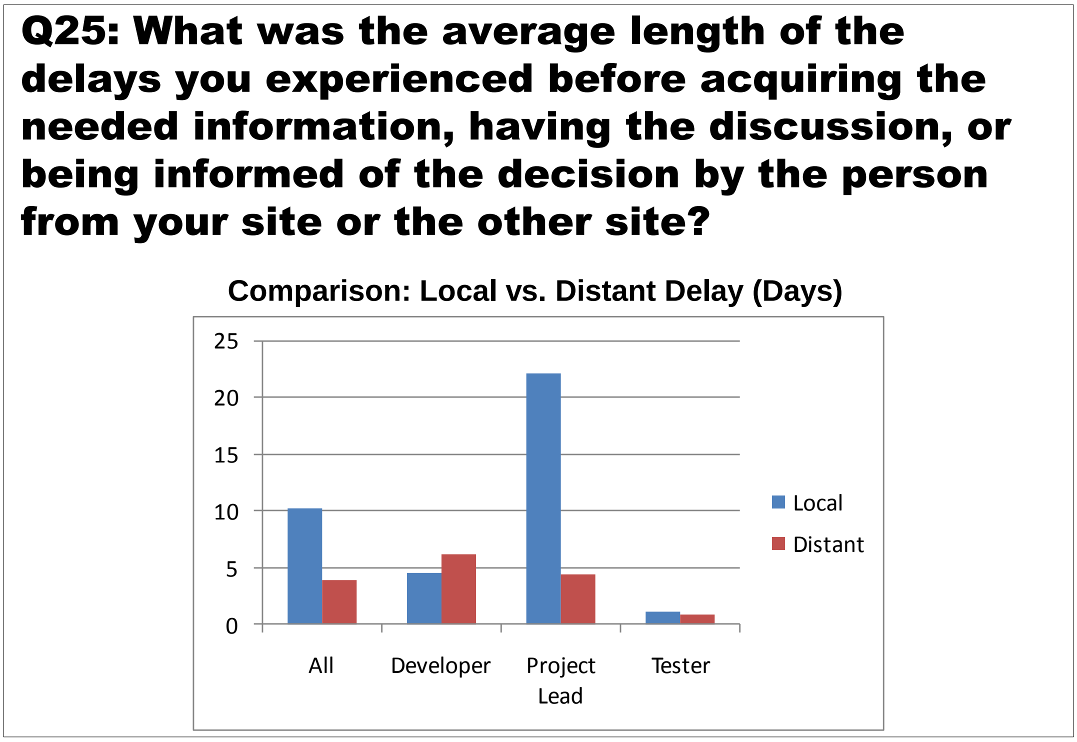


Figure C.31: Question Q25

more delays with respect to the distant site than within the local site. This is confirmed in case of Tester (50% more distant delays than local delays) as well as Developer (32% more distant delays than local delays). However, in case of the role group Project Lead, the opposite was noticed: There were 62% more local delays than distant delays. This result hints at local problems that cause significant delays in the Project Lead's work. Q25 provides more details on the duration of the delays: *What was the average length of the delays you experienced before acquiring the needed information, having the discussion, or being informed of the decision by the person from your site or the other site?* The role Developer on average experienced 4.5 days of delay in the local case and 6.3 days of delay in the distant case. For the role Project Lead the average local delay was 22.1 days and 4.4 days of distant delay. With respect to the role Tester, there were 1.2 days of local delay and 0.8 days of distant delay. Overall, the average delay was 10.3 days in the local case and 3.9 in the distant case. The duration of delays of the group Project Lead shows a striking result, since the average local delay is about 5 times higher than the distant delay, while the opposite behavior would be expected. This result again hints at local problems in particular within the group Project Lead.

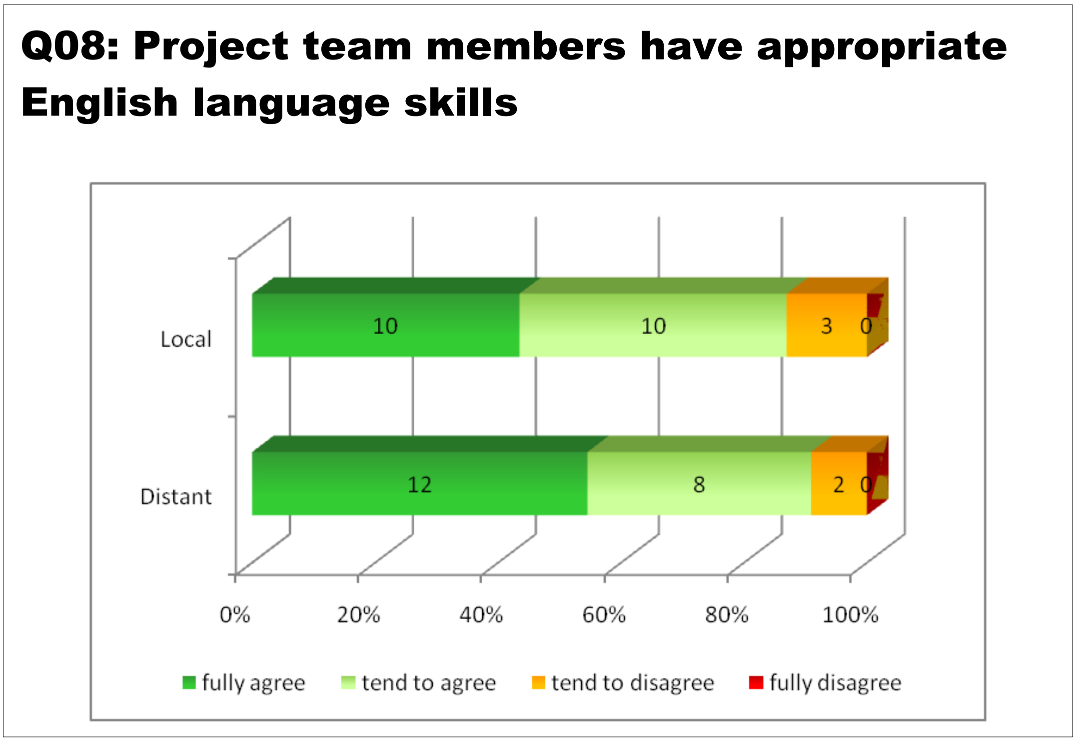


Figure C.32: Question Q08

Conclusions

Although it is a well-know effect of GSE that delay occurs in a distributed setting, the local delays in case of the role Project Lead outweigh that. The conclusion is that there must be local problems that cause significant delays in the Project Lead's work.

C.1.9 Culture and Language

Questions Q08 and Q16 dealt with English language skills and the influence of culture on the project team.

Survey Results

Figure C.32 and C.33 show the answers to the questions *Project team members have appropriate English language skills* (Q08) and *Diversity in the project team (e.g. national cultures) is valuable for team collaboration* (Q16). English language skills seem not to be a problem in Project X: With respect to the local site, 87% confirmed that project team members have appropriate English language skills (43% fully agree, 43% tend to agree), while only 13% didn't confirm it (13% tend to disagree, 0% fully disagree). For the distant site, 91% agreed with the

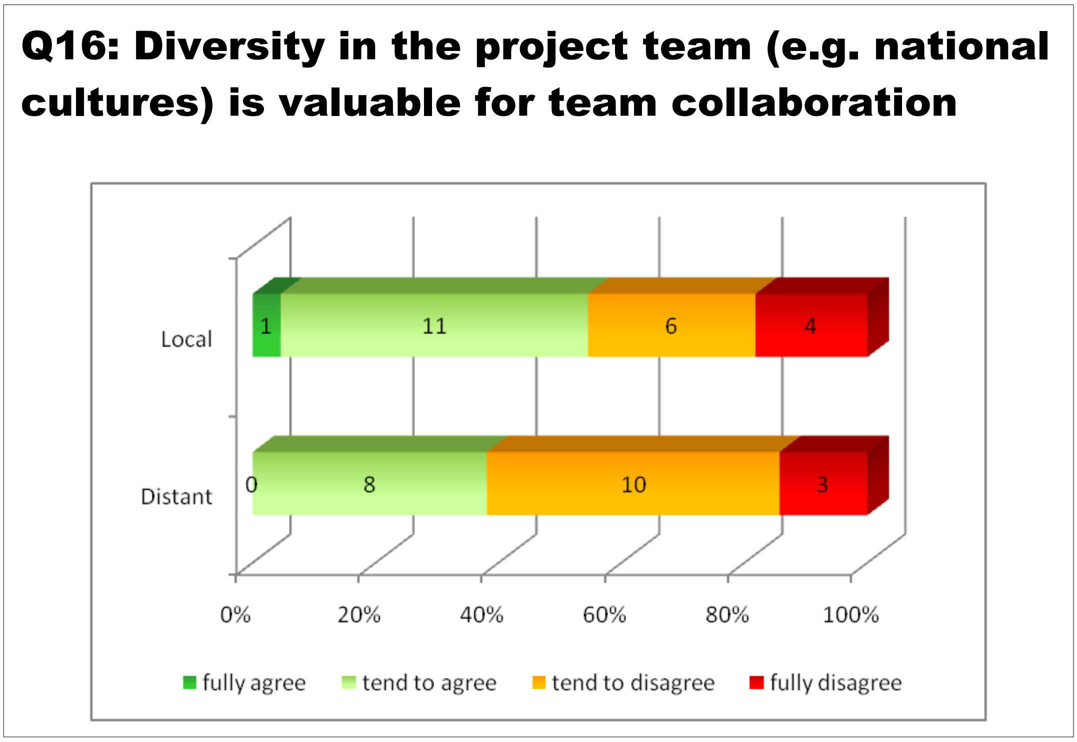


Figure C.33: Question Q16

statement (55% fully agree, 36% tend to agree), only 9% disagreed (9% tend to disagree, 0% fully disagree). Q16 on diversity yielded mixed results: For the local site, 55% agreed that diversity is valuable for team collaboration (5% fully agree, 50% tend to agree), while 45% disagreed (27% tend to disagree, 18% fully disagree). Related to the distant site, 38% confirmed the statement (0% fully agree, 38% tend to agree) and 62% didn't confirm it (48% tend to disagree, 14% fully disagree).

Conclusions

There were no significant problems noticeable with respect to Culture and Language, e.g. English language skills are no problem.

C.1.10 Management and Strategy

In questions Q17-Q23 we investigated the team members' opinion related to management and strategy.

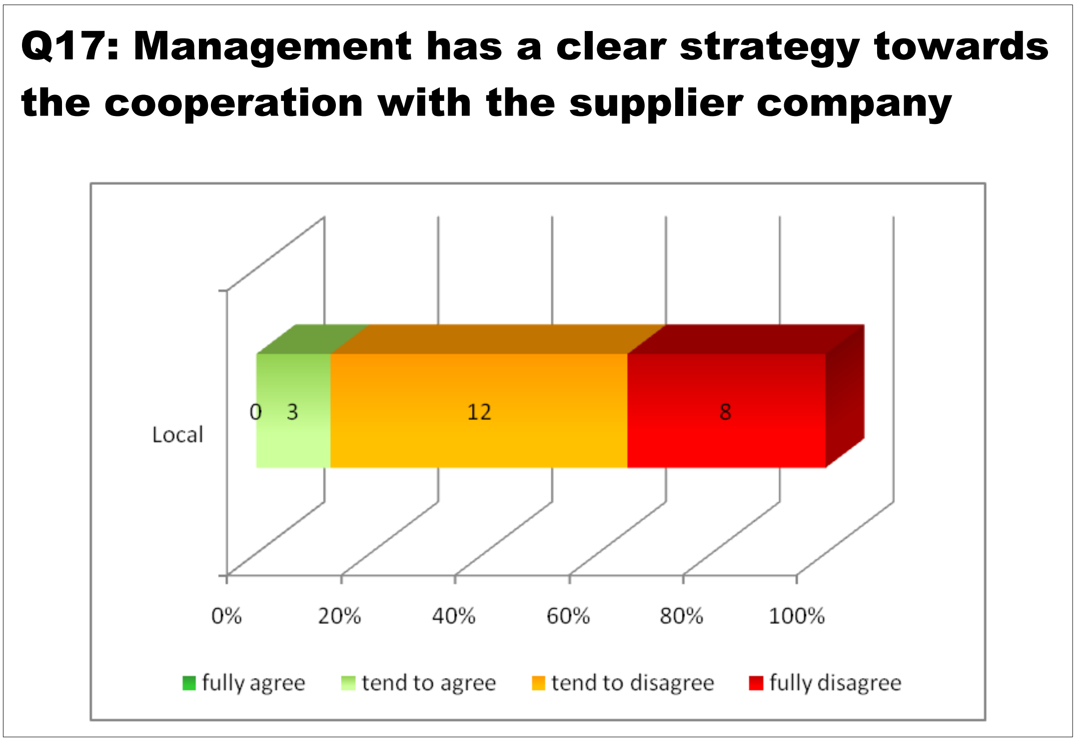


Figure C.34: Question Q17

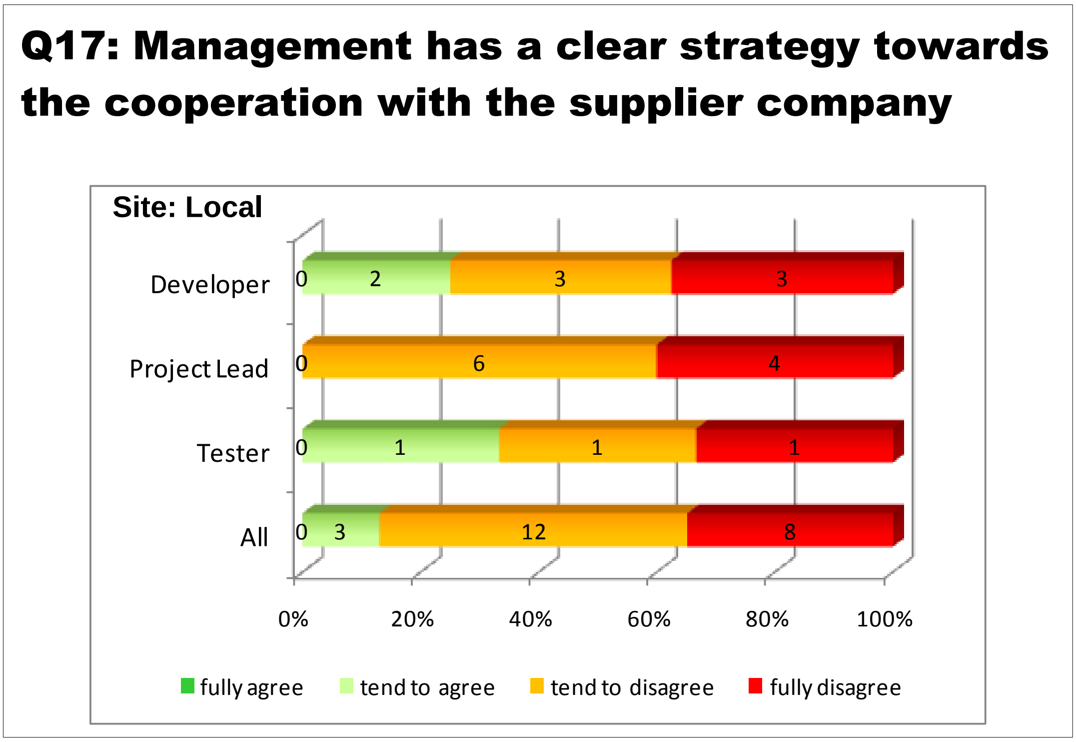


Figure C.35: Question Q17 per role, local

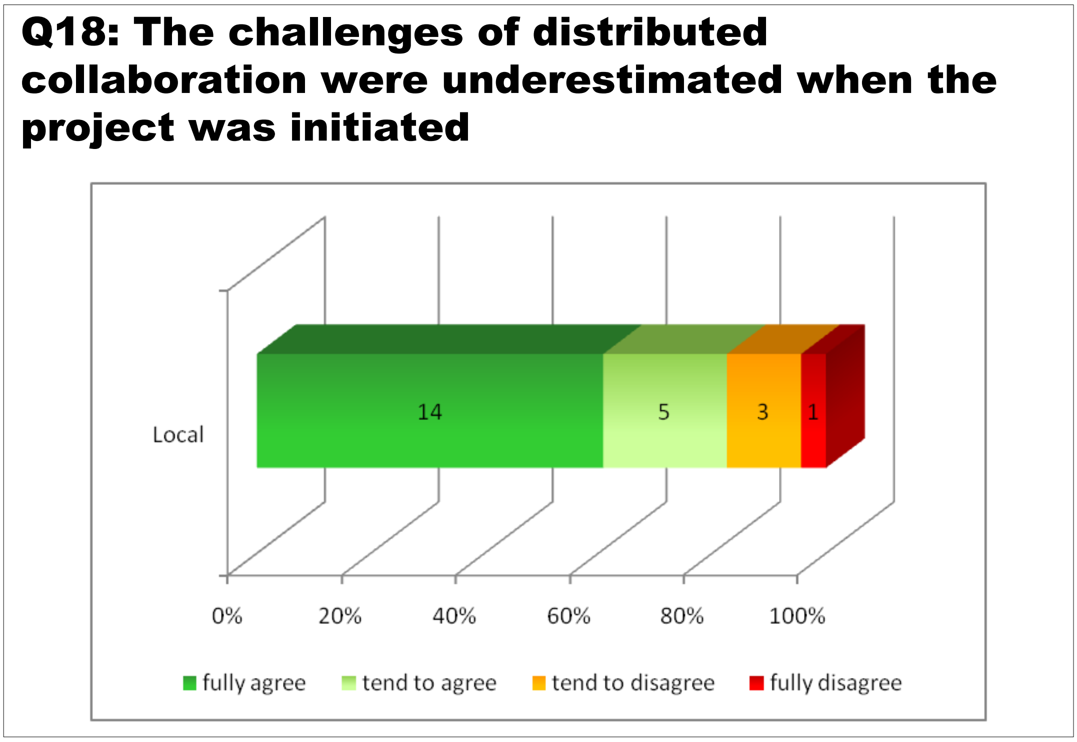


Figure C.36: Question Q18

Survey Results

Q17 asked the team members to evaluate the sentence *Management has a clear strategy towards the cooperation with the supplier company*, see Figure C.34. The result was a strong disagreement with the statement: there were no answers with 'fully agree' (0%) and only 13% with 'tend to agree', 87% disagreed with 52% tend to disagree and 35% fully disagree. Figure C.35 shows the role breakdown of this question. It is noticeable that the role Project Lead which is supposed to be closest to management answered this question in the most negative way compared to the other roles (0% fully agree, 0% tend to agree, 60% tend to disagree, 40% fully disagree). We conclude that from the perspective of the team and in particular seen from the role Project Lead, management has no clear strategy towards the cooperation with the supplier company.

In Q18 we wanted to know if *The challenges of distributed collaboration were underestimated when the project was initiated*, see Figure C.36. There was a strong confirmation of this statement: 83% expressed their agreement (61% fully agree, 22% tend to agree), only 17% showed disagreement (13% tend to disagree, 4% fully disagree).

Question Q21 asked if *Project successes are adequately appreciated by management*. The results are shown in Figure C.37. 33% of respondents agreed to this statement (10% fully agree, 24% tend to agree) and 67% expressed disagreement

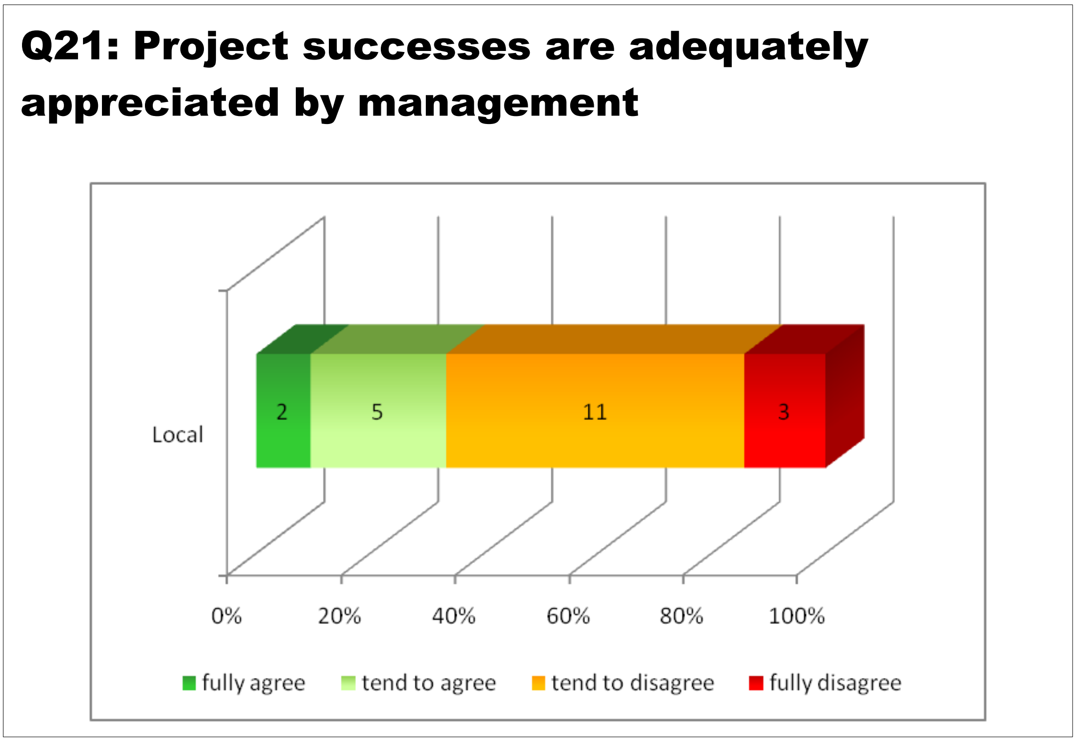


Figure C.37: Question Q21

(52% tend to disagree, 14% fully disagree). One of the respondents commented: "I believe e-mails are no perceptible sign." In summary, appreciation for project successes seems to be inadequate from perspective of the team members.

Being asked to evaluate the sentence *Quality problems is one of the major root-causes for project delay* (Q22, see Figure C.38), there was a strong agreement of 83% (43% fully agree, 39% tend to agree), while only 17% disagreed (13% tend to disagree, 4% fully disagree). Figure C.39 shows the role breakdown of question Q22. It can be seen that the role Tester 100% fully agreed with the statement. This is explainable by the fact that all quality problems are dealt with in test and system integration. Respondents of the role Developer 50% fully agreed with the statement, while the role Project Lead only 20% fully agreed and even 10% fully disagreed and another 10% tended to disagree. We conclude that quality problems are less visible with the roles Project Lead compared to the other roles.

Also the statement *Requirements are clearly defined and communicated* (Q19) showed a very clear tendency (see Figure C.40): While 33% expressed agreement (4% fully agree, 29% tend to agree), 67% showed disagreement (42% tend to disagree, 25% fully disagree). This is an alarming situation, as unclear requirements have the potential of serious impact on project success. (One respondent added the note: "Sometimes there was a daily change of decisions".)

Question Q23 asked for evaluation of the sentence *I feel powerless to change*

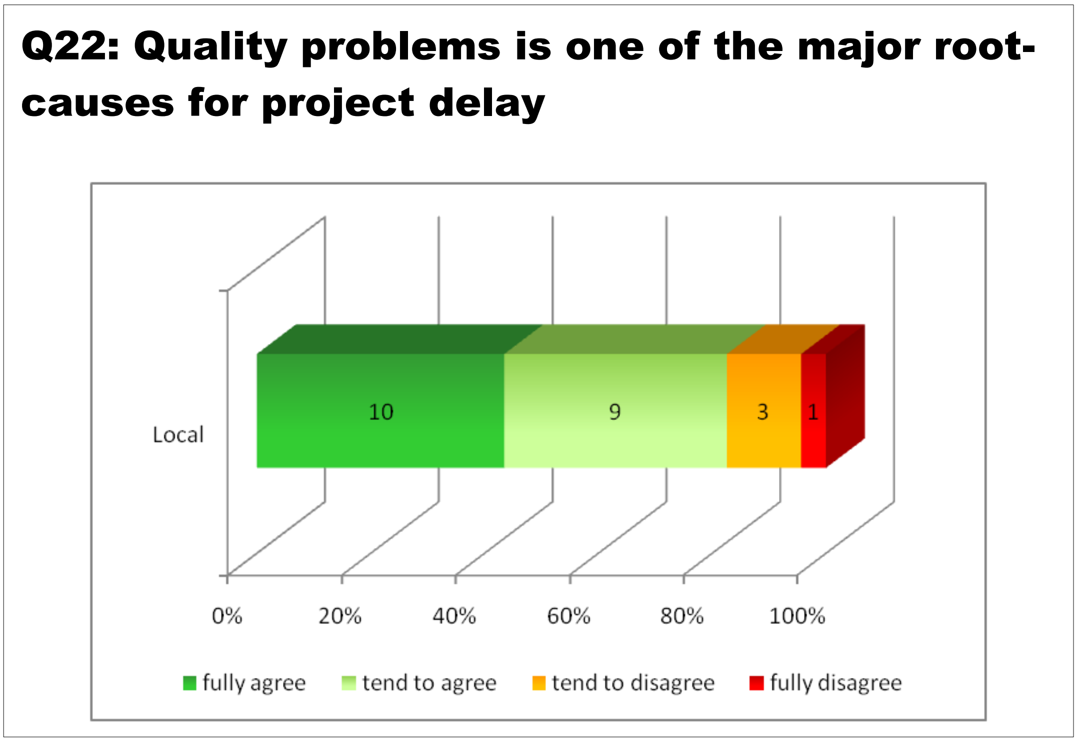


Figure C.38: Question Q22

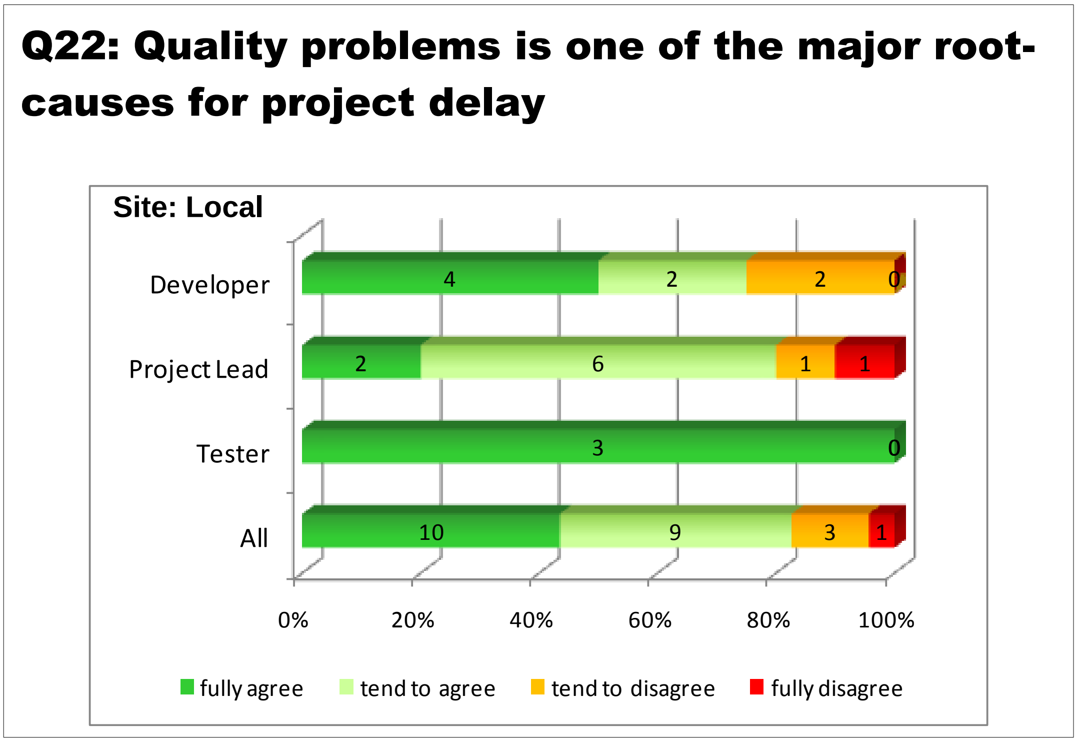


Figure C.39: Question Q22 per role, local

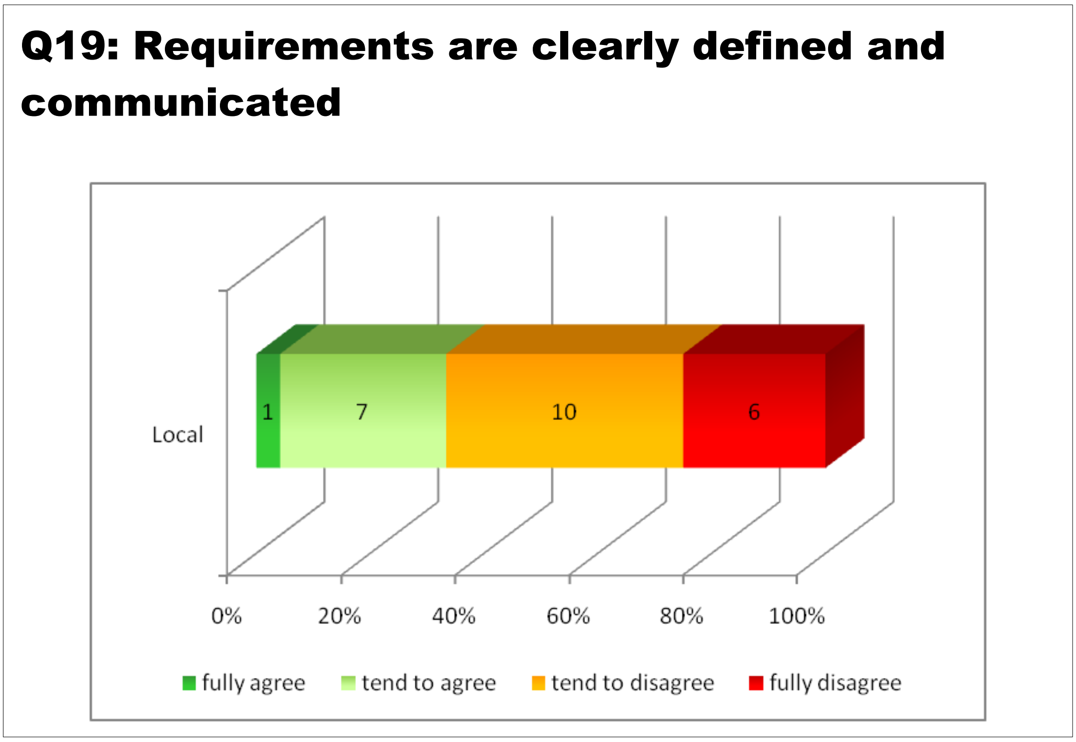


Figure C.40: Question Q19

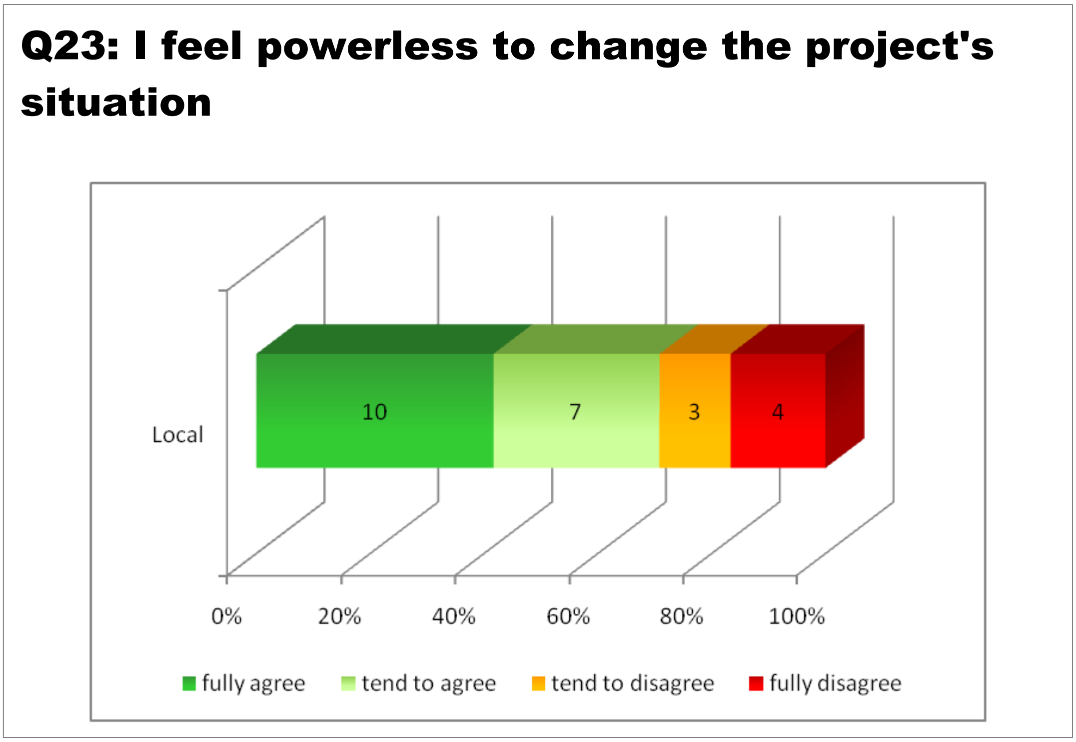


Figure C.41: Question Q23

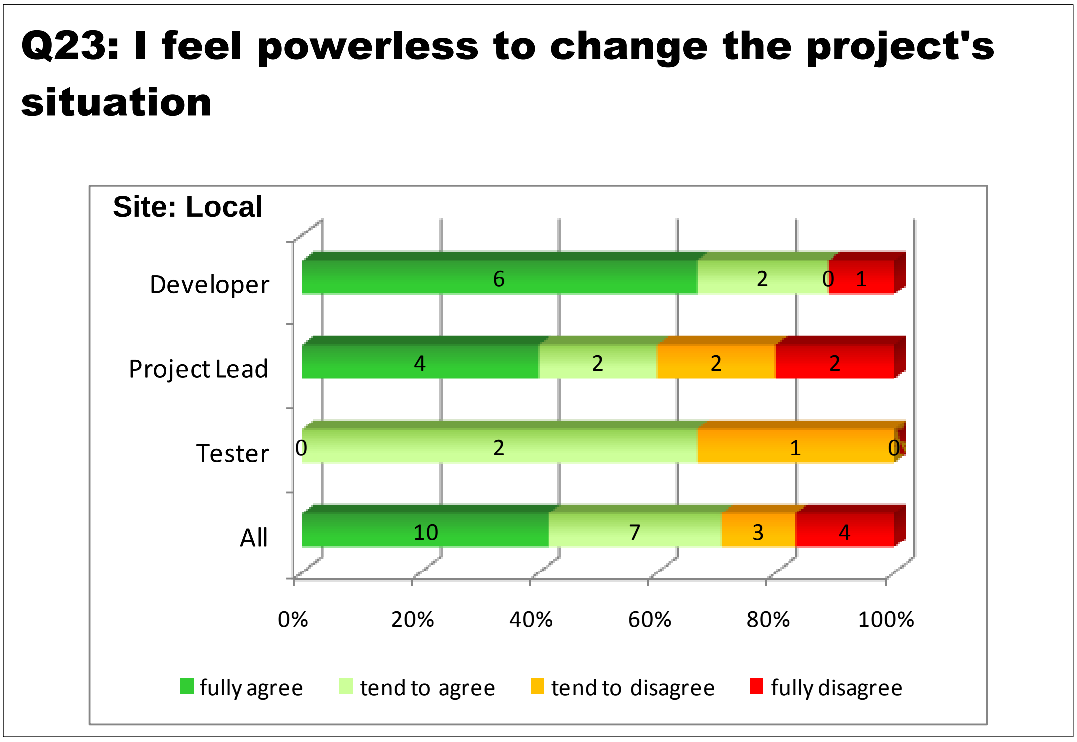


Figure C.42: Question Q23 per role, local

the project's situation. As shown in Figure C.41, most of the respondents confirmed this statement: 71% expressed agreement (42% fully agree, 29% tend to agree), 29% disagreement (13% tend to disagree, 17% fully disagree). Figure C.42 shows the role-breakdown of Q23. People in the role Developer are most negative in their answers: 89% agreed to the statement (67% fully agree, 22% tend to agree), followed by Project Lead with 60% agreement (40% fully agree, 20% tend to agree) and then Tester. The pessimistic view of Developer is understandable due to the fact that they have less influence. Noticeable is the situation of the Project Lead role, as they are supposed to be in the position to change something in the project, however, the frame conditions under which they operate seems to make them feel powerless to change the project's situation.

Conclusions

Summarizing the results related to management and strategy, the situation seem critical with respect to multiple aspects:

- From the team's point of view, there is no clear strategy towards the cooperation with the supplier company
- Challenges of distributed collaboration were underestimated

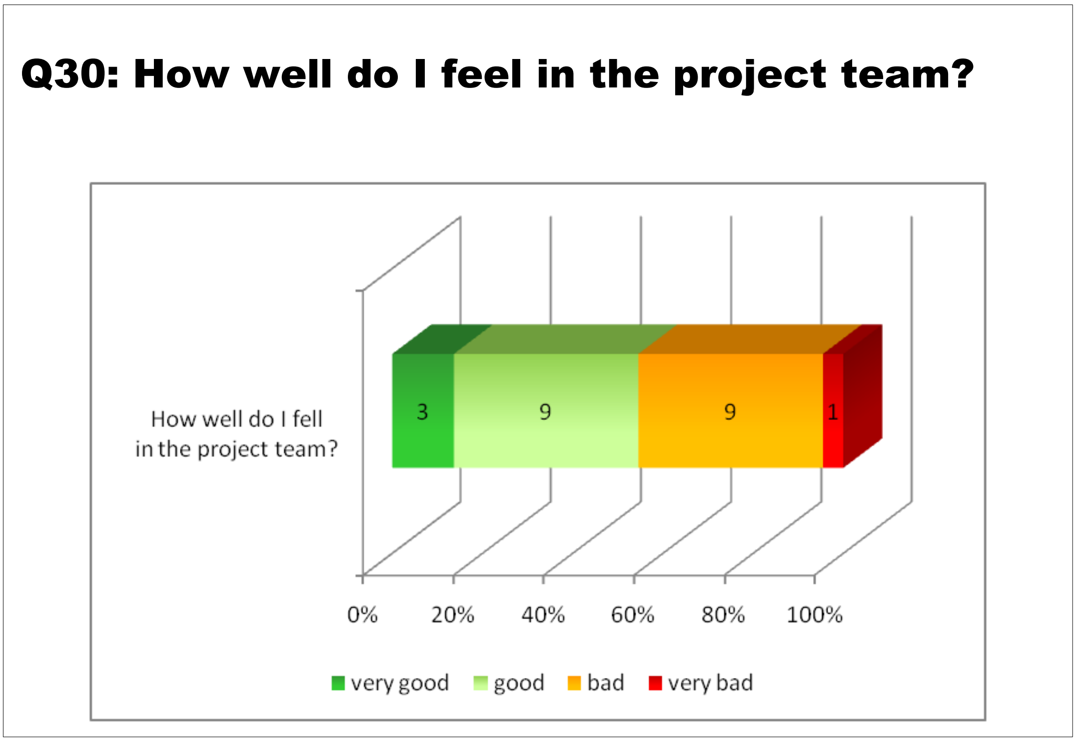


Figure C.43: Question Q30

- Project successes are not adequately appreciated by management
- Requirements are not clearly defined
- Quality is a major root cause for project delay (interrelated to unclear requirements)
- People feel powerless to change the situation

C.1.11 Team Satisfaction

Finally, we captured the situation of team satisfaction in question Q30 and Q31.

Survey Results

Regarding team satisfaction, we asked the project team members *How well do I feel in the project team?* (Q30) and *How satisfied am I with the project's achievements so far?* (Q31) Figure C.43 shows the results of question Q30: 55% replied that they feel well in the project team (14% very good, 41% good), while 45% gave a negative answer (41% bad, 5% very bad). From the role-breakdown in Figure C.44 it can be seen that almost all of the negative answers originate from

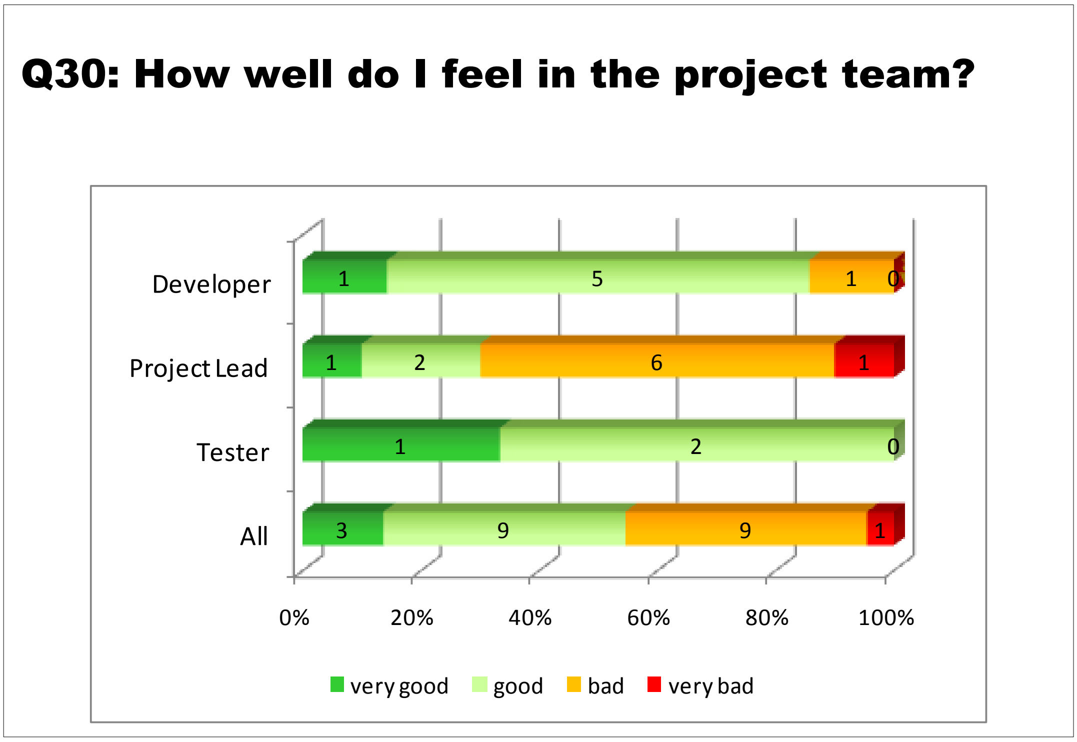


Figure C.44: Question Q30 per role

Project Lead: 70% of this role answered bad or very bad (60% bad, 10% very bad). Obviously people of the role Project Lead don't feel well in their situation.

With respect to the question *How satisfied am I with the project's achievements so far* (see Figure C.45), only few respondents stated that they are satisfied: 0% are very satisfied, 30% are satisfied, while 61% are unsatisfied and 9% very unsatisfied. If we look again at the role-breakdown in Figure C.46, again the role Project Lead gave the most negative answers: 80% unsatisfied or very unsatisfied as opposed to 63% unsatisfied or very unsatisfied in case of Developer and 67% unsatisfied or very unsatisfied of Tester.

The text comments reveal further insights:

> "The Sub-PMs are not able to detect gaps and issues inside the sub-project. Partially the unqualified Sub-PMs are responsible for the sub-projects. It brings huge problems in the understanding and communication. Normally they try to transmit all technical problems to developers and do a pure administrating work."

> "Individual goals are given more importance as team goals. The skills of many persons (developers) are not properly utilized. Task planning is weak. Despite all this, we are moving forward."

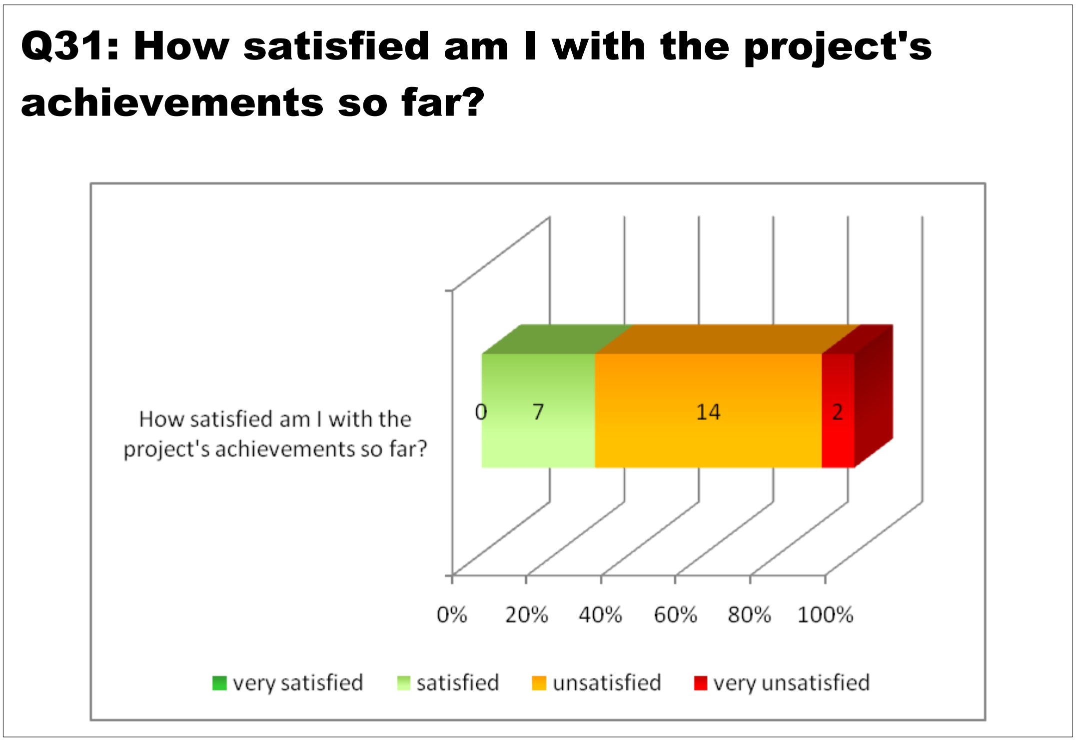


Figure C.45: Question Q31

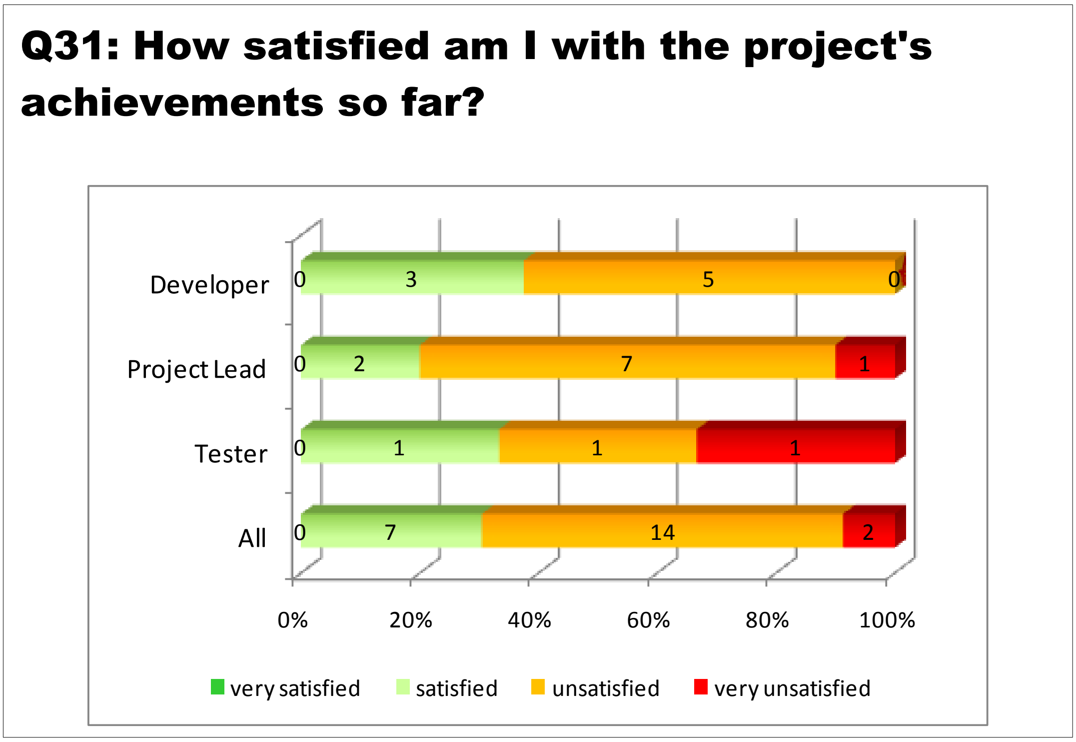


Figure C.46: Question Q31 per role

> "With better internal collaboration and a better contact of all team members to the remote site the project would be in a better shape."

Conclusions

In summary, nobody is really satisfied with the current situation of the project. In particular team members of the role Project Lead don't feel well.

C.2 Summary of Conclusions

In summary, the following aspects turned out to be positive:

- Overall team situation not too bad
- Good English skills
- Exchange, traveling

We identified the following major problem areas:

- Local problems with respect to the role Project Lead
- Competition between teams
- Management and Strategy
- Handling of requirements

One of the respondents aptly summarized the situation:

> "The way how is communicated is important. Respect for everyone from everyone would be helpful. Only together as one team the challenges could be managed."

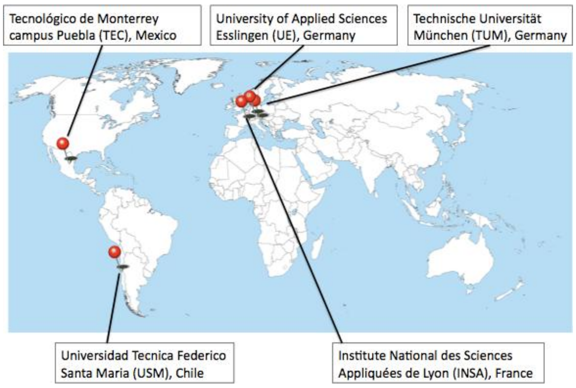
Appendix D

Validation Projects: Sample Report

This appendix presents a sample report as it was prepared for the project teams in the validation phase.

NEREID Collaboration Analysis
Team: Carpooling and Ride Sharing

Status: Dec. 9th, 2010



Christian Lescher
lescher@in.tum.de

Executive Summary

- The overall status of collaboration has improved compared to the first survey
- Communication frequency, reachability and awareness across sites now better
- Exception: Munich team seems to have internal problems (communication, reachability and awareness at Munich site worse compared to other sites)
- Ratings improved for *I can rely on my colleagues* in particular regarding the distant sites
- Shorter and fewer delays (however, there are still significant delays)
- Team is much more satisfied now with the project's achievements

Technische Universität München

Suggestions for the team

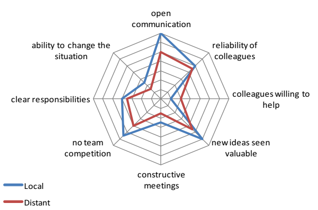
Discuss with the results your team and your supervisors, in particular:

- What were the success factors for improving the overall situation in team collaboration? Which practices would you keep or change in future projects?
- What is the situation of the Munich team?
- Discuss the outlier in *(Virtual) meetings are constructive*: 50% of the Lyon team gave a very negative rating – why?
- What are the root causes for the remaining delays due to waiting for information from other sites and how to improve the situation in the future?

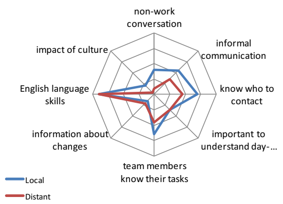
Technische Universität München

Measurement Cockpit for Collaboration

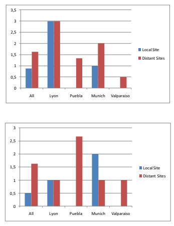
Team Cohesion



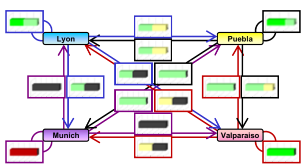
Communication & Awareness



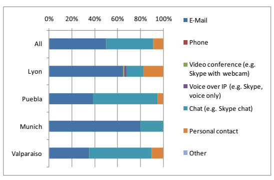
Work Delays



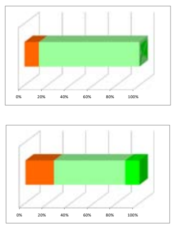
Communication Network

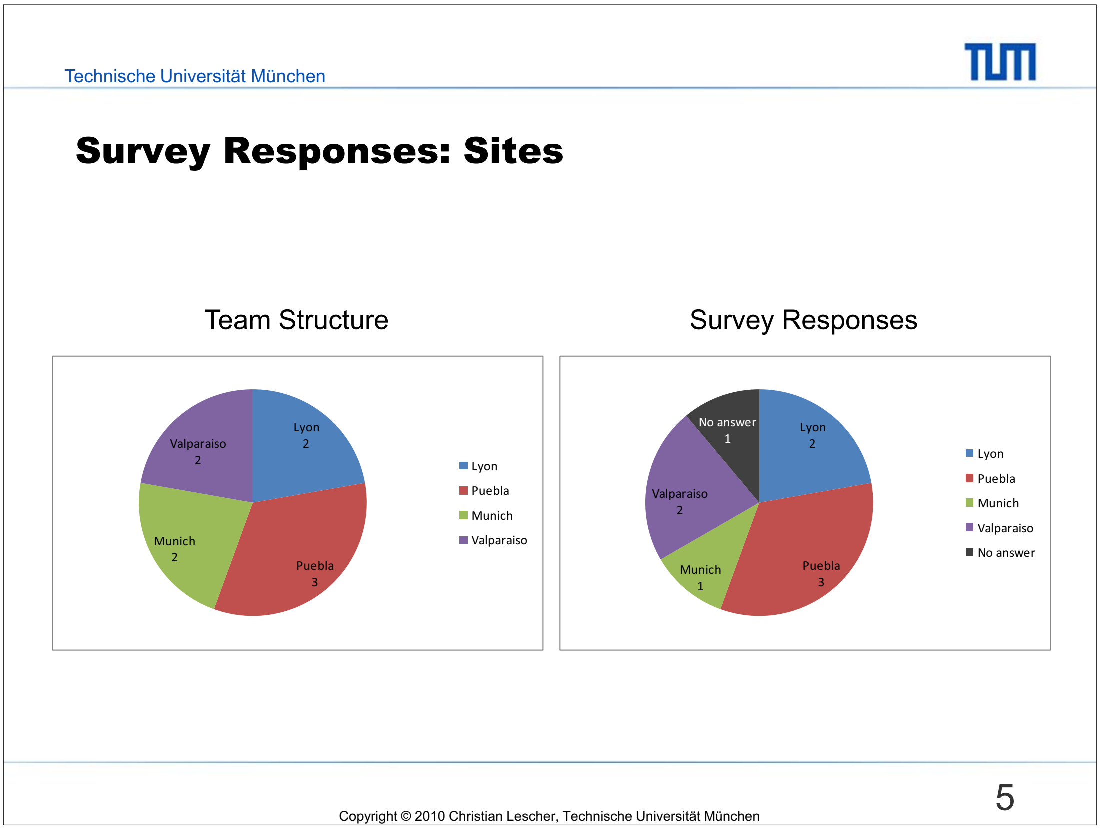


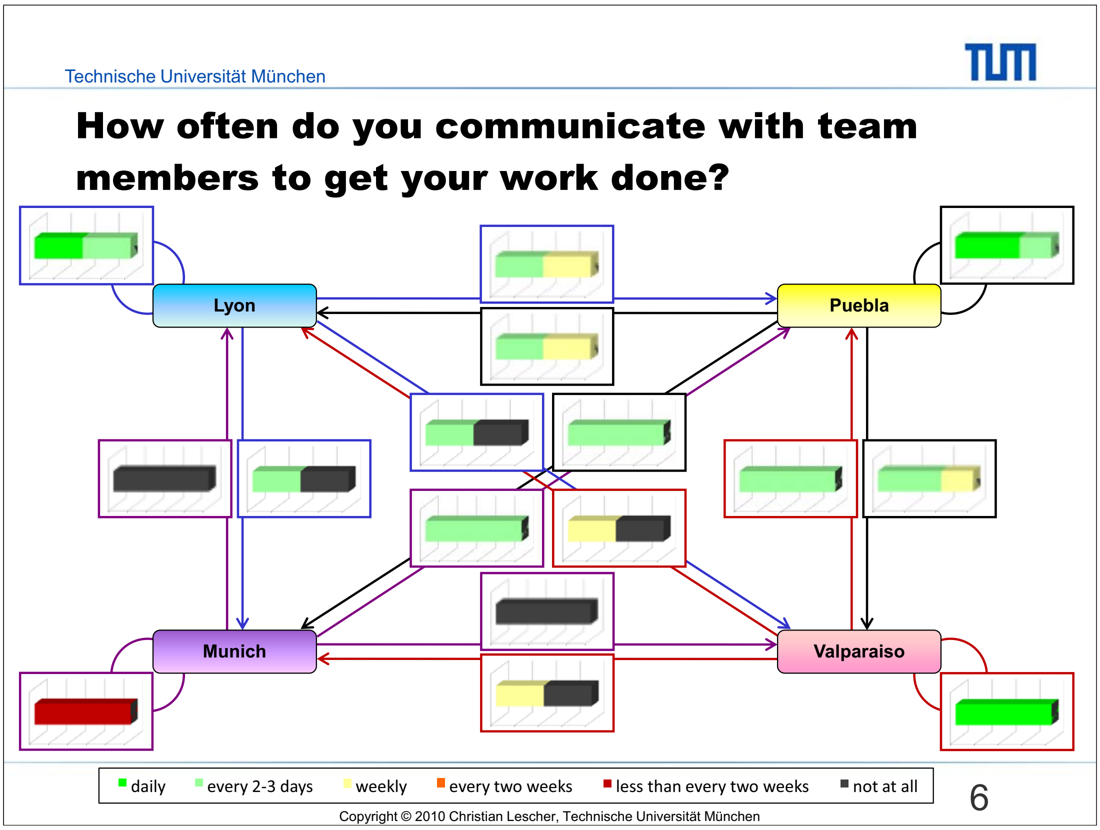
Communication Media

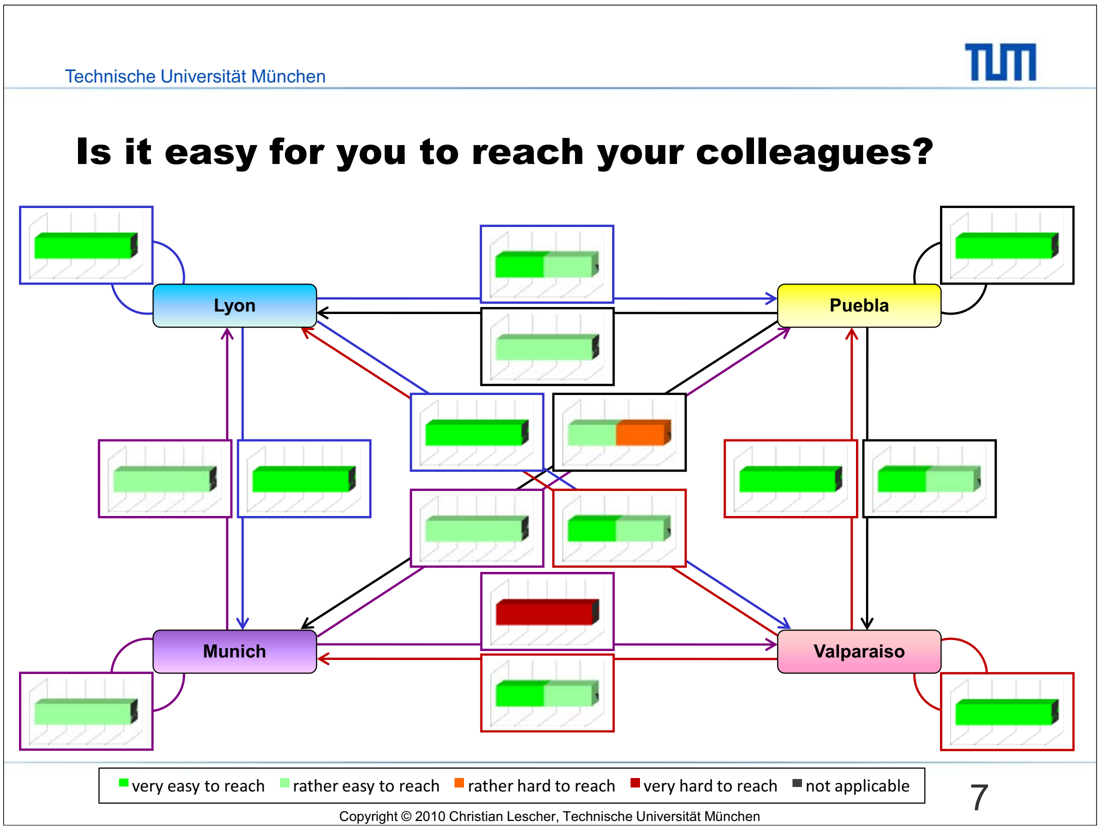


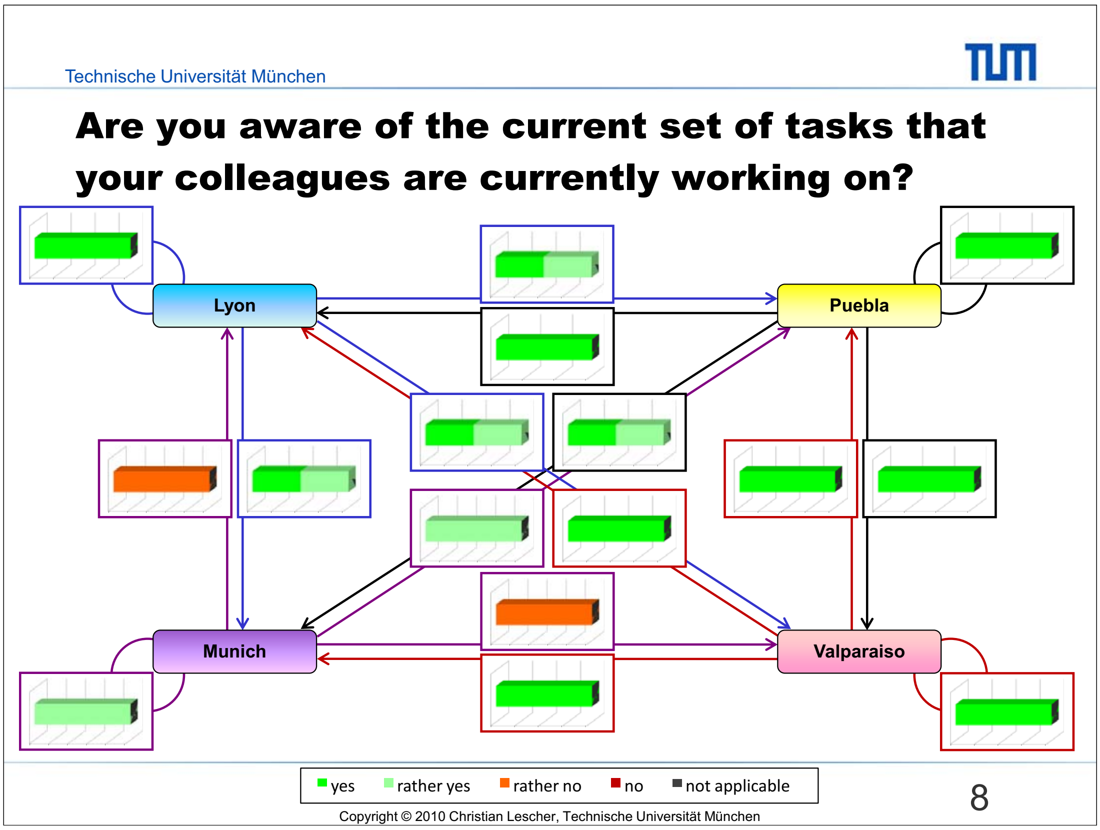
Team Satisfaction




Technische Universität München
Survey Responses: Sites
Team Structure
Survey Responses
Lyon 2
Puebla 3
Munich 2
Valparaiso 2
No answer 1
Munich 1
Lyon
Puebla
Munich
Valparaiso
No answer
Copyright © 2010 Christian Lescher, Technische Universität München
5

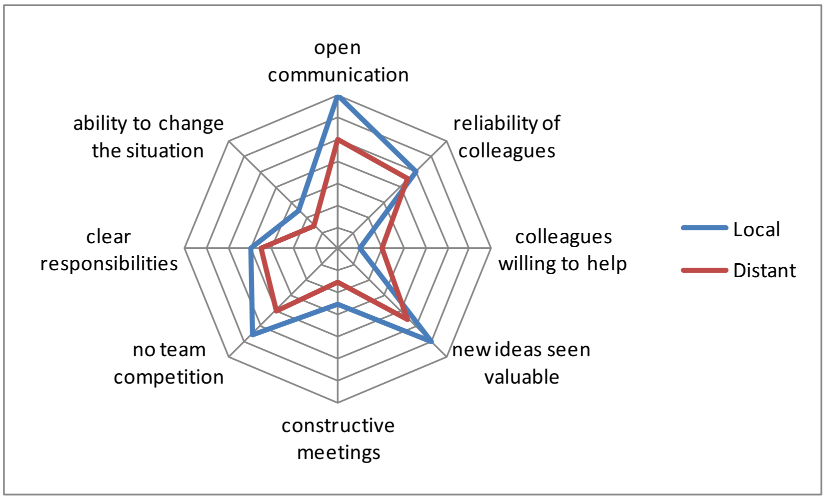

Technische Universität München
How often do you communicate with team members to get your work done?
Lyon
Puebla
Munich
Valparaiso
daily
every 2-3 days
weekly
every two weeks
less than every two weeks
not at all
Copyright © 2010 Christian Lescher, Technische Universität München
6


Technische Universität München
Is it easy for you to reach your colleagues?
Lyon
Puebla
Munich
Valparaiso
very easy to reach
rather easy to reach
rather hard to reach
very hard to reach
not applicable
7
Copyright © 2010 Christian Lescher, Technische Universität München


Technische Universität München
Are you aware of the current set of tasks that your colleagues are currently working on?
Lyon
Puebla
Munich
Valparaiso
yes
rather yes
rather no
no
not applicable
8
Copyright © 2010 Christian Lescher, Technische Universität München

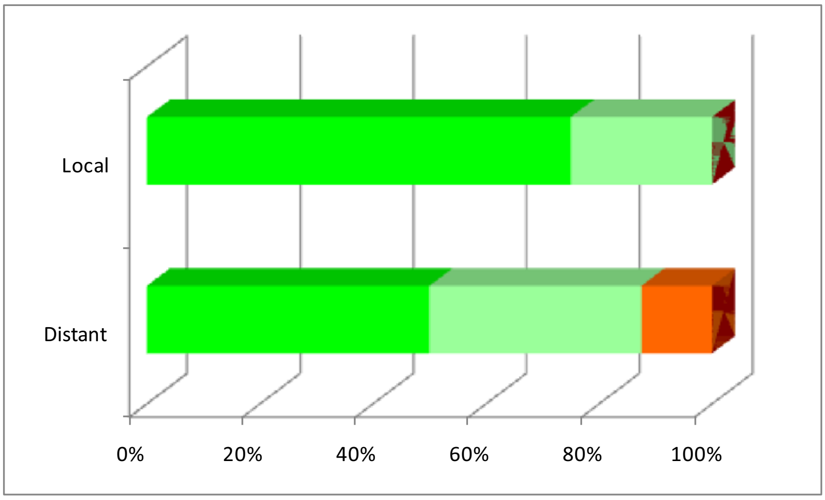
Technische Universität München

Overview: Team Cohesion



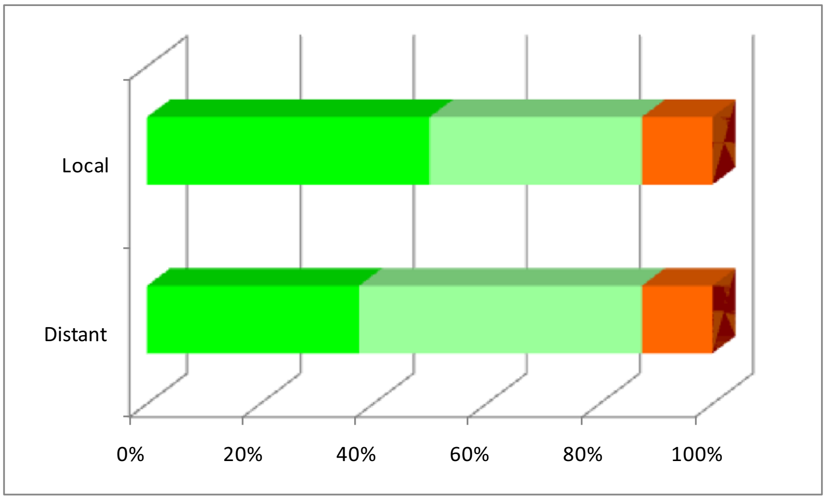
Technische Universität München

Communication inside the project team is honest and open



Technische Universität München

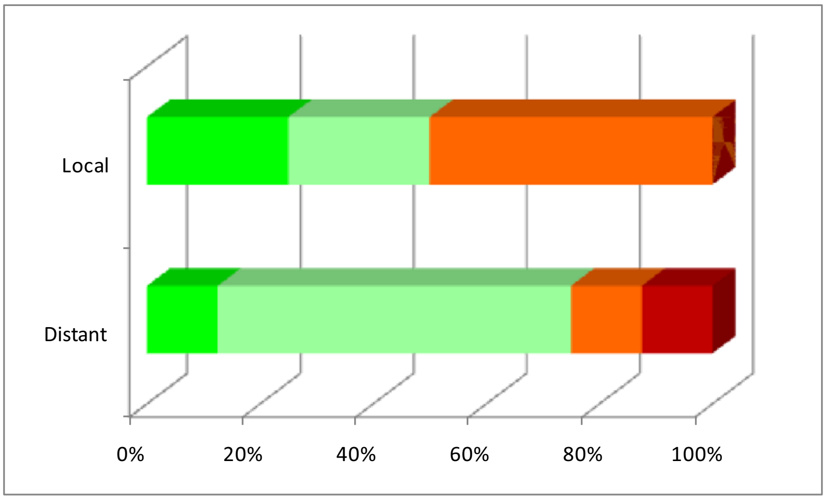
I can rely on my colleagues



fully agree | tend to agree | tend to disagree | fully disagree

Technische Universität München

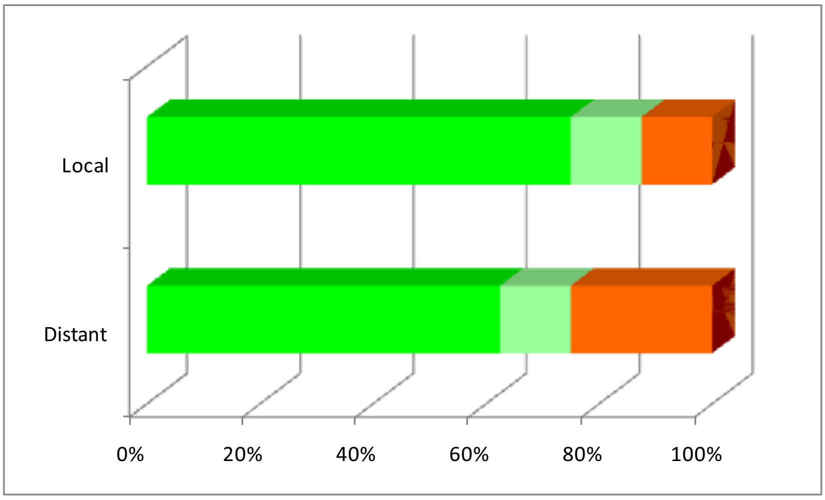
My colleagues assist me with heavy workloads, beyond what they are required to do



fully agree | tend to agree | tend to disagree | fully disagree

Technische Universität München

New ideas of colleagues from the local/distant site are seen as valuable

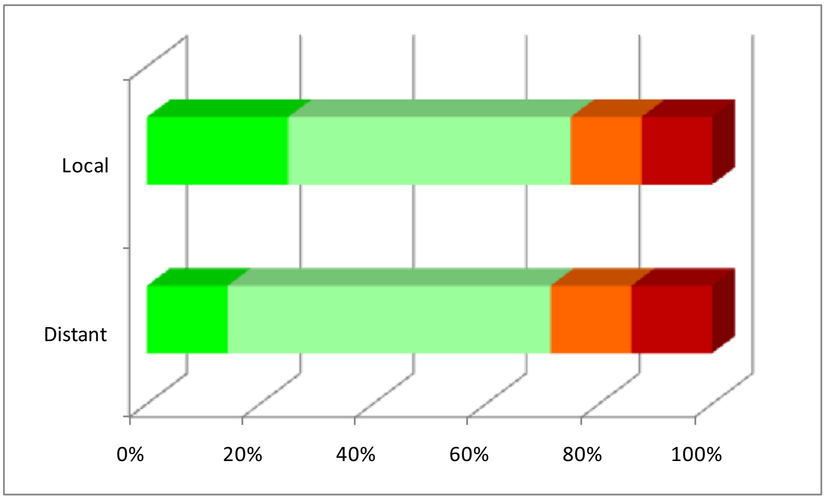


fully agree | tend to agree | tend to disagree | fully disagree

13

Technische Universität München

(Virtual) meetings are constructive

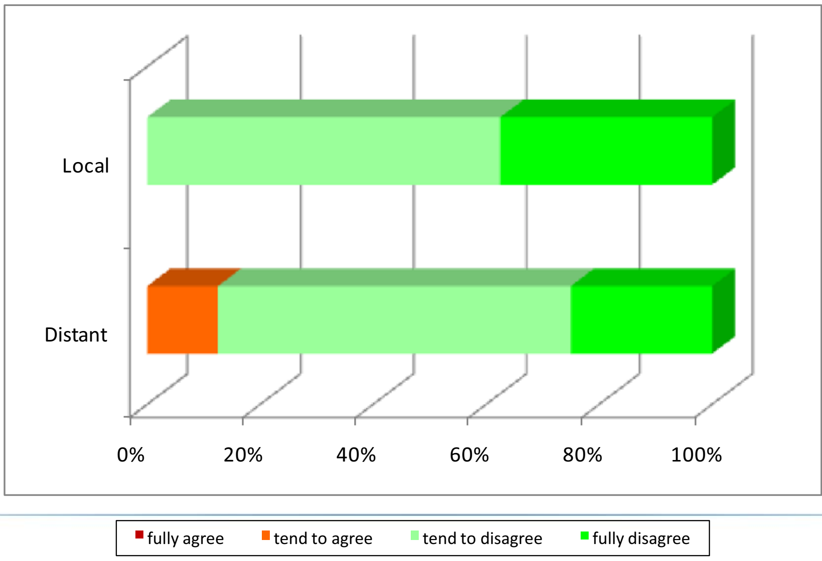


fully agree | tend to agree | tend to disagree | fully disagree

14

Technische Universität München

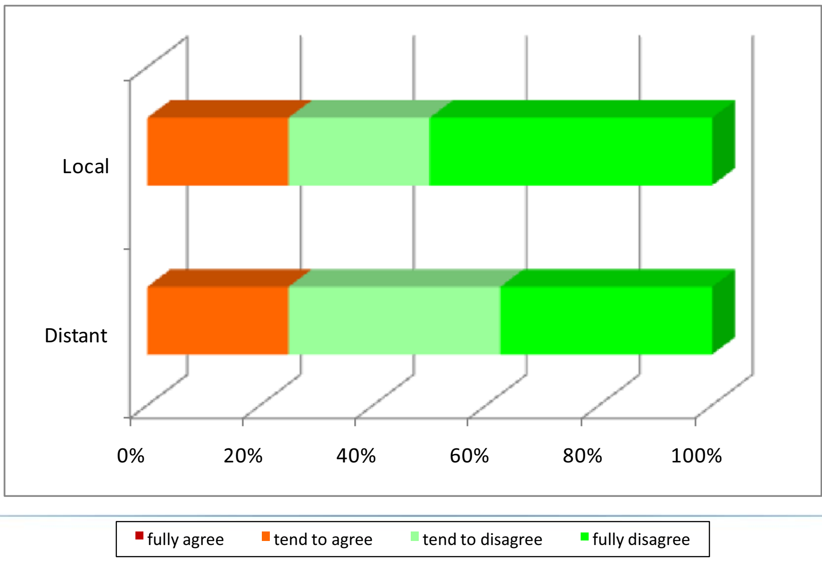
There is individual competition between sub-project teams



15

Technische Universität München

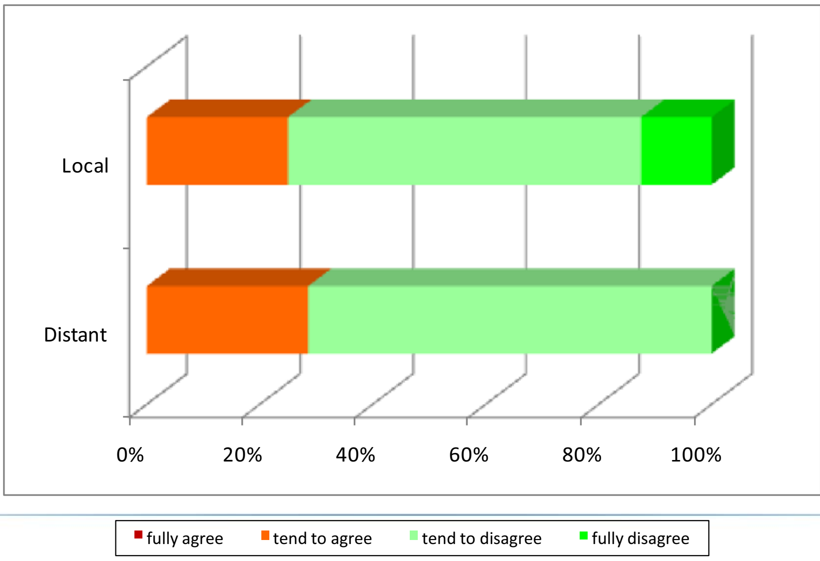
There are many discussions over who has a particular responsibility



16

Technische Universität München

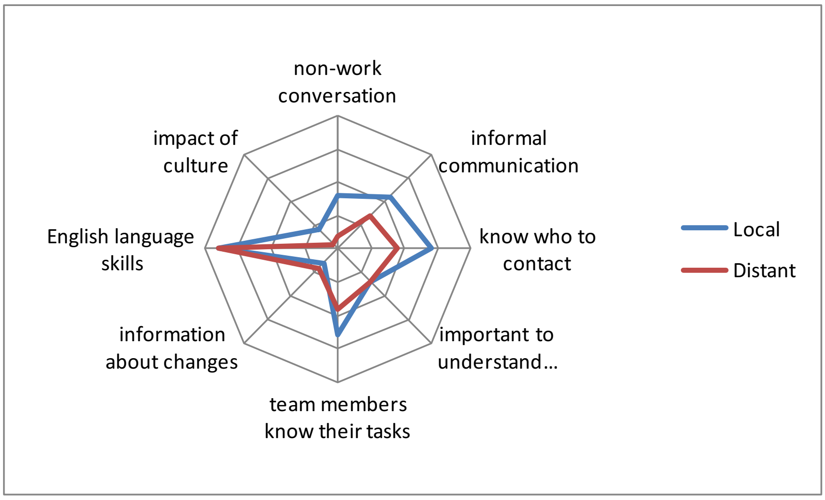
I feel powerless to change the project's situation



17

Technische Universität München

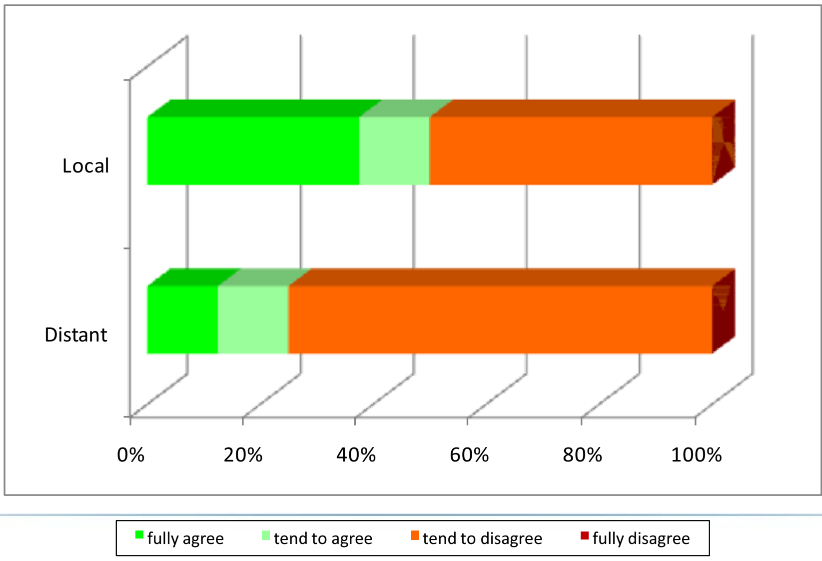
Overview: Communication & Awareness



18

Technische Universität München

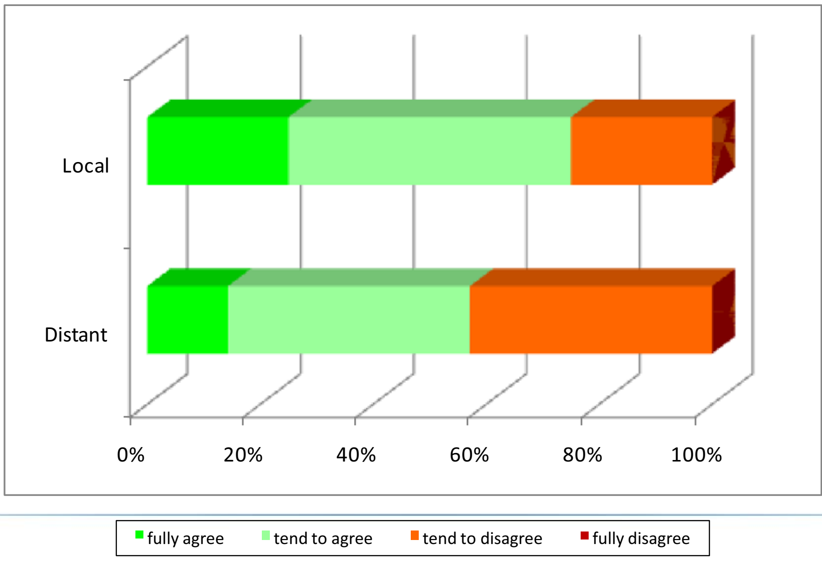
I discuss non-work related matters with other project team members



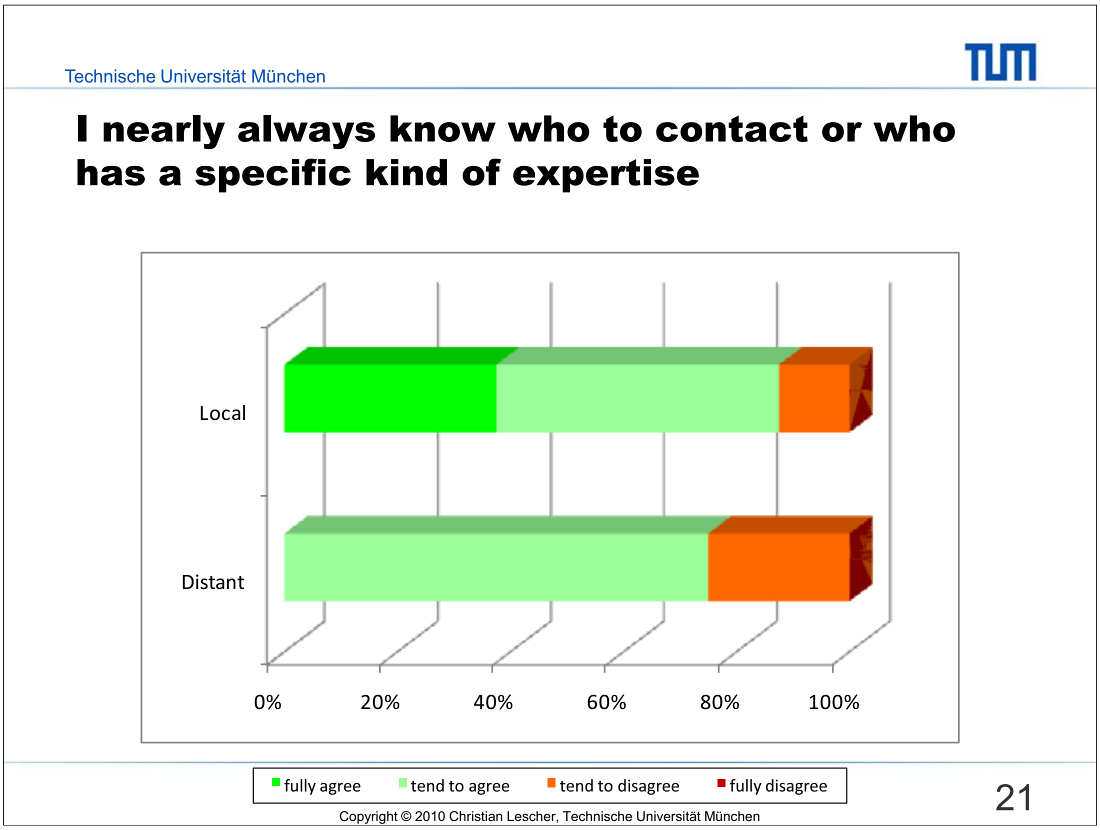
19

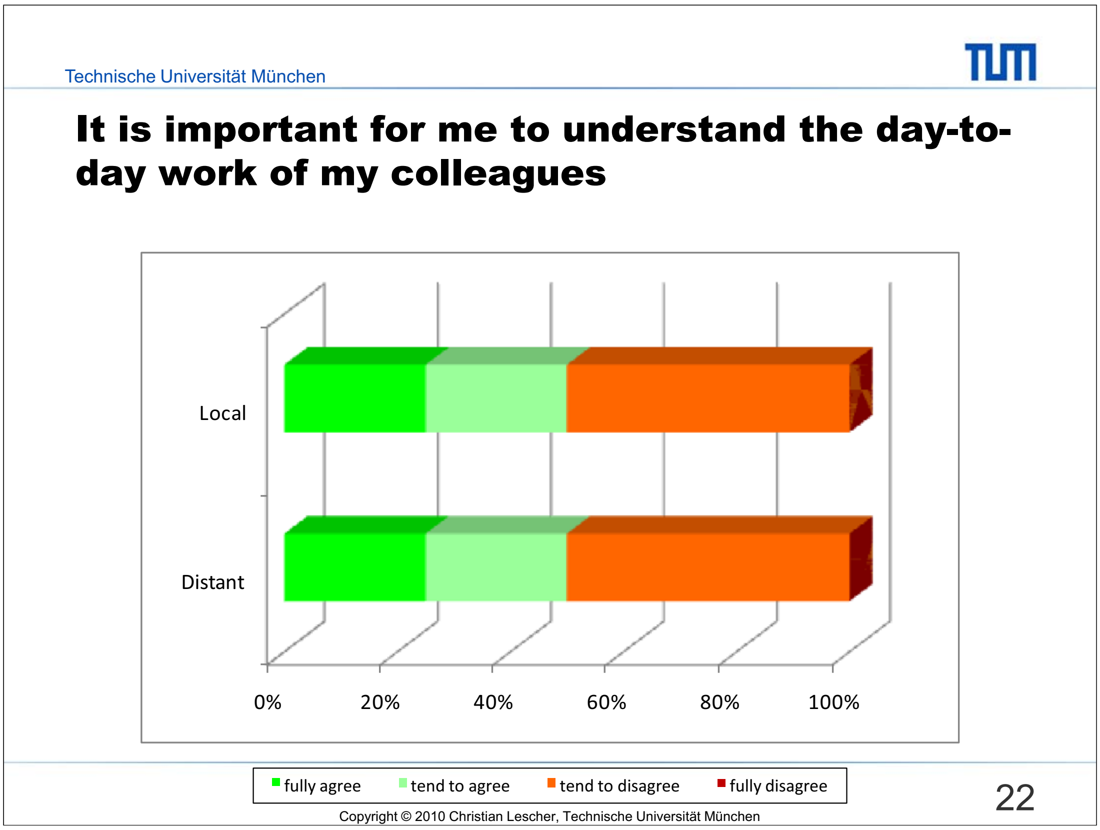
Technische Universität München

I often get useful work-related information through informal communication



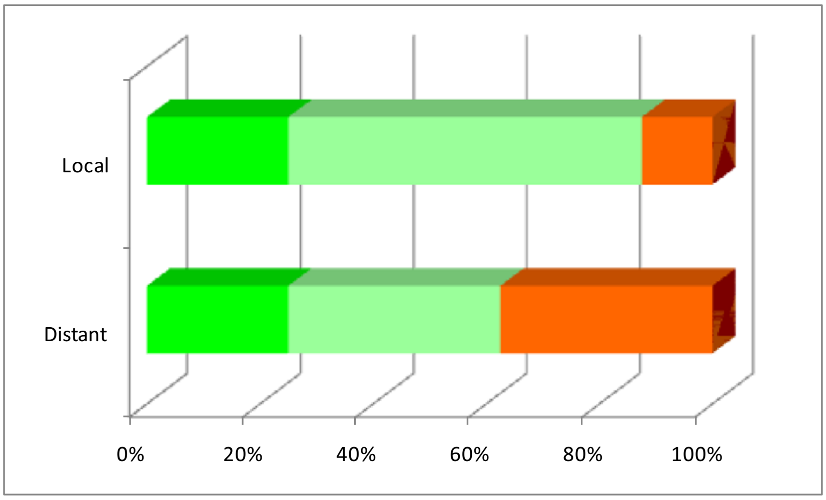
20


Technische Universität München
I nearly always know who to contact or who has a specific kind of expertise
Local
Distant
0%
20%
40%
60%
80%
100%
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
21


Technische Universität München
It is important for me to understand the day-to-day work of my colleagues
Local
Distant
0%
20%
40%
60%
80%
100%
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
22

Technische Universität München

When work is assigned, everyone is clear about his or her task

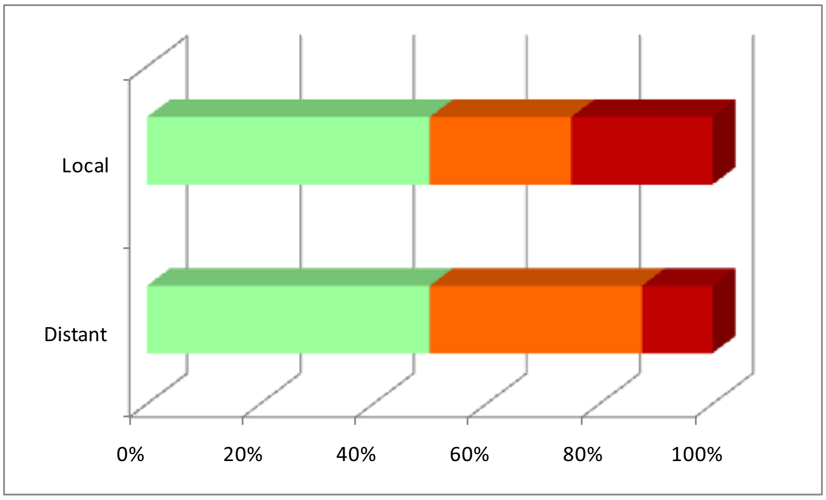


fully agree | tend to agree | tend to disagree | fully disagree

23

Technische Universität München

My colleagues provide timely information about changes in current plans

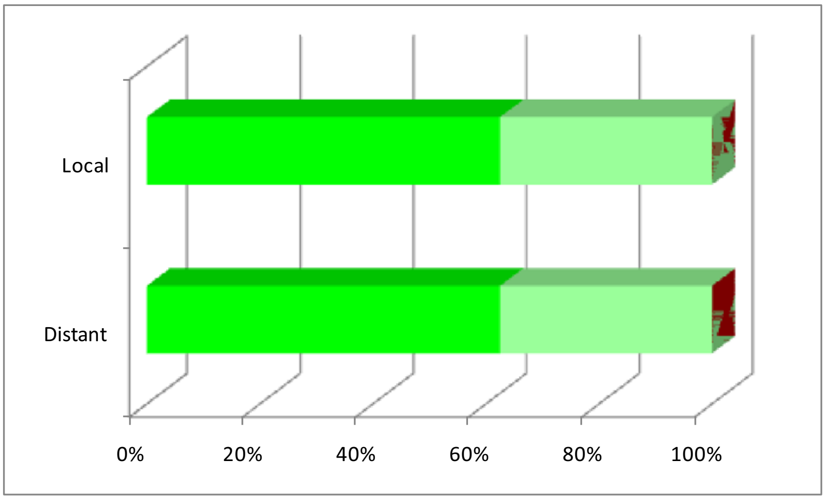


fully agree | tend to agree | tend to disagree | fully disagree

24

Technische Universität München

Project team members have appropriate English language skills

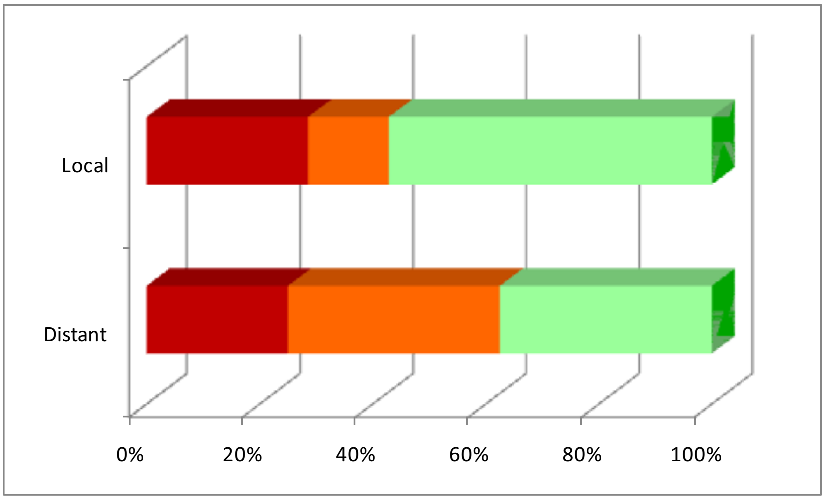


fully agree tend to agree tend to disagree fully disagree

25

Technische Universität München

National culture has significant influence on team collaboration

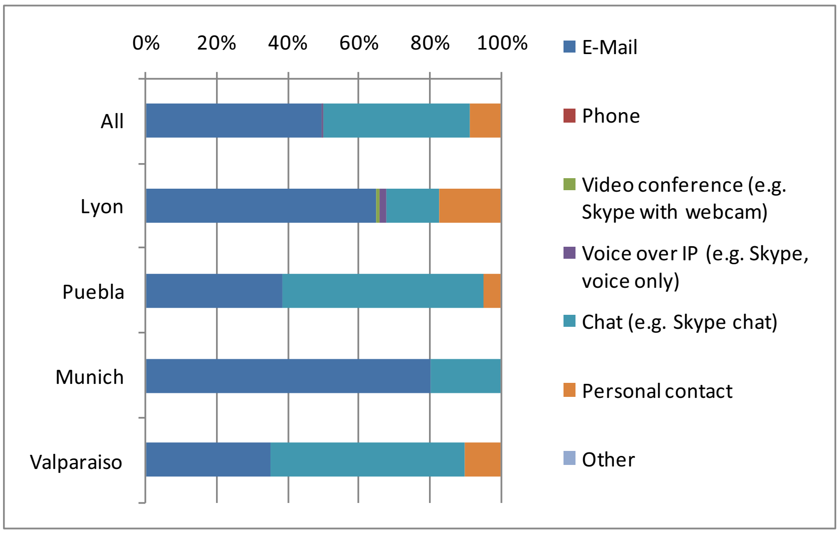


fully agree tend to agree tend to disagree fully disagree

26

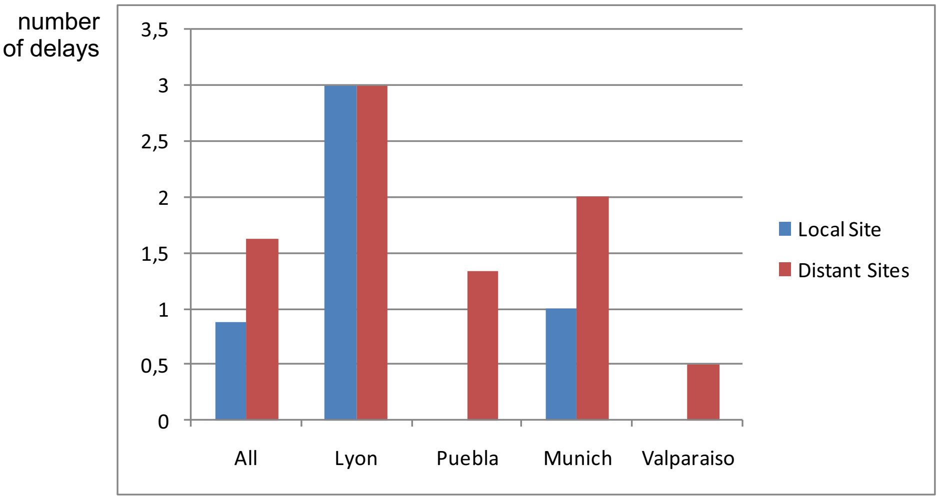
Technische Universität München

How many percent of your communication is done via the following media?



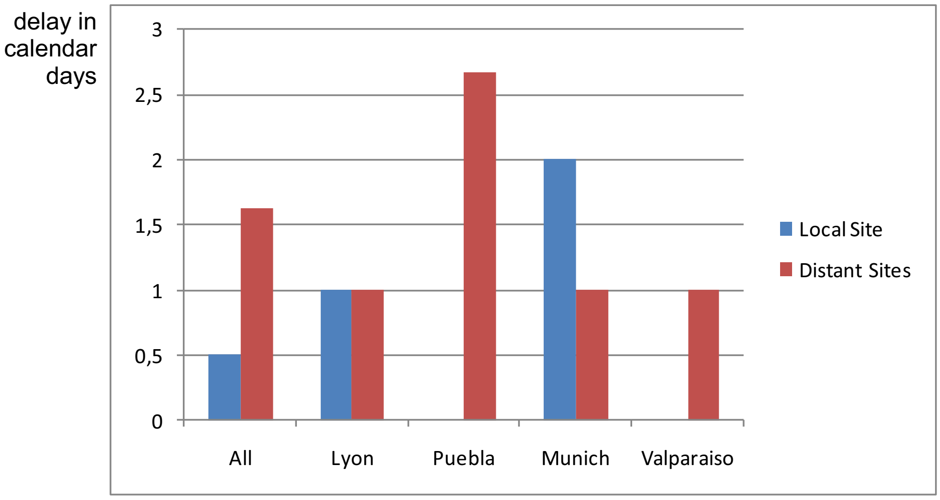
Technische Universität München

How many times in the past two weeks was your own work delayed?



Technische Universität München

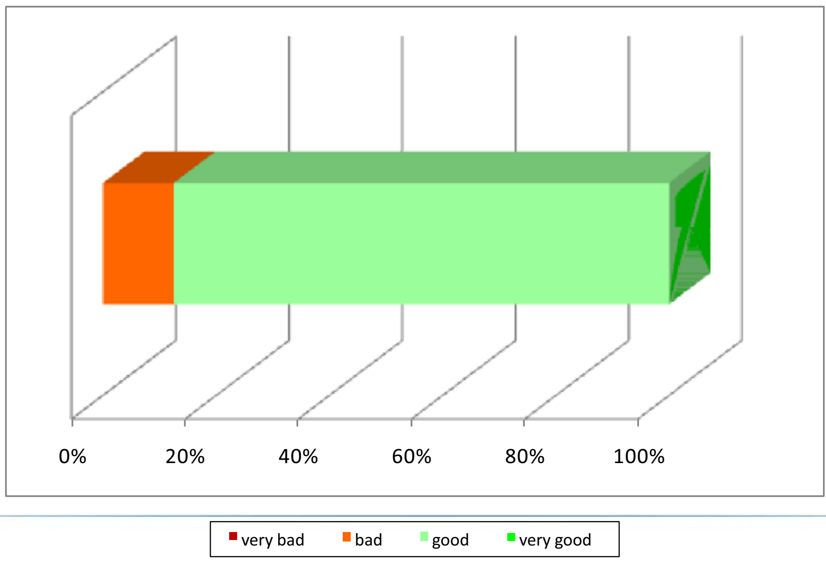
What was the average length of delays you experienced?



29

Technische Universität München

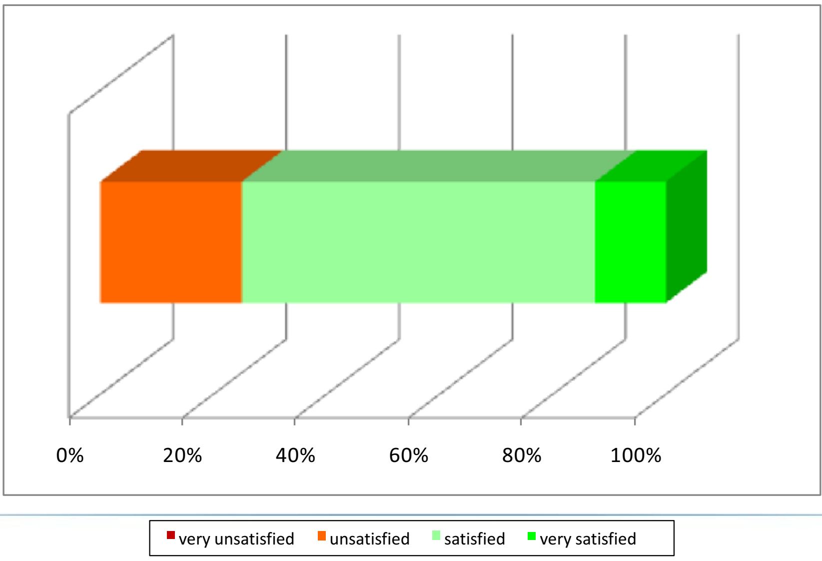
How well do you feel in your project team?



30

Technische Universität München

How satisfied are you with the project's achievements so far?



Technische Universität München

Comments from the survey

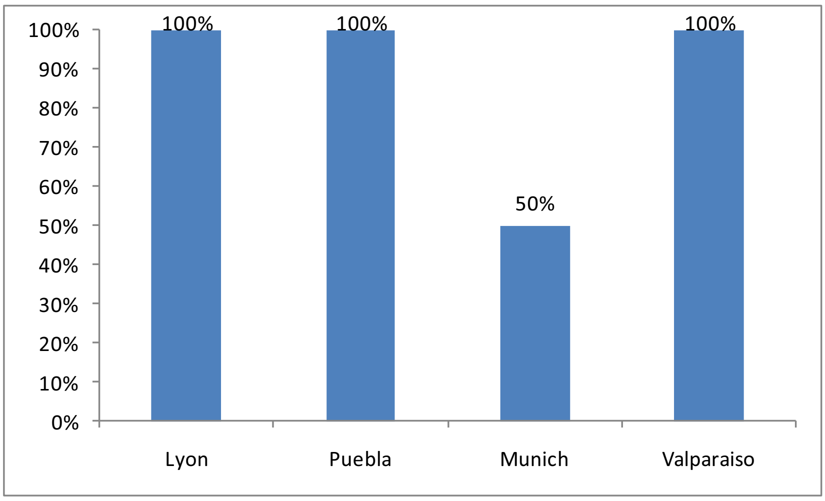
- It goes much better than at the moment of the last survey...
- It is a valuable experience for me, not only in aspect technique, but also in project management.

Technische Universität München

Backup

Technische Universität München

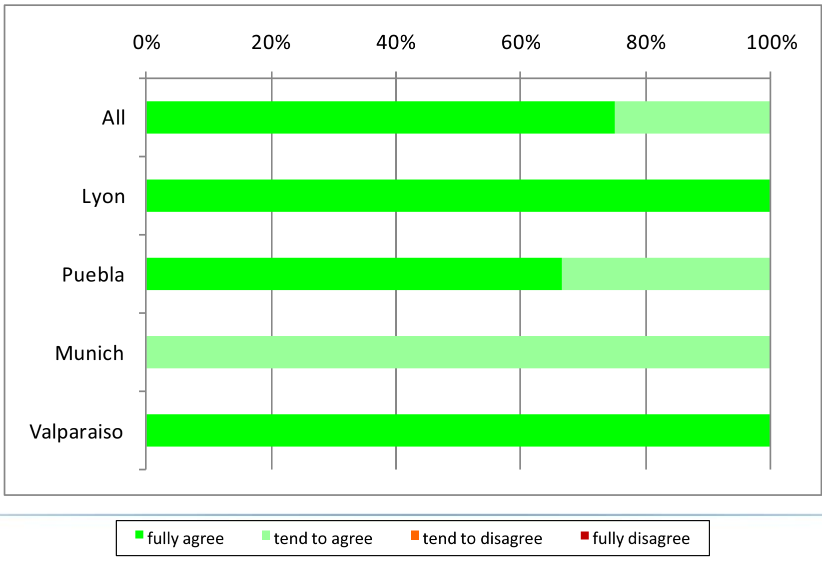
Survey Responses: Response Rate



Technische Universität München

Communication inside the project team is honest and open

Breakdown of 'Local Site' answers

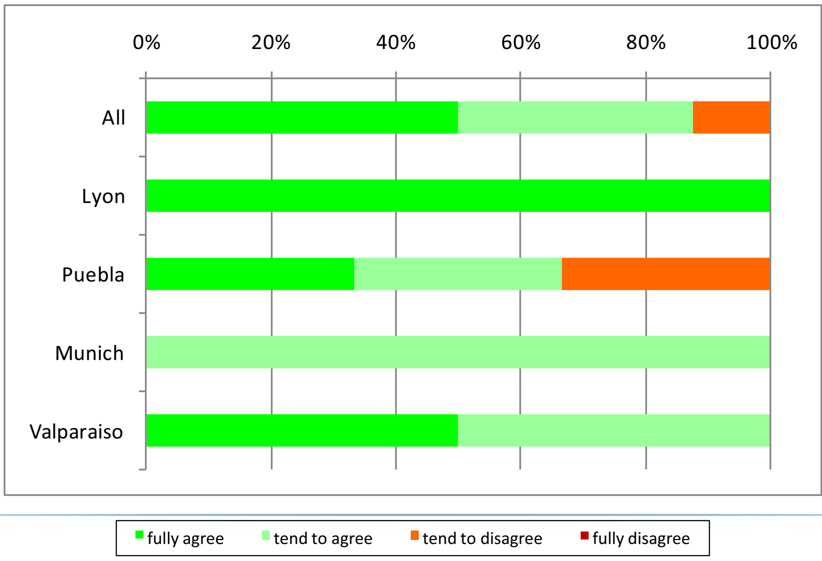


35

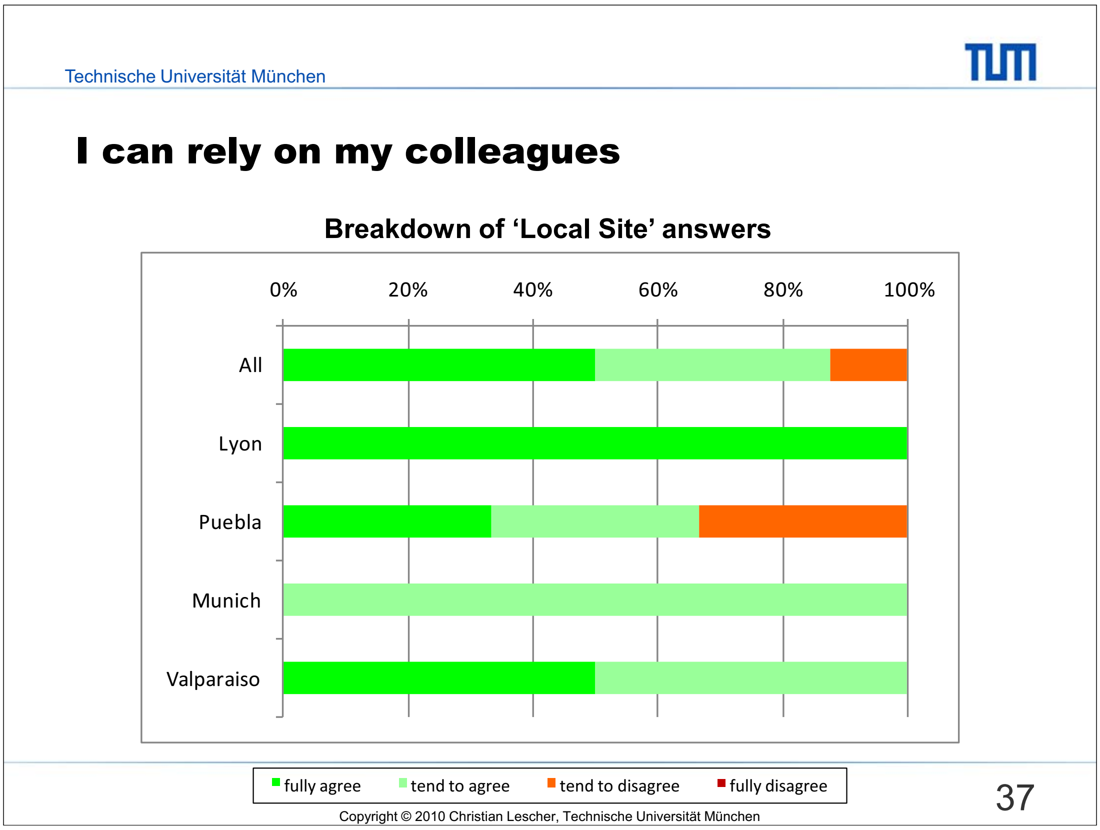
Technische Universität München

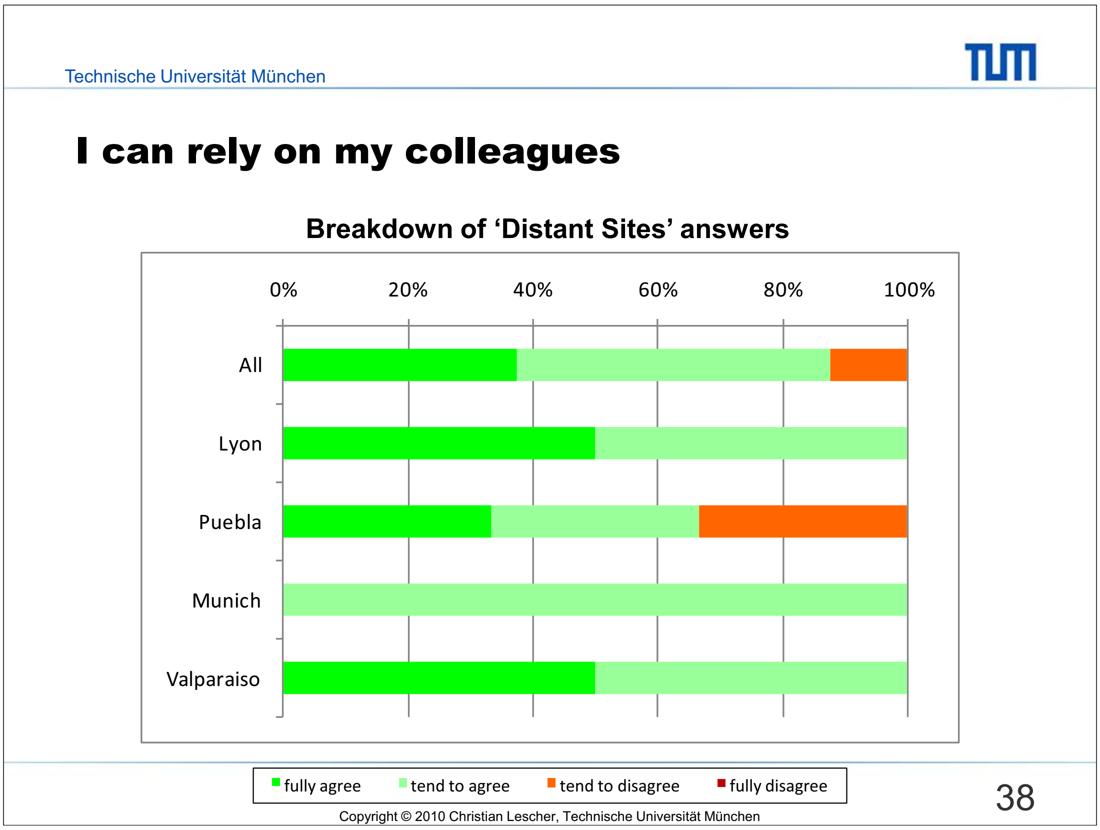
Communication inside the project team is honest and open

Breakdown of 'Distant Sites' answers



36

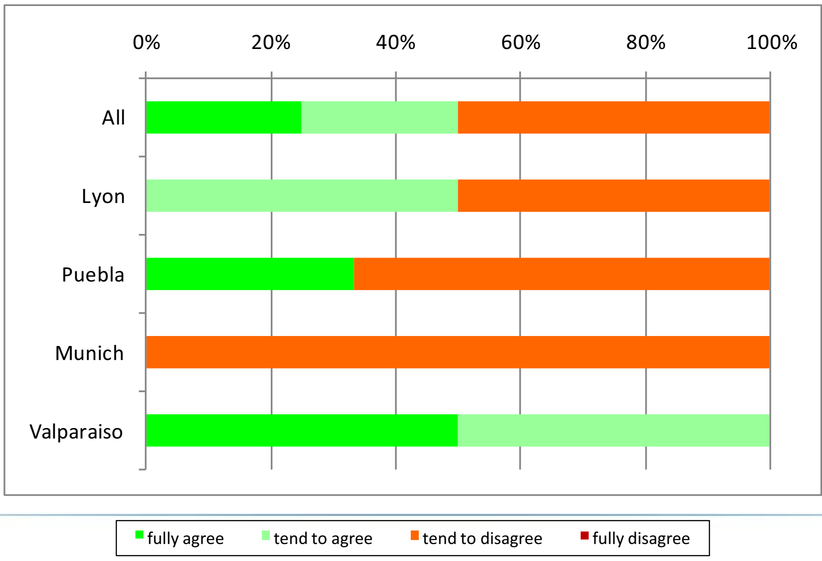

Technische Universität München
I can rely on my colleagues
Breakdown of 'Local Site' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
37


Technische Universität München
I can rely on my colleagues
Breakdown of 'Distant Sites' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
38

Technische Universität München

My colleagues assist me with heavy workloads, beyond what they are required to do

Breakdown of 'Local Site' answers

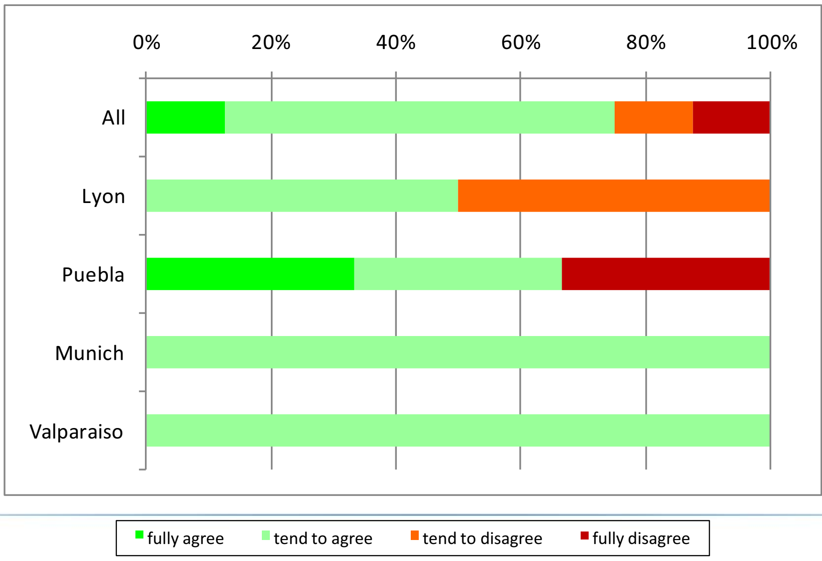


39

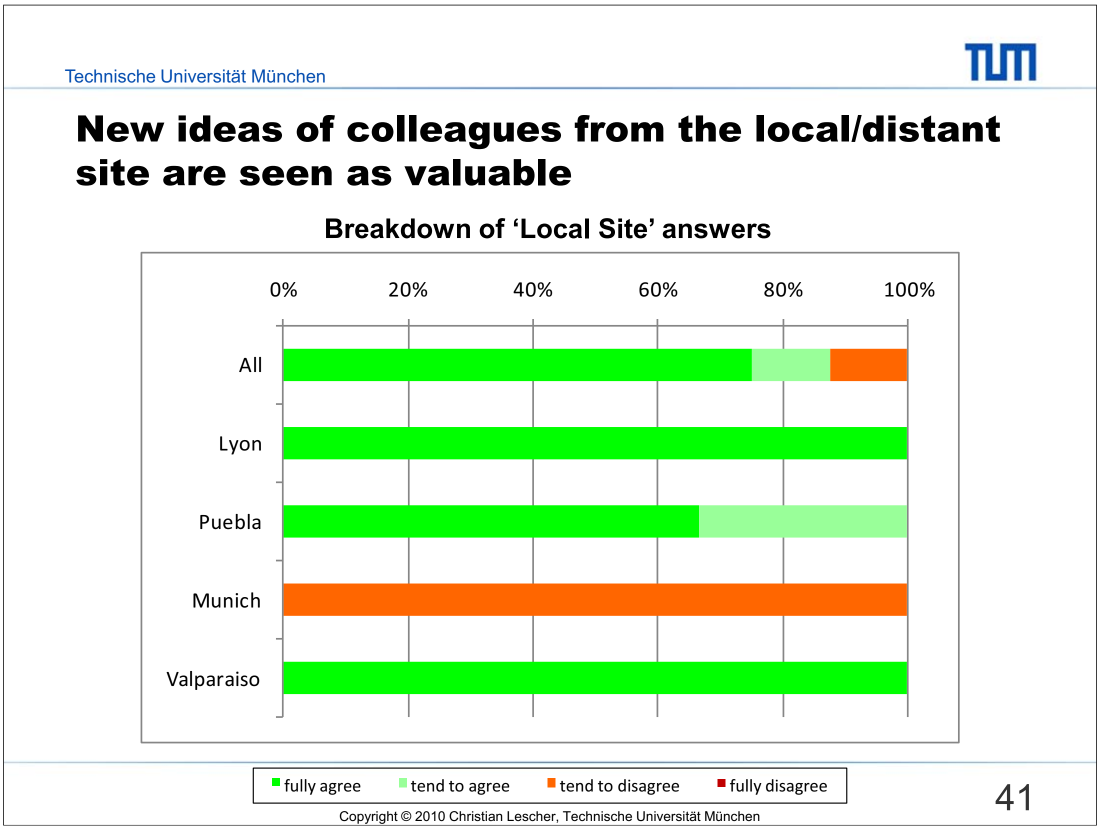
Technische Universität München

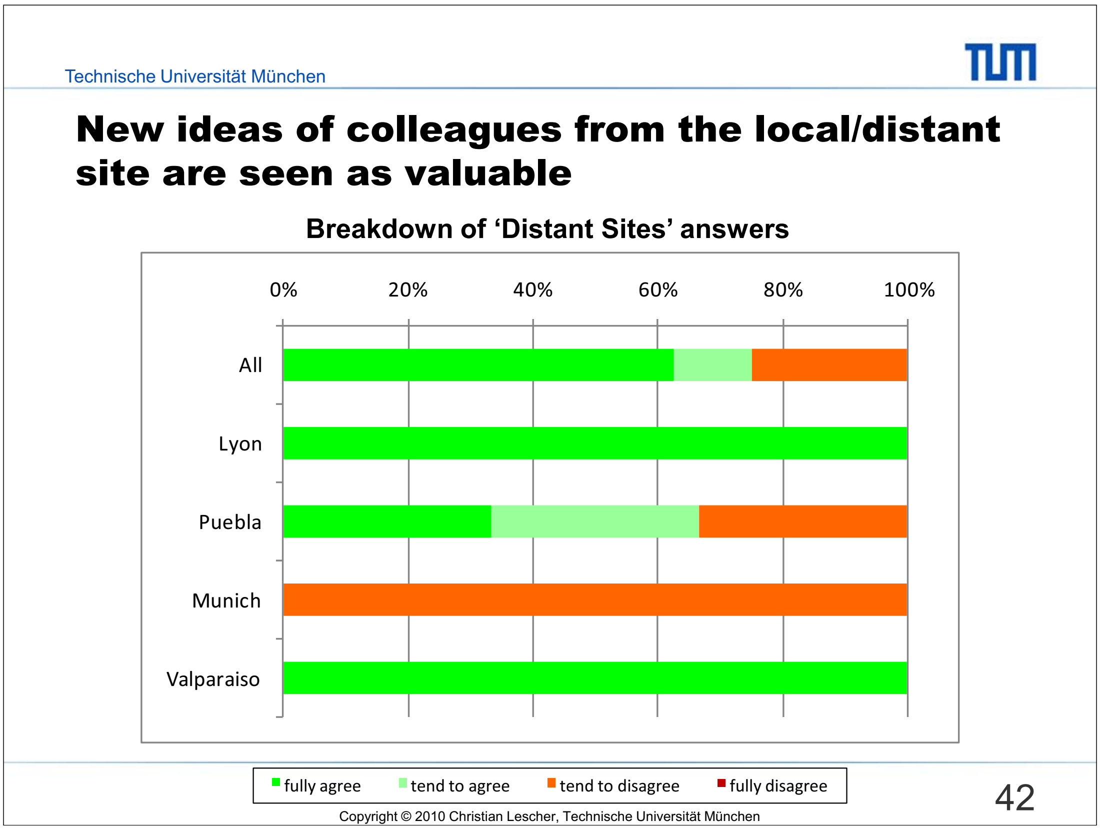
My colleagues assist me with heavy workloads, beyond what they are required to do

Breakdown of 'Distant Sites' answers



40

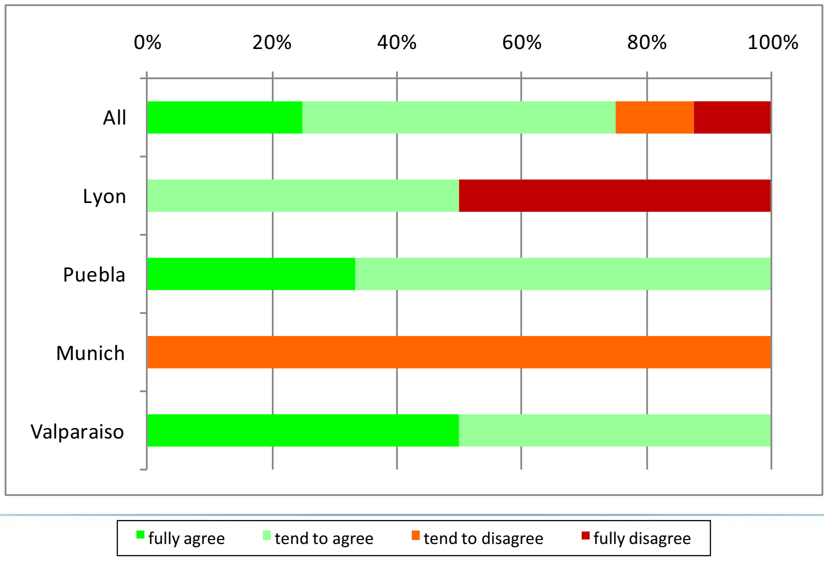

Technische Universität München
New ideas of colleagues from the local/distant site are seen as valuable
Breakdown of 'Local Site' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
41


Technische Universität München
New ideas of colleagues from the local/distant site are seen as valuable
Breakdown of 'Distant Sites' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
42

Technische Universität München

(Virtual) meetings are constructive

Breakdown of 'Local Site' answers

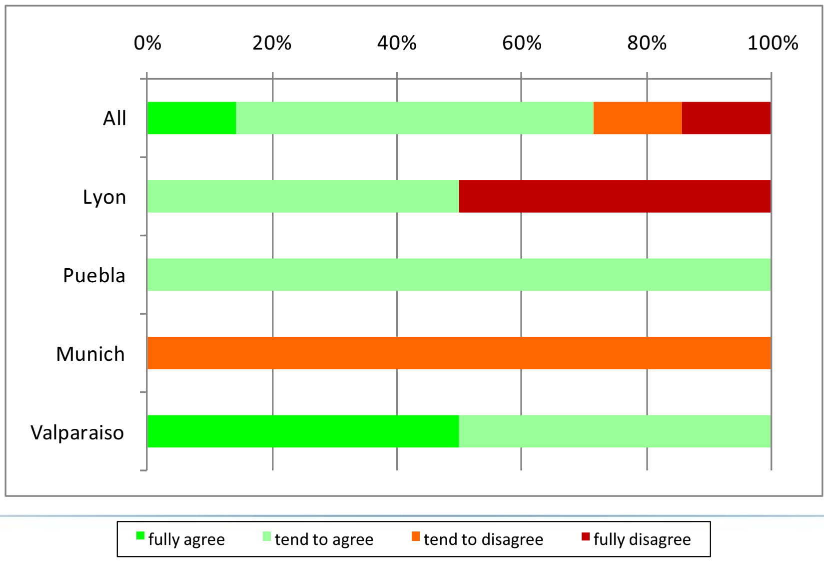


43

Technische Universität München

(Virtual) meetings are constructive

Breakdown of 'Distant Sites' answers

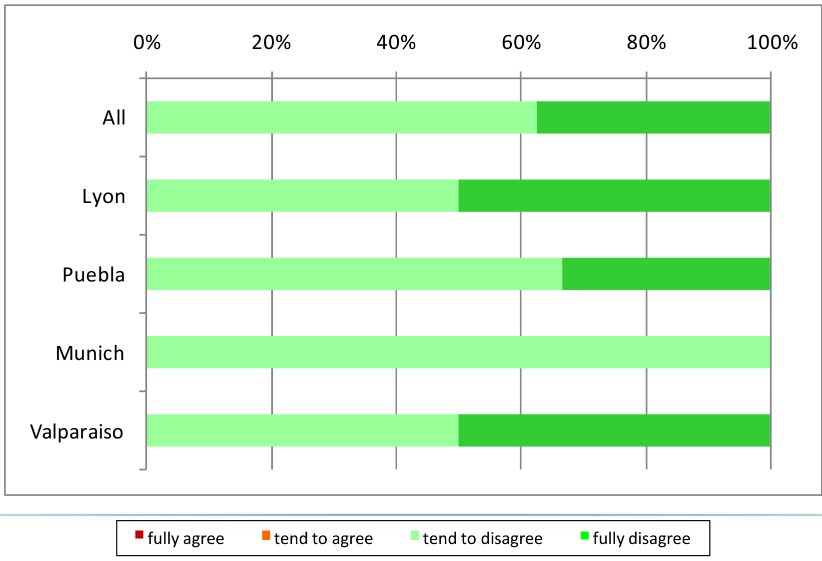


44

Technische Universität München

There is individual competition between sub-project teams

Breakdown of 'Local Site' answers

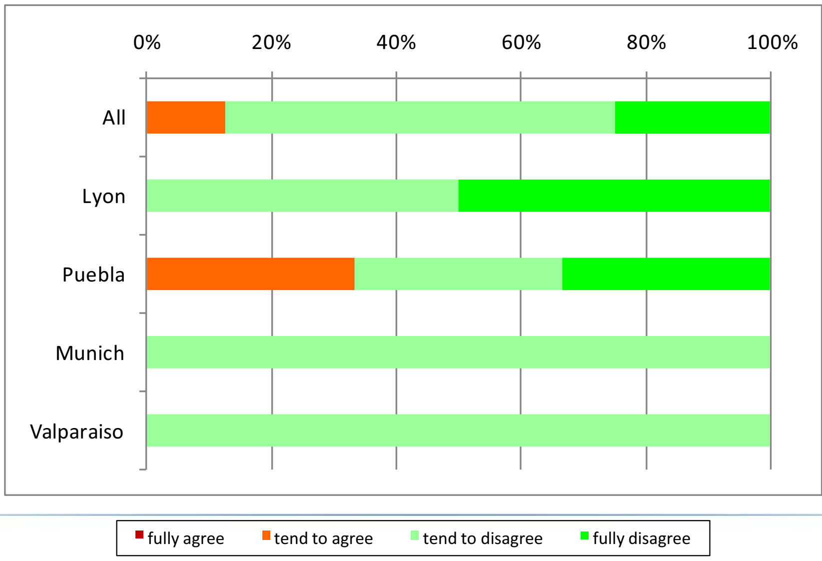


45

Technische Universität München

There is individual competition between sub-project teams

Breakdown of 'Distant Sites' answers

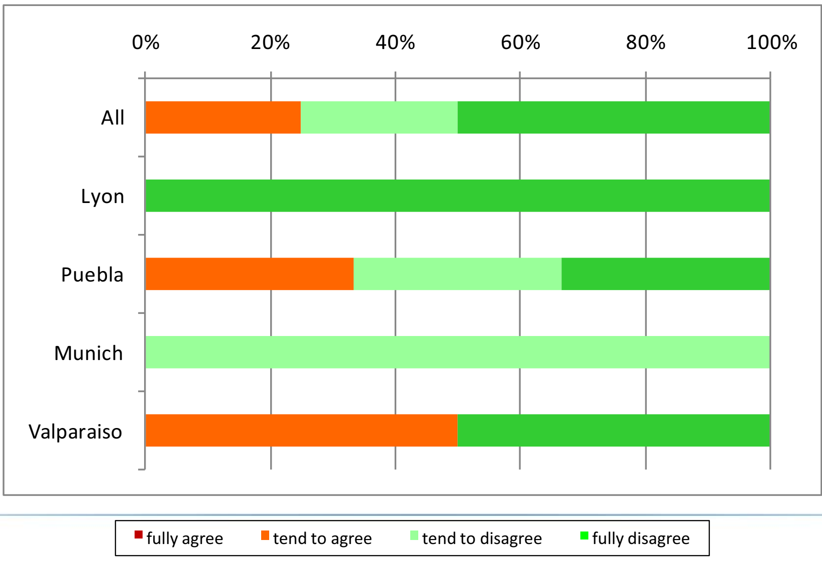


46

Technische Universität München

There are many discussions over who has a particular responsibility

Breakdown of 'Local Site' answers

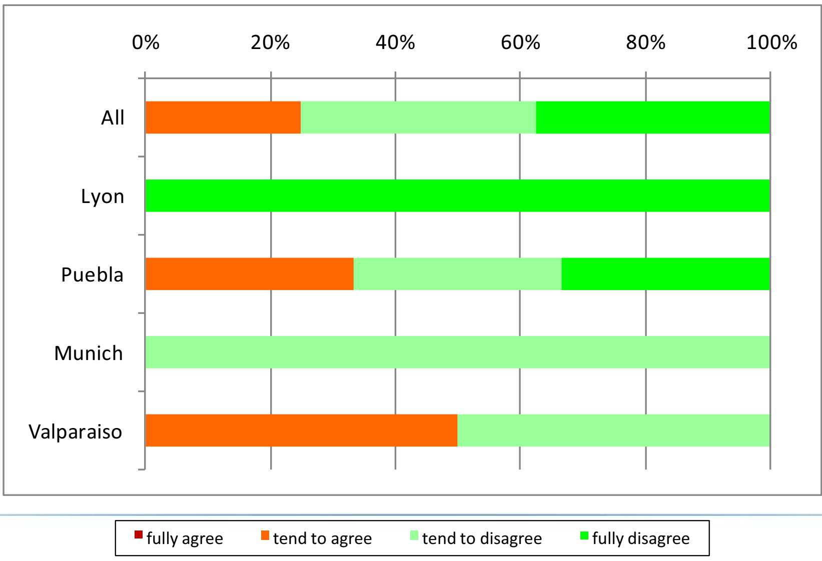


47

Technische Universität München

There are many discussions over who has a particular responsibility

Breakdown of 'Distant Sites' answers

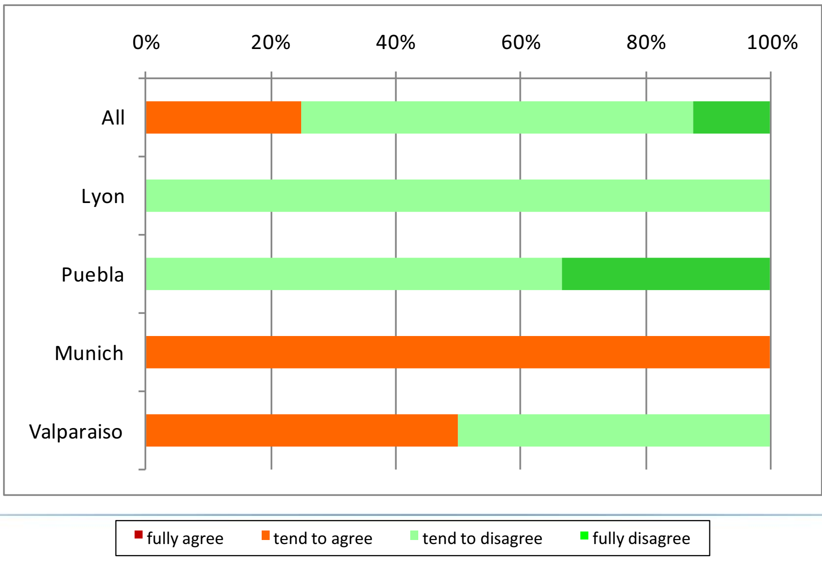


48

Technische Universität München

I feel powerless to change the project's situation

Breakdown of 'Local Site' answers

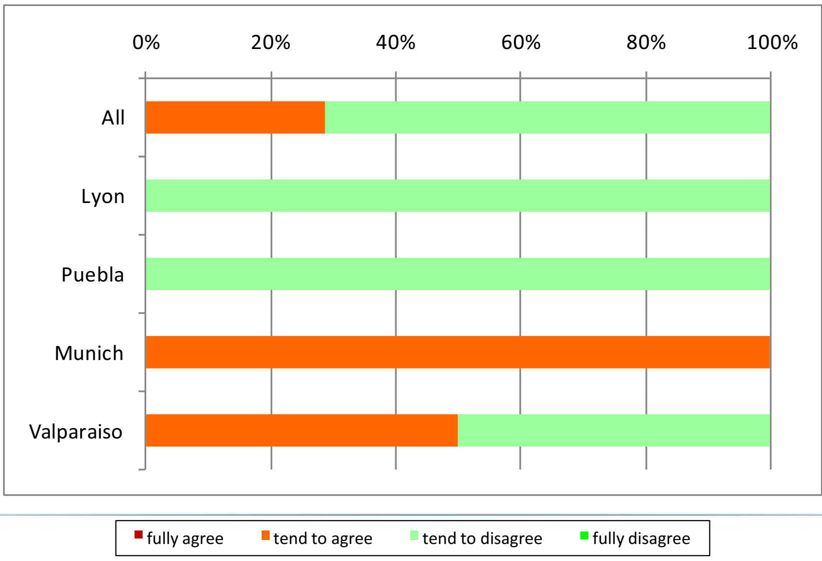


49

Technische Universität München

I feel powerless to change the project's situation

Breakdown of 'Distant Sites' answers

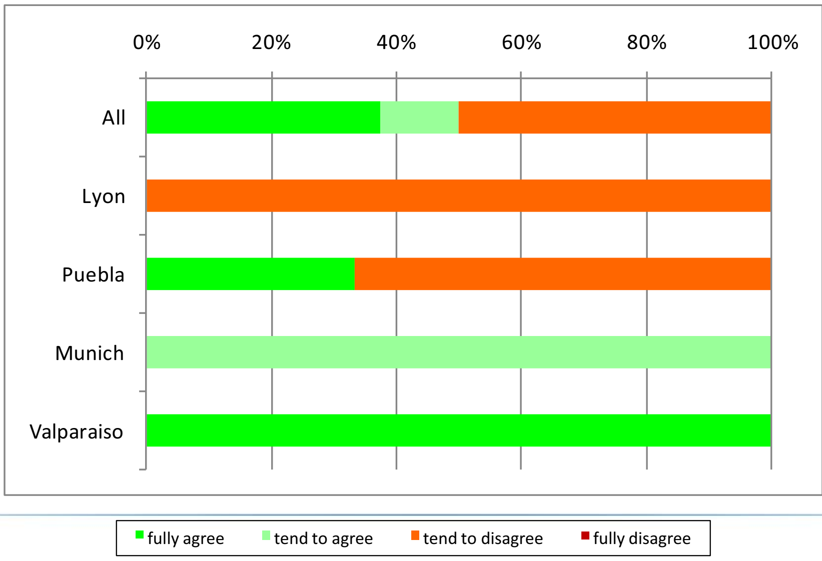


50

Technische Universität München

I discuss non-work related matters with other project team members

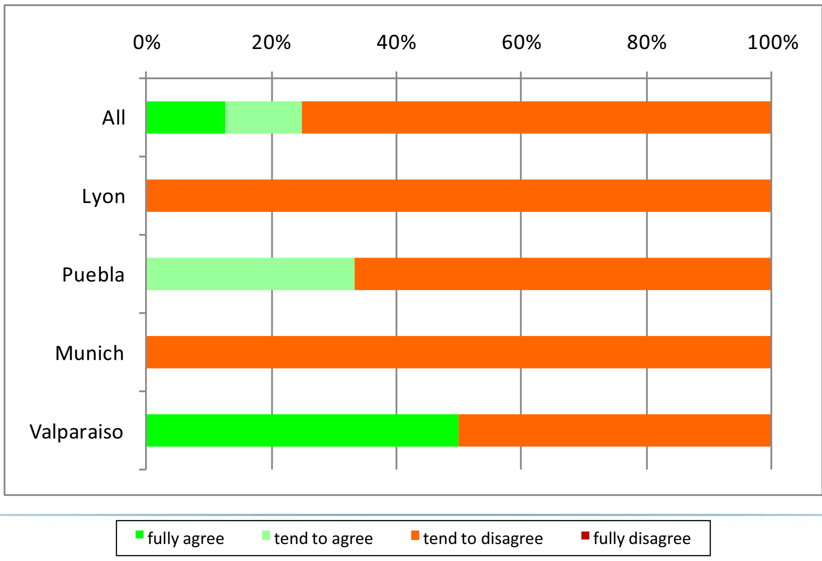
Breakdown of 'Local Site' answers



Technische Universität München

I discuss non-work related matters with other project team members

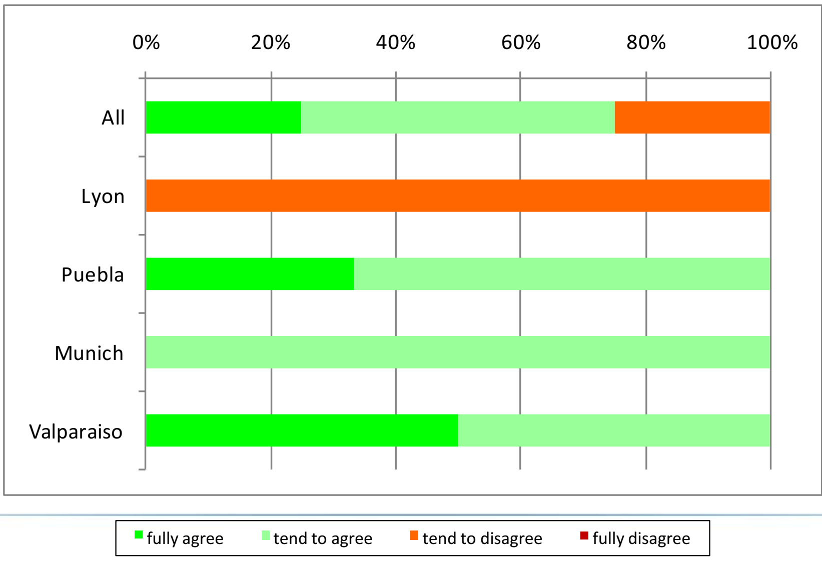
Breakdown of 'Distant Sites' answers



Technische Universität München

I often get useful work-related information through informal communication

Breakdown of ‘Local Site’ answers

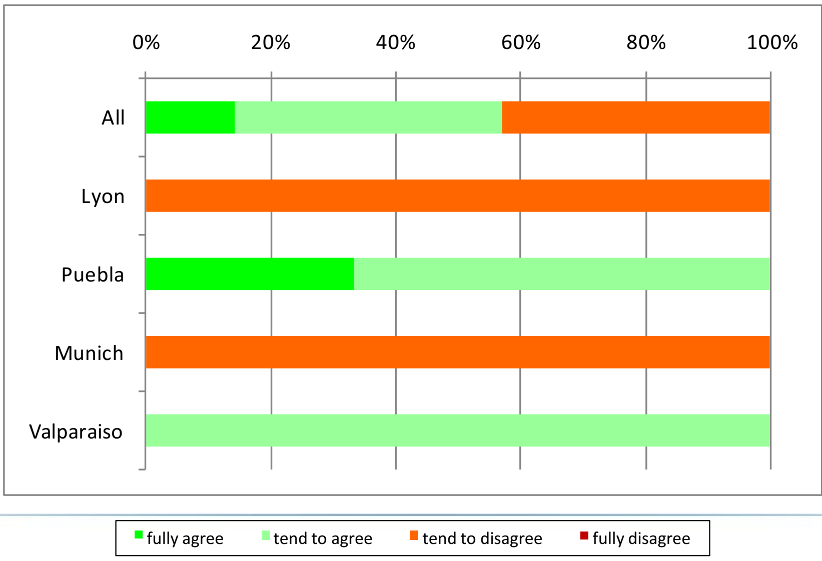


53

Technische Universität München

I often get useful work-related information through informal communication

Breakdown of ‘Distant Sites’ answers



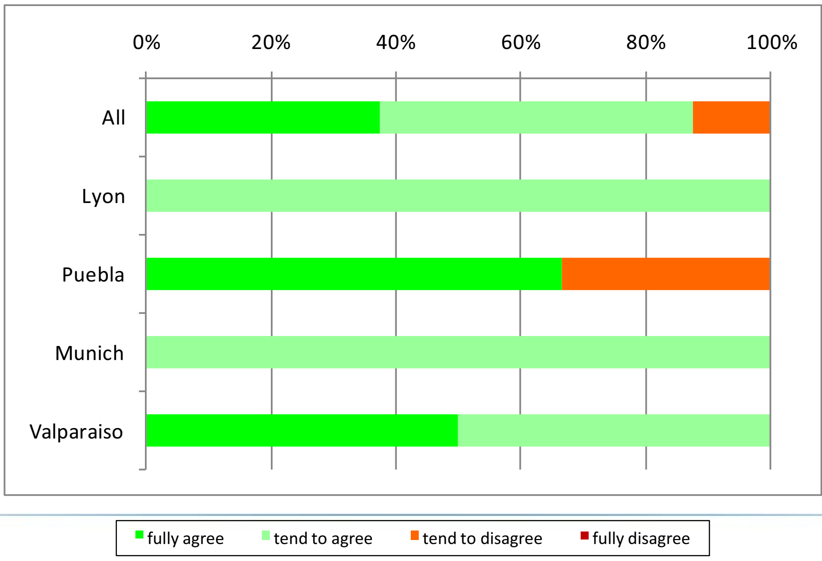
54

Copyright © 2010 Christian Lescher, Technische Universität München

Technische Universität München

I nearly always know who to contact or who has a specific kind of expertise

Breakdown of 'Local Site' answers

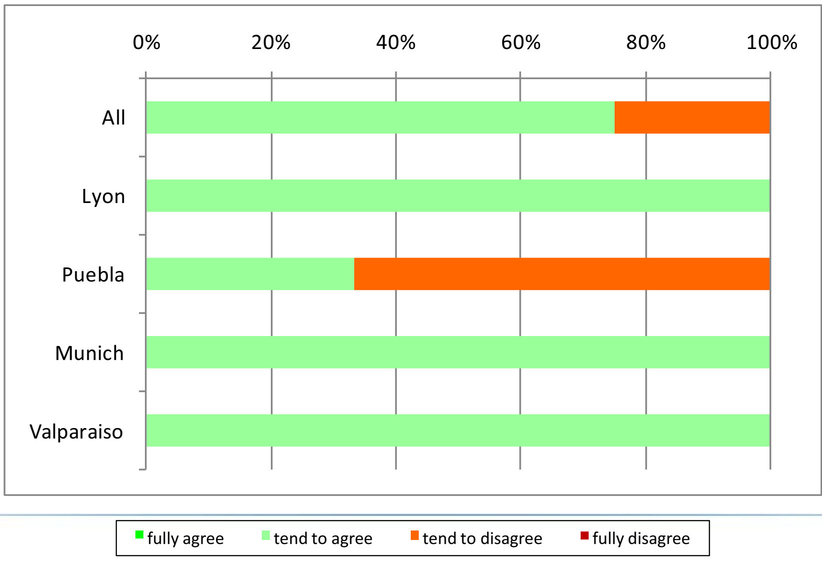


55

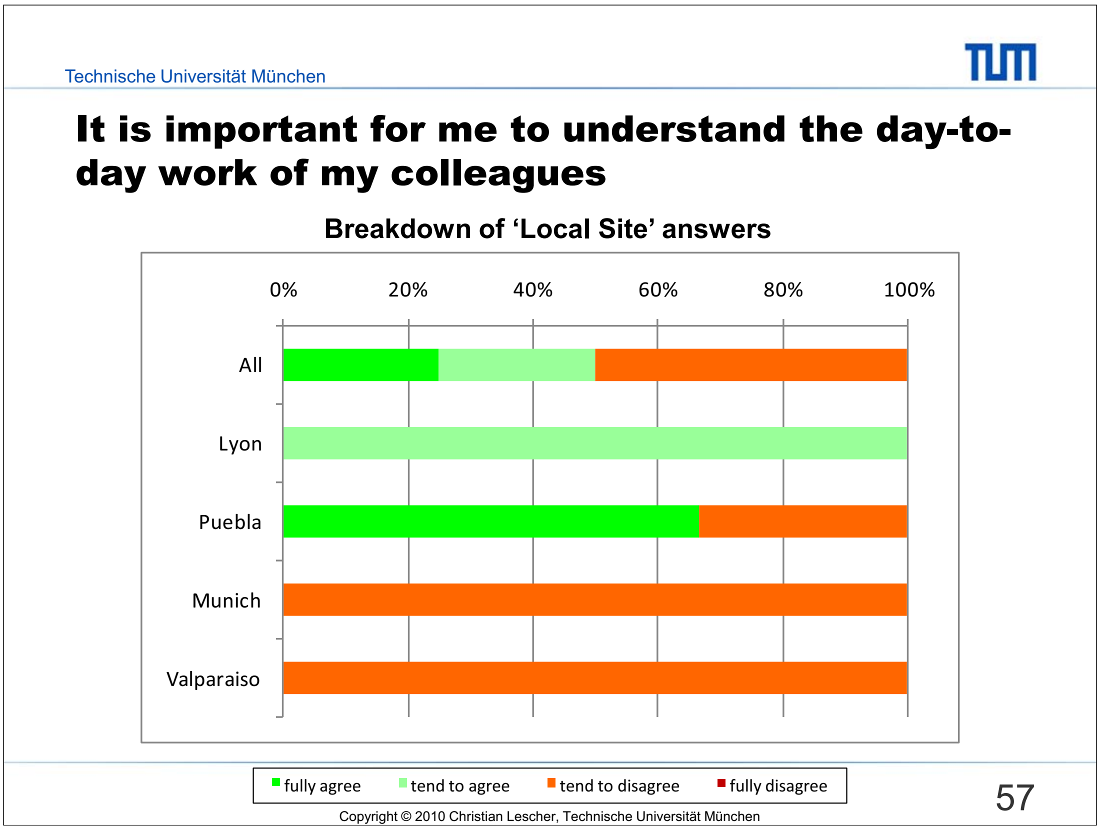
Technische Universität München

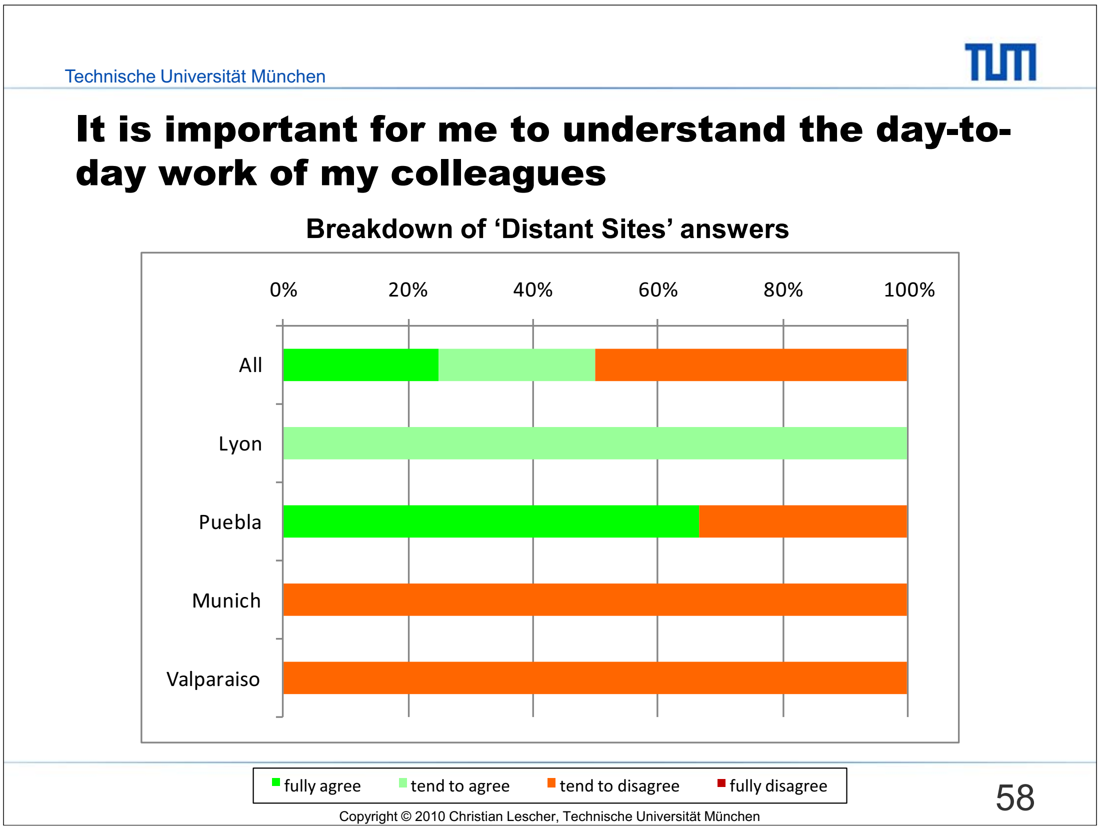
I nearly always know who to contact or who has a specific kind of expertise

Breakdown of 'Distant Sites' answers



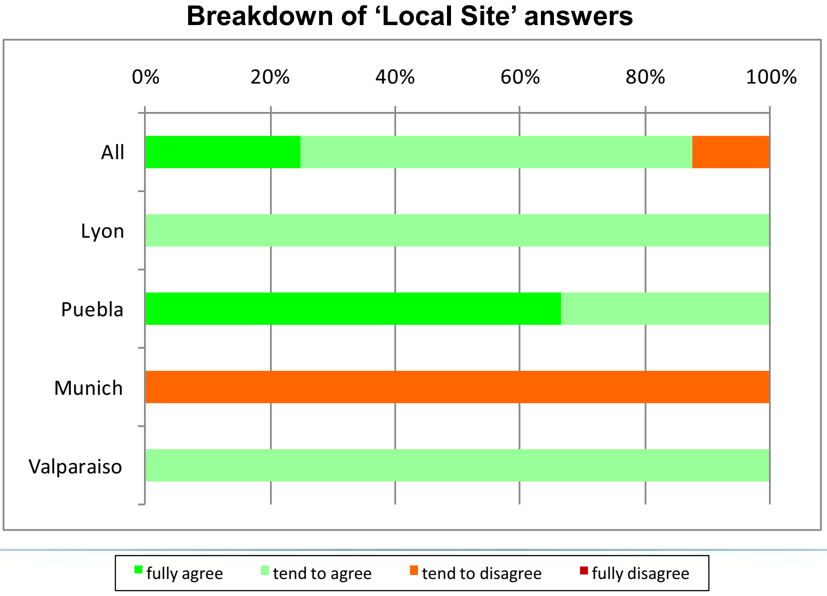
56


Technische Universität München
It is important for me to understand the day-to-day work of my colleagues
Breakdown of 'Local Site' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
57


Technische Universität München
It is important for me to understand the day-to-day work of my colleagues
Breakdown of 'Distant Sites' answers
0%
20%
40%
60%
80%
100%
All
Lyon
Puebla
Munich
Valparaiso
fully agree
tend to agree
tend to disagree
fully disagree
Copyright © 2010 Christian Lescher, Technische Universität München
58

Technische Universität München

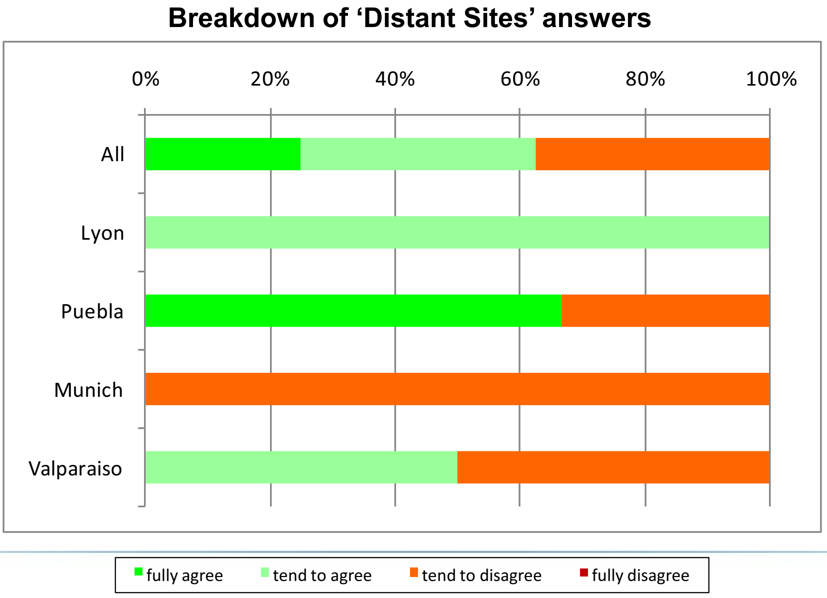
When work is assigned, everyone is clear about his or her task



59

Technische Universität München

When work is assigned, everyone is clear about his or her task

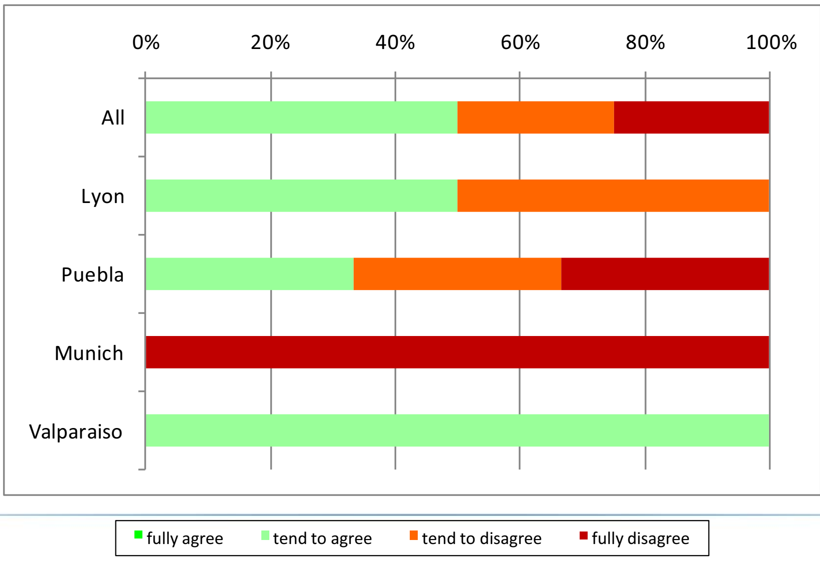


60

Technische Universität München

My colleagues provide timely information about changes in current plans

Breakdown of 'Local Site' answers

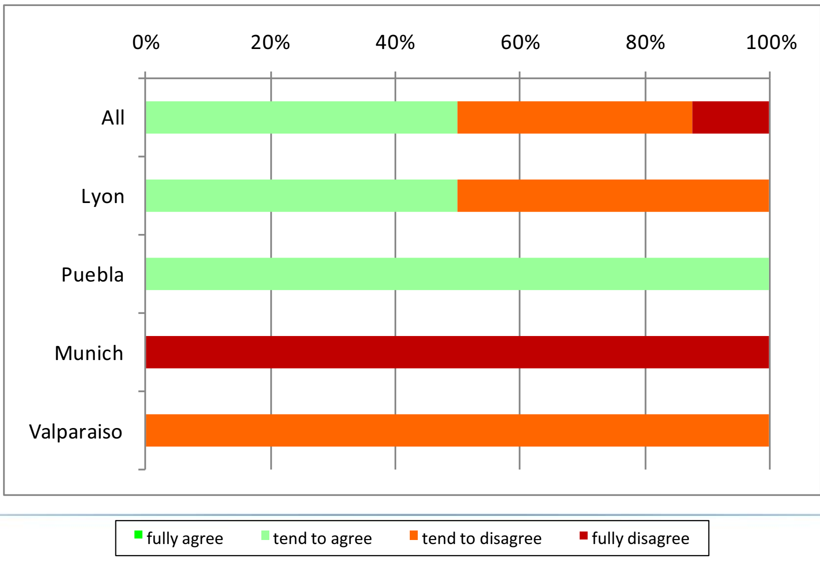


61

Technische Universität München

My colleagues provide timely information about changes in current plans

Breakdown of 'Distant Sites' answers

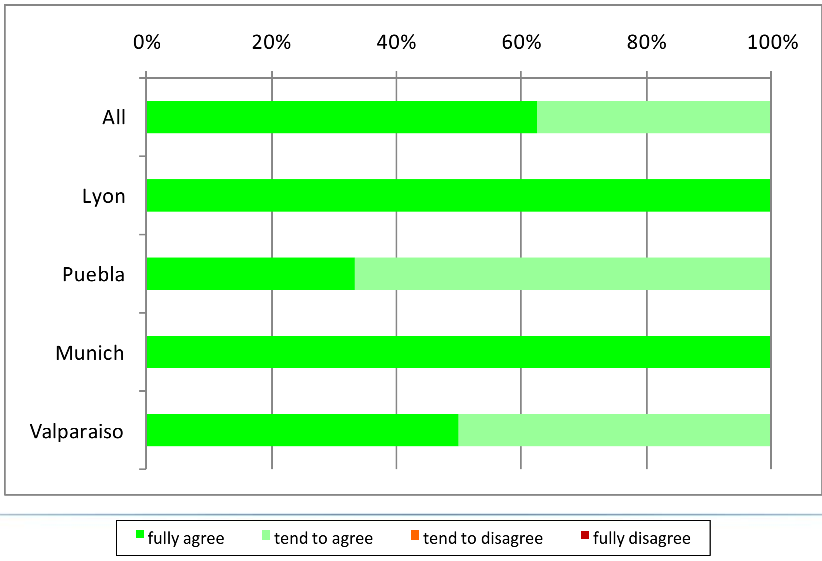


62

Technische Universität München

Project team members have appropriate English language skills

Breakdown of ‘Local Site’ answers

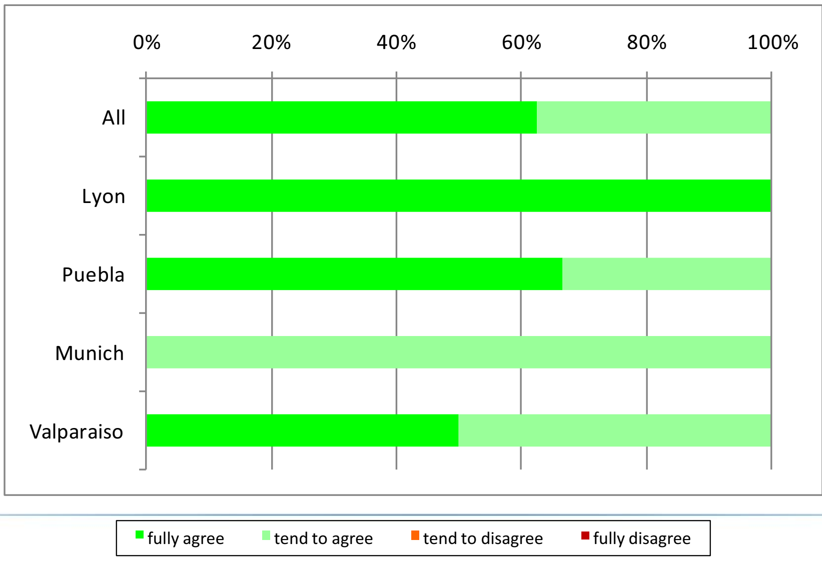


63

Technische Universität München

Project team members have appropriate English language skills

Breakdown of ‘Distant Sites’ answers

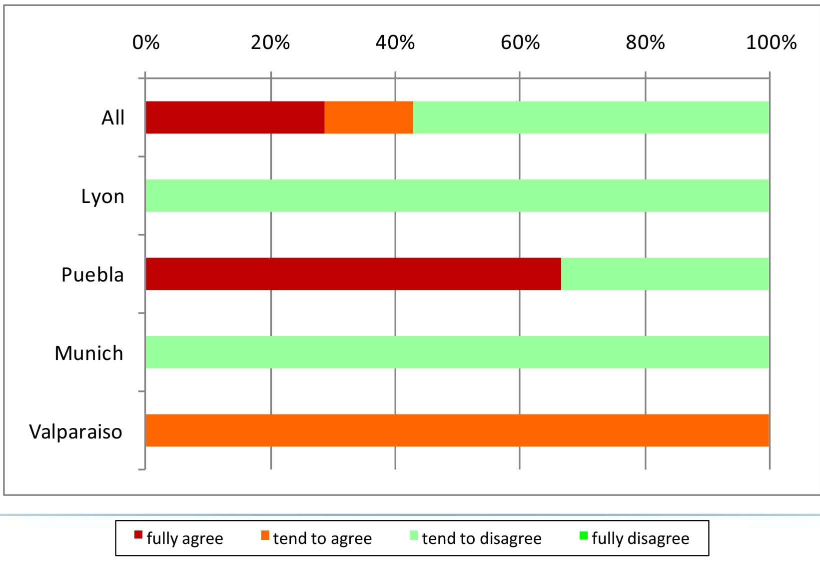


64

Technische Universität München

National culture has significant influence on team collaboration

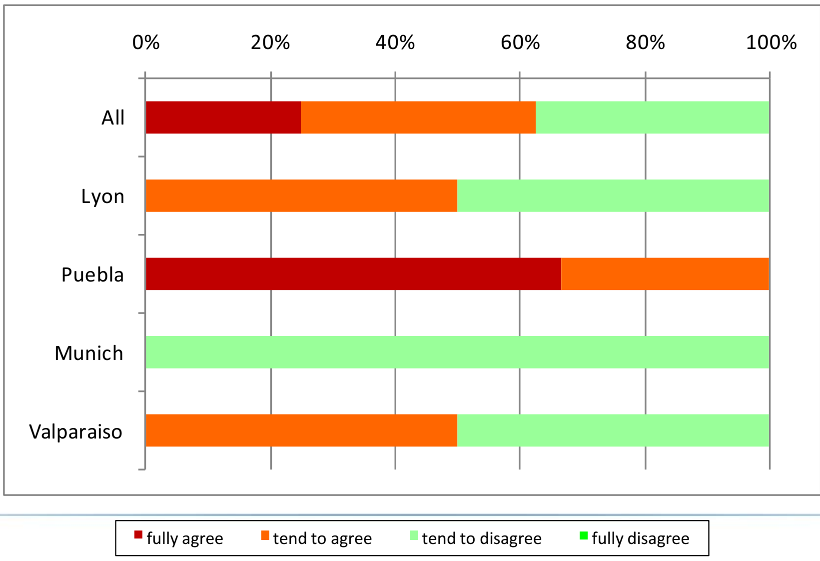
Breakdown of 'Local Site' answers



Technische Universität München

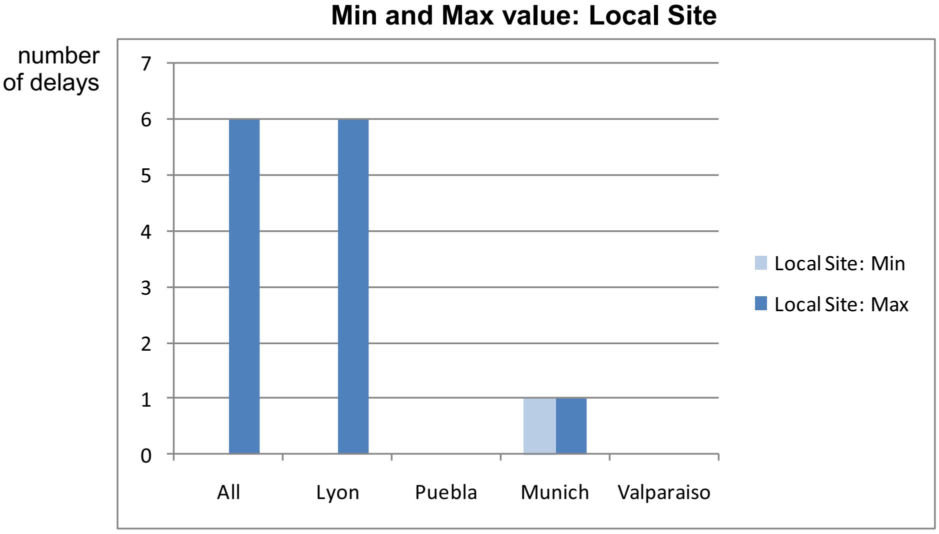
National culture has significant influence on team collaboration

Breakdown of 'Distant Sites' answers



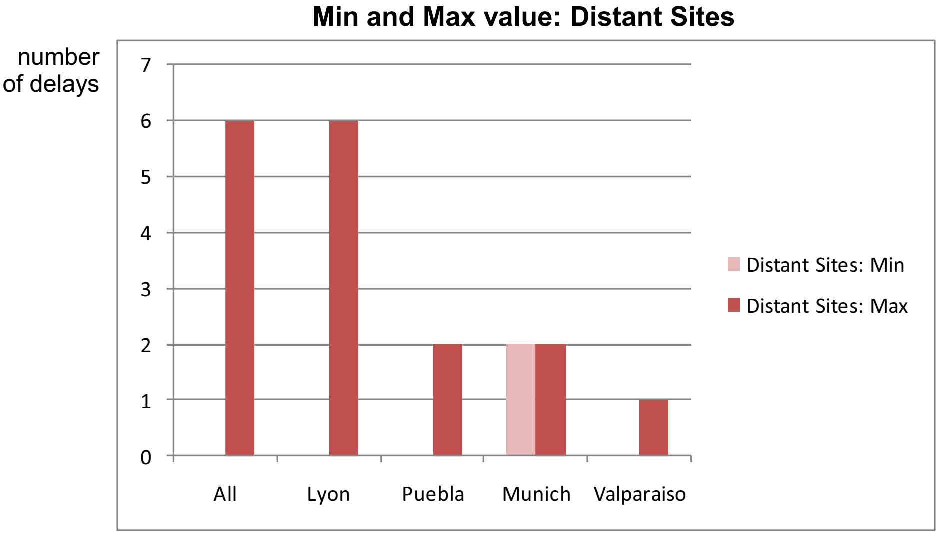
Technische Universität München

How many times in the past two weeks was your own work delayed?



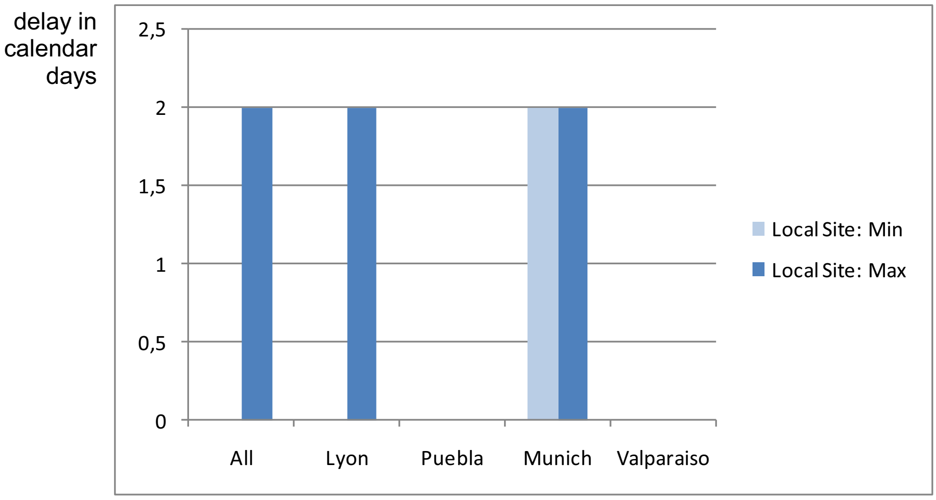
Technische Universität München

How many times in the past two weeks was your own work delayed?



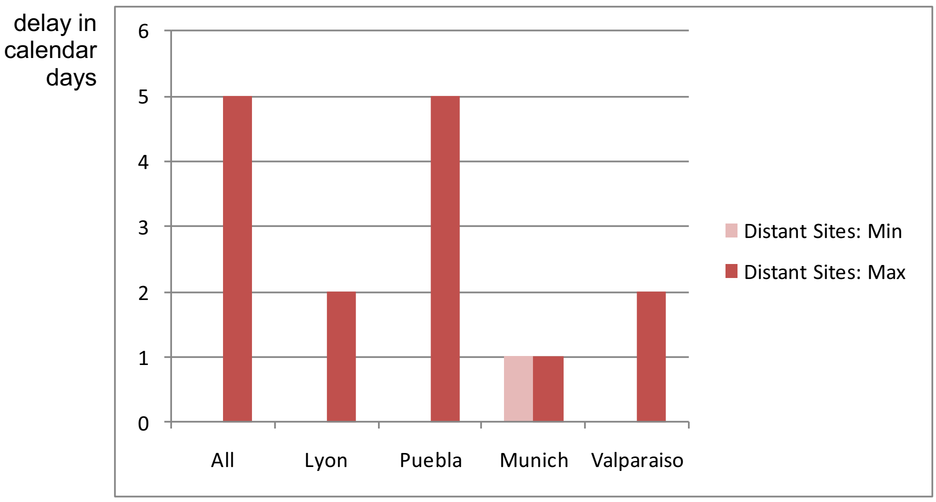
What was the average length of delays you experienced?

Min and Max value: Local Site



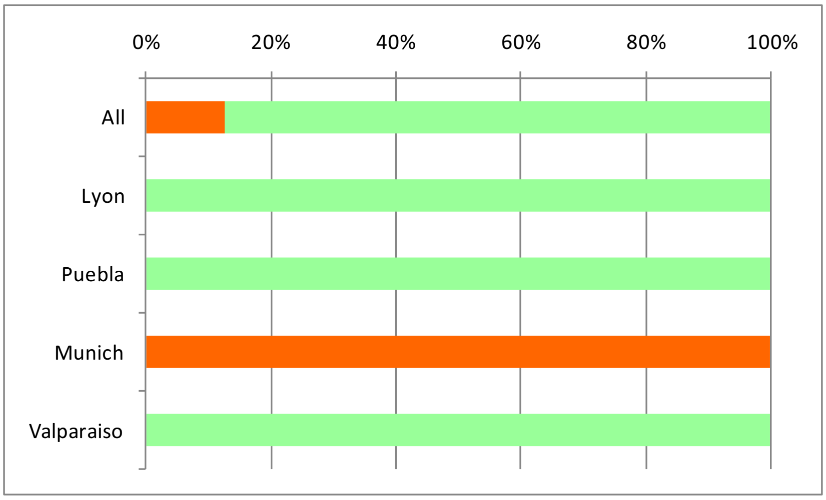
What was the average length of delays you experienced?

Min and Max value: Distant Sites



Technische Universität München

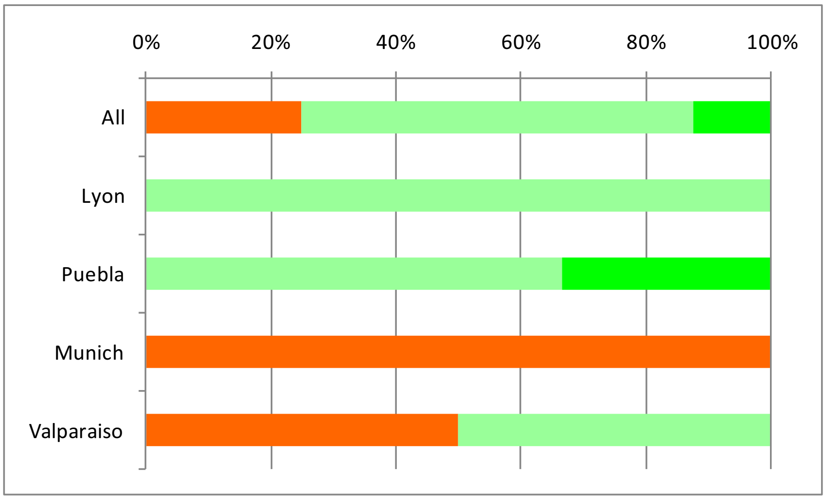
How well do you feel in your project team?



very bad bad good very good

Technische Universität München

How satisfied are you with the project's achievements so far?



very unsatisfied unsatisfied satisfied very satisfied

Copyright © 2010 Christian Lescher, Technische Universität München

Technische Universität München

Contact Information

Christian Lescher

Doctoral Researcher at
Technische Universitaet Muenchen
Department of Informatics
Chair for Applied Software Engineering
Boltzmannstr. 3
D-85748 Garching, Germany

mailto:lescher@in.tum.de

Appendix E

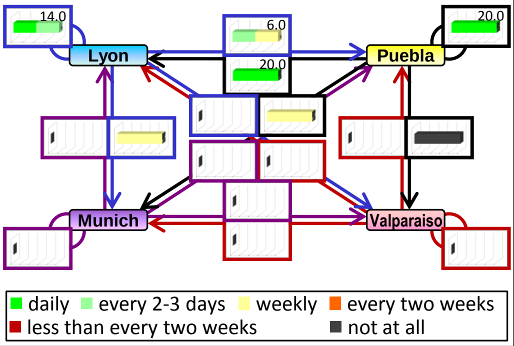
Validation Projects: GSE Metrics Results

In the following the survey results of the four NEREID projects are shown which we observed in the validation phase. The results of the DOLLI4 project were already shown in Chapter 9, see Figure 9.13 and Figure 9.14 on page 154 and 155.

Team: Tricia Google Translate
Status: Nov 14th, 2010

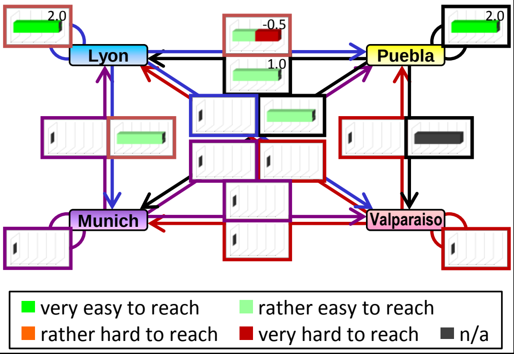
Communication Frequency

How often do you communicate with each site?

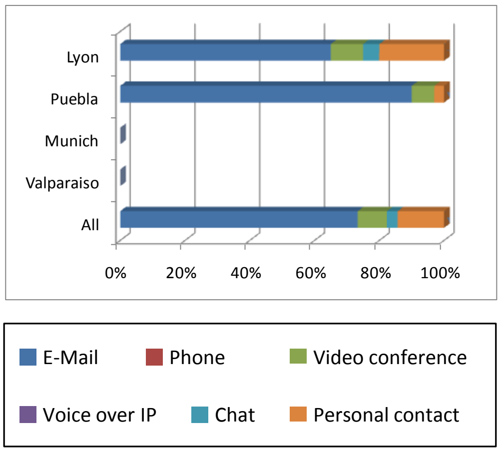


Reachability

Is it easy for you to reach colleagues at these sites?

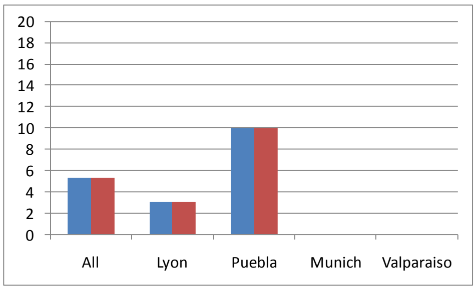


Communication Media

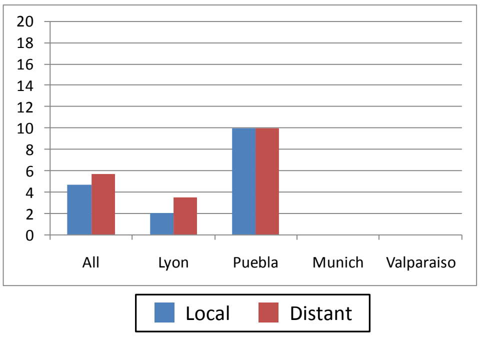


Delay

Number of delays in past two weeks



Duration of delays in past two weeks (days)

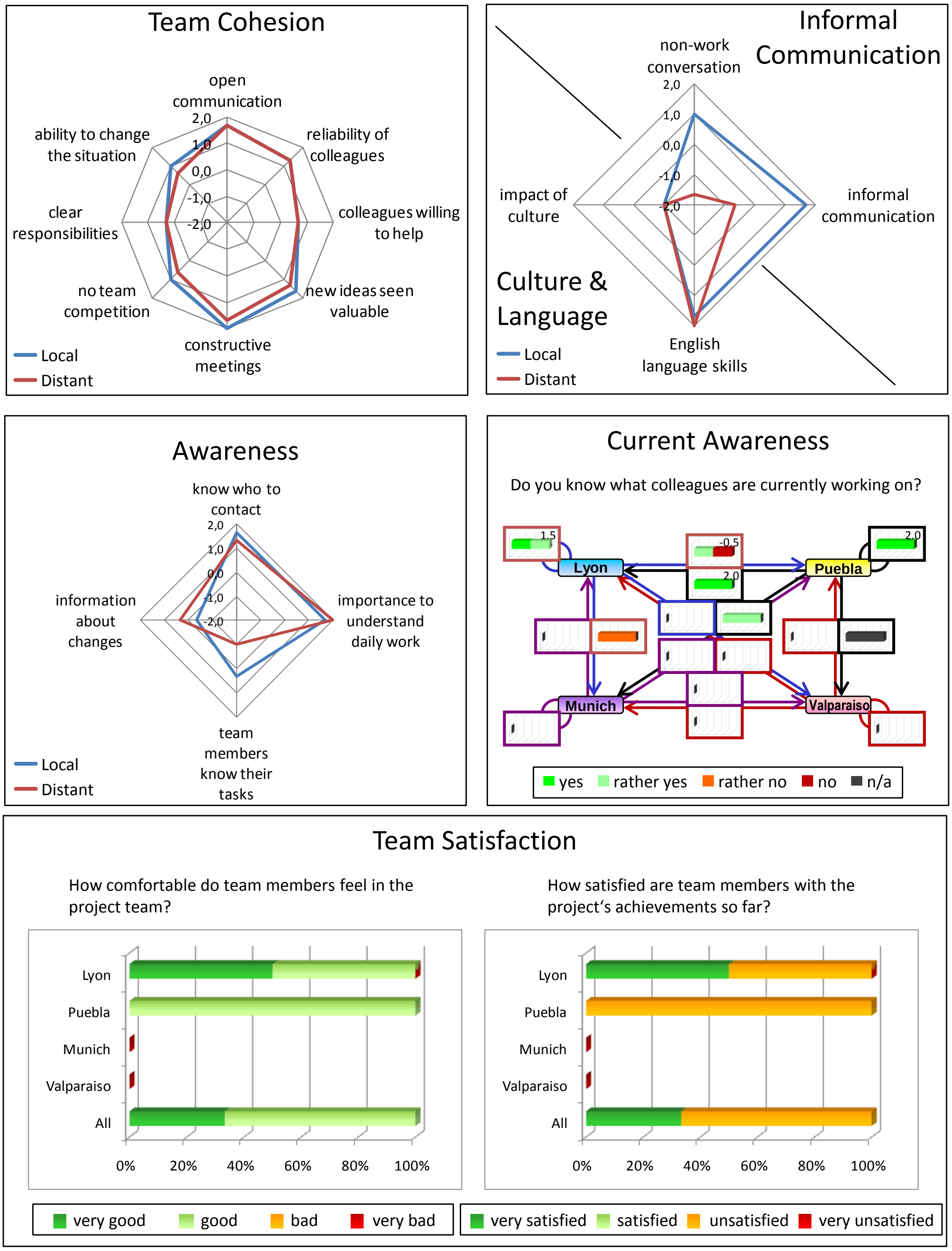


Traveling

None

communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Team: Tricia Google Translate
Status: Nov 14th, 2010

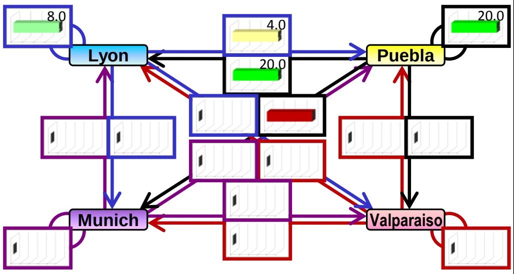


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Tricia Google Translate
Status: Dec. 8th, 2010

Communication Frequency

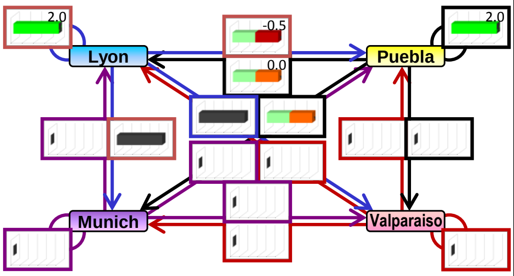
How often do you communicate with each site?



daily, every 2-3 days, weekly, every two weeks, less than every two weeks, not at all

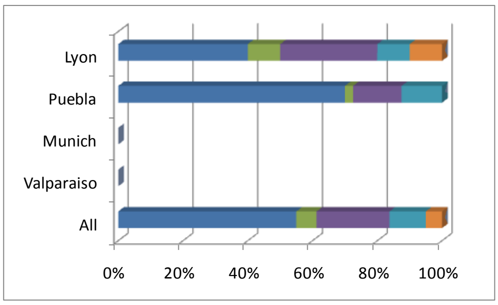
Reachability

Is it easy for you to reach colleagues at these sites?



very easy to reach, rather easy to reach, rather hard to reach, very hard to reach, n/a

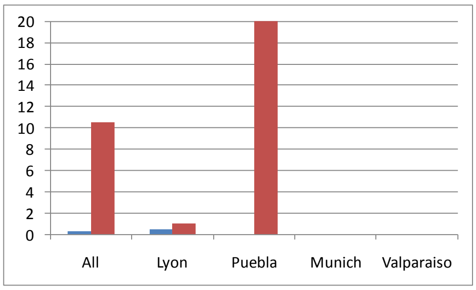
Communication Media



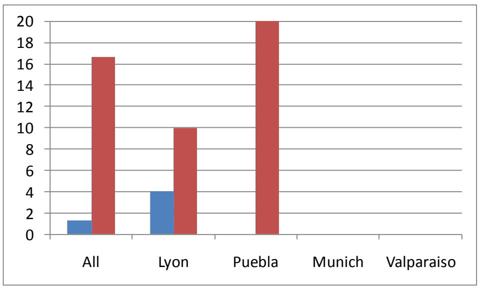
E-Mail, Phone, Video conference, Voice over IP, Chat, Personal contact

Delay

Number of delays in past two weeks



Duration of delays in past two weeks (days)



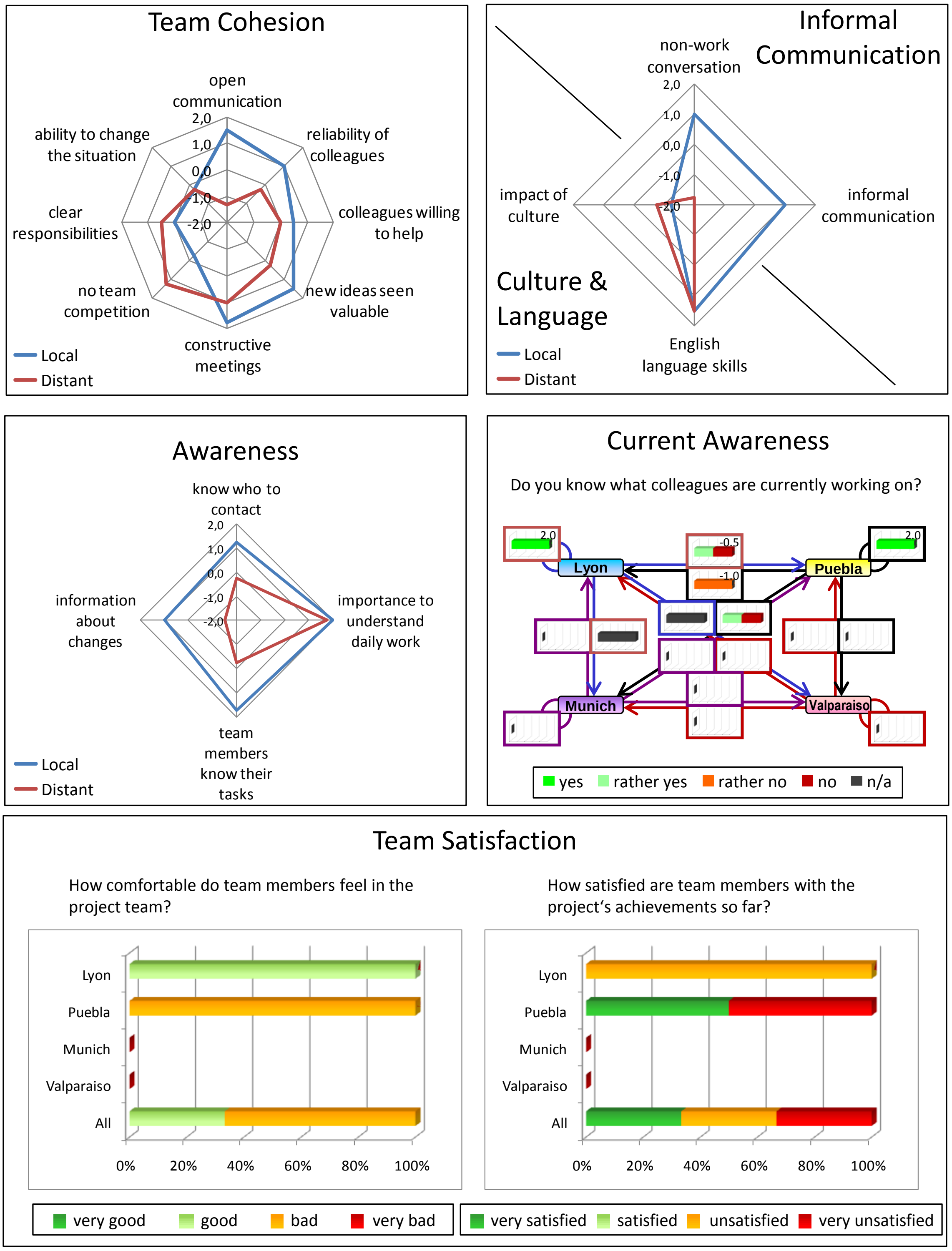
Local, Distant

Traveling

None

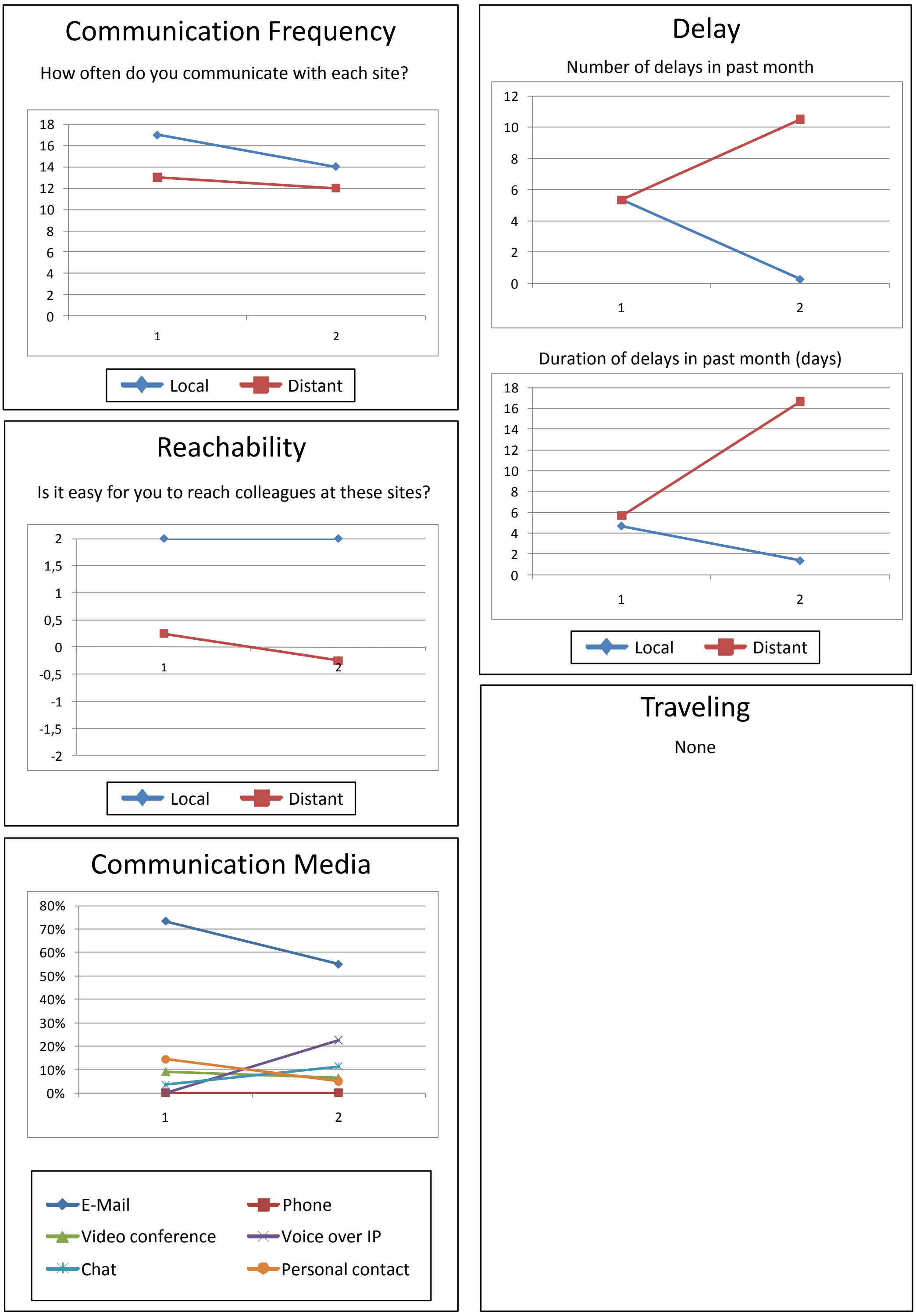
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Team: Google Translate
Status: Dec. 8th, 2010



values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

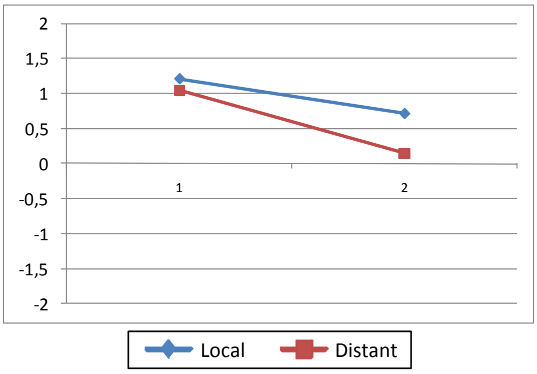
Team: Google Translate
Status: Nov. 14th and Dec. 8th, 2010



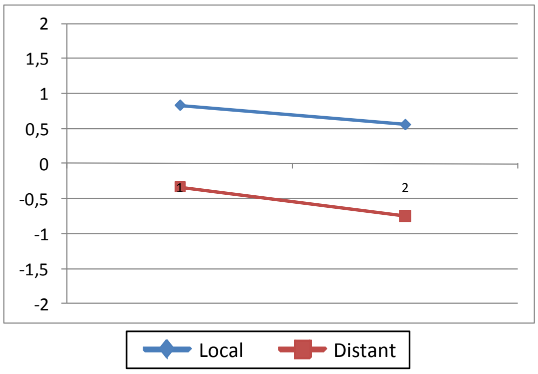
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Team: Google Translate
Status: Nov. 14th and Dec. 8th, 2010

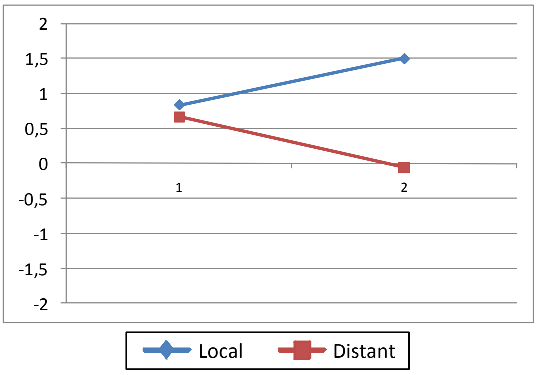
Team Cohesion



Informal Communication, Culture & Language

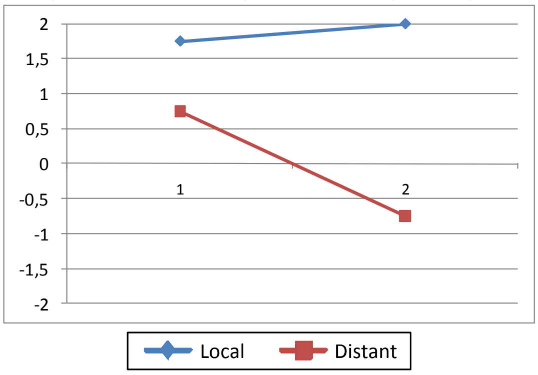


Awareness



Current Awareness

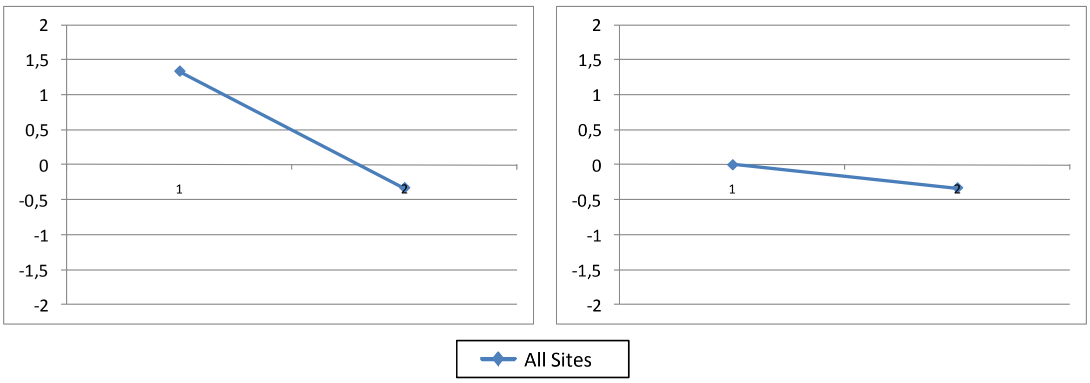
Do you know what colleagues are currently working on?



Team Satisfaction

How comfortable do team members feel in the project team?

How satisfied are team members with the project‘s achievements so far?

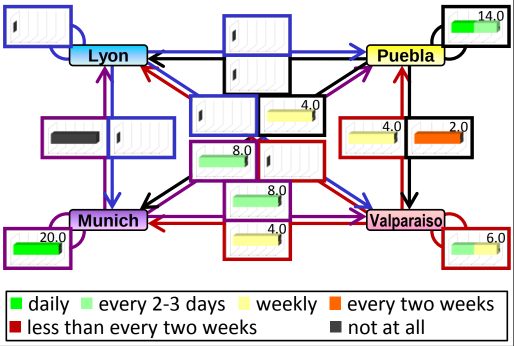


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Touch Screen Tablet
Status: Nov. 14th, 2010

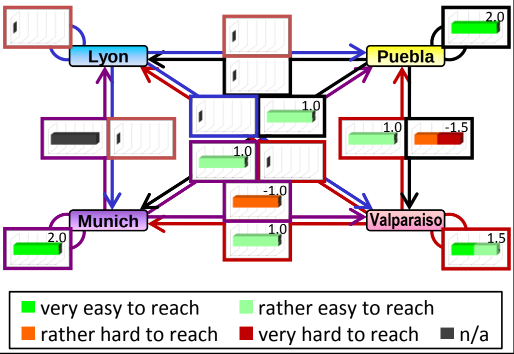
Communication Frequency

How often do you communicate with each site?

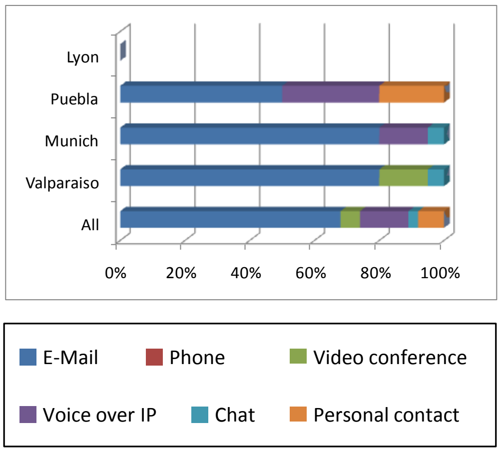


Reachability

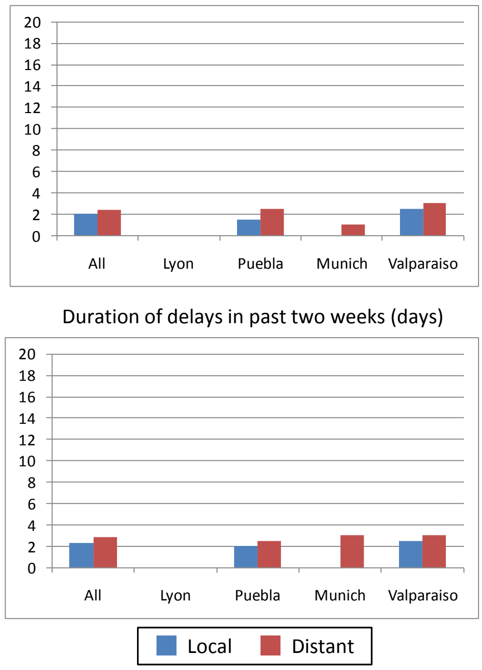
Is it easy for you to reach colleagues at these sites?



Communication Media



Delay



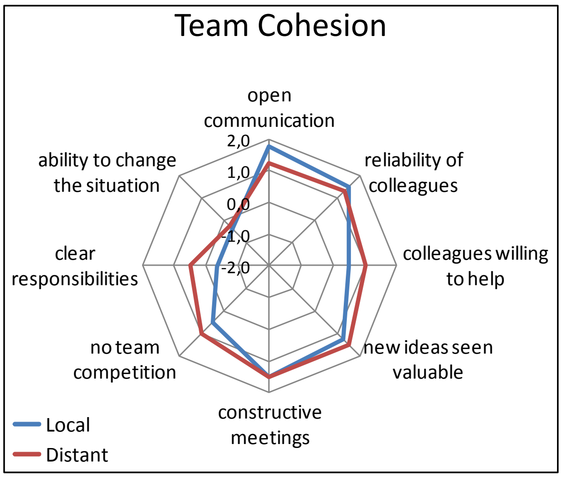
Traveling

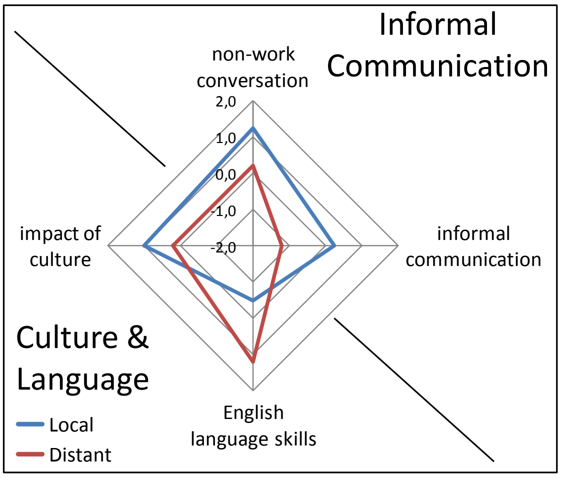
None

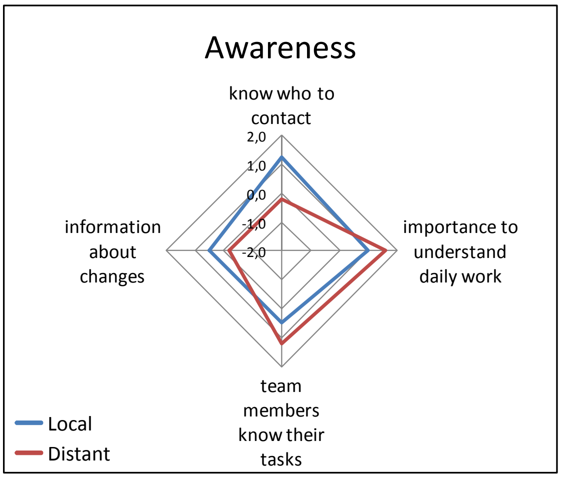
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

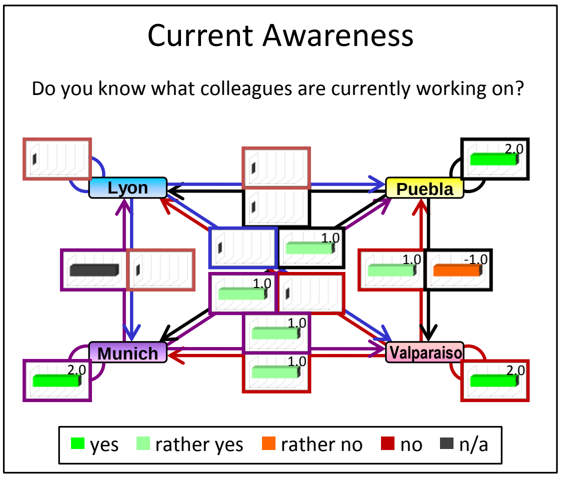
Team: Touch Screen Tablet

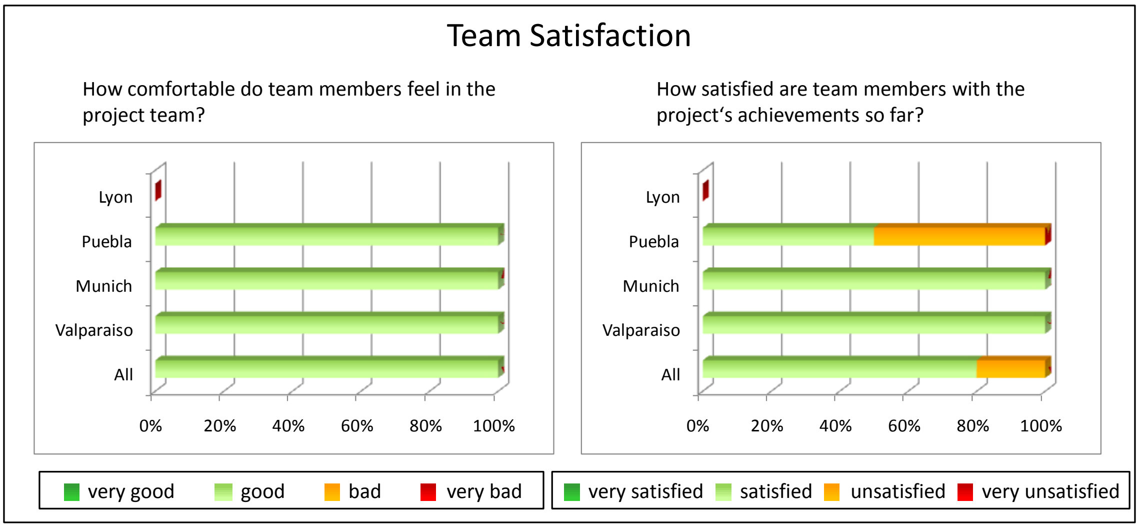
Status: Nov. 14th, 2010









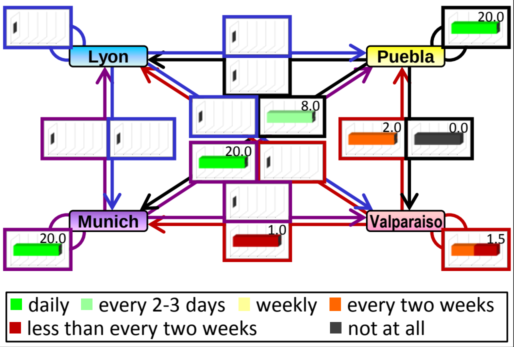


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Touch Screen Tablet
Status: Dec. 8th, 2010

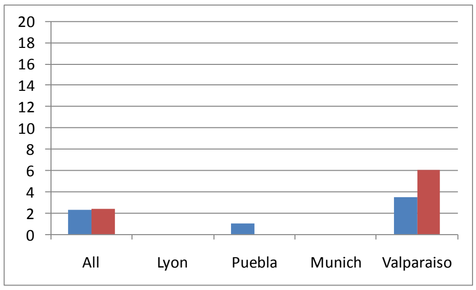
Communication Frequency

How often do you communicate with each site?

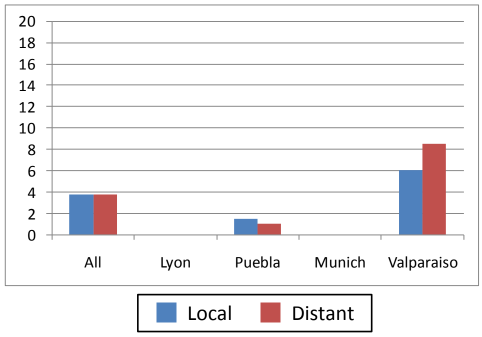


Delay

Number of delays in past two weeks

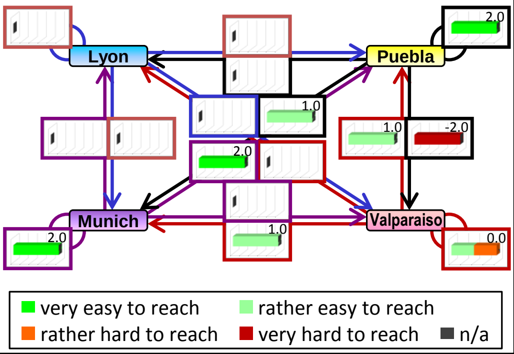


Duration of delays in past two weeks (days)



Reachability

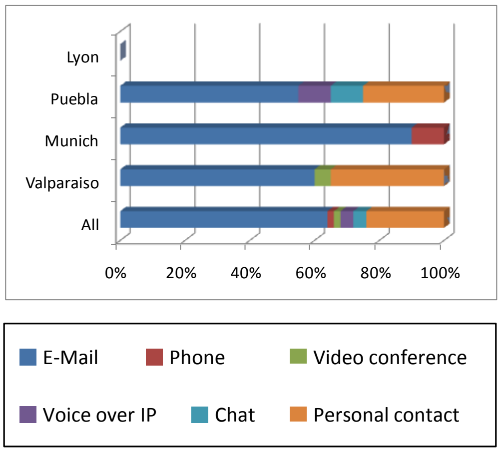
Is it easy for you to reach colleagues at these sites?



Traveling

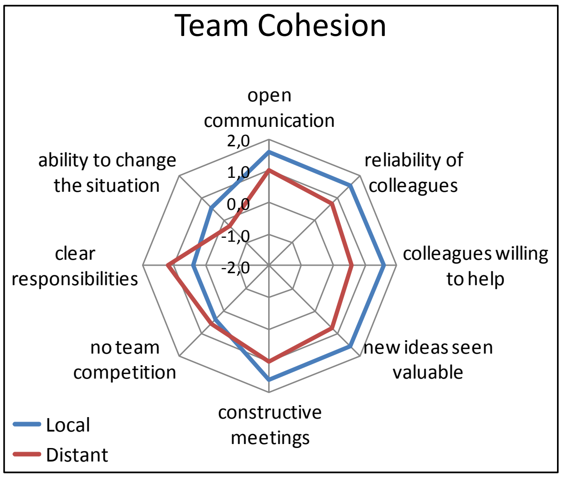
None

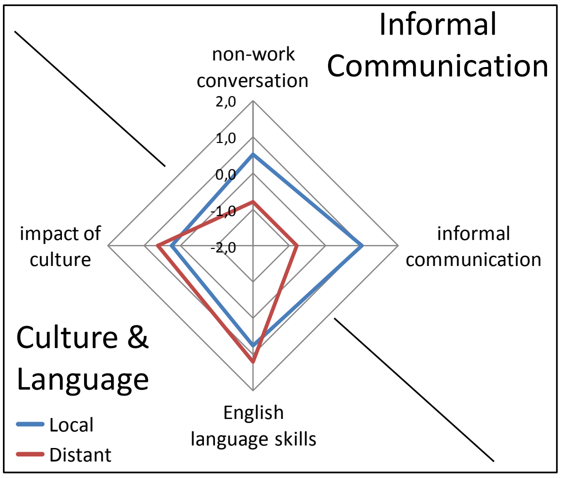
Communication Media

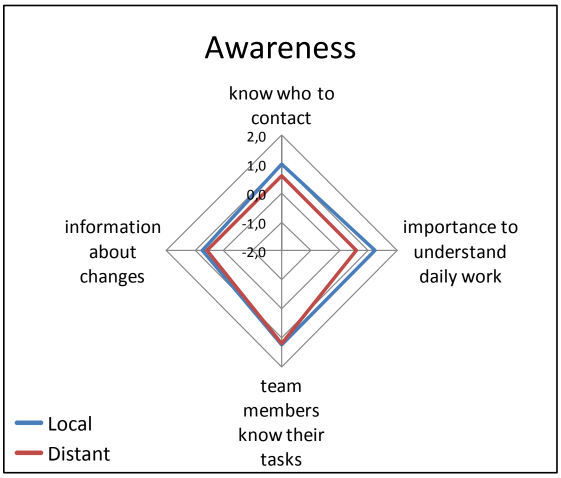


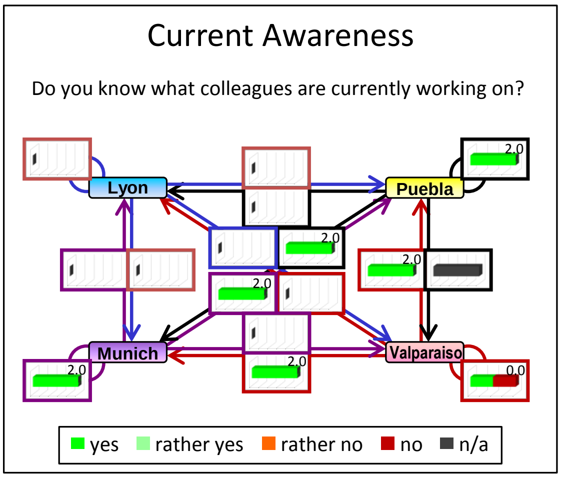
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

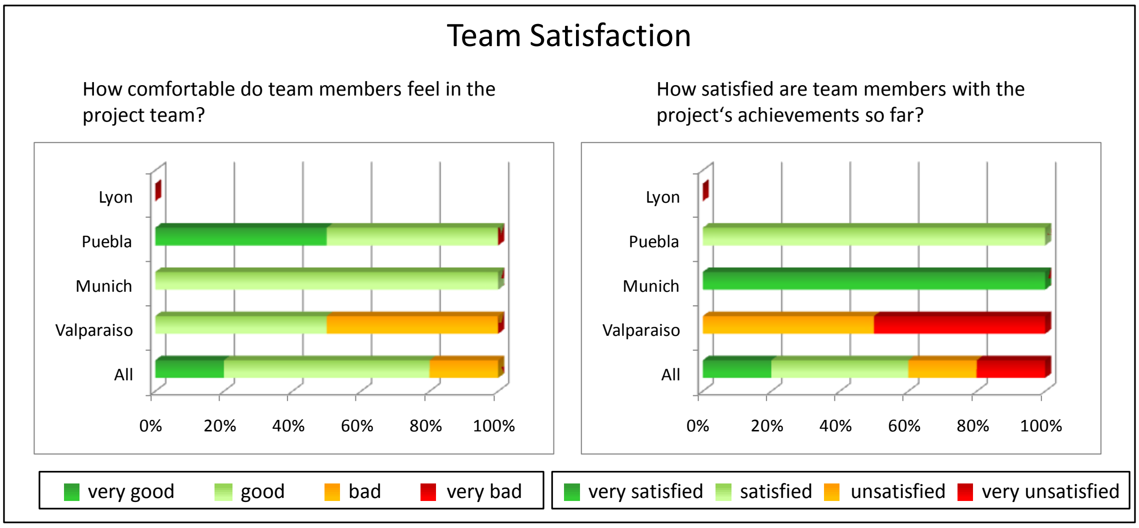
Team: Touch Screen Tablet
Status: Dec. 8th, 2010









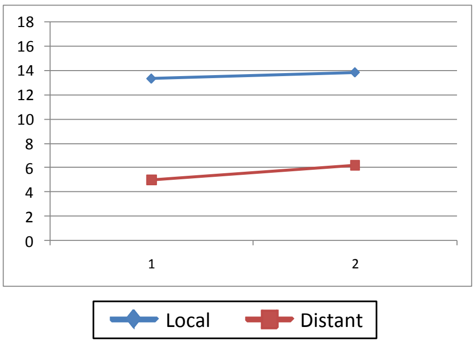


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Touch Screen Tablet
Status: Nov. 14th and Dec. 8th, 2010

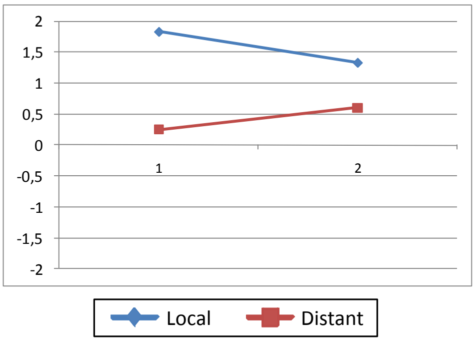
Communication Frequency

How often do you communicate with each site?

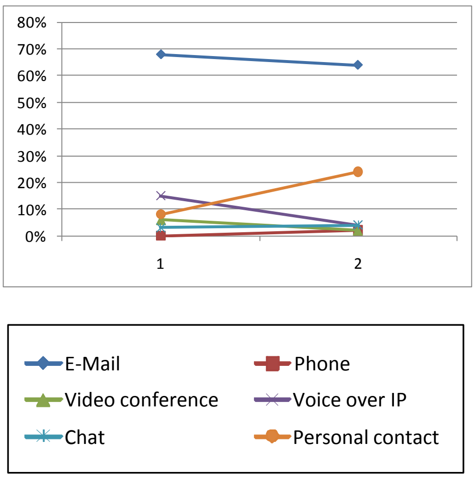


Reachability

Is it easy for you to reach colleagues at these sites?

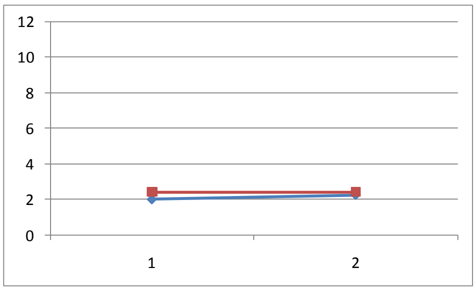


Communication Media

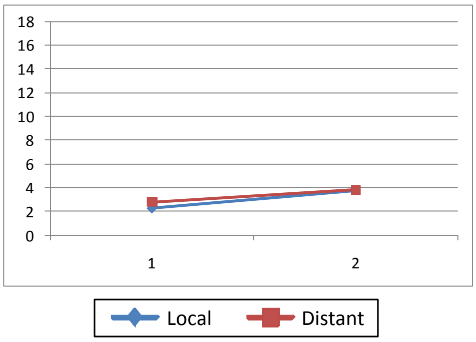


Delay

Number of delays in past month



Duration of delays in past month (days)

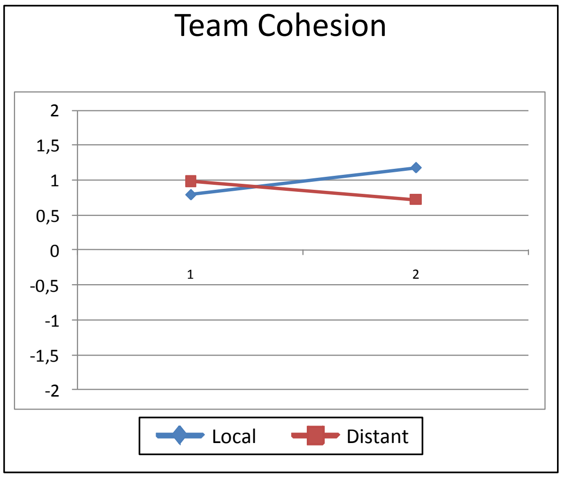


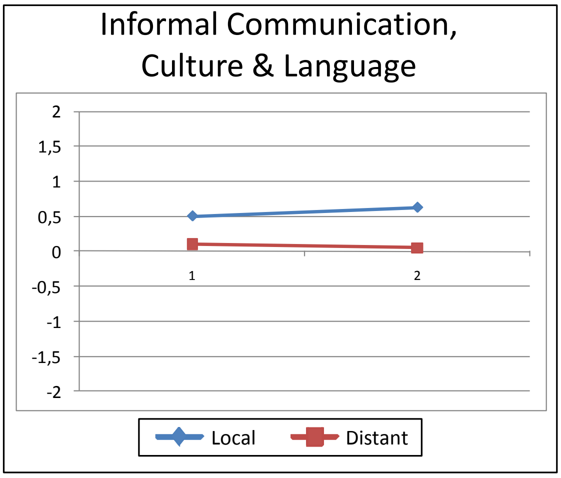
Traveling

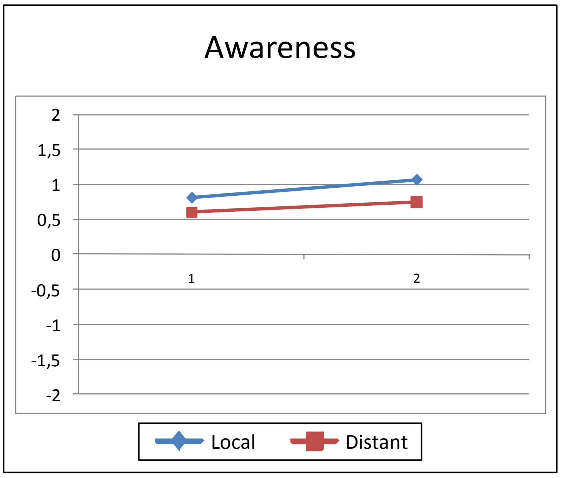
None

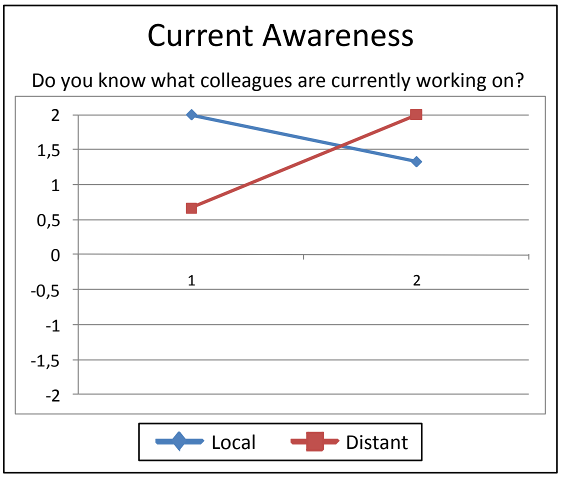
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

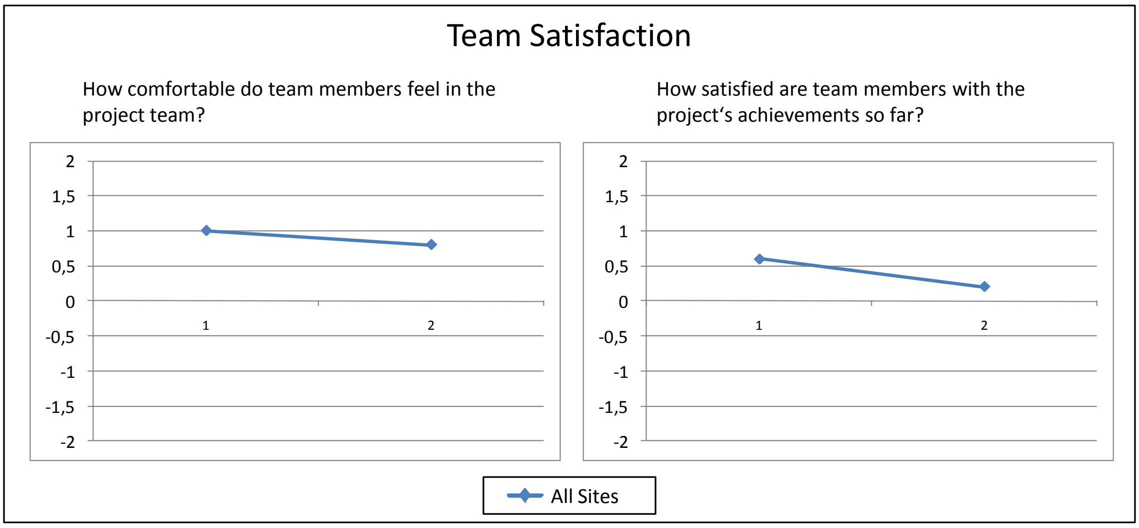
Team: Touch Screen Tablet
Status: Nov. 14th and Dec. 8th, 2010









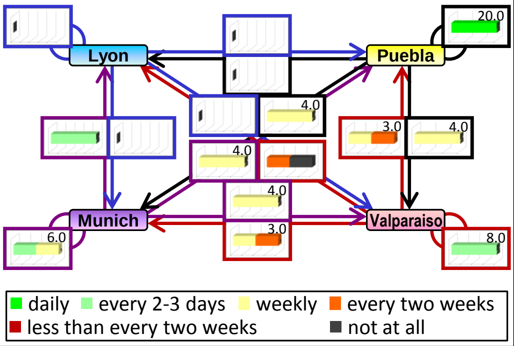


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Electronic Billing
Status: Nov. 14th, 2010

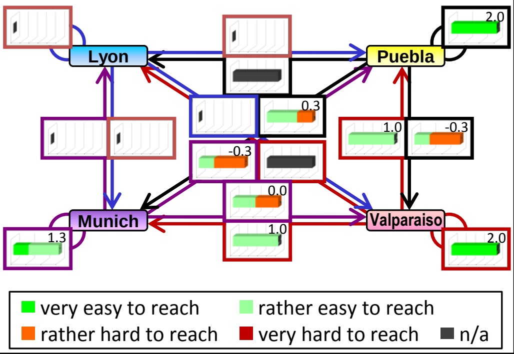
Communication Frequency

How often do you communicate with each site?

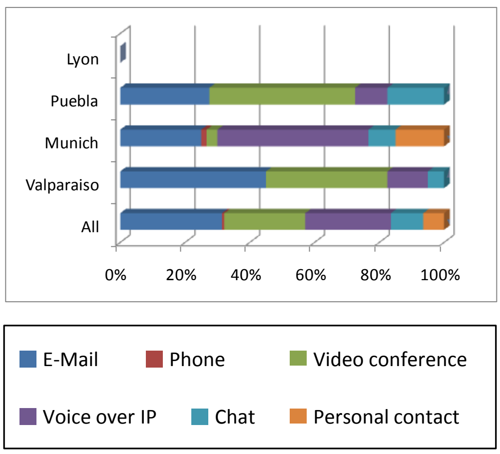


Reachability

Is it easy for you to reach colleagues at these sites?

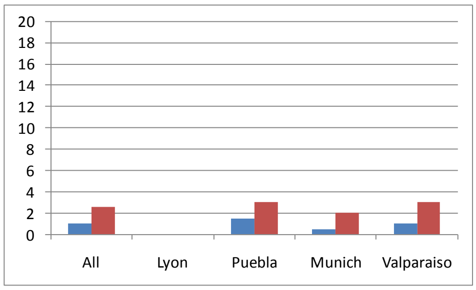


Communication Media

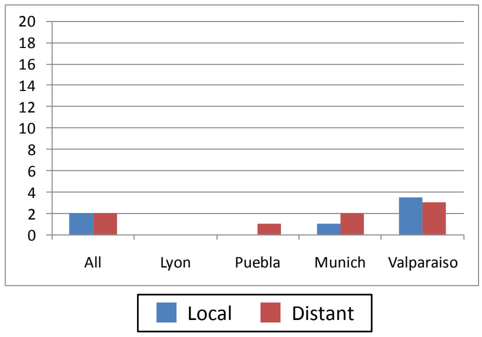


Delay

Number of delays in past two weeks



Duration of delays in past two weeks (days)

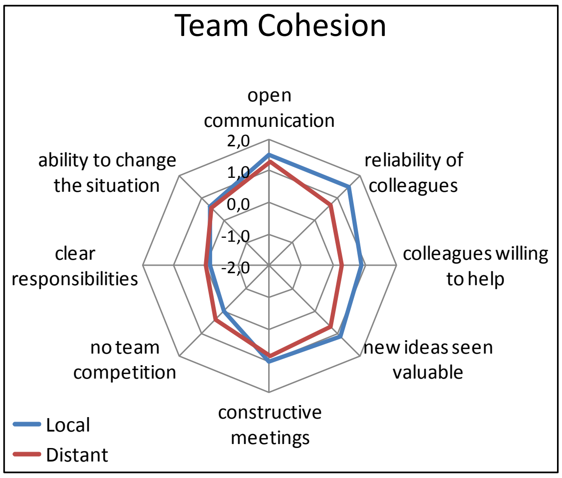


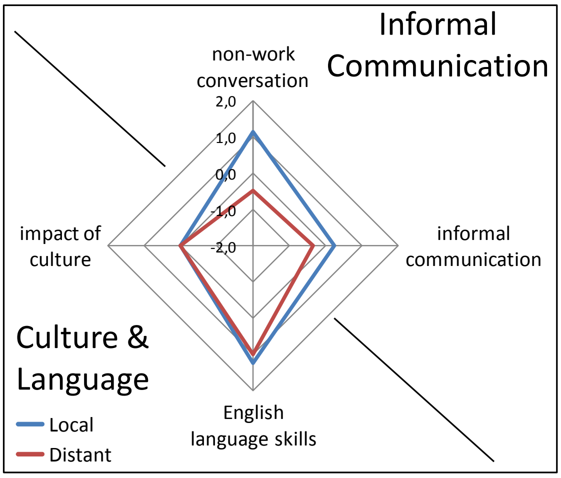
Traveling

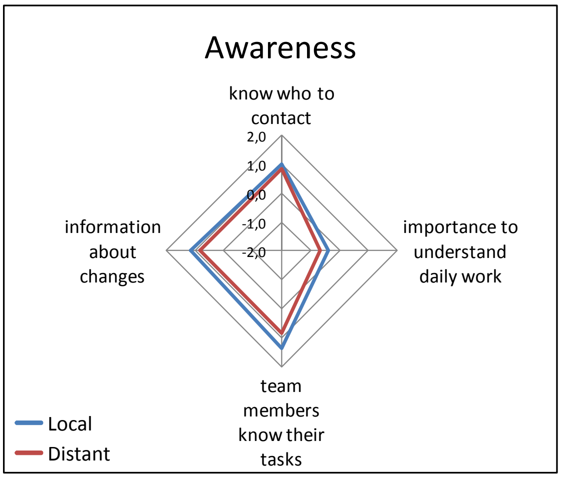
None

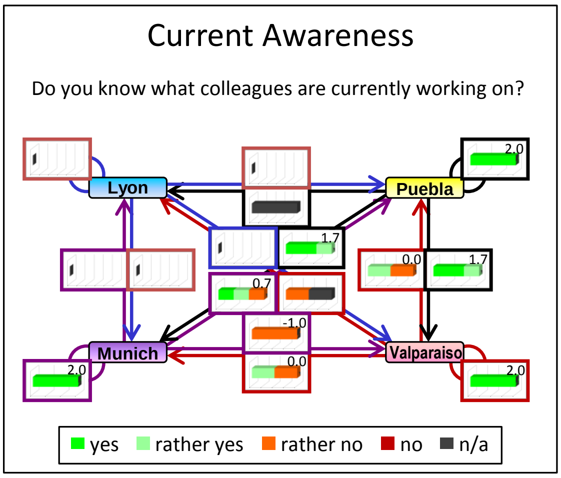
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

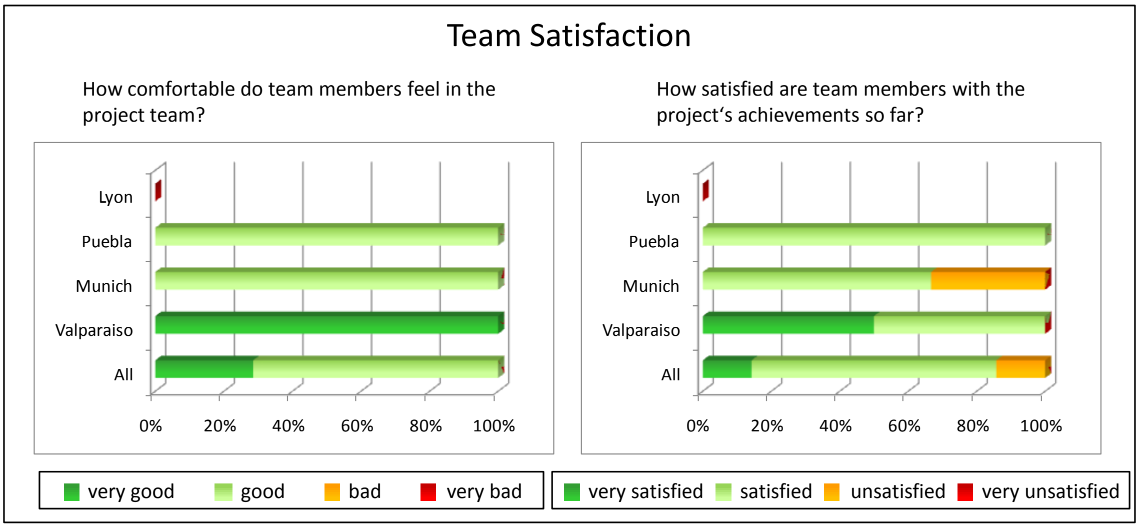
Team: Electronic Billing
Status: Nov. 14th, 2010





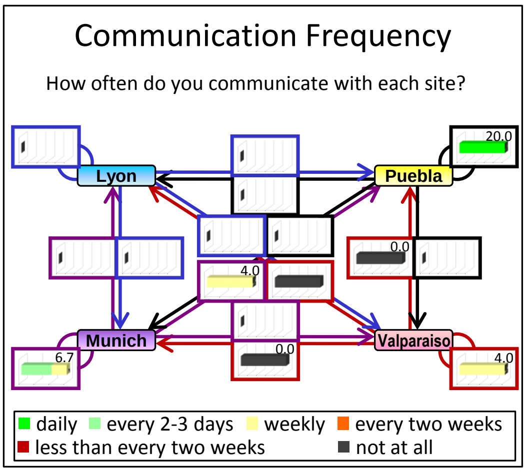


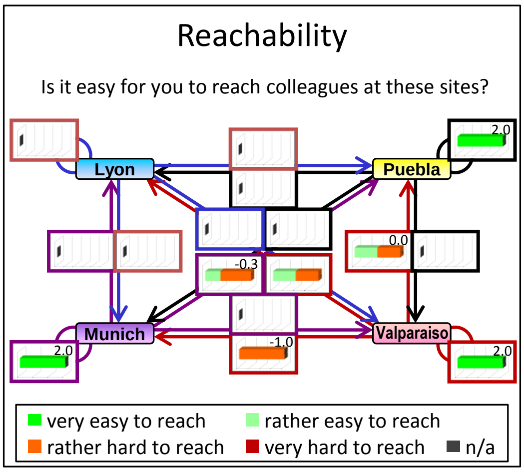




values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

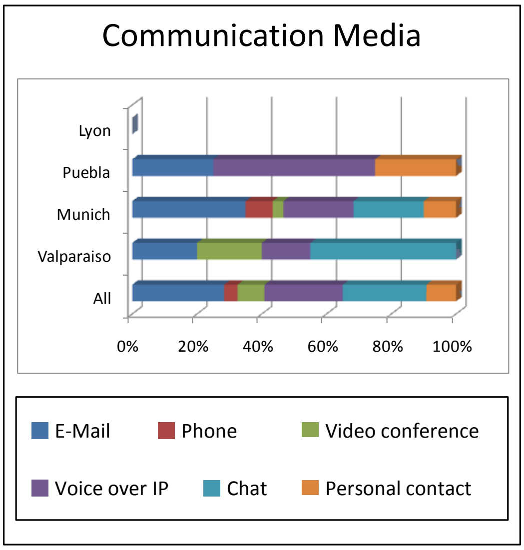
Team: Electronic Billing
Status: Dec. 8th, 2010





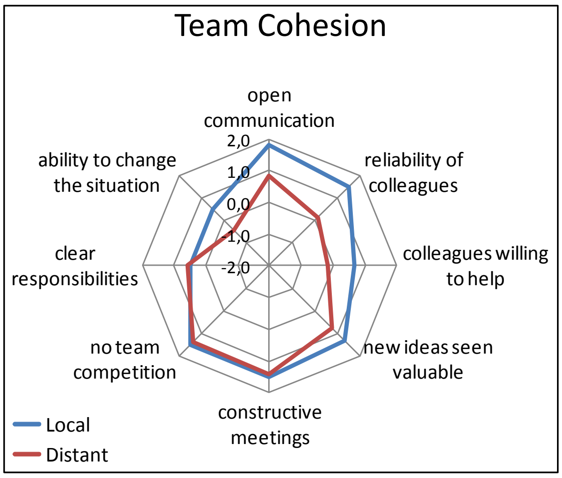
Traveling

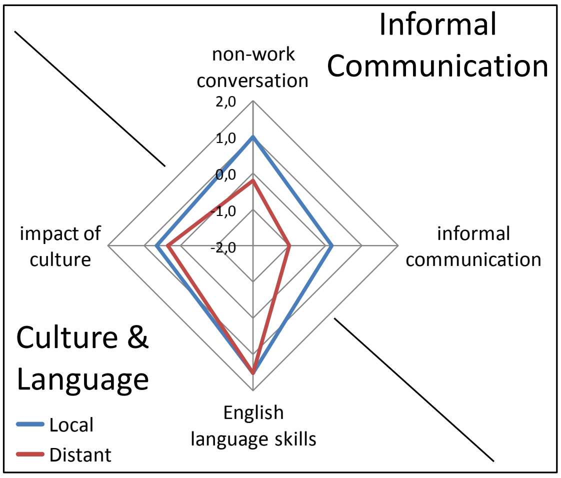
None

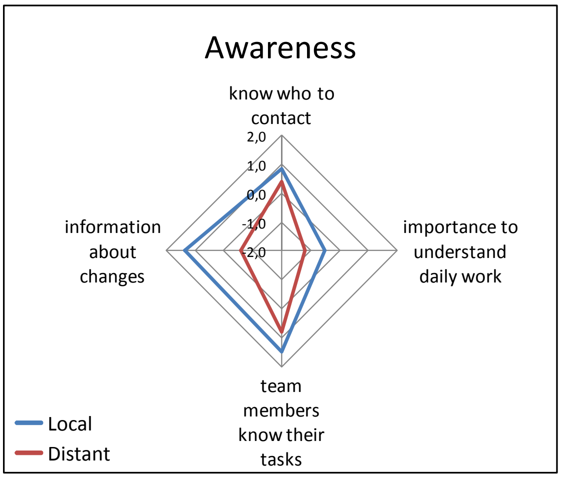


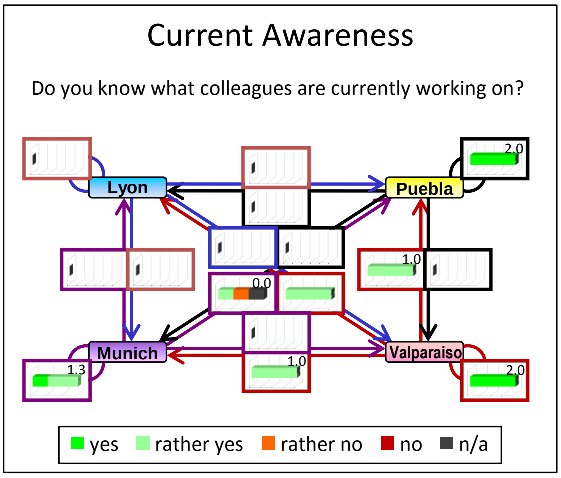
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

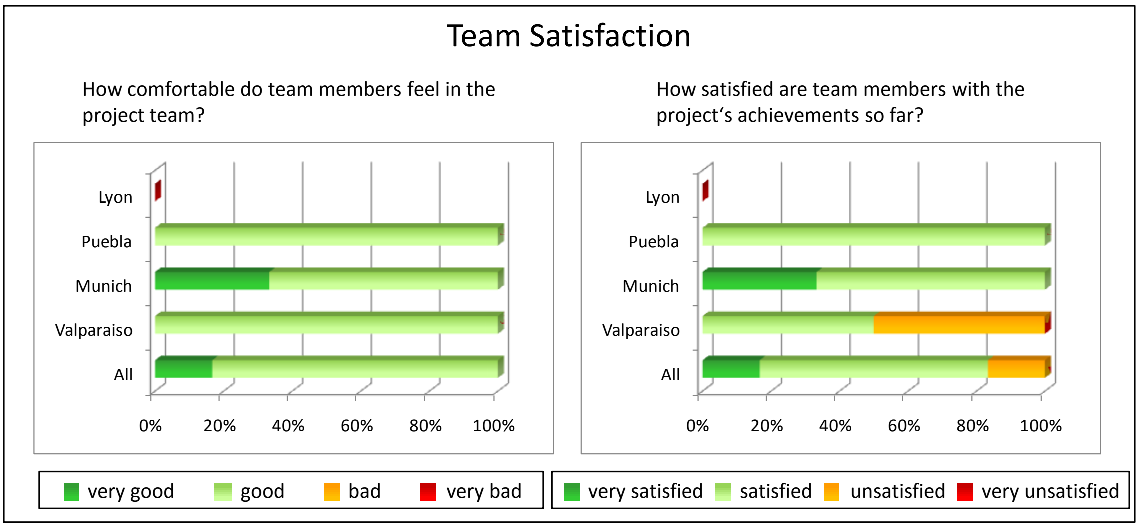
Team: Electronic Billing
Status: Dec. 8th, 2010





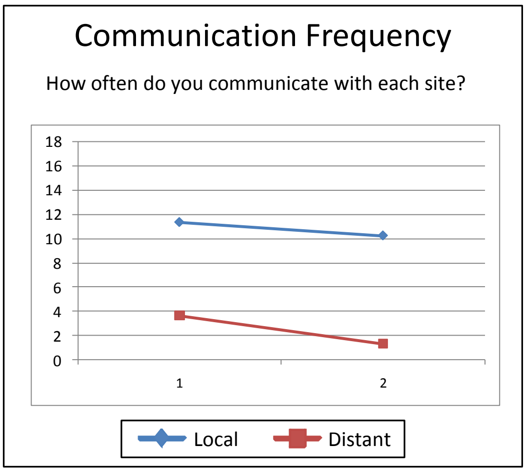


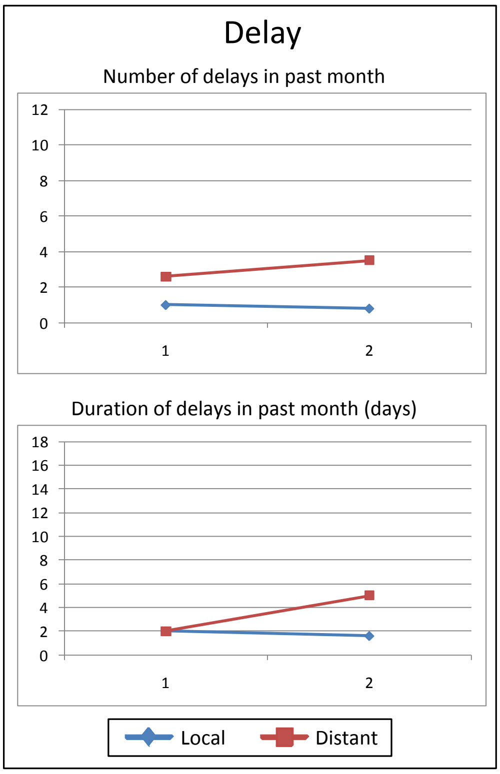


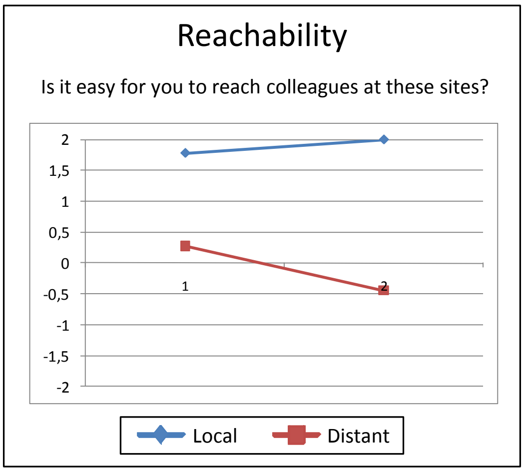


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Electronic Billing
Status: Nov. 14th and Dec. 8th, 2010

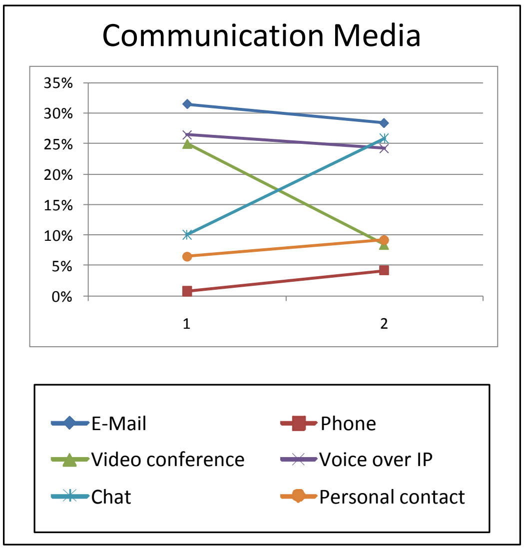






Traveling

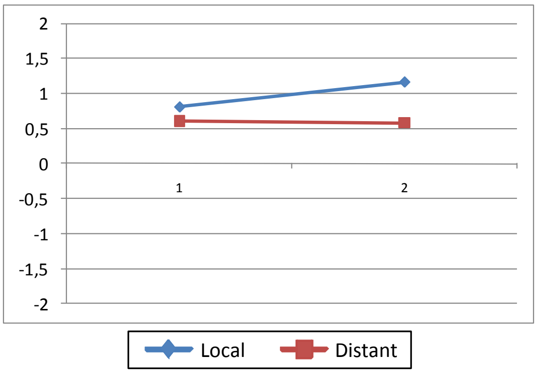
None



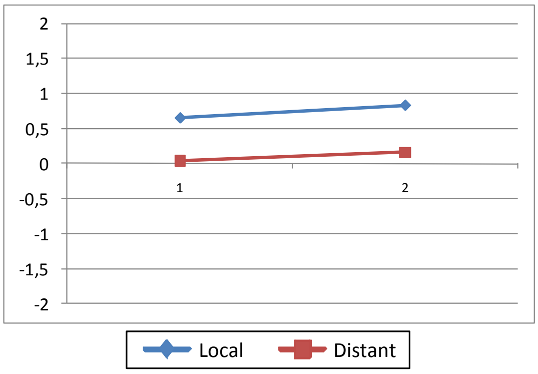
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Team: Electronic Billing
Status: Nov. 14th and Dec. 8th, 2010

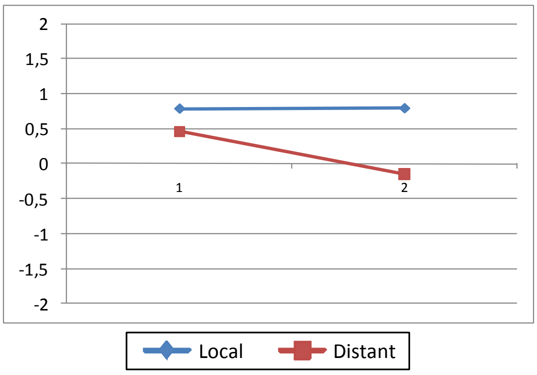
Team Cohesion



Informal Communication, Culture & Language

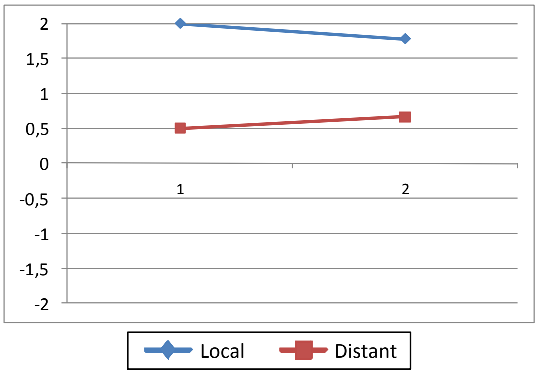


Awareness



Current Awareness

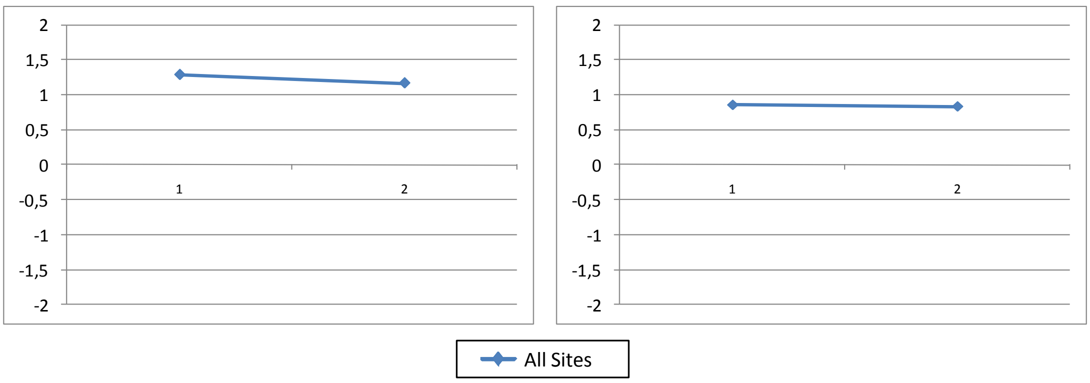
Do you know what colleagues are currently working on?



Team Satisfaction

How comfortable do team members feel in the project team?

How satisfied are team members with the project‘s achievements so far?

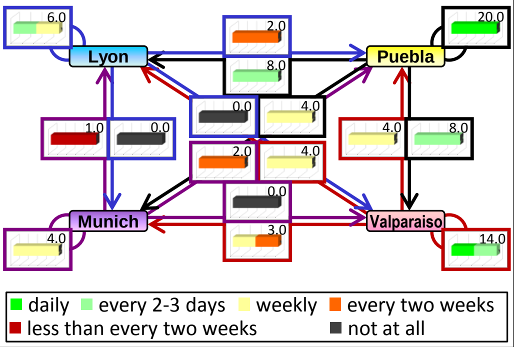


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Carpooling and Ride Sharing
Status: Nov. 14th, 2010

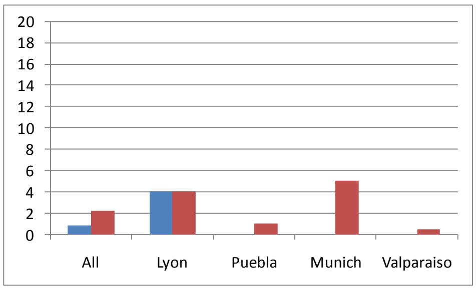
Communication Frequency

How often do you communicate with each site?

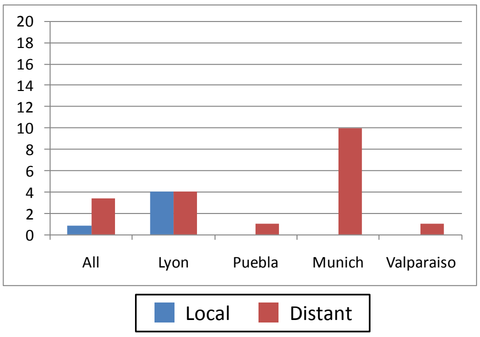


Delay

Number of delays in past two weeks

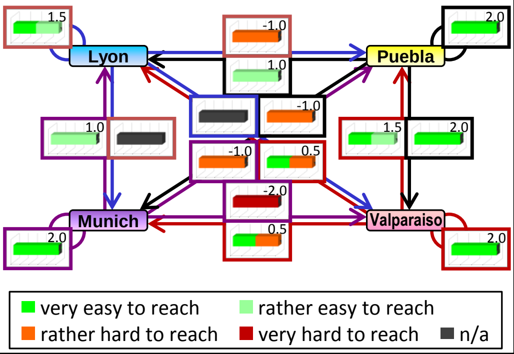


Duration of delays in past two weeks (days)



Reachability

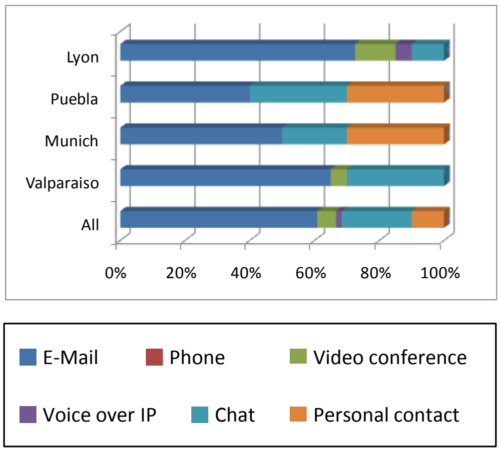
Is it easy for you to reach colleagues at these sites?



Traveling

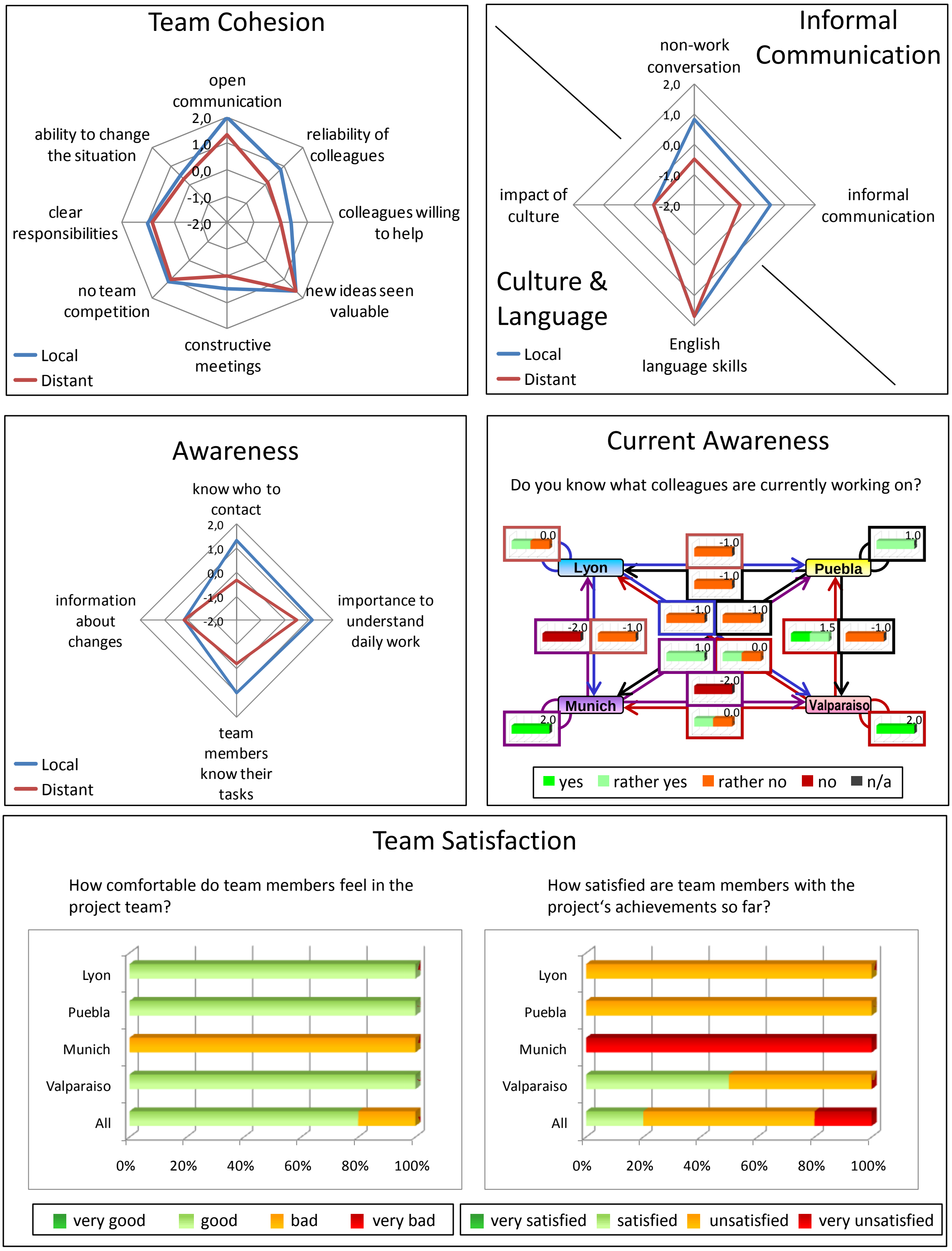
None

Communication Media



communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

Team: Carpooling and Ride Sharing
Status: Nov. 14th, 2010

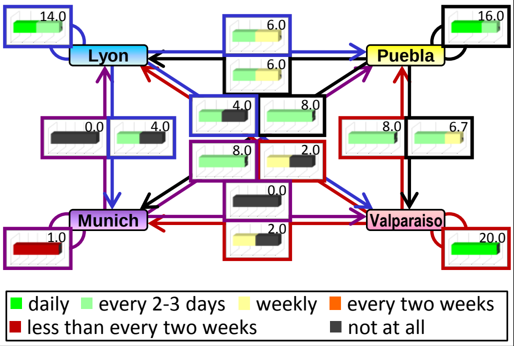


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Carpooling and Ride Sharing
Status: Dec. 8th, 2010

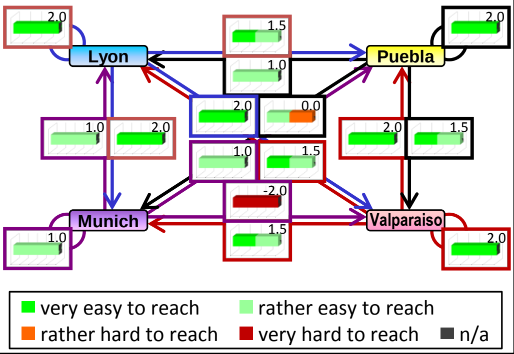
Communication Frequency

How often do you communicate with each site?

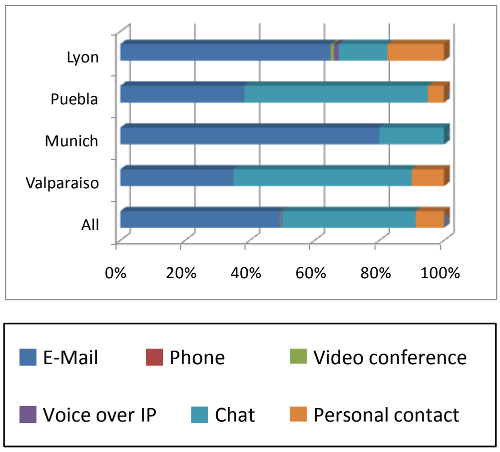


Reachability

Is it easy for you to reach colleagues at these sites?

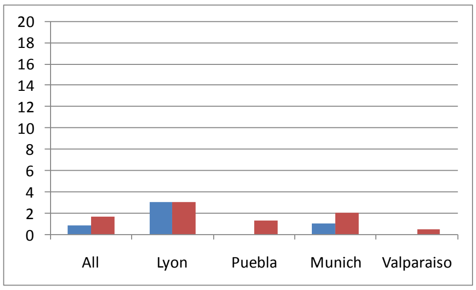


Communication Media

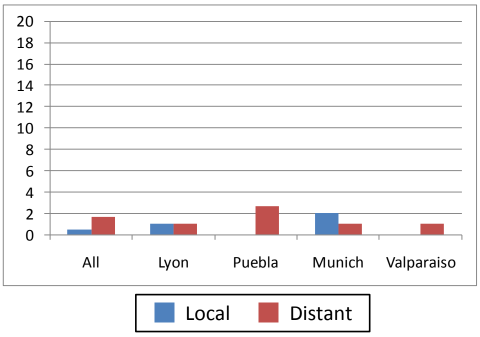


Delay

Number of delays in past two weeks



Duration of delays in past two weeks (days)

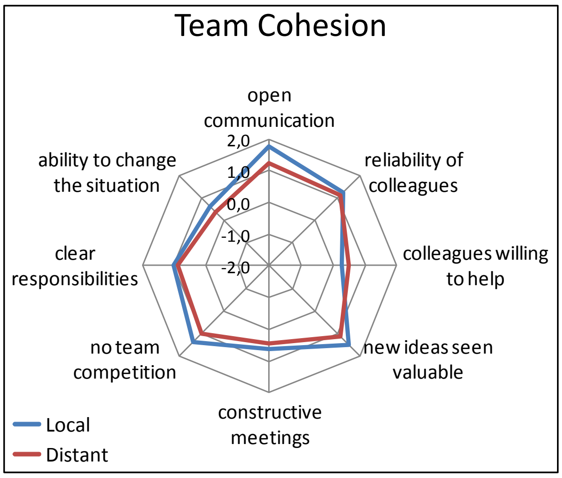


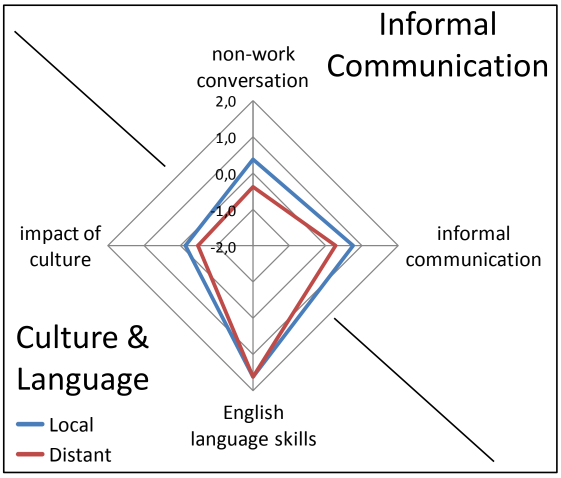
Traveling

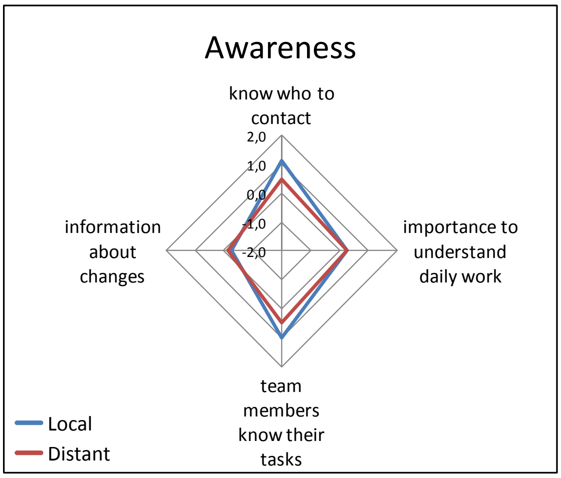
None

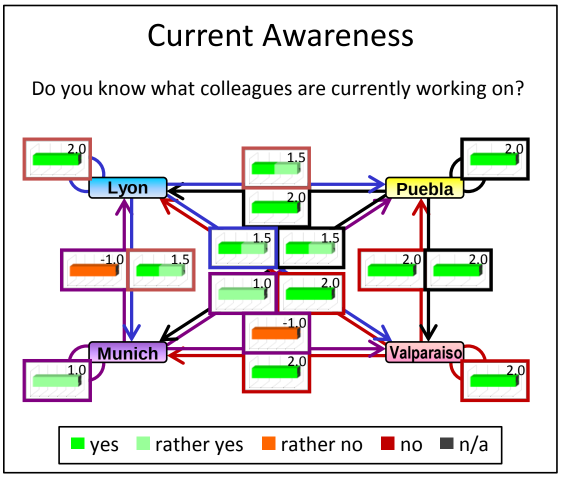
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

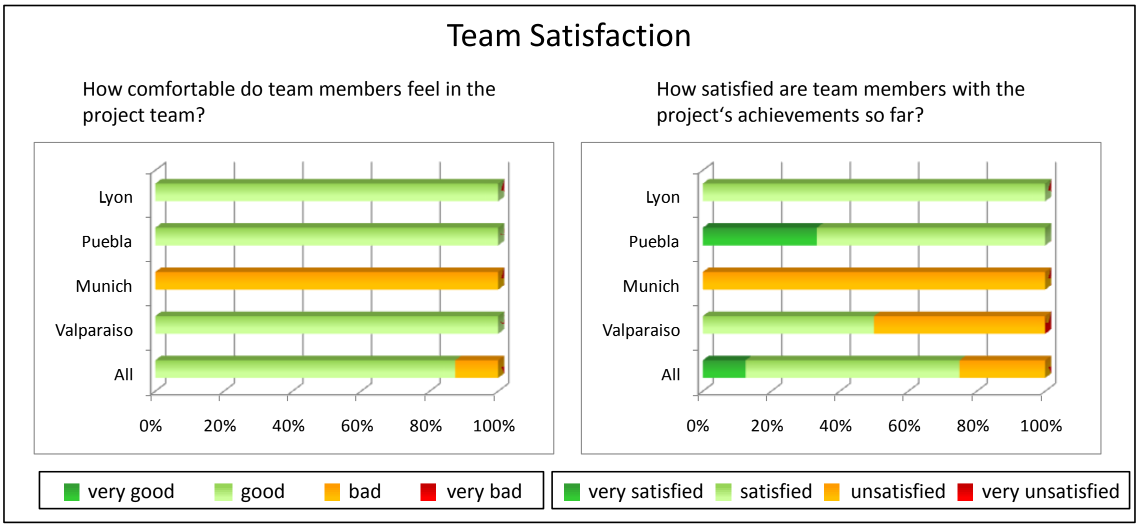
Team: Carpooling and Ride Sharing
Status: Dec. 8th, 2010









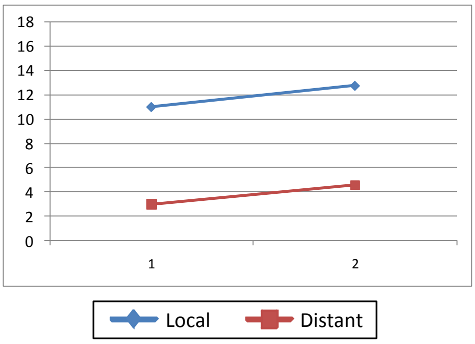


values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Team: Carpooling and Ride Sharing
Status: Nov. 14th and Dec. 8th, 2010

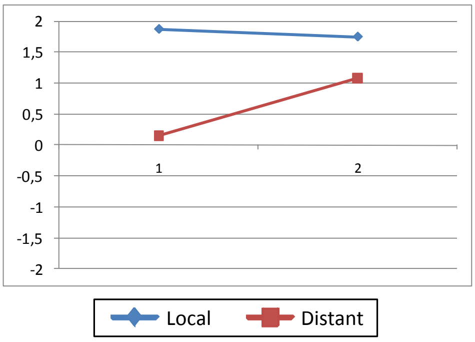
Communication Frequency

How often do you communicate with each site?

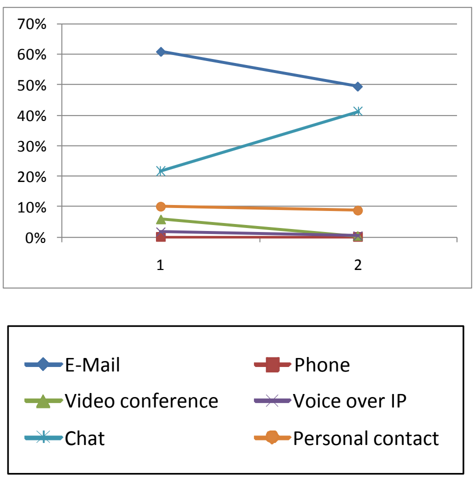


Reachability

Is it easy for you to reach colleagues at these sites?

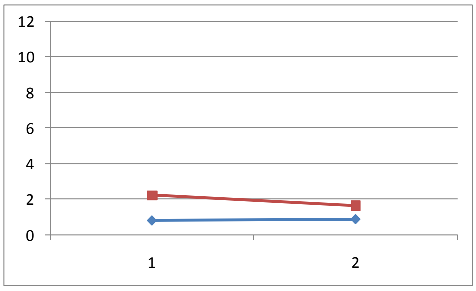


Communication Media

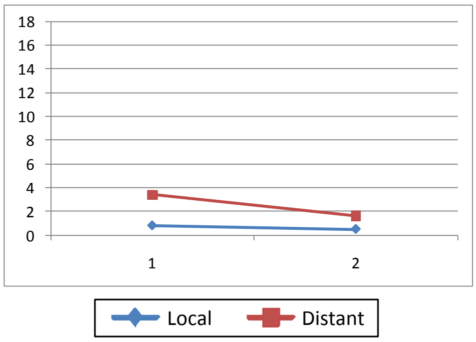


Delay

Number of delays in past month



Duration of delays in past month (days)

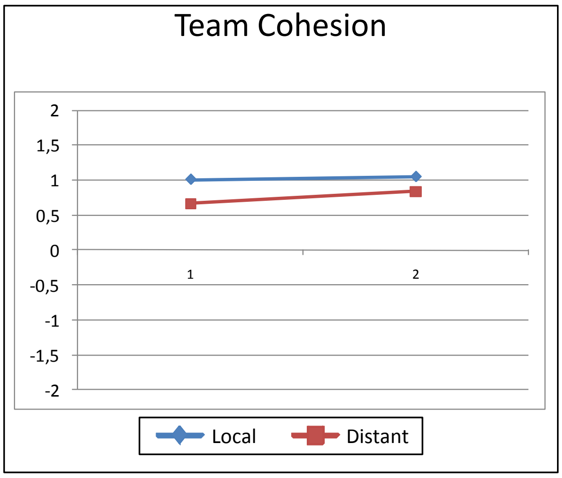


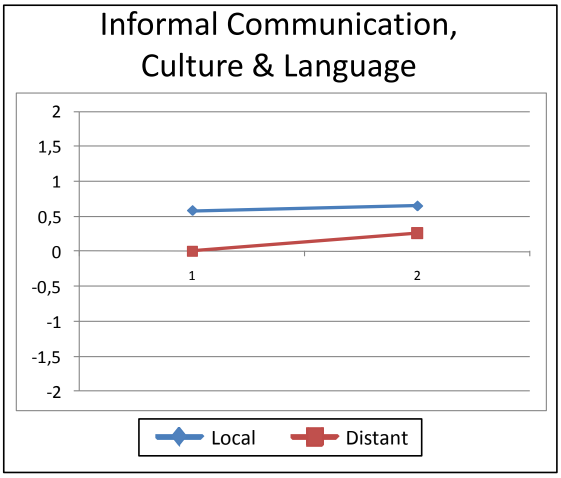
Traveling

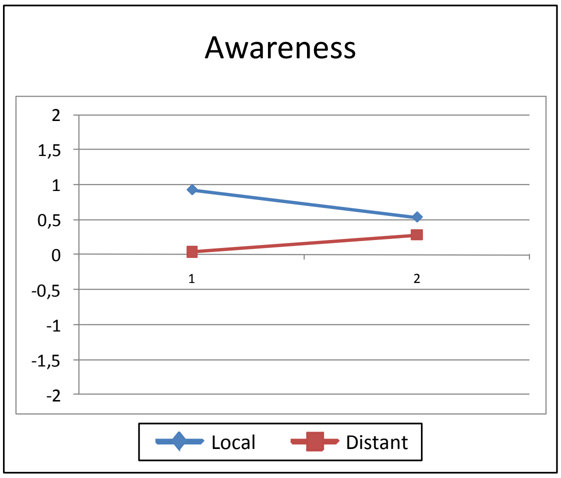
None

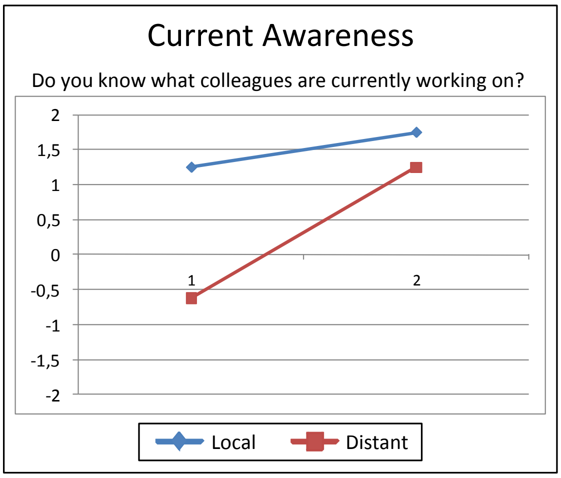
communication frequency: daily = 20, every 2-3 days = 8, weekly = 4, every 2 weeks = 2, less than every 2 weeks = 1, not at all = 0
reachability: very easy = 2, rather easy = 1, rather hard = -1, very hard = -2

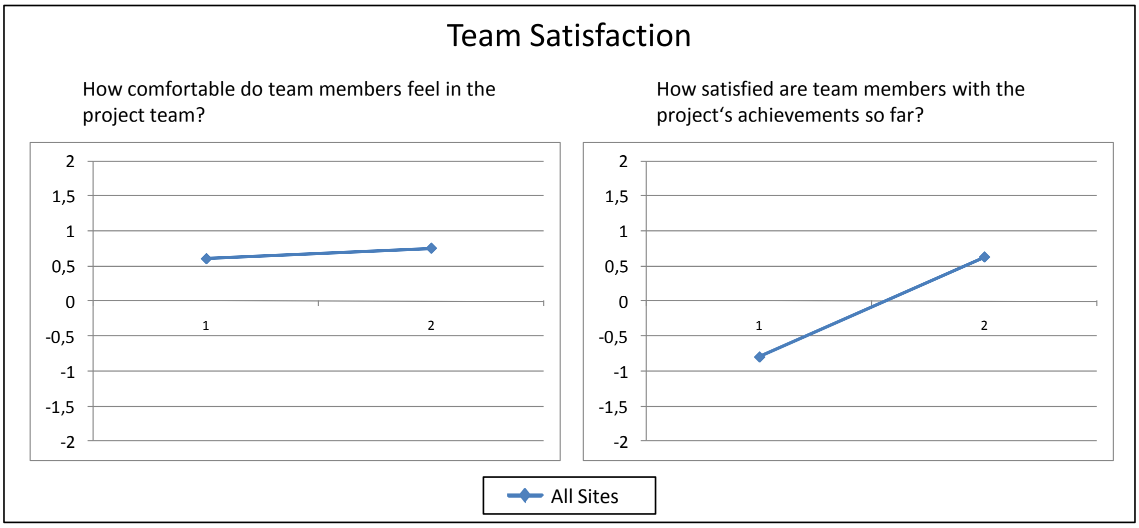
Team: Carpooling and Ride Sharing
Status: Nov. 14th and Dec. 8th, 2010











values: very good/fully agree = 2, good/tend to agree = 1, bad/tend do disagree = -1, very bad/fully disagree = -2

Bibliography

[Ale79] Christopher Alexander. *The Timeless Way of Building.* Oxford University Press, New York, 1979.

[All84] Thomas J. Allen. *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization.* The MIT Press, January 1984.

[BBvB+01] Kent Beck, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, Jim Highsmith, Andrew Hunt, Ron Jeffries, Jon Kern, Brian Marick, Robert C. Martin, Steve Mellor, Ken Schwaber, Jeff Sutherland, and Dave Thomas. Principles behind the Agile Manifesto, retrieved May 2, 2011 from `http://agilemanifesto.org/principles.html`, 2001.

[BD10] Bernd Bruegge and Allen H. Dutoit. *Object-Oriented Software Engineering Using UML, Patterns and Java: International Version.* Prentice Hall, third edition, 2010.

[Ber06] Brian Berenbach. Impact of Organizational Structure on Distributed Requirements Engineering Processes: Lessons Learned. In *Proceedings of the 2006 international workshop on Global software development for the practitioner (GSD'06)*, pages 15–19, 2006.

[Ber08] David Berreby. *Us and Them: The Science of Identity.* University Of Chicago Press, 2008.

[BGR94] Victor R. Basili, Caldiera Gianluigi, and Dieter Rombach. Goal Question Metric Paradigm. *Encyclopedia of Software Engineering*, 1:528–532, 1994.

[BHL07] Matthew Bass, James D. Herbsleb, and Christian Lescher. Collaboration in Global Software Projects at Siemens: An Experience Report. In *Proceedings of the IEEE International Conference on Global Software Engineering (ICGSE'07)*, 2007.

[BHL09] Matthew Bass, James D. Herbsleb, and Christian Lescher. A Coordination Risk Analysis Method for Multi-Site Projects: Experience Report. In *Proceedings of the IEEE International Conference on Global Software Engineering (ICGSE'09)*, 2009.

[Blo07] J. Michael Blocher. Formative v.s. Summative Evaluation, retrieved April 30, 2011 from `http://jan.ucc.nau.edu/edtech/etc667/proposal/evaluation/summative_vs._formative.htm`, 2007.

[BLR+10] Victor R. Basili, Mikael Lindvall, Myrna Regardie, Carolyn Seaman, Jens Heidrich, Jurgen Munch, Dieter Rombach, and Adam Trendowicz. Linking Software Development and Business Strategy Through Measurement. *Computer*, 43:57–65, 2010.

[Boe81] Barry W. Boehm. *Software Engineering Economics*. Prentice-Hall, 1981.

[Bro86] Roger Brown. *Social Psychology: The Second Edition*. The Free Press, 1986.

[Car99] Erran Carmel. *Global Software Teams – Collaborating Across Borders and Time Zones*. Prentice Hall, 1999.

[CH04] James O. Coplien and Neil B. Harrison. *Organizational Patterns of Agile Software Development*. Prentice Hall, 2004.

[CHM09] Bill Curtis, William E. Hefley, and Sally A. Miller. *The People CMM: A Framework for Human Capital Management*. Addison-Wesley, 2009.

[CKS06] Mary B. Chrissis, Mike Konrad, and Sandy Shrum. *CMMI(R): Guidelines for Process Integration and Product Improvement (2nd Edition) (SEI Series in Software Engineering)*. Addison-Wesley Professional, 2 edition, November 2006.

[Con68] Melvin E. Conway. How Do Committees Invent? *Datamation*, 14(5):28–31, April 1968.

[Dam07] Daniela Damian. Stakeholders in Global Requirements Engineering: Lessons Learned from Practice. *IEEE Software*, 24(2):21–27, 2007.

[Dav93] Alan M. Davis. *Software Requirements: Objects, Functions, and States*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[DB98] Allen H. Dutoit and Bernd Bruegge. Communication Metrics for Software Development. *IEEE Transactions on Software Engineering*, 24(8):615–628, 1998.

[DM82] Tom De Marco. *Controlling Software Projects: Management, Measurement and Estimation*. Yourdon Press, New York, NY, USA, 1982.

[Doh08] Sharon Doherty. *Heathrow's Terminal 5: History in the Making*. John Wiley & Sons, 2008.

[DvdWH07] Roger Derksen, Huub van der Wouden, and Paul Heath. Testing The Heathrow Terminal 5 Baggage Handling System – Before It Is Built. *STAR Tester*, April 2007.

[DZ03] Daniela Damian and Didar Zowghi. Requirements Engineering Challenges in Multi-Site Software Development Organizations. *Requirements Engineering Journal*, 8:149–160, 2003.

[Ebe06] Christof Ebert. Understanding the Product Life Cycle: Four Key Requirements Engineering Techniques. *IEEE Software*, 23:19–25, 2006.

[EC06] Kate Ehrlich and Klarissa Chang. Leveraging Expertise in Global Software Teams: Going Outside Boundaries. In *ICGSE '06: Proceedings of the IEEE International Conference on Global Software Engineering (ICGSE'06)*, pages 149–158, Washington, DC, USA, 2006. IEEE Computer Society.

[EC11] Tricia Ellis-Christensen. What Is a Controlled Experiment?, retrieved April 30, 2011 from `http://www.wisegeek.com/what-is-a-controlled-experiment.htm`, 2011.

[Eck10] Jutta Eckstein. *Agile Software Development with Distributed Teams*. Dorset House Publishing, 2010.

[ED07] Christof Ebert and Reiner Dumke. *Software Measurement: Establish – Extract – Evaluate – Execute*. Springer, 2007.

[FU91] Roger Fisher and William L Ury. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin (Non-Classics), 1991.

[Gha06] Jamshid Gharajedaghi. *Systems Thinking – Managing Chaos and Complexity: A Platform for Designing Business Architecture*. Elsevier, second edition, 2006.

[GHKdJCG+00] Arne Gillert, Mohamed Haji-Kella, Maria de Jesus Cascao Guedes, Alexandra Raykova, Claudia Schachinger, and Mark Taylor. *Intercultural Learning T-Kit No. 4*. Council of Europe publishing, November 2000.

[GK04] John S. Gero and Udo Kannengiesser. The Situated Function–Behaviour–Structure Framework. *Design Studies*, 25(4):373–391, 2004.

[GRV09] Ashok Goel, Spencer Rugaber, and Swaroop Vattam. Structure, Behavior & Function of Complex Systems: The SBF Modeling Language. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 23(1):23–25, 2009.

[hea08] Unions had warned BA that Terminal 5 move would cause chaos. *The Times*, March 29, 2008.

[Her06] James D. Herbsleb. Overcoming the Challenges of Global Development. Presentation at OOP Conference, 2006.

[Her07] James D. Herbsleb. Global Software Engineering: The Future of Socio-technical Coordination. In *FOSE '07: 2007 Future of Software Engineering*, pages 188–198, Washington, DC, USA, 2007. IEEE Computer Society.

[HH04] Geert Hofstede and Gert Jan Hofstede. *Cultures and Organizations – Software of the Mind: Intercultural Cooperation and Its Importance for Survival*. McGraw-Hill, 2004.

[HL09] Anne Hoffmann and Christian Lescher. Collaboration and Intercultural Issues on Requirements: Communication, Understanding and Softskills (CIRCUS). In *Proceedings of the RE'09 International Workshop on Collaboration and Intercultural Issues on Requirements: Communication, Understanding and Softskills (CIRCUS'09)*, 2009.

[HLRP99] Detlev J. Hoch, Sandro K Lindner, Cyriac R. Roeding, and Gert Purkert. *Secrets of Software Success: Management Insights from 100 Software Firms around the World*. Harvard Business School Press, 1999.

[HM03] James D. Herbsleb and Audris Mockus. An Empirical Study of Speed and Communication in Globally Distributed Software Development. *Software Engineering, IEEE Transactions on*, 29(6):481–494, 2003.

[Hof11] Geert Hofstede. Dimensions of National Cultures, retrieved April 22, 2011 from http://www.geerthofstede.nl/culture/dimensions-of-national-cultures.aspx, 2011.

[Hou08] House of Commons, Transport Committee. *The Opening of Heathrow Terminal 5, Twelfth Report of Session 2007-08*. The Stationary Office Limited, 2008.

[IBM07] IBM Corporation. IBM Rational Unified Process, retrieved April 22, 2011 from ftp://public.dhe.ibm.com/software/rational/web/datasheets/RUP_DS.pdf, 2007.

[JKG11] Yucheng Jin, Suleman Khan, and Simon Grötzinger. Final Report: Team Electronic Billing, retrieved April 22, 2011 from http://wwwmatthes.in.tum.de/file/attachments/wikis/nereid/nereid-student-team-3/110118_Khan_Groetzinger_Darren_NEREID_Electronic%20Billing%20-%20Final%20Report.pdf, 2011.

[Ker01] Norman Kerth. *Project Retrospectives: A Handbook for Team Reviews*. Dorset House Publishing, 2001.

[KK09] Spencer Kagan and Miguel Kagan. *Kagan Cooperative Learning*. Kagan Publishing, 2nd edition edition, 2009.

[KL08] Rüdiger Kress and Julia Leendertse. Der Horror von Heathrow. *Wirtschaftswoche*, April 14 2008.

[KM88] Stephen Kemmis and Robin McTaggart. *The Action Research Planner*. Deakin University Press, third edition, 1988.

[Kol05] Gina Kolata. Enron Offers an Unlikely Boost to E-Mail Surveillance. *The New York Times*, May 22, 2005.

[KRH06] Ralph L. Keeney, Howard Raiffa, and John S. Hammond. The Hidden Traps in Decision Making. *Harvard Business Review*, 118:120–126, January 2006.

[Les10] Christian Lescher. Patterns for Global Development: How to Build One Global Team? In *Proceedings of the 15th European Conference on Pattern Languages of Programs (EuroPLoP)*, 2010.

[lim] LimeSurvey homepage, retrieved April 22, 2011 from http://www.limesurvey.org.

[lon08] BA chief Willie Walsh brands Heathrow's Terminal 5 a 'national embarrassment'. *London News – London Evening Standard*, May 9 2008.

[LP07] Christian Lescher and Frances Paulisch. Establishing a Common Software Measurement System at Siemens: An Experience Report. In *Business Process Engineering, Proceedings of the CONQUEST 2007, 10th International Conference on Quality Engineering in Software Technology*, pages 1–15. dpunkt Verlag, 2007.

[MCJ+01] John McGarry, David Card, Cheryl Jones, Beth Layman, Elizabeth Clark, Joseph Dean, and Fred Hall. *Practical Software Measurement: Objective Information for Decision Makers*. Addison-Wesley, 2001.

[MDSS08] Sabrina Marczak, Daniela Damian, Ulrike Stege, and Adrian Schröter. Information Brokers in Requirement-Dependency Social Networks. In *RE '08: Proceedings of the 2008 16th IEEE International Requirements Engineering Conference*, pages 53–62, Washington, DC, USA, 2008. IEEE Computer Society.

[MNS+11] Florian Matthes, Christian Neubert, Christopher Schulz, Christian Lescher, Jose Contreras, Robert Laurini, Beatrice Rumpler, David Sol, and Kai Warendorf. Teaching Global Software Engineering and International Project Management: Experiences and Lessons Learned from Four Academic Projects. In *Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU)*, 2011.

[Mor34] Jacob L. Moreno. *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama*. Nervous and Mental Disease Publishing Co., Washington, DC., 1934.

[MW11a] Merriam-Webster. Medical Dictionary, retrieved April 22, 2011 from `http://www.merriam-webster.com/medlineplus/psychology`, 2011.

[MW11b] Merriam-Webster. Medical Dictionary, retrieved April 22, 2011 from `http://www.merriam-webster.com/medlineplus/social%20psychology`, 2011.

[MW11c] Merriam-Webster. Medical Dictionary, retrieved April 22, 2011 from `http://www.merriam-webster.com/medlineplus/sociology`, 2011.

[MW11d] Merriam-Webster. Medical Dictionary, retrieved May 4, 2011 from http://www.merriam-webster.com/medlineplus/control%5B2%5D, 2011.

[MW11e] Merriam-Webster. Merriam-Webster Dictionary, retrieved April 22, 2011 from http://www.merriam-webster.com/dictionary/collaboration, 2011.

[MW11f] Merriam-Webster. Merriam-Webster Dictionary, retrieved May 14, 2011 from http://www.merriam-webster.com/dictionary/group, 2011.

[MW11g] Merriam-Webster. Merriam-Webster Dictionary, retrieved May 14, 2011 from http://www.merriam-webster.com/dictionary/team, 2011.

[MW11h] Merriam-Webster. Merriam-Webster Dictionary, retrieved May 3, 2011 from http://www.merriam-webster.com/dictionary/cohesion, 2011.

[MW11i] Merriam-Webster. Merriam-Webster Dictionary, retrieved May 3, 2011 from http://www.merriam-webster.com/dictionary/culture, 2011.

[MW11j] Merriam-Webster. Merriam-Webster Dictionary, retrieved May 3, 2011 from http://www.merriam-webster.com/dictionary/miscommunication, 2011.

[NAS] NASA. Lost on the Moon Worksheet from NASA, retrieved April 22, 2011 from http://www.kathimitchell.com/lost.htm.

[NB10a] Helmut Naughton and Bernd Bruegge. DOLLI4 Kickoff Meeting Presentation, retrieved April 22, 2011 from http://www1.in.tum.de/static/lehrstuhl/files/dolli4-kickoff.pdf, October 2010.

[NB10b] Helmut Naughton and Bernd Bruegge. Projekt DOLLI4 - Problembeschreibung, retrieved April 22, 2011 from http://www1.in.tum.de/static/lehrstuhl/files/dolli4_problem_statement.pdf, October 2010.

[NS11] Dipak Narayanan Sriram. Final Report: Team Carpooling and Ride Sharing, retrieved April 22, 2011 from http://wwwmatthes.in.tum.de/file/attachments/wikis/nereid/nereid-student-team-4/110118_Sriram_NEREID%20Carpooling%20-%20Final%20Report.pdf, January 2011.

[O'B01] Rory O'Brien. An Overview of the Methodological Approach of Action Research. *Theory and Practice of Action Research*, 2001.

[PDL09] Maria Paasivaara, Sandra Durasiewicz, and Casper Lassenius. Using Scrum in Distributed Agile Development: A Multiple Case Study. In *Proceedings of the IEEE International Conference on Global Software Engineering (ICGSE'09)*, 2009.

[Pro11] Project Management Institute. Project Management Institute: What is Project Management?, retrieved April 22, 2011 from `http://www.pmi.org/About-Us/About-Us-What-is-Project-Management.aspx`, 2011.

[RBP+91] James Rumbaugh, Michael Blaha, William Premerlani, Frederick Eddy, and William Lorensen. *Object-Oriented Modeling and Design*. Prentice Hall, 1991.

[SBM+06] Raghvinder Sangwan, Matthew Bass, Neel Mullick, Daniel J. Paulish, and Juergen Kazmeier. *Global Software Development Handbook*. Auerbach Publications, 2006.

[Sch04] Juergen Schaaf. Offshoring: Globalisation Wave Reaches Services Sector, retrieved May 14, 2011 from `http://www.dbresearch.de/PROD/DBR_INTERNET_DE-PROD/PROD0000000000179790.pdf`, Deutsche Bank Research, September 27, 2004.

[Sch11] Sebastian Schmidt. Final Report: Team Touch Screen Tablet, retrieved April 22, 2011 from `http://wwwmatthes.in.tum.de/file/attachments/wikis/nereid/nereid-student-team-5/110115_Schmidt_NEREID%20Digital%20Menu%20-%20Final%20Report.pdf`, January 2011.

[scr] Scrum homepage, retrieved April 22, 2011 from `http://www.scrum.org/`.

[She88] Muzafer Sherif. *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. Wesleyan, 1st wesleyan ed edition, February 1988.

[Sie11a] Siemens AG. *Siemens Annual Report 2010*. January 2011.

[Sie11b] Siemens AG. Software & Engineering – The Invisible Foundation of Business Success, retrieved February 27, 2011 from `http://www.siemens.com`, 2011.

[SvT81] Friedemann Schulz von Thun. *Miteinander reden 1: Störungen und Klärungen. Allgemeine Psychologie der Kommunikation.* rororo, 1981.

[tha09] London's Heathrow Airport voted worst airport in the world. *Thaindian News*, October 24 2009.

[Toc06] William M.K. Tochim. Research Methods Knowledge Base: Introduction to Evaluation, retrieved April 30, 2011 from `http://www.socialresearchmethods.net/kb/intreval.htm`, 2006.

[UP11] Fatih Ulusoy and Meyyar Palaniappan. Final Report: Tricia Google Translate (Student Team 2), retrieved April 22, 2011 from `http://wwwmatthes.in.tum.de/file/attachments/wikis/nereid/nereid-student-team-2/110118_Ulusoy_Palaniappan_NEREID%20Google%20Tricia%20Translate%20-%20Final%20Report.pdf`, January 2011.

[Wie03] Karl E. Wiegers. *Software Requirements.* Microsoft Press, Redmond, WA, USA, 2003.

[Wik11] Wikipedia. NASA-Weltraumspiel, retrieved April 22, 2011 from `http://de.wikipedia.org/wiki/NASA-Weltraumspiel`, 2011.

[WJB67] Paul Watzlawick, Don D. Jackson, and Janet Beavin Bavelas. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes.* W. W. Norton & Company, 1967.

[WR08] Stefan Wagner and Melanie Ruhe. A Systematic Review of Productivity Factors in Software. In *Proceedings of the 2nd International Workshop on Software Productivity Analysis and Cost Estimation (SPACE 2008)*, 2008.

[WSDN09] Timo Wolf, Adrian Schröter, Daniela Damian, and Tanh Nguyen. Predicting Build Failures using Social Network Analysis on Developer Communication. In *Proceedings of 31st International Conference on Software Engineering (ICSE'09)*, Vancouver, Canada, 2009.