

Using the Delphi Method

Kari K. Lilja, Kimmo Laakso, Jari Palomäki
Tampere University of Technology, Pori, Finland

Abstract--The reliability and validity of the selected research method are subjects to which every researcher is bound to address himself when representing the findings and conclusions of his/her work. In this study we will discuss the reliability, validity and philosophical aspects of the Delphi method, first with a small literature review and then by representing two different surveys conducted using the Delphi method. The point of view in our report is the usability of Delphi in collecting qualitative data for software engineering research. The most significant features of the Delphi method are its recursion and the possibility to get immediate feedback and evaluate one's own answer. Although there are many forms of Delphi techniques, these features exist in one form or the other in all Delphi variations. In Delphi-based surveys, the minimum number of participants is smaller and participants are not selected at random but because of their particular expertise. Among the traditional research methods this is seen to cause the risk of bias and endangering both reliability and validity. In the Delphi method, the recursion produced by three or more rounds, the expertise and – in most cases – anonymity of participants, and the opportunity to evaluate and argue one's own answer after having seen the other opinions and arguments are thought to guarantee the quality of well-planned and well-conducted research.

I. INTRODUCTION

The aim of this study is to explore the usability of the Delphi method to collect complex qualitative data from small groups of experts and advisors to be used as source material in scientific research in relation to software engineering. The key points in this work are the validity and reliability of the collected data. Firstly, we will present the method and discuss its philosophical and theoretical background. In the second section we will introduce two different research studies, one made with the traditional Delphi method and one using modified Delphi research. The third part of this paper will discuss the validity and reliability of data collected by these methods, identifying the problems found when using the method and listing the factors to be taken into consideration when using the Delphi method to collect data in scientific research.

II. DELPHI METHOD

Delphi (or Delphoi, as it is often called) is a method in which the experiences, knowledge, and presumptions of expert panelists on an issue or development process under study are collected in an interactive process, normally by interview or survey. As a data collection method, the Delphi can fall in the category of both a quantitative and qualitative study. It is useful when the phenomenon under study is complex or when the topic is somehow delicate – difficult to

define, awkward to talk about, politically delicate, etc – or the number of members in the focus group is relatively small [9].

The method got its name from Delphoi, Greece, where the priestess, Pythia, also called the Oracle, brought messages from the ancient god Apollo and answered people's questions. These virgin priestesses became famous for their ability to see the future and forecast it. Forecasts were told by muttered mumbling and in a lyrical form, and only the priests of Apollo had the right to interpret the answers [9].

As a scientific method for data collection, the method was developed and first used by Olaf Helmer in the USA in the 1950s and 1960s [6]. At first the method was used as a tool in creating strategies for the army. The Delphi process was originally meant to define the future of a certain phenomenon with the help of experts. The goal was to gain experts' unanimity on how they saw the future of the issue in question. Consensus was the ultimate target, and it was reached by iterating the opinions and the grounds for them among the experts so many times that it was reached – everybody agreed to think in the same way at the end [9] [1].

Due to the topic of the first notable Delphi study, it took time to bring Delphi to the attention of individuals outside the Army and defense industries. Finally, the "Report on a Long-Range Forecasting Study," by T. J. Gordon and Olaf Helmer, was published as a "Rand paper" in 1964 [11]. At the beginning, the Delphi method was received very positively, but over time the results started to arouse doubts and criticism, especially for gaining too simple results [1]. Consensus in itself, particularly a forced one, is strange to futures studies, where one of the main principles of the field of science is that no one can have certain knowledge of what the future will be. Secondly, consensus as a goal leads to the idea of there being just one possible future. The second principle of futures studies is that there are many equally possible futures, among which one will become true. Instead, we can only study possibilities and with certain methods, establish the different levels of probabilities for those plausible future states [9].

The heavy criticism it received caused the Delphi method to be forgotten for close to 20 years. In the 1980s, some researchers returned to the method and started to develop it. Their question was: How could it give better answers to the needs of a rapidly changing society? Especially thanks to the work of U.S. researchers Harold Linstone and Murray Turoff (1975, 2002), the method's reputation was rehabilitated [12, 14, 24]. One of the new developers of the method is Dr. Osmo Kuusi (1999) from Finland [9] [8].

Instead of aiming at consensus, the new Delphi concentrates on new and different knowledge, especially tacit knowledge, and the method's target is to bring this

knowledge under the evaluation and comments of other experts. The new Delphi, argumentative or policy Delphi, as it is often called in contrast to the older version, consensus Delphi, is democratic and equal by its nature. Young, unconventional, unknown in their field of expertise, or somehow and in some other way suppressed experts can also raise their opinions and thoughts and because of the anonymity principle of the new Delphi, no one knows who is behind which answer. All the arguments and points are treated objectively and on an equal basis. Another richness of the new Delphi is its ability to reveal and utilize tacit knowledge [9].

The keywords of Delphi as a method are experts, (small) focus groups or panels, anonymity and iteration [13]. In the next chapters we will discuss all of these elements in more detail.

A. Experts

The Delphi method can be seen at its best as an expert method. The most knowledgeable people on account of expertise are often also thought to be ahead of the others in their ideas about the future because of their exceptional understanding concerning the technical, sociological, medical, political development etc., or, at least, their potential capability to imagine it.

An expert fit for a Delphi panel should be [8]:

1. At the top of her/his field of technical or scientific knowledge;
2. Interested in a wide range of knowledge, not only in her/his own field but everything around it;
3. Able to see connections between national and international, present and future development as well as connections between different fields of science;
4. Able to disregard the traditional viewpoints and regard problems not only from the known and safe but also from unconventional angles;
5. Interested in creating something new.

This viewpoint of the requirements of a good panelist also reflects the modern idea of expertise [9]. Furthermore, the experts selected for the panel should be able to express their ideas and visions in such a manner that others, not specialized as such in this field, are able to pick up their ideas for upgrading. They should also be motivated and committed to being a member of the panel.

B. Panel

The panel consists of a group of selected experts. Delphi as a method does not set limits on the size of focus group. The main task is to include in the group people who have the greatest knowledge and / or experience in the area of science / technology. Due to this, the size of the group remains in most cases small. The interactivity and recursivity, which are essential features of the process, might suffer if the group grows too much. The final size of the panel group is always a

compromise between practical needs and the requirements set by reliability and scientific principles.

C. Anonymity

The anonymity of the panelists is essential for the Delphi method. Anonymity supports independence by avoiding the limits and problems of expression and listening to one another, which are always present in face-to-face expert groups. The official position or unofficial status of a panelist does not affect the opinion or its formation and expression. Also, a member of the panel can be free from the fear of losing face, even if she/he gives an answer or comment that others might find to be wrong or inaccurate. A panelist also does not need to be wary of attitudes, which her/his employer might find inappropriate to be stated in public. In interest or value conflicts, issues do not become personalized in the same way as in face-to-face communication. [9] Avoiding face-to-face communication between the members of the group also avoids impacts of mimicking and other forms of inarticulate communication that occurs as a problem in other kind of focus group methods [11]. Furthermore, anonymity provides safety for panelists in cases where the panelists or their employers are competitors, especially if some or all of the panelists come from the business world. Anonymity also gives safety to focus groups when the subjects studied are experienced as "hot" politically, or incorrect in some other way. In short, anonymity guarantees more objective answers and results.

There might also be cases where anonymity is not necessary, or where it can even be an obstacle to potential results [20, 21, 22]. In studies where expert panelists are needed as representatives of their specific group of interest, or where a group of experts is combined by consensus on the development of the study subject, anonymity might encourage the panelists to give personal opinions, while the aim is to obtain knowledge of their specific background group. Delphi of this kind is called Disaggregative Policy Delphi. It is used in societies which are largely institutionalized and structured and in which it is possible to nominate a representative group for each relevant line of thinking. Then the tacit and/or social knowledge is brought to light and presented to the others – to wait for their comments. Each information producer is set at the same level from the point of view of the study. [9]

D. Iteration and feedback

The fundamental difference between ordinary surveys and the Delphi method is the iteration and feedback used in Delphi. In contrast to Gallup-type surveys, opinions are not only collected for analysis, but information on the answers will be fed back to the panelists for comments and / or as a basis for the next round. With the help of this feedback, the respondents are obliged to give grounds for their choices. The building up of information proceeds round by round so that the previous round forms the basis for the next one, which is essential for the Delphi process. [9]

The Delphi process normally consists of four phases [11]. The first round questionnaire starts the study process by orientating the panelists to position themselves regarding the Delphi process, the subject and each other [9]. In the first phase, each participant contributes the additional information he/she feels is pertinent to the issue [11]. The next phase involves reaching an understanding of how the group sees the issue, where the members agree or disagree, and what kind of meaning they give to relative terms such as importance, desirability, or feasibility. If there is any significant disagreement, it will be explored in the third phase to bring out the underlying reasons for the differences and possibly to evaluate them [11]. During the commenting and arguing of the second and third rounds, the panelists clarify their opinions and views and try to convince the others [9]. The panelists are able to clarify their answers and comments during the phases. If this happens, it is a positive signal of listening and ongoing dialogue. Between the phases, the manager (researcher) analyzes the results and formulates the arguments given as new claims for the panel to vote on in the next round [9].

In Internet-based Delphi, there is an opportunity for synchronic dialogue between participants. It is vitally important to promote communication and problem-solving in the focus group. The panelists do not necessarily have to answer all the claims, only those about which they feel they have something relevant to say. Expert evaluations have been shown to improve when the panelists are able to reflect on the credibility of their answer [9] [25].

III. THE PHILOSOPHY BEHIND THE DELPHI PROCESS

In this section we will briefly outline the characteristics that concern us here, which are of particular importance in validating the Delphi method. To support the different criteria people usually use to validate a method, see e.g. Kvale, S.: Issues of Validity in Qualitative Research [26].

Firstly, every method involves a repeatable sequence of operations bringing about, when performed, a repeatable event called the *objective* of the method. We may call the range of the scientific method to be the class of all objects to which this method could possibly be applied. In the case of the Delphi method, we should make it very clear what the objective is, i.e. whether we are basically interested in the experiences and/or opinions of the panelists, or the past and/or future situations of which the panelist have experience / opinions. Especially, are we validating the opinions of the panelists, or the future situation? – If we are primarily interested in future situations, we are basically making a prediction. All predictions of the future involve the concept of probability, and it is well known that there are controversies over the correct interpretation of the meaning of probability. The Delphi method is based on the supposition that experts know more facts and have more experience behind their prediction and thus their opinion has more weight than people not accustomed to the area in question.

Moreover, when predicting the future, we must admit that future situations are unique, and, most probably, not repeatable events.

Secondly, every method requires a person or a class of persons who are said to be familiar with it, or, in other words, who are *competent to apply it*. The same person may apply the same method several times, and several persons may attain their objectives by applying the same method. In the Delphi method applied to a particular situation, we may think that the more experienced the person using it, i.e. the more he/she has used it in previous situations, the more confident he/she will also be in that particular situation. However, since every situation is unique, predicting the future will always be based on probability only. Even the professional people who predict the future, like meteorologists, sometimes make wrong predictions about the weather, and the very same holds for economists as well. Moreover, it is assumed that whoever used the Delphi method for the same panelists would have the same results. This is not a very plausible supposition, since the user of the Delphi method has in fact a very active role when applying it, which depends on the personal character as well as the knowledge of the researcher.

Thirdly, the operations involved in the method may require the handling of a specific set of objects, which does not include the object to which the method is applied. Such a set of objects is called the *instrumental equipment* of the method, which must meet the requirements of the method. Of course, the objects may vary from application to application, i.e. the very same instrument does not have to be used to attain the objective of the method. In the Delphi method, the instrumental equipment consists of the objects used when the questionnaire is distributed to the panelists. One essential requirement for the instrument used is to guarantee the anonymity of the panelists. Fortunately, nowadays we use computers and web-based questionnaires having the technical possibilities for the anonymous processing of a panelist's answers.

Fourthly, an individual application of the method may fail in a certain fraction of its application to realize the objective of the method on some particular occasion. The probability that a single application of a given method chosen at random will successively produce the objective of the method may be called the *reliability* of the method, i.e. the reliability is the ratio of all successful applications of the method to all its applications. The probability that a particular application of a given method will be successful may be affected by the special circumstances under which the method is applied. This probability will then differ from the reliability of the method. The rules that specify the circumstances left undetermined by the description of the method, such as the choice of equipment or the way of handling it, constitute its *technique*. More particularly, for example, the successful application of the Delphi method may depend upon the choice of the person who applies the method, upon the physical, psychological, and social conditions prevailing in his/her environment, upon the choice of the instrumental

equipment and questionnaires, and the way of handling them. The reliability of the Delphi method, which is one of the main issues of this paper, will be considered in more depth in the next section.

Fifthly, in some cases, there is a special theory determining the degree of reliability of a given method and how this reliability may possibly be enhanced through the use of a suitable technique. A theory of this kind may be called the *theory underlying the method*. Again, one of the aims of this paper is to formulate an approach towards developing an underlying theory for the Delphi method.

IV. RELIABILITY AND VALIDITY OF THE RESULTS OF THE DELPHI PROCESS

The Delphi process differs from traditional surveys in two ways: Firstly, the respondents are not picked randomly but are selected because of their knowledge and experience – that is: due to their expertise [15]. Secondly, the number of respondents can be much smaller than what is traditionally thought to be sufficient to guarantee the reliability of a survey [15]. This is why there has been a lot of discussion, occasionally even strong disagreement, on the scientific reliability of the results assessed by Delphi. The critics argue that the number of respondents in an average Delphi research study is too small to guarantee the reliability of the work [15, 17]. The second argument presented by critics is that the method by which the respondents are selected for the Delphi panel is not objective or based on probability, and therefore the answers cannot be considered reliable in the scientific meaning [15, 17]. The fact that results obtained from different panels may differ from each other has also been seen as a sign of the unreliability of the Delphi method [15].

There are many sampling methods available in traditional research surveys, and not all of them are based on probability. For example, in small populations the whole population may be a sample [18]. Other non-randomized sampling methods are for example quota-sampling and haphazard methods [18]. We must also remember that in surveys, even if we use a randomized sampling method, the research question always limits the population to a target group from which the sample is picked [19]. The size of the sample in traditional surveys is in many cases large. However, size itself is not significant for reliability but the representativeness of the sample is – how well it represents the whole population [19]. A greater sample reduces sampling error and enhances representativeness, but does not guarantee it [18].

Delphi is an expert method. This is a fact accepted by many writers [15, 17]. It is a method used to collect experts' opinions, knowledge, and experiences concerning a certain limited problem or research question. The reliability of this kind of method is condensed into three items: selection of experts, size of panels, and conducting of the process, including setting the questions and reaching consensus – if consensus is a goal.

For example, during an advanced course on the scientific method in autumn 2010, we illustrated the principles of the Delphi process with a half-hour study among students. The aim of this study was to get the students themselves to think how many experts could guarantee reliable results, and who would be a good expert. Participants in the course were both experienced researchers (6 persons) and novices (4 persons). The research started with two questions: "How many experts should there be for you to believe their answer?" and "Who is a good expert?" The answers to the first question varied from 1 to 9 experts and a good expert was defined as a person who has sufficient knowledge of the branch and enough experience. In the second phase, the panel members were instructed to evaluate their answers and one additional question was set: "Who should define the expertise of candidates?" The number of experts needed was evaluated at 3 – 9. It was suggested that even numbers be omitted to avoid a 50/50 result. The definition of an expert was "someone with years of the experience needed". The expertise of the panelist should be recognized by someone other than the researcher. In the third round, participants evaluated their answers and the final questions were refined. Although none of them knew at that moment that consensus could be the best situation achieved, the answers came closer to each other and the results of this small Delphi work achieved the following form:

"To guarantee a reliable result, an expert panel should consist of 3-9 members as a minimum, even numbers should be allowed, experts should have sufficient knowledge of the discipline gathered via education and / or experience, and their expertise should be recognized by colleagues or some third party with the capability of evaluating expertise in this field."

It should be remembered that – although all of them knew some research methods - none of these students had earlier knowledge or experience of Delphi, and even so, the answers are quite similar to those found in the literature, with one exception: In the literature, no one has been afraid of an even number of participants and the possibility of a dead heat.

In the literature, the number of panelists is mostly set at between 15 and 30 [e.g. 15], the minimum is said to be 13 [16], but also smaller and bigger panels have been seen [17, 2]. An expert is defined – as described above – as a person with excellent and recognized knowledge in the field, a wide interest in knowledge outside his own discipline, long experience in the branch, and willingness to create something new without being tied to traditional viewpoints [9, 17]. It is also recommended that a panel should be as heterogeneous as possible [17] to ensure discourse and a real achievement of consensus – if that is the goal. In homogenous groups there is the risk of axiomatic consensus: People with the same background, education and experience seldom find new approaches or solutions to a problem.

Key qualities ensuring the features of the Delphi method are anonymity [4] and the recursive and iterative nature [3] of

the Delphi process with the possibility of bringing up new ideas that are perhaps not so politically correct, and the possibility of evaluating and comparing one's own knowledge, opinions and answers to those of others.

The validity of the answers and results is mostly seen to be in the researcher's hands. How well have the questions been formed and set, does the panel consist of precisely those experts who have the best knowledge and experience, and are the answers correctly collected and analyzed? Ensuring this requires careful planning and testing of research settings [7].

V. TWO DIFFERENT DELPHI STUDIES

After the original Delphi process was presented by Rand Corporation in the late 1950s, the number of different Delphi versions has grown continuously. The basic idea in all of them is the same as in the first: Gathering data from a limited group of experts through an iterative and in most cases anonymous process enables the evaluation of the answers and reaching of consensus if so desired.

In this chapter we introduce two different Delphi-based studies conducted within the period 2009-2010, both having the same goal: To assess both the known and tacit knowledge of experienced experts concerning history and the future.

A. Study A: *The impact of the organizational culture of customer and supplier on a common IT project*

The aim of the study below was to collect the experiences of experienced project managers concerning cultural differences and their impacts. The purpose of the research was to evaluate the findings presented in PICMET 2010 [10] and to find more accurate questions and means of becoming aware of risky cultural differences. The Delphi-based research with 35 respondents and 5 phases was conducted during 2010.

Before we were able to start, we had to determine what was meant by expertise i.e. who the experts are in this case. We wanted to find out if people involved with projects other than those investigated in the study named above had similar experiences. This meant that suitable persons should have been working in several projects as project manager on the customer's or supplier's side or as consultant and that they should have such an education, either formal or informal, that they were able to answer the questions. Respondents to this study were selected from multinational IT suppliers, consultants and their Finnish customers so that each company named a person whose experience and knowledge was at a level that could be called expert. The average experience of the respondents was estimated to be nearly 20 years, and the oldest had been in the business since the early 70s. The education level of most of the participants varied from Bachelor's degree to Master of Sciences, and their disciplines were Technology, Information Technology or Financial / Economic Science. Suppliers' and consultants' representatives were project managers and customers'

representatives either CFOs or CEOs with responsibility for IT projects.

They were first asked to describe in their own words what kind of problems caused by differences in customer and supplier business culture they had met during their career. From their answers the keywords were elicited and standardized. The most common keyword was "Requirements assessment." Together with that the term "Language" was very often found, with the meaning that the opposite sides did not have a common language. The total number of different terms found in the answers was 110 terms occurring 220 times.

The next step in the study was to find out the priority of the terms. A report concerning the results of first questionnaire was delivered to the respondents who were allowed to comment on the results and/or their own answers. The findings of the first round were formulated into new questions in other words. For example the question concerning the relation between requirements assessment and a common language was formulated as follows: Is it difficult to find common terms /language within the requirements assessment process? The alternative answers were: "It is difficult", "It is not difficult" and "Cannot say."

The members of the panel commented on the findings of the first round with expressions like "This is what I have seen all the time but have not been able to describe", "The soft aspects in projects are underestimated by engineers" and so on.

In the third round, where the results of the second phase were published, comments on the questions were that it might have been a good idea to have more alternatives, for example in levels "very difficult", "difficult", "a little difficult" and "not difficult at all". With hindsight we must agree that these comments were right. On the other hand, we had experiences from earlier studies that if a Finnish person is allowed to answer in a neutral way, he will do so, and the clear divisions were selected advisedly. The second round received more criticism from the panelists than the other phases due to its length. The number of terms in the first round was greater, and although single terms meaning the same were combined if possible, the questionnaire was long.

The aims of the third and fourth rounds was to find out what kind of impact the cultural differences and problems identified within the first two rounds had had, according to the respondents' experience, and how they had tried to manage, avoid and/or cope with these impacts. Questions were now divided into two groups between the third and fourth rounds depending on their subject. The respondents were also asked to suggest questions they would set to find out cultural differences that might cause problems during a project. The possibility to evaluate one's own answers and comment on others was maintained all the time.

The fifth round was a converging phase. The questions and means presented in the third and fourth rounds were grouped according to their subject into 4 groups and respondents were told to select 2-4 items from each group,

which they would ask or check before a project to assess risky cultural differences. In addition to this, the respondents were told to evaluate the impact of differences between customer and supplier in the subjects gathered within the first round on a common IT project.

35 persons were invited to join the panel. All of them accepted the invitation. The number of participants that answered in each round was:

	Number of Participants
First round	35
Second round	20
Third round	19
Fourth round	17
Fifth round	17

It is easy to see that after the first round those who remained involved were interested in the subject. The biggest loss of respondents was after the first round. The number of active respondents stayed at over 15, and all of them answered all the questions.

The environment in which the work was conducted was Webropol, a www-based questionnaire tool with possibilities to log the users, identify them and give immediate feedback and send e-mail to all panelists at once. Due to the fact that the respondents were each other's customers, competitors and suppliers, we had to guarantee complete anonymity to the respondents, meaning that no one was able to recognize an individual respondent. Complete anonymity also means that Webropol's logging, identifying and feedback functions were not used, which made it impossible to trace who answered and who did not.

The environment, questions and questionnaires were tested before the first round with a smaller test group, and the errors found were corrected before the research started. However, the diversity of Internet browsers and Webropol's own update during the fifth phase caused a small problem: Two questions with radio-buttons did not work correctly. Luckily, this was found out after just two respondents had answered, and was corrected at once.

The validity of the questions and answers can at this moment be evaluated via comments given on the feedback sent to the respondents after each phase. Feedback consisted of summary results and conclusions of the latest completed round. Results were seen to be useful in the respondents' daily projects, accurate and exactly what had been needed but not received until this point. Critical opinions wished for a deeper approach to the questions, more alternative answers, and also shorter questionnaires. One third-party representative also wondered if any work of this kind – no matter how good it is – could lead in practice to any usable conclusions.

The reliability of the results can be assessed from the number of panelists and respondents, their expertise and

commitment. The number of panelists that answered all the questions was 17 whereas some kind of minimum is thought to be 15 [15]. The expertise was estimated by each respondent's superior or colleague, i.e. someone other than the researcher. All the panelists had long experience in IT projects and all of those who answered all the questions seemed to have a strong commitment to research. We wanted to get information about problems caused by collisions between different cultures. Every one of the respondents had experiences of their own, and told them openly in their answers to open questions trusting the anonymity, and they also answered the structured questions. There are no signs of manipulation or hiding the facts in answers. From comments given to the feedback of earlier rounds we could see that problems were quite common but people working with them do not necessarily see them until someone speaks about them openly. This is the way that the Delphi process helps bring tacit knowledge into the spotlight.

B. Study B: The role of regulation in the mobile operator business in Finland

The second study concentrates on the mobile operator business. The purpose of this study was to collect both historical data and forecasts concerning the role of regulation in the mobile operator business. The study was originally presented in PICMET 2010 [9]. In this paper we want to assess the reliability and the validity of the study. Mobile communication has grown beyond its previous scope and scale, and at the same time it has become one of the most influential factors of change in society and the way people interact with each other. Since the mobile operator business is highly regulated, the effects of regulation on the industry have been analyzed. The potential effects in the years up to 2015 were also considered. The method chosen for this study was Delphi, because expert opinions on the most important changes in regulation and their impacts on the business was the target of research interest in the study. The challenge was that the Delphi method was originally created to assess experts' opinions about the course of development of a certain technology or phenomenon in the future and then, by using for example a scenario technique, to draw conclusions about its possible futures. In this case the Delphi method was also used to estimate past development, i.e. experts' opinions of the causes and effects of laws and other regulations in the past few decades.

The research started with a desk study on the mobile operator industry in Finland, which analyzed the changes in regulation in a period of over two decades, starting from the second half of the 1980s and ending in 2009. Then a number of companies were analyzed and various data from the industry as a whole was collected. Among other things, the collected data included the services offered, the number of subscribers, as well as the usage and average prices of services. The final research questions were then created on the basis of the research problem. A three-round iterative Delphi study was conducted in 2009–2010.

According to e.g. Gordon [5] and Kuusi [8], the selection of experts for the Delphi panel is one of the most critical phases of a Delphi study. For the needs of this specific study, the successful realization of Delphi also required the design of a panel structure which allowed many knowledgeable individuals from different disciplines or specialties, and having a different working background and experience, to contribute information or judgments to a problem area that is much broader in scope than the knowledge which any single individual can possess. Therefore, the objective of research could not have been achieved if all the parties involved in the Delphi had been drawn from the same specialized interest group [5]. For that reason, before selecting any individuals for the panel, attention was paid to selecting companies and other organizations that were considered likely to possess the desired knowledge on the regulation of mobile operator industry. Therefore it was decided that the panel should represent the following interest groups: a) Authorities, b) Mobile operators, and c) Other stakeholders.

It was also considered that the experts on the panel should have personal competences and working experience covering: a) Operations (including management and product development), b) Law, and c) Research and development of the industry. In the next stage, senior level persons in the selected organizations were contacted in order to find the right individuals for the panel.

The actual size of a Delphi panel is not limited but the literature often recommends that the panel should have at least 10-15 members [5, 8, 20, 23]. There were 14 experts on this panel, consisting of 12 men and 2 women. The majority of the panelists were male since the topic area is such that even today there are fewer female experts. The panelists shared a wide range of understanding of the telecommunications industry. The average age of the panelists was 48.1 and the average working experience in the telecommunication sector 20.9 years. Their personal competences and organizational interests can be expressed as follows Fig. 1:

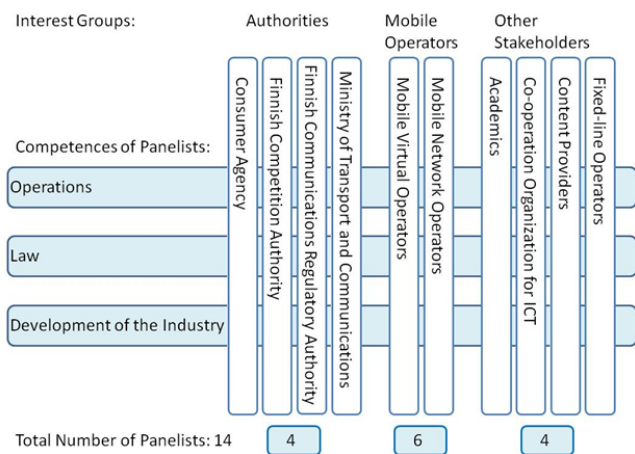


Fig. 1 Panelists in the Competence/Interest Group matrix

Since the panel was relatively small, it was decided to carry out the first Delphi round by means of recorded interviews. Panelists were interviewed for 1.5 – 2 hours. The interviews were recorded, transcribed, and documented. The interviewees were given a comprehensive list of changes in regulation and they were also asked to talk about and describe the changes that, in their opinion, had been the most significant. The effects of those changes on the industry were also the focus of the interviews. Through personal contact with the panelists we also wished to increase the commitment of the panelists for the next rounds.

The second and third Delphi rounds were carried out using Internet-based Delphi software. This software has been developed by the Finnish Delphi Community (see <http://www.edelphi.fi>). The main features of the software are user administration, questionnaire creation, and organization of answers and comments. In addition, the software has various reporting possibilities. All 14 panelists participated in the second Delphi round and 13 also participated in the final third round.

In this Delphi application, the complete anonymity of panelists was considered unnecessary [20, 21]. It would also have been difficult to put into practice because all three mobile network operators in the country were represented on the panel and the representatives were persons with comprehensive experience in telecommunications. Therefore, at the beginning of the first Delphi round, the panelists were given the names of all the other panelists participating on the panel, but nevertheless they were promised that the answers and arguments would be handled and shown to the panel anonymously.

To ensure the validity of the study, a proper desk study on the subject had been conducted before starting the Delphi process. The first Delphi round was done by interviewing all panelists personally. Interviewees had the possibility to express their opinions on the research subject freely. Questions and claims for the next two Internet-based questionnaires were then formulated based on the desk study and the information from the interviews. This was intended to increase the validity of the study. Questions and claims for the third round questionnaire were formulated iteratively based on information from the earlier rounds. Panelists gave focused answers and comments on the first Delphi round, which they shared with the other panelists in the second and the third Delphi rounds. Their answers and comments demonstrated the validity of the study, since the answers concerned the core topic of the study.

From the reliability point of view, a great deal of attention was paid to the selection of experts for the panel. First of all, it was carefully considered in which organizations there might be the best understanding of the topic of the study. Secondly, top-level managers of these organizations were asked to recommend an expert for the panel. They were asked to recommend persons with the best possible expertise in the research area. The selected experts on the panel have an average of more than 20 years' experience of the industry,

2011 Proceedings of PICMET '11: Technology Management In The Energy-Smart World (PICMET)

which was considered to be adequate for the objectives of the study.

Thirdly, in order to ensure the experts' suitability for this research, they were asked to assess their own skills in the

desired competence areas. Responses were on a scale of 0 to 5. The averages of the responses by interest group were then calculated and are shown in Table 1.

TABLE 1: PANELISTS' SELF ASSESSMENT OF THEIR COMPETENCES

Competence Area:	Authorities	Interest Group:		Average
		Mobile Operators	Other Stakeholders	
Operations	3.50	4.67	4.50	4.22
Law	4.25	4.00	4.00	4.08
Development of the Industry	3.50	4.00	4.00	3.83
Average	3.75	4.22	4.17	4.05

The answers ranged between 3.50 and 4.67. Of course, not all of the panelists were expected to have expertise in all of the desired competence areas, but in any case, the results varied relatively little. This was thought to increase the reliability of the research results. Fourthly, there was only one dropout during the whole Delphi process. All 14 experts participated in the first two Delphi rounds and 13 of them also in the final third round. It was considered that the personal interviews at the beginning of the Delphi process increased the commitment of the experts to the research.

VI. EXPERIENCES AND IDEAS OBTAINED WHILE CONDUCTING THE DELPHI PROCESSES

The first round interviews proved here at least to be a good way to get panelists committed to the whole Delphi process since there was only one dropout during the whole Delphi process in study B. On the other hand, the fact that the employers selected the people to participate in the panel did not guarantee commitment. In study A, out of the 35 panelists taking part in the first phase, only 17 were involved in all 5 phases. Due to the guaranteed anonymity, we were not able to study the reasons for this, but thanks to the messages sent to the researcher at least three different reasons are known: Firstly, many of the panelists were working with customers and/ or plants located both in Finland and abroad. The strikes at airports in Finland and the volcanic ash clouds from Iceland in 2010 disturbed travel during the second, third and fourth phase so that people were overworked and gave up everything that did not directly belong to their job. Only those who were personally interested in the subjects of the research remained. Secondly, although people were told before their approval was asked that the research had 5 different phases and that it would last from February to November, some panelists who gave their acceptance had not realized this. And thirdly, some of the participants found that the second phase was much harder than expected. The advance information could have been more accurate and direct. A face-to-face meeting with each panelist might have clarified the situation and increased commitment. However,

for reasons of force majeure, it might still not have been possible to eliminate dropouts.

In both studies, many of the panelists said in their comments that these subjects were so mundane that they had no idea that there might be aspects they had never recognized, heard of, or thought about until someone else spoke out about them. This is the way that the Delphi process helps us to gather tacit or hidden knowledge. Each expert has knowledge and experience of his/her own. This intellectual property may in some cases even be a critical part of his/her business, a competitive advantage that must not be given to competitors. In the safety of anonymity, it is easier to express opinions and share experiences and lessons gained over the years. Experts are motivated to participate when realizing that they might not only share their own knowledge but also learn from the experiences of others, even their competitors. To maximize this, the Delphi panel should be as heterogeneous as possible.

The capability of the Delphi process to gather complex and qualitative data, tacit knowledge and experiences as well as ideas and opinions about the future makes it very usable in many areas. Until now it has been mostly used in futurology, where it was first applied, and in social, health, and medical disciplines. In management and technological disciplines, its use is continuously growing. According to our experience, it is well suited to studies where the population and / or sample would be relatively small, and giving the information the researchers wish to collect would require expert knowledge and experience. This kind of setting will occur increasingly in technical sciences and software engineering. The bias caused by a small sample is minimized by the iteration and anonymity of the respondents.

In Delphi studies, the biggest probability of bias comes from issues that the researcher has a major impact on. For instance, how well the criteria for panelists are defined and prepared. How well the panelists are selected and familiarized with the technique. How well the questions are set and written and how well the questionnaire and the technical solution to conduct the study are tested. And finally, how well the analyzing methods are planned and tested. Thus

advance defining, planning and testing are the keys to reliable and valid results. In some cases, like in study 1, the technique deployed may lead to situations where we are bound to act ad-hoc. In such a situation good advance planning is even more important. We cannot test everything but we should be prepared for anything.

The reliability of both studies presented is based on four arguments:

1. The number of panelists was adequate (minimum of 17 in the first, 13 in the second study) vs. the minimum presented in the literature, which varies from 10 to 15 panelists.
2. The iterative nature of the method – each panelist had the possibility to evaluate his answer and to comment on other answers.
3. The anonymity of the method – none of the panelists knew who had said what. There was no fear of losing face, being laughed at or being identified as a representative of a certain stakeholder or company.
4. The quality of the panelists. In the first study, each company nominated a panelist according to experience and knowledge; in the second, almost all of the perceived stakeholders were involved.

The validity of the research can be assessed by evaluating the answers and results: Do the answers really answer the question, are they logical and well-formed, are the results logical, do they answer the research question and are they applicable in practice? In both studies, the logic and the applicability of the answers were recognized by the panelists. The results answered the research question but also produced a lot more information and data.

VII. CONCLUSIONS

Although mainly used in futurology and social, health, and medical disciplines, the Delphi method can be applied to certain types of research in technical sciences, software engineering and related disciplines as well. In studies where the aim of the method is to collect qualitative data and information from a limited group of specialists or group of people that can be seen to be expert because of their knowledge and / or experience, some variation of the Delphi method is one alternative for the researcher to consider. As inbuilt properties, it has iteration and (in most versions) anonymity that helps guarantee reliability even with small samples.

However, the researcher has a great impact on the quality of the results. By planning the work carefully, letting an external evaluator evaluate the expertise when choosing the panelists, and testing the environment, advice, questions, and questionnaires with an external testing group will help avoid bias and improve the reliability of the results. While testing the questionnaires and questions it could be a good practice to compare them with the research question: Do the results obtained via the questionnaire really answer the research

question and are the results usable in the rest of the research or in practice? If the answer to these questions is “yes,” the results are more probably valid enough to conduct the study.

There are certain circumstances for which the Delphi methods proved to be very practicable. Tacit and hidden knowledge, sensitive information, undocumented practices, and facts and opinions that were thought to be incorrect to utter were made known from the safety of anonymity. Anonymity also seemed to make it easier for the panelists to evaluate both their own and other participants' answers and comments. Due to the features described above, some variation of the Delphi process might be a good data gathering tool for purposes such as collecting requirements for a new system, defining and improving business processes, and defining and improving software engineering processes.

ACKNOWLEDGMENTS

The authors would like to thank the Finnish Cultural Foundation, the Foundation for Economic Education, the Ulla Tuominen Foundation, The Teliasonera Finland Foundation for Research and Education and the High Technology Foundation of Satakunta for partly funding our research projects.

REFERENCES

- [1] Bell W., “Foundations of futures studies”. *Human Science for a New Era. Vol. 1 History, Purposes, and Knowledge*. Transaction Publishers, New Brunswick and London. pp. 261-272. 1997.
- [2] Dagenais, F., “The reliability and convergence of the Delphi technique”, *The Journal of General psychology*, 1978, 98, pp. 307-308
- [3] Dietz, T., “Methods for analyzing data from Delphi panels: Some evidence from a forecasting study”, *Technological Forecasting and Social Change* 31(1), 1987, pp.79-85.
- [4] Goodman, C. M., “The Delphi technique: a critique”, *Journal of Advanced Nursing* 12(6), 1987, pp.729-734.
- [5] Gordon, T., J., “The Delphi method in futures research methodology-V3.0.” *The Millennium Project*. Eds. J. C. Glenn and T. J. Gordon. Retrieved 10/26/09 World Wide Web, <http://www.millennium-project.org/millennium/FRM-V3.html>
- [6] Gordon T.J. and O. Helmer, “Report on a long-range forecasting study”. *RAND-paper P-2982*, Santa Monica, CA. 1964.
- [7] Hasson, F., S. Keeney, H. McKenna, “Research guidelines for the Delphi survey technique”, *Journal of Advanced Nursing* 32(4), 2000, pp.1008-1015
- [8] Kuusi, O., “*Expertise in the future use of generic technologies. Epistemic and methodological considerations concerning Delphi studies*”, HeSE Print, Helsinki. 1999.
- [9] Laakso, K., A. Rubin, H. Linturi, “Delphi method analysis: The role of regulation in the mobile operator business in Finland”, in *Papers presented at PICMET*, July 2010
- [10] Lilja, K. K., and H. Jaakkola, “The differences between the supplier's and the customer's business cultures and their impact on the result of an IT project”, in *Papers presented at PICMET*, July 2010
- [11] Linstone, H. A., M. Turoff, O. Helmer, (eds.); “*The Delphi method techniques and applications*”, 2002, CA, USA
- [12] Linstone H. and M. Turoff, (eds.); “*The Delphi method: techniques and applications*”, Addison Wesley Publishing Co, Advanced Book Programme. Retrieved 10/26/09 World Wide Web, <http://is.njit.edu/pubs/delphibook/>

2011 Proceedings of PICMET '11: Technology Management In The Energy-Smart World (PICMET)

- [13] "Millennium project". Retrieved 10/11/09 World Wide Web, <http://millennium-project.org/>
- [14] Mitroff I. and M. Turoff, "Philosophical and methodological foundations of Delphi" in *The Delphi Method: Techniques and Applications*. Eds. H. Linstone and M. Turoff. Addison Wesley Publishing Co, Advanced Book Programme. Retrieved 10/26/09 World Wide Web, <http://is.njit.edu/pubs/delphibook/>
- [15] Loo, R., "The Delphi-method: a powerful tool for strategic management", *Policing: International Journal of Police Strategy & Management*, Vol 25, No. 4, 2002, pp. 762-769
- [16] Philip B., "Mission, initiatives, and obstacles to research in agricultural education: a national Delphi using external decision-makers", *Journal of Agricultural Education*, winter 1989, pp. 14-23 (p.15)
- [17] Powell, C., "The Delphi technique: myths and realities", *Journal of Advanced Nursing* 41(4), 2003, pp. 376-382
- [18] Sapsford, R., "Survey research", Sage Publications Ltd, 1999, London, UK, pp. 49 - 100
- [19] Sapsford, R., "Survey research", Sage Publications Ltd, 1999, London, UK, pp. 1 - 48
- [20] Tapio, P., "Limits to traffic volume growth. The content and procedure of administrative futures studies on Finnish transport CO2 policy". *Acta Futura Fennica*, no 8. Finnish Society for Futures Studies. 2002.
- [21] Tapio, P., "Disaggregative policy Delphi: Using cluster analysis as a tool for systematic scenario formation". *Technological Forecasting and Social Change*, Turku. 2002.
- [22] Tapio, P., V. Varho, M. Vinnari, "Finding a balance for qualitative and quantitative information in Delphi processes", in *Future of the Consumer Society. Proceedings of the Conference "Future of the Consumer Society," 28-29 May 2009, Tampere, Finland*. Eds. M. Koskela and M. Vinnari. FFRC eBooks 7/2009. Finland Futures Research Centre, Turku School of Economics. 2009.
- [23] Turoff, M., "Delphi for foresight", Lecture given in the Millennium seminar at the University of Helsinki, Finland, on the 6th of November 2009. Retrieved 11/19/09 World Wide Web, <http://edelfoi.ning.com/video/delphi-for-foresight>
- [24] Turoff, M. and H. A. Lindsay, "Introduction" in *The Delphi method: techniques and applications*. Eds. H. Linstone and M. Turoff. Addison-Wesley, London. 1975.
- [25] Turoff, M. and R. H. Starr, "Computer based Delphi processes", in *Gazing Into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*. Eds. Adler and Ziglio. London, Kingsley Publishers. 1996.
- [26] Kvale, S. (ed.): "Issues of Validity in Qualitative Research", Lund: Student Litteratur, 1989.