Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review

Alan Davis[1], Oscar Dieste[2,*], Ann Hickey[1], Natalia Juristo[2], Ana M. Moreno[2]

[1] *University of Colorado at Colorado Springs*
College of Business
Colorado Springs, CO 80918 USA
Email:{adavis, ahickey}@uccs.edu

[2] *Universidad Politécnica de Madrid*
Facultad de Informática
28660 Boadilla del Monte, Madrid, Spain
Email:{odieste, natalia, ammoreno}@fi.upm.es

Abstract

This paper reports a systematic review of empirical studies concerning the effectiveness of elicitation techniques, and the subsequent aggregation of empirical evidence gathered from those studies. The most significant results of the aggregation process are as follows: (1) Interviews, preferentially structured, appear to be one of the most effective elicitation techniques; (2) Many techniques often cited in the literature, like card sorting, ranking or thinking aloud, tend to be less effective than interviews; (3) Analyst experience does not appear to be a relevant factor; and (4) The studies conducted have not found the use of intermediate representations during elicitation to have significant positive effects. It should be noted that, as a general rule, the studies from which these results were aggregated have not been replicated, and therefore the above claims cannot be said to be absolutely certain. However, they can be used by researchers as pieces of knowledge to be further investigated and by practitioners in development projects, always taking into account that they are preliminary findings.

Keywords: Systematic review, software engineering, elicitation techniques, empirical studies

1. Introduction

It is a reasonably well documented fact that software requirements definition has a big impact on final product quality [8,83,87]. Three primary categories of problems affecting the correct definition of software requirements are defined in the literature: acquisition, comprehension and volatility [18,88]. This paper focuses on the first of the above problems, namely, information or knowledge elicitation.

A variety of methods or techniques have been proposed to acquire information for the purposes of elicitation, such as: interviewing [58], protocol analysis [31], repertory grid [56], work groups [89], etc. Apart from these basic techniques, there are a host of variations upon these methods (such as [34,54,65]), a great many not very commonly used techniques (like, for example, [2,48]) and combinations of techniques (like [35,86]), etc. Similarly, there are a number of studies that have gathered and analysed elicitation techniques (like, for example, [9,19,38,44,46]).

In view of the diversity of techniques and the fair few summary papers there are, one might think that there is a sound scientific foundation for carrying out elicitation. This could not be further from the truth, however, as there is absolutely no agreement among experts on how best to elicit information or knowledge [50,91]. These differences of opinion are fuelled by the shortage of comparative studies analyzing the potential of one technique against the capabilities of others from either a theoretical or empirical viewpoint [69] [16,19,85]. In this context, a number of authors have

* This author has been partially sponsored by the Spanish MECD/Fulbright Grant FU2004-0589.

pointed out the need for further investigation of along these lines [6,16,17,46].

One step towards this goal is to identify what we can ascertain empirically from the studies already conducted on elicitation techniques. In actual fact, we are interested in studies that have analysed the effectiveness of elicitation techniques with a view to identifying under what circumstances it would be best to apply one particular elicitation technique or, alternatively, find out which is likely to be the best elicitation technique for a specific situation.

To achieve this goal, the article will be divided into five sections. Section 2 will describe the motivation of this study in more detail and will introduce the respective background. The research methodology used will be described in section 3. Section 4 will briefly list the key milestones in the application of the selected methodology, leading on to section 5 where the results of the review will be explained. Finally, section 6 will discuss the findings and the limitations of the approach taken.

2. Motivation and related work

As already mentioned, there are a lot of elicitation methods and techniques published in the literature. This naturally raises questions as to when one technique is preferable over another.

Experience has shown that for simple and well-defined problems, elicitation techniques are more or less equivalent [45,81]. Outside the toy-problem universe, however, a number of studies demonstrate that elicitation techniques are not interchangeable and that there are far-reaching differences between methods, chiefly as regards the type of knowledge or information each technique can be used to acquire [1,6,10,21,33,37,42,46,47,64,70,90]. This is what is termed "Differential Access Hypothesis". However, a wide variety of other issues, such as the quantity of information or elicitation efficiency, also serve to differentiate elicitation techniques (see, for example, [13,16,80]).

In this respect, some authors have proposed criteria about when to use a particular elicitation technique [40,55,58,61,76]. These criteria are, at the very least, open to discussion because of the weakness of the evidence on which they are grounded. There can be no doubt that the "Differential Access Hypothesis", like any other differential aspect, is hard to address theoretically, especially taking into account that the applicable theory (basically, premises laid down by cognitive psychology [4]) is not mature enough to be able to predict the effectiveness of elicitation techniques.

A reasonable approach to be used to undertake a study of the effectiveness of the different elicitation techniques would therefore be based on carrying out empirical studies that compare the application of different techniques in more than one setting. Note that this approach is by no means new. Knowledge engineering has been tackling the question of elicitation technique effectiveness [49] since as far back as 1987 [13,80]. It was aped some years later by the software engineering/information systems field [3,41].

The empirical studies most commonly referenced in the literature for researching the effectiveness of elicitation techniques are case studies [23], used for example in [6,28,39,86] and experiments, like [3,63,74,78,94]. We were surprised to find, however, that the results yielded by these studies have not generally been given due consideration in summary papers, like [9,19,38,46], to mention just the best known. An exception is [62], as its authors have participated in several empirical studies [13,15,16,20,78]. Irrespective of the authors' proven merit, this paper is nevertheless still too immature because many of its findings are based on arguments supported by general cognitive psychology and not on empirical evidence.

This is a far from satisfactory state of affairs. On the one hand, a good many recommendations on the effectiveness of elicitation techniques are debatable, as they are not based on empirical proofs. On the other, there are a number of empirical studies on elicitation that have yielded valuable results of which proper account has not been given in the literature.

All this goes to justify the objective of this research, namely, to aggregate those empirical results there are in the literature that can be used to help to determine the effectiveness of requirements elicitation techniques.

3. Research Method

No special-purpose method would, a priori, appear to be necessary to achieve the above-mentioned objective. As there is unlikely to be many empirical studies, all that we would need to do would be to locate the works and aggregate the results or evidence they contain. This would finalize this research. However, this is an oversimplification:

1) It is not easy to locate empirical articles of interest, because there are not very many and due to the variety of descriptors used in their indexing in bibliographical databases.
2) The types of empirical studies used (randomized experiments, pseudo-randomized experiments,

surveys, case studies, etc.) can differ greatly from one article to another.
3) The design, the factors and the response variables used in the studies are very wide ranging.
4) Irrespective of the methodology used, the execution of the empirical studies can pose validity threats to the results output.
5) Down to all the above, it is not easy to aggregate all the evidence gathered in such a way as to ensure that missing evidence or researcher biases, caused by their absolute freedom of action unconstrained by any protocol, do not threaten the synthesis.
6) It is worth mentioning that not having a specific method or procedure for undertaking this aggregation more or less assures the impossibility of replicating the results of this review.

Consequently, it is clear that the research method to be used should assure the repeatability of the review and the correctness of the aggregation carried out. The best method for this purpose is systematic review [52], which is specifically designed to integrate current best evidence from research [29]. This method demands that, before any review work at all is done, a review protocol should be defined specifying as accurately as possible all the important aspects of the review and aggregation. These are basically the aspects listed earlier: identification and selection of primary studies, data collection, quality assessment and, finally, aggregation strategy.

The definition of these aspects in advance assures the repeatability of the review, because all the important decisions (what empirical studies to consider, what data to collect, etc.) are taken beforehand. This constrains the researcher's freedom to improvise and, consequently, prevents biases from being introduced. It also assures the correctness of the aggregation, as the disclosure of the protocol means that its soundness and, ultimately, the certainty of the conclusions reached can be evaluated.

The following sections very briefly describe the basic aspects of the review protocol used in this research. For a much more detailed version of the protocol, see [26].

3.1. Research question

As mentioned in [52],[29] a systematic review should answer a clearly stated research question. An example of a clearly stated research question is, "Are more requirements elicited using interviews or thinking aloud techniques?" Another possible example would be "In which domains is laddering better than card sorting?" A clearly stated research question eases the acquisition of data from the selected empirical studies, the aggregation of evidence and, also importantly, reduces the time it takes to complete the systematic review.

The above examples are highly specific research questions. The specificity of the research question is likely to vary depending on the objective of the research. In this case, the research question would be, "What elicitation technique is most efficient in a particular setting?" In disciplines where there is ample and extensive empirical knowledge, answering a specific question can be of interest because it takes knowledge further. As there is generally little empirical knowledge in the field of software engineering, not only would it be difficult to choose a specific research question, but also its benefit would be debatable. On this ground, we opted to use an open research question.

3.2. Identifying and selecting primary studies

The empirical studies were identified from the SCOPUS, IEEEXPLORE and ACM DL databases. The search deadline was March 2005. The search string used was:

(elicitation OR "requirements gathering" OR "requirements acquisition")
AND
(capture OR empirical OR experiment OR study OR review OR evaluation)

Additionally, the bibliographical references of the studies originally identified by means of keyword search were examined with a view to locating more useful studies. Regarding this strategy, in a discipline where there are a lot of empirical studies (such as medicine, for example), the search for empirical studies using bibliographical references would introduce inadmissible biases in the review. The use of bibliographic references for the purposes of search in the field of software engineering does not introduce biases and, simply, increases the amount of evidence available for aggregation.

Each identified empirical study was evaluated by at least two work team members to decide on whether or not it should be selected. Any discrepancies were settled by consensus. Those studies that met the following conditions were selected:

- Studies should be genuine empirical studies and not just the validation of an approach or technique proposed by the authors.
- The elicitation techniques should be simple individual techniques, that is, should be neither group elicitation techniques nor computer-aided

methods. This constraint should obviously not be taken to mean that we do not consider these two technique types appropriate for elicitation purposes. Quite the contrary, they are omitted on strictly pragmatic grounds as a way of confining the number of studies for consideration to those that use elicitation techniques common in development projects. The excluded techniques could, however, be considered in future research efforts.

- The empirical study design should compare two different techniques, the same technique in two different situations, or gather information from expert opinions. As mentioned earlier, this constraint prevents the inclusion of a good many case studies that are difficult to generalize.

3.3 Acquiring data and assessing quality

Two forms, shown in [26] for reasons of space, were used to acquire the data. The first is a data collection form. This form is used to state the significant differences between the elicitation techniques, as well as other features of the empirical study like the number of experimental units, the factors and the response variables, the experimental problem, etc.

The objective of the second form is to assess the quality of the empirical studies. Quality was assessed in two different ways. On the one hand, each study was assigned a level in the evidence hierarchy, as indicated in [52]. On the other, each study was subjected to an evaluation of its internal quality. This evaluation was based on the criteria specified in [53].

For reasons of efficiency, the data from each study was gathered by one member of the work team. Because these data were not manipulated or interpreted in any way, no specific protocol was established to control this activity.

3.4 Aggregation strategy

Having reached this point, the best thing would have been to be able to summarize the aggregation strategy to be followed in just a few words. Unfortunately, things are not as simple as all that in the case of this review, for several reasons:

- There are likely to be hardly any replications of the selected studies. This rules out the use of meta-analysis techniques [73].
- The studies of interest mainly fit the category of quantitative comparative studies. The extent to which the important aspects of these studies are perfectly defined, without the need for too much interpretation or elaboration, appears to advise against interpretative strategies, such as narrative, subject or content analysis [27], as well meta-ethnographical techniques [71].
- The field of study, that is, information or knowledge elicitation, is a relatively well-defined area. The elicitation techniques, as well as the response variables and other factors that the empirical studies test, have no need of elaboration, and the data collected fall into relatively well-defined categories. This appears to advise against the use of primarily inductive methods, like grounded theory [84].
- There are likely to be few empirical studies. This is an enormous obstacle to synthesis by means of techniques like case surveys [92].
- Having called the above-mentioned methodologies into question, we could always turn to weak methods, such as meta-synthesis [72] or realistic synthesis [27]. However, the very weakness of these methods can lead to the results of aggregation not being as reliable as they should be.

Looking at the key factor then, namely, that there are no replications, the most reasonable alternative that we are left with is to use comparative analysis [75] or, at least, an adapted version of this technique (see [25] for further details). Comparative analysis is a well suited methodology because it assumes that there is a cause-effect relationship between factors and contextual aspects, and the results observed in the response variables. Therefore, there is no need for a specified number of observations to achieve statistical significance.

Almost any causal relationship could be concluded from such an underlying hypothesis, because one observation would be enough. To overcome this weakness, comparative analysis introduces two concepts to refine and assess the quality of the observed relationships: logical simplification and confidence. Logical simplification involves eliminating factors from a relationship when they are not logically necessary to explain the observed effect. Confidence is a similar concept to statistical significance. It simply reflects the fact that the more often the same effect is observed, the more confidence there is in the causal relationship being true. Obviously, when the number of observations is over a given threshold, comparative analysis could be abandoned and statistical techniques used to analyze the observations.

The departure from comparative analysis is a generalization prior to the aggregation of the elicitation techniques and response variables used in the empirical

studies. This procedure yields "generalized" evidence to be used for the purpose of aggregation.

Let us take an example to illustrate this point. Suppose that one response variable V1 is defined as "Customer needs", whereas another response variable V2 is defined as "Number of requirements". If one piece of evidence A is gathered using V1 and another piece of evidence B is gathered using V2, then the two pieces of evidence A and B would obviously not be able to be aggregated. Now, if

- V1 and V2 refer to aspects that are so alike (user needs and number of requirements
- And there appears to be a clearly positive correlation between V1 and V2

it is at least worth considering whether the two variables are not measuring the same underlying fact, albeit using different measurement scales and procedures. If this is true, V2 could also be used in all the studies where V1 is used, and, ideally, all the statistically significant effects should hold using both V1 and V2. A similar line of reasoning could be applied with respect to elicitation techniques.

Although generalization may appear to be an arbitrary procedure within the stated strategy of aggregation, it should be understood to be a mechanism very like common sense that does no more than formalize the reasoning typically used in works of synthesis. It should be noted, however, that the aggregation strategy used in this paper is much less restrictive than most classical techniques, like meta-analysis, and is, consequently, much less reliable. Even so, the stated strategy allows what little evidence there is to be aggregated, thereby yielding extremely valuable findings that are backed empirically.

4. Application of the review protocol

The application of the review protocol yielded the results specified below (for further details, see [24]).

- The bibliographic database search identified 74 potentially relevant publications.
- The bibliographic review identified another 484 potentially relevant publications, yielding a total of 564 publications that needed to be thoroughly reviewed.
- The thorough review of the above-mentioned publications, according to the criteria mentioned in section 3.2, led to the selection of 26 relevant empirical studies. These studies were: [1,3,5,10,11,13,16,20,22,32,34,36,41,43,48,50,57, 63,67,68,74,77,78,80,82,94].
- Some of the above-mentioned empirical studies had been published in more than one medium. The identified cases were: [2,12,14,59,60,66,93]. These publications were not taken into account for obvious reasons.
- At the time of writing, we had not been able to get the full text of a number of publications for thorough review. These include a set of 28 potentially relevant publications, that is, publications that are likely to describe an empirical study. We hope to get the full text of these remaining publications and add the evidence they provide to our aggregation in the near future.
- We were able to gather around two hundred pieces of empirical evidence from the selected studies (among many other data shown in [24]). A piece of evidence should be construed as a relationship between the elicitation techniques that were tested in the empirical study. One possible example of a piece of evidence would be *"Novices using a structured interview elicit more knowledge than experienced analysts using a unstructured interview"* [3]. Evidence is the key material used by the comparative analysis for aggregating the empirical evidence.

5. Results

For reasons of space, not all the results can be shown. Therefore, Table 1 shows a portion of the results yielded during aggregation, specifically, those results that are directly supported by observation in more than one empirical study. These are the results that are most relevant for routine requirements engineering practice, because the aspects they address (interviews, prototypes and the effect of experience) are often top-ranking concerns in development projects. Which studies provide evidence for and against, as well as any relevant aggregation-related issues, are listed for each result.

Additionally, we have made other findings, which albeit not as relevant as the results shown in Table 1, are just as interesting, referring, for example, to the time efficiency or the completeness of interviewing as compared with other elicitation techniques, the efficiency of repertory grid or laddering, or the subjective effects caused by techniques on clients and analysts. These results are shown in [24].

The results shown in Table 1 can be divided into two major groups: expected and unexpected.

Results #1, #2 and #3, all referring to interviews, fall into the first group. Although we have no knowledge of any paper stating the primacy of interviewing over any other elicitation technique, it is true that interviews are the elicitation technique most

Table 1. Most relevant aggregation results

#	Aggregation result	(1)	(2)	Comments
1	Structured interviews gather more information than unstructured interviews	[3,11,63,67]	---	---
2	Unstructured interviews gather more information than sorting and ranking techniques	[10,16,20,80]	[5]	---
3	Unstructured interviews appear to gather more information than thinking aloud techniques	[13,16,20]	[22]	• The evidence given in [16] is confusing, but suggests that interviews are better than thinking aloud techniques. • The quality of the study [22] can be qualified as being on the low side
4	Elicitation techniques do not appear to provide specific types of information, that is, there is not enough evidence to support differential information access depending on what elicitation technique is used	[10,11,13,22,78]	[16]	• The quality of the study [22] can be qualified as being on the low side.
5	Analyst experience does not appear to be a relevant factor during information acquisition, at least using interviews as an elicitation technique.	[3,63,74]	[34]	---
6	The use of visual aids or prototypes focuses the discussion on the displayed artifact and does not generally help to discover new requirements.	[41,68]	---	• Not a lot of evidence is available as yet, although other studies (not covered by this review), like [30], support this finding.

(1) = Studies that provide evidence in favor of the aggregation
(2) = Studies that provide evidence against the aggregation

commonly used in practice. Therefore, the results serve to back the belief in its effectiveness.

The second group, unexpected results, encompasses #4, #5 and #6:

- Result #4 claims that there is not enough evidence to support the "Differential access hypothesis". Translated into everyday language, this result means that all the elicitation techniques have more or less the same power to elicit information. Although further research is needed before this result can be classed as confirmed, if it were true, its effect would be enormously positive, as, contrary to what appears to be a widely held opinion today, this would mean that just one technique or a small subset of techniques would be enough to undertake elicitation, irrespective of the problem to be solved.
- Result #5 claims that analyst experience is not a relevant factor during information acquisition using interviews or, alternatively, a novice analyst is capable of eliciting information just as well as an experienced analyst. The empirical studies show, in fact, that the careful preparation of interviews has a much more marked effect than analyst experience. In other words, a novice analyst who prepares the interview well beforehand is even capable of eliciting more information than an experienced analyst. This result is really surprising, because it contradicts decade-old requirements engineering beliefs and, although it warrants further corroboration, should lead both requirements engineering practitioners and educators to think matters over.
- Finally, result #6 claims that prototypes do not help to discover new requirements. This is another controversial result and has need of some explanation. What actually appears to be happening is that prototypes tend to focus the discussion on the artifact (mock-up, prototype). This then prevents information that cannot be reflected (visualized, experienced) in the displayed artifact from being gathered. User experience appears to be an important factor here, as novice users tend to focus the discussion more on the artifact than experienced users [30]. Obviously, the evidence there is to date is very weak, but everything points in the specified direction. As in the case of result #5, the effect of prototype-

induced focus should be taken into account both in practice and in the education of future engineers.

6. Discussion and conclusions

The findings of this review can be summarized in four key points:

- Interviews, preferentially structured, appear to be one of the most effective elicitation techniques in a wide range of domains and situations.
- Several techniques often cited in the literature, like card sorting, ranking or thinking aloud, tend to be less effective than interviews.
- Analyst experience does not appear to be a relevant factor, at least using interviews as an elicitation technique.
- The studies conducted have not found the use of prototypes during software requirements elicitation to have significant positive effects.

It should be noted that the selected studies have not been replicated, and therefore the above claims cannot be said to be certain. However, this was something we had counted on from the very moment this research was planned. In other areas of software engineering knowledge, like software testing, the shortage of empirical studies and the almost complete absence of replications is, as in the field of elicitation, the dominant tendency [51].

Furthermore, we should not overlook the fact that there are a range of constraints, caused by the primary studies, that influence the aggregation, as indicated in [24]. These effects should be considered before using these results both in theory and in practice. The most relevant are:

- Contextual factors (problem type, experimental environment, subjects' background, etc.) are very wide ranging. This need not necessarily be a constraint, as it helps to increase the external validity of the empirical studies [7]. However, the absence of replications stops the influence of these factors on the observed effects being properly taken into account and, therefore, leads to any attempts at extrapolation having only limited validity. For example, one might want to know whether results #3 and #6 would still hold if the experimental problem involved determining the user interface instead of acquiring more general requirements. Unfortunately, this fact cannot be clarified without future empirical studies.
- Sample sizes used in most experiments are quite small, considering the variety of contextual factors present in the empirical studies. In most cases, there are around twenty experimental units, albeit with exceptions, like [5,10,94], where the experimental units were of the order of a hundred., Save for cases like [41,50], where the sample size is 7 and 8, respectively, the number of experimental units is, however, enough to yield reasonably reliable results.
- There are big differences between the elicitation techniques used in the experimental studies examined. Even when several studies, like [3],[11,63] nominally use the same technique (interviews in these cases), there are sizeable differences between the studies that can affect the results if the techniques in question are generalized as proposed in the aggregation strategy. One possible way of overcoming this shortcoming would be to use the idea proposed in [79]. This paper suggests decomposing the elicitation techniques into facets. Instead of considering a technique as a whole, it would be possible to analyze the method as being composed of several parts that interact to acquire information. Such decomposition would introduce a new set of factors that could be considered using the selected aggregation strategy.
- The response variables used in the different empirical studies differ largely as regards both their definition (what is measured) and the measurement procedure (how it is measured). Consequently, and as in the case of elicitation techniques, such differences can affect the results if these response variables are generalized.

Even so, the coincidences between the results of many empirical studies, despite there being profound, both methodological and contextual differences among them, are surprising. At least some confidence in the reliability of the results can be derived from this. We trust, therefore, that the results yielded by this research, as well as the findings explained in [24], will serve to support the routine activities of practitioners and above all as an incentive to researchers to do further investigation in the field of knowledge elicitation.

7. References

[1] Adelman, L., Measurement issues in knowledge engineering *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 483-488, 1989.

[2] Adelman, L., Sticha, P. J., and Donnel, M. L., An experimental investigation of the relative effectiveness of two techniques for structuring multiattributed hierarchies *Organizational Behavior and Human Decision Processes*, vol. 37, pp. 188-196, 1986.

[3] Agarwal, R. and Tanniru, M. R., Knowledge Acquisition Using Structured Interviewing: An Empirical

Investigation *Journal of Management Information Systems*, vol. 7, pp. 123-141, Summer, 1990.
[4] J.R. Anderson. *Cognitive Psychology and Its Implications*, Worth Publishers, 2004.
[5] Bech-Larsen, T. and Nielsen, N. A., A comparison of five elicitation techniques for elicitation of attributes of low involvement products *Journal of Economic Psychology*, vol. 20, pp. 315-341, 1999.
[6] Blandford, A. and Rugg, G., A case study on integrating contextual information with analytical usability evaluation *International Journal of Human-Computer Studies*, vol. 57, pp. 75-99, 2002.
[7] Bloomfield, R. and Littlewood, B., "Multi-legged arguments:the impact of diversity upon confidence in dependability arguments," *Proceedings of the International Conference on Dependable Systems and Networks,* pp. 25-34, 2002.
[8] Boehm, B. W., McClean, R. K., and Urfrig, D. B., Some experience with automated aids to the design of large-scale reliable software *IEEE Transactions on Software Engineering*, vol. 1, pp. 125-133, Mar, 1975.
[9] Boose, J. H. , A survey of knowledge acquisition techniques and tools *Knowledge Acquisition*, vol. 1, pp. 3-37, 1989.
[10] Breivik, E. and Supphellen, M., Elicitation of product attributes in an evaluation context: A comparison of three elicitation techniques *Journal of Economic Psychology*, vol. 24, pp. 77-98, 2003.
[11] Browne, G. J. and Rogich, M. B., An Empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques *Journal of Management Information Systems* , vol. 17, pp. 223-249, Spring, 2001.
[12] Burton, A. M. and Shadbolt, N., "A formal evaluation of knowledge elicitation techniques for expert systems: domain 1," *Proceedings of a SERC Workshop on knowledge Acquisition for Engineering Applications 20-28,* pp. 20-28, 1987.
[13] Burton, A. M., Shadbolt, N. R., Hedgecock, A. P., and Rugg, G. A formal evaluation of knowledge elicittion techniques for expert systems: Domain 1. In: *Research and development in expert systems IV: proceedings of Expert Systems '87, the seventh annual Technical Conference of the British Computer Society Specialist Group on Expert Systems*, ed. Moralee, D. S. Cambridge, UK: Cambridge University Press, 1987.
[14] Burton, A. M., Shadbolt, N. R., Hedgecock, A. P., and Rugg, G. A., "A formal evaluation of knowledge elicitation techniques for expert systems: Domain 1," *Proceedings of the 1st European Workshop on Knowledge Acquisition for Knowledge-Based Systems,* Reading University, UK, 1987.
[15] Burton, A. M., Shadbolt, N. R., Rugg, G., and Hedgecock, A. P., "Knowledge elicitation techniques in classification domains," *Proceedings of the 8th Conference in Artificial Intelligence ECAI-88,* 1988.
[16] Burton, A. M., Shadbolt, N. R., Rugg, G., and Hedgecock, A. P., The efficacy of knowledge acquisition techniques: A comparison across domains and levels of expertise *Knowledge Acquisition*, vol. 2, pp. 167-178, 1990.
[17] Byrd, T. A. , Cossick, K. L., and Zmud, R. W., A synthesis of research on requirements analysis and knowledge acquisition techniques *MIS Quarterly*, vol. 16, pp. 117-138, 1992.
[18] Anonymous, CMU/SEI-92-TR-012, Sep 1992.
[19] Cooke, N. J., Varieties of knowledge elicitation techniques *International Journal of Human-Computer Studies*, vol. 41, pp. 801-849, 1994.
[20] Corbridge, B., Rugg, G., Major, N. P., Shadbolt, N. R. , and Burton, A. M., Laddering - technique and tool use in knowledge acquisition *Knowledge Acquisition*, vol. 6, pp. 315-341, 1994.
[21] Coughlan, J. and Macredie, R. D., Effective Communication in Requirements Elicitation: A Comparison of Methodologies *Requirements Engineering Journal*, vol. 7, pp. 47-60, 2002.
[22] Crandall, B., A comparative study of think-aloud and critical decision knowledge elicitation methods *SIGART Newsletter*, vol. 108 , pp. 144-146, 1989.
[23] J.W. Creswell. *Research design: qualitative, quantitative, and mixed method approaches*, Thousand Oaks, CA, USA: SAGE Publications, 1994.
[24] Davis. A.M. , Dieste, O., Hickey, A., Juristo, N., López, M., and Moreno, A. M. Evidencias extraídas de los estudios empíricos. http://is.ls.fi.upm.es/udis/i_d/educcion.html TR3. 2006.
[25] Davis. A.M. , Dieste, O., Hickey, A., Juristo, N., López, M., and Moreno, A. M. Protocolo de agregación de estudios empíricos. http://is.ls.fi.upm.es/udis/i_d/educcion.html TR4. 2006.
[26] Davis. A.M. , Dieste, O., Hickey, A., Juristo, N., and Moreno, A. M. Protocolo de revisión de estudios empíricos. http://is.ls.fi.upm.es/udis/i_d/educcion.html TR2. 2006.
[27] Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., and Sutton, A., Synthesising qualitative and quantitative evidence: a review of possible methods *Journal of Health Services Research and Policy*, vol. 10, pp. 45-53B(9), 2005.
[28] Dong-Gil Ko , "Information requirements analysis and multiple knowledge elicitation techniques: experience with the pricing scenario system," *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences,* pp. 10 pp., 1999.
[29] Dyba, T., Kitchenham, B. A., and Jorgensen, M., Evidence-based software engineering for practitioners *IEEE Software*, vol. 22, no. 1, pp. 58-65, 2005. ./methods/6.pdf.
[30] Engelbrektsson, P., Effects of product experience and product representations in focus group interviews *Journal of Engineering Design*, vol. 13, pp. 215-221, 2002.
[31] K.A. Ericson and H.A. Simon. *Protocol Analysis: Verbal Reports as Data*, Cambridge, MASS: Bradford Books/ MIT Press, 1984.
[32] Eva, M., Requirements acquisition for rapid applications development *Information & Management*, vol. 39, pp. 101-107, Dec, 2001.
[33] Fischhoff, B., Eliciting knowledge for analytical representation *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 448-461, 1989.
[34] Fowlkes, J. E., Salas, E., and Baker, D. P., The utility of event-based knowledge elicitation *Human Factors*, vol. 42, pp. 24-35, 2000.
[35] Frankel, S. , NGT+MDS: An adaptation of the nominal group technique for ill-structured problems *Journal of Applied Behavioral Science*, vol. 23, pp. 543-551, 1987.

[36] Freeman, L. A., "The effects of concept maps on requirements elicitation and system models during information systems development," *Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping,* Pamplona, Spain, 2004.
[37] Gammack, J. G. and Anderson, A., Constructive interaction in knowledge engineering *Expert Systems*, vol. 7, pp. 19-26, 1990.
[38] Gammack, J. G. and Young, R. M. Psychological techniques for eliciting expert knowledge. In: *Knowledge Acquisition for Expert Systems*, ed. Kidd, A. New York: Plenum Press, 1985.pp. 137-163.
[39] Go, K., Takamoto, Y., and Carroll, J. M., "Designing a mobile phone of the future: requirements elicitation using photo essays and scenarios," *18th International Conference on Advanced Information Networking and Applications, 2004. AINA 2004,* pp. 475-480, 2004.
[40] Goguen, J. A. and Linde, C., " Techniques for requirements elicitation," *Proceedings of IEEE International Symposium on Requirements Engineering,* pp. 152-164, 1993.
[41] Goodrich, V. and Olfman, L., "An experimental evaluation of task and methodology variables for requirements definition phase success," *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences,* pp. 201-209, 1990.
[42] Gordon, S. E., Schmierer, K. A., and Gill, R. T., Conceptual graph analysis: Knowledge acquisition for instructional system design *Human Factors*, vol. 35, pp. 459-481, 1993.
[43] Griffin, A. and Hauser, J. R., The voice of the customer *Marketing Science*, vol. 12, pp. 1-27, 1993.
[44] A. Hart. *Knowledge Acquisition for Expert Systems*, Tiptree, Essex, UK: Anchor Brendon Ltd., 1986.
[45] Hoffman, R. R., The problem of extracting the knowledge from experts from the perspective of experimental psychology *The AI Magazine*, vol. 8, pp. 53-67, Summer, 1987.
[46] Hoffman, R. R., Shadbolt, N. R., Burton, A. M., and Klein, G., Eliciting knowledge from experts: A methodological analysis *Organizational Behavior and Human Decision Processes*, vol. 62, pp. 129-158, 1995.
[47] Holsapple, C. W. and Raj, V. S., An exploratory study of two KA methods *Expert Systems*, vol. 11, pp. 77-87, 1994.
[48] Hudlicka, E., "Requirements elicitation with indirect knowledge elicitation: Comparison of Three Methods," *Proceedings of the Second International Conference on Requirements Engineering,* pp. 4-11, 1996.
[49] Johnson, L. and Johnson, N. E. Knowledge elicitation involving teachback interviewing. In: *Knowledge Acquisition for Expert Systems: A Practical Handbook*, ed. Kidd, A. L. New York, NY, USA: Plenum, 1987.pp. 91-108.
[50] Jones, S. R. and Miles, J. C., The Use of a Prototype System for Evaluating Knowledge Elicitation Techniques *Expert Systems*, vol. 15, pp. 83-97, 1998.
[51] Juristo, N., Moreno, A. M., and Vegas, S., Reviewing 25 Years of Testing Technique Experiments *Empirical Software Engineering Journal*, vol. 9, pp. 7-44, 2004.
[52] Kitchenham, B. A., Keele University, TR/SE-0401, 2004.
[53] Kitchenham, B. A., Pfleeger, S. L., Pickard, L. M., Jones, P. W., Hoaglin, D. C., El Emam, K., and Rosenberg, J., Preliminary guidelines for empirical research in software engineering *IEEE Transactions on Software Engineering*, vol. 28, pp. 721-734, 2002.
[54] Korhonen, L. J. Nominal group technique. In: *Adult Learning Methods*, ed. Galbraith, N. W. Krieger Publishing Company, 1990.pp. 247-259.
[55] G. Kotonya and I. Sommerville. *Requirements Engineering*, Wiley, 1998.
[56] LaFrance, M., "The knowledge acquisition grid: A method for training knowledge engineers," *Proceedings of the Knowledge Acquisition for Knowledge-Based Systems Workshop,* Banff, AB, Canada, 1986.
[57] Larsen, T. J. and Naumann, J. D., An experimental comparison of abstract and concrete representations in systems analysis *Information & Management*, vol. 22, pp. 29-40, 1992.
[58] S. Lauesen. *Software Requirements: Styles and Techniques*, Addison-Wesley, 2002.
[59] Lines, R., Breivik, E., and Supphellen, M., "Elicitation of attributes: A comparison of preference model structures derived from two elicitation techniques," *Proceedings of the 24th European Marketing Academy Conference,* pp. 641-656, 1995.
[60] Lines, R., Breivik, E., and Supphellen, M. Method Effects in Attribute Elicitation: A Multidimensional Comparison of Four Techniques. 95.
[61] L. Macaulay. *Requirements Engineering*, Springer, 1996.
[62] Maiden, N. A. M. and Rugg, G., ACRE: selecting methods for requirements acquisition *Software Engineering Journal*, vol. 11, pp. 183-192, 1996.
[63] Marakas, G. M. and Elam, J. J., Semantic structuring in analyst and representation of facts in requirements analysis *Information Systems Research*, vol. 9, pp. 37-63, 1998.
[64] Martinez-Bejar, R., Benjamins, R., Martin, F., and Castillo, V., "Deriving formal parameters for comparing knowledge elicitation techniques based on mathematical functions," *Proceedings of the Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop ,* 1996.
[65] Massey, A. P. and Wallace, W. A., Focus groups as a knowledge elicitation technique: an exploratory study *IEEE Transactions on Knowledge and Data Engineering*, vol. 3, pp. 193-200, 1991.
[66] Moody, J. W., Blanton, J. E., and Cheney, P. H., A theoretically grounded approach to assist memory recall during information requirements determination *Journal of Management Information Systems*, vol. 15, pp. 79-89, 1998.
[67] Moody, J. W., Will, R. P., and Blanton, J. E., Enhancing knowledge elicitation using the cognitive interview *Expert Systems with Applications*, vol. 10, pp. 127-133, 1996.
[68] Moore, J. M. and Shipman, F. M. I., "A comparison of questionnaire-based and GUI-based requirements," *Proceedings of the Fifteenth IEEE International Conference on Automated Software Engineering ,* pp. 35-43, 2000.
[69] Munro, M. C. and Davis, G. B., Determining management information news: A comparison of methods

MIS Quarterly, vol. 1, pp. 55-67, 1977.
[70] Musen, M. A., Fagan, L. M., Combs, D. M., and Shortliffe, E. H., Use of a domain model to drive an interactive knowledge-editing tool *International Journal of Man-Machine Studies*, vol. 26, pp. 105-121, 1987.
[71] G.W. Noblit and R.D. Hare. *Meta-Ethnography: Synthesising Qualitative Studies*, Newbury Park, CA: Sage, 1988.
[72] B.L. Paterson, S.E. Thorne, C. Canam, and C. Jillings. *Meta-Study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis* , Thousand Oaks, CA, USA: SAGE Publications, 2001.
[73] Pickard, L. M., Kitchenham, B. A., and Jones, P. W., Combining empirical results in software engineering *Information and Software Technology*, vol. 40, pp. 811-821, Dec 1, 1998.
[74] Pitts, M. G. and Browne, G. J., Stopping Behavior of Systems Analysts During Information Requirements Elicitation *Journal of Management Information Systems*, vol. 21, pp. 203-226, 2004.
[75] C.C. Ragin. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, Berkeley, CA: University of California Press, 1987.
[76] Robertson, S., Requirements trawling: Techniques for discovering requirements *International Journal of Human-Computer Studies*, vol. 55, pp. 405-421, 2001.
[77] Rowe, A. L. , Cooke, N. J., Hall, E. P., and Halgren, T. L., Toward an on-line knowledge assesment methodology: Building on the relationship between knowing and doing *Journal of Experimental Psychology: Applied*, vol. 2, pp. 31-47, 1996.
[78] Rugg, G., Corbridge, C., Major, N. P., Burton, A. M., and Shadbolt, N., A comparison of sorting techniques in knowledge acquisition *Journal of Knowledge Acquisition*, vol. 4, pp. 279-291, 1992.
[79] Rugg, G., McGeorge, P., and Maiden, N., Method fragments *Expert Systems*, vol. 17, pp. 248-257, 2000.
[80] Schweickert, R., Burton, A. M., Taylor, N. K., Corlett, E. N., Shadbolt, N. R., and Hedgecock, A. P., Comparing knowledge elicitation techniques: A case study *Artificial Intelligence Review*, vol. 1, pp. 245-253, 1987.
[81] Shadbolt, N. R. and Burton, A. M., The empirical study of knowledge elicitation techniques *SIGART Newsletter*, vol. 108, pp. 15-18, 1989.
[82] Silver, J. A. and Thompson, J. C., Understanding customer needs-a systematic approach to the "voice of the customer" 1991. Thesis (M.S.)--Massachusetts Institute of Technology, Sloan School of Management.
[83] Standish Group. The CHAOS Report. http://www.standishgroup.com/sample_research/PDFpages/chaos1994.pdf . 2005.
[84] A. Strauss and J. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Thousand Oaks, CA : Sage Pubs, 1998.
[85] Sutcliffe, A., "Requirements rationales: integrating approaches to requirements analysis," *Proceedings of Designing Interactive Systems DIS'95,* pp. 33-42, 1995.
[86] Sutcliffe, A., "A technique combination approach to requirements engineering," *Proceedings of the Third IEEE International Symposium on Requirements Engineering,* pp. 65-74, 1997.
[87] Tavolato, P. and Vincena, K. A Prototyping Methodology and Its Tool. In: *Approaches to Prototyping*, ed. Budde, R. Berlin: Springer Verlag, 1984.
[88] Valusek, J. R. and Fryback, D. G. Information requirements determination: Obstacles within, among and between participants. In: *Information Analysis: Selected Readings*, ed. Gallers, R. Reading, MA, USA: Addison Wesley, 1987.pp. 139-151.
[89] J. Wood and D. Silver. *Joint Application Development*, New York, NY, USA: Wiley, 1995.
[90] Wood, L. E. , Davis, T. C., Clay, S. L., Ford, J. M., and Lammersen, S., Evaluation of interviewing methods and mediating representations for knowledge acquisition *International Journal of Expert Systems*, vol. 8, pp. 1-23, 1995.
[91] Wooten, T. C. and Rowley, T. H., Using anthropological interview strategies to enhance knowledge acquisition *Expert Systems with Applications*, vol. 9, pp. 469-482, 1995.
[92] Yin, R. K. and Heald, K. A., Using the Case Survey Method to Analyze Policy Studies *Administrative Science Quarterly*, vol. 20, pp. 371-381, 1975.
[93] Zmud, R. W. , Anthony, W. P., and Stair, R. M. The use of mental imagery as a requirements analysis technique. In: *Challenges and Strategies for Research in Systems Development*, eds. Cotterman, W. W. and Senn, J. A. New York, NY, USA: john Wiley & Sons, 1992.pp. 337-350.
[94] Zmud, R. W. , Anthony, W. P., and Stair, R. M., The use of mental imagery to facilitate information identification in requirements analysis *Journal of Management Information Systems*, vol. 9, pp. 175-191, 1993.