

R. J. Wieringa · J. M. G. Heerkens

The methodological soundness of requirements engineering papers: a conceptual framework and two case studies

Received: 26 September 2005 / Accepted: 11 May 2006 / Published online: 20 June 2006
© Springer-Verlag London Limited 2006

Abstract This paper was triggered by concerns about the methodological soundness of many RE papers. We present a conceptual framework that distinguishes design papers from research papers, and show that in this framework, what is called a research paper in RE is often a design paper. We then present and motivate two lists of evaluation criteria, one for research papers and one for design papers. We apply both of these lists to two samples drawn from the set of all submissions to the RE'03 conference. Analysis of these two samples shows that most submissions of the RE'03 conference are design papers, not research papers, and that most design papers present a solution to a problem but neither validate this solution nor investigate the problems that can be solved by this solution. We conclude with a discussion of the soundness of our results and of the possible impact on RE research and practice.

1 Introduction

This paper was triggered by concerns about the methodological soundness of many requirements engineering (RE) papers. As we argue in this paper, many of these papers describe techniques but do not report on any research. The techniques reported on are intended for use in RE practice: for example, how to improve the process of negotiating requirements, or how to build use case models, how to customize information systems

based on user requirements, etc. Our concern is not that techniques such as that are described in RE papers. The RE conference is a platform for such papers. Our concern is that there are very few *other* papers in RE conferences, such papers that investigate the properties of these techniques, or that investigate the problems to be solved by these techniques. Investigation of properties of techniques, or of problems to be solved by techniques are the examples of research papers.

We think that the absence of such research prevents the transfer of the results of requirements engineering research to practice. Companies will hesitate to adopt techniques of which the properties are not investigated thoroughly or for which it has not been investigated which problems they solve, and under which conditions. Secondly, without a proper research method, there cannot be a growth of knowledge that builds upon previous results produced by others. This creates the risk that new techniques in fact do not improve already existing techniques. Hence, if our concern about the methodological soundness of RE papers is valid, then it is relevant to do something about it. However, we must first investigate whether our concern is valid, i.e., whether there really is a problem with the methodological soundness of RE papers.

The research problem to be investigated in this paper is, then, what is the methodological structure of RE papers, and to what extent do they satisfy the criteria for sound methodological structure? This is an evaluation question. We investigate the actual structure and compare this with a norm (which we present and motivate in this paper too). One way to answer this question is to survey a representative sample of RE papers, observe the methodological structure of paper in this sample, and draw conclusions from this about the set of all RE papers. Tichy [22], Zelkowitz and Wallace [26], and Glass et al. [8] have done such surveys for software engineering, and their observations show that validation is lacking in a large proportion of software engineering papers. Here, our concern is with requirements engineering papers, not software engineering papers. And we

R. J. Wieringa (✉)
Faculty of Electrical Engineering, Mathematics,
and Computer Science, University of Twente,
P.O. Box 217, Enschede 7500AE, The Netherlands
E-mail: roelw@cs.utwente.nl
<http://www.cs.utwente.nl/~roelw>

J. M. G. Heerkens
Faculty of Business, Public Administration,
and Technology, University of Twente,
Enschede, The Netherlands
E-mail: j.m.g.heerkens@utwente.nl

will choose a case study approach, by analyzing two samples drawn from one population, namely the set of all submissions to the 11th IEEE requirements Engineering Conference (RE'03). The advantage of using a case study approach is that one can do a more in-depth analysis of the structure of papers in the sample than is possible with surveys. A disadvantage is that generalization of the conclusions beyond the case study is tenuous. The outcome of our research can at most be a hypothesis that should be validated across the entire population of RE papers before they can be accepted as true about that entire population.

However, another outcome is a normative framework for evaluating RE papers. We claim that this is applicable to all RE papers, past and future, and our case study shows at least that it is important to publish this normative framework in the RE community.

In Sect. 2, we set out our conceptual framework, in which we distinguish problems in which we want to change something in the world from problems in which we want to change our knowledge about the world. We sketch the engineering and the research cycles as the rational structure of world changes and knowledge changes, respectively. We derive from this the criteria for evaluation of papers that report about performing such changes, namely design papers and research papers. Design papers report about the result of applying the engineering cycle, and research papers report about the result of applying the research cycle. We summarize the criteria in Appendices 1 and 2. In Sect. 3 we then restate our research problem in terms of our conceptual framework. Sections 4 and 5 describe our research design and discuss its validity. We present our measurements in Sect. 6 and analyze the results in Sect. 7. Section 8 discusses the wider implications of our results.

2 Conceptual framework

2.1 Knowledge problems and world problems

We base our conceptual framework for analyzing RE papers on frameworks for systems engineering, product engineering, and software engineering presented elsewhere in the literature [1, 4, 15, 18, 23]. The basic distinction in our framework is between studying the world and changing the world. This allows us to distinguish knowledge problems from world problems.

- *Knowledge problems* consist of a lack of knowledge about the world. To solve a knowledge problem, we need to change the state of our knowledge, and when we do that, we try not to change the world. For example, when we study the behavior of people in a business, we try not to intervene in what we study; when monitoring the performance of a software system we try not to influence the performance of the software we are monitoring. There is only one

criterion to evaluate an answer to a knowledge problem, namely by evaluating the validity of the answer. We discuss validity later. All *research problems* are knowledge problems. Calling a knowledge problem a research problem expresses our intention to solve the knowledge problem by following a sound research method.

- *World problems* consist of a difference between the way the world is and the way we think it should be. We solve a world problem by trying to change the state of the world: we change an organization, we build a device, we implement a program, etc. Each world problem contains its own criteria for evaluating proposed solutions. For example, part of the problem to change an organization is the identification of the criteria by which to evaluate a change proposal; part of the problem to build a device is the identification of the criteria to evaluate the device; etc. All *engineering problems* are world problems. Calling a world problem an engineering problem expresses our intention to solve the problem by following an engineering method.

We discuss research methods and engineering methods in the next section. Here, we discuss the preceding definitions. First, common to both of them is the concept of a *problem*, defined as a difference between what is and what we would like to have [7, page 49]. Knowledge problems differ from world problems in the location of the problem: in our minds or in the world outside our minds. Even when the researcher may share a lot of knowledge with the subjects investigated, such as in research problems about the social world, the research problem still exists in the mind of the researcher, not in the world studied.

Second, the definitions of knowledge problem and world problem may create the false impression that it is always possible to gather knowledge without changing the world or to change the world without gathering knowledge. In fact, it is extremely difficult to gather knowledge without changing the world, and some research methods, like action research and laboratory experiments, gather knowledge by manipulating the world, i.e., by changing it. It must then be shown by the researchers that these changes in the world do not invalidate the knowledge claims generated by the research. At the other end of the spectrum, it is also extremely difficult to change the world without learning something from it. Solving a world problem usually generates knowledge. The distinction between world problems and knowledge problems is a distinction between goals: when our goal is to solve a knowledge problem, we do something to get knowledge and, if necessary, manage the changes that create in the world in such a way that the acquired knowledge is valid. And when our goal is to solve a world problem, we change the world and are usually happy with any knowledge we gain from this.

Third, there are some knowledge problems, such as the problem how our galaxy evolved, that are purely curiosity-driven. The resulting knowledge is not expected to change something in the world, although unexpectedly the research may yield a revolutionary new observation technology that has an impact on the world outside the research project. But producing revolutionary new observation technology that changes something in the world outside the research project is not the goal of this research.

In many cases, there *is* a world problem in the background of a knowledge problem. This is an example of the fact that there is a mutually recursive relation between knowledge problems and world problems. For example, suppose we want to improve requirements negotiation practice. This is a world problem. We want to change requirements negotiation practice. When working on this problem, we may encounter the knowledge problem to find out what the current structure of requirements negotiations is. This is a knowledge problem because we want to know something about requirements negotiations that have taken place in the world. If this knowledge is not readily available somewhere, we have to do research to solve it. Doing research, in turn, is a world problem, because research is an activity in the world. We will for example write, test, and send out questionnaires, and perform interviews, all of which are changes of the world. To solve the problem of doing research, we may have to gather knowledge about possible research methods applicable in this particular case—a knowledge problem. The mutual recursion of world problems and knowledge problems ends when all the knowledge needed is available, or when the desired state of the world already exists.

Note that in this mutual recursion, knowledge problems exist in the context of world problems, and world problems exist in the context of knowledge problems. Figure 1 represents the mutual recursion in our example. All engineering researches take place as part of such a mutual recursion. At the top, there is some intended change of the world, such as the desire to solve problems that people have to explore the possible uses of a new technology, to reach a business goal, or in general to realize some vision of the future. At the bottom, there are knowledge questions that are not problematic because we already know the answer, or desired world states that are not problematic because they already exist.

Usually, the mutual recursion is a lot more complex than the one shown in Fig. 1. For example, we may also have the knowledge problem of learning about current negotiation theory. And our research may lead to the realization that we must restrict our initial world problem to a particular domain such as for example requirements negotiation for ERP systems, etc.

The reason for distinguishing world problems from knowledge problems is that the rational methods to solve them are different. We discuss these in the following two sections.

2.2 The engineering cycle

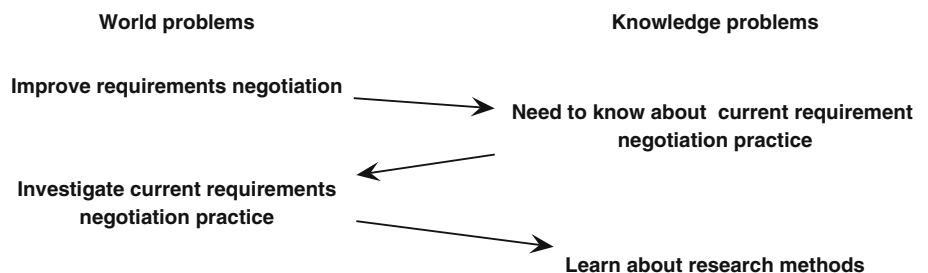
The *engineering cycle* is a collection of tasks that a rational problem solver would follow to solve a world problem. To solve world problem rationally, we would investigate the current situation, generate possible actions, rank these actions on their problem-solving effects, choose one, do it, and then investigate the result to see if further actions need be taken [9]. It turns out that this is the structure of engineering as identified in a variety of disciplines, ranging from architecture to product engineering [1, 4, 15, 18, 23]. More in detail, the structure of the engineering cycle is as follows.

1. *Problem investigation* This is a knowledge problem, because the engineer wants to acquire knowledge about the problem. Examples of relevant knowledge questions include the following:
 - Who are the stakeholders?
 - What are their goals?
 - What are the observable phenomena?
 - What are the causal relations among these phenomena?
 - Why are some phenomena problematic? In other words, which criteria are used to decide that certain phenomena are problematic?

For example, if we want to improve requirements negotiation practice, we need to know who the stakeholders are in requirements negotiation processes, what they want, and what actually happens during requirements negotiation, why this happens, and what is good or bad about this.

2. *Solution design* We use the word “design” in its dictionary sense as “to conceive and plan out in the

Fig. 1 Mutual recursion between world problems and knowledge problems



mind". Solution design may be a radically creative task in which the engineer comes up with a design never seen before; more usually, it will be an adaptation of known designs to the current problem. For example, once the engineer knows what the problem is with current ERP requirements negotiation practices, she may select one of the available multi-party negotiation techniques and propose it as a solution to the problem. Alternatively, the engineer may invent a radically new negotiation technique, and propose it as a solution to the problem. In both cases, the engineer should specify the technique clearly and completely so that different people can test it, and also give an argument why this would solve the problem.

3. *Design validation* An argument that a solution S solves the problem with phenomena P must have the following form: If S is implemented, then P will be changed into P' , and P' is better than P . The argument contains a causal claim and a value claim. The causal claim is that implementation of S causes P' . The value claim is that P' is better than P . One way to justify these claims is to just implement the solution and observe what happens. However, in any real world problem, this is not possible. Before stakeholders want to implement a solution, they want sufficient proof that S does indeed cause P' and that P' is indeed better for them than P . Validation is the activity of providing such a sufficient proof. It is a knowledge problem. Questions to be asked in solving it include the following.

- What properties will the implemented solution have?
- What effects will the implemented solution have on the phenomena in the problem domain? Under which conditions?
- Are the expected effects better than the current situation? According to which criteria?
- Does the solution introduce new problems? In other words, are their additional effects with a negative value?

One is reminded here of research of the effects of new medicines [6].

4. *Choose a solution* In a rational engineering process, several solutions must be specified, and validated on their problem-solving effects. A solution can then be chosen that has the preferred problem-solving effects. The choice is not made by the engineer but by a stakeholder such as the sponsor who pays the engineer to solve the world problem.

5. *Implement the chosen solution* "Implementation" is an overloaded word. For a software engineer it means to write software; for a software system manager it means to introduce software in an organization; and for the business manager it means to introduce a new way of working, software that supports it, and software management procedures in an organization, etc. In general, what counts as an implementation

depends on what specification must be implemented: a software specification, a system specification, a business solution specification, etc. And what counts as a solution specification depends on what the problem is. So indirectly, what counts as an implementation depends on what the original world problem is. But in all these cases, implementation is the activity of realizing the chosen solution specification in the real world.

6. *Implementation evaluation* The solution specification is implemented with the intention to solve the world problem with which we started. However, our solution may have the intended effects only to some extent, or not have them at all. Additional unforeseen effects may occur, the world may have changed even when we were designing the solution, etc. So the rational engineer will monitor the implemented solution in order to discover its actual effects, and evaluate these effects. This is again a knowledge problem, called the implementation evaluation problem. Questions to be asked include:

- Does the solution really behave as predicted earlier?
- Does it really solve the problem as we predicted?
- What unexpected properties does the implemented solution have?
- Are any new problems introduced?

This is a classification of six engineering tasks with a certain justification relationship. The choice of solution to implement is justified by the solution validations and the problem investigation.

Real-world decisions contain all these six tasks, but with varying degrees of concurrency. Experienced engineers for example start solving an engineering problem with only a little problem investigation, just enough for them to recognize the problem class, and then select a solution direction that they know usually works for this problem class. From then on they perform problem investigation, solution design, and solution validation in parallel [5]. Complex management decision processes also exhibit this parallelism [9, 24]. In very complex management problems, all six tasks occur in an interleaved way [14]. However, even if these tasks occur in parallel, they can still be distinguished as separate tasks. And in order to justify the result rationally, the engineer has to refer to a problem investigation, a solution design, and a solution validation.

A report about an engineering process should follow the logical structure of the rational engineering cycle because this is the way in which other engineers can understand what was done, i.e., understand the problem, the solution design and its validation, and the actual performance of the implementation. By presenting the engineering process as if the rational process was followed, accountability is created. This is no different from Parnas and Clements' idea of faking a rational software engineering process [16], Lakatos'

idea of rational reconstruction of scientific discovery [12, 13], or Suchman's idea of reconstructing office procedures from actual practice [20, 21]. In all cases, constructibility of the rational process from what was actually done creates accountability. The rational engineering cycle is the structure of justification and not necessarily the structure of the actual engineering process.

RE researchers act as engineers of the RE process, because they propose solutions to problems in this process. RE papers should therefore report on one or more tasks in the engineering cycle followed by the RE researcher. The engineering cycle is therefore a source of criteria to be used in evaluating the structure of RE papers.

2.3 Criteria for the presentation of a solution to a world problem

When a paper presents a solution to a world problem, it should answer the following questions:

- Which world problem is solved?
- How was the problem solved?
- Is the solution relevant?

Appendix 1 gives a checklist based on these questions. Here we discuss each of the criteria in turn.

Which world problem is solved? World problems consist of problematic phenomena. Now, being problematic is not a property of a phenomenon itself, but of the relationship between a phenomenon and some norm. For example, the phenomenon that a development project took 3 years is not good or bad in itself. It is good with respect to the norm that it should not exceed 4 years, but it is bad with respect to the norm that it should be less than 2 years. Operationalized in terms of values for observable variables, a problem consists of the difference between observed values and desired values of these variables. The desired values are called *norms* in this paper. Stakeholders often refer to their norms as *goals*. We regard these terms as synonyms.

It is not always clear how exactly the relevant norms apply to the phenomena. If certain groups of stakeholders receive less attention than others in ERP implementation projects, does this contradict the norm that all stakeholder groups should get a fair hearing? What is "fair" in this case? Should all stakeholders get equal attention, or are some truly more important than others? If a paper calls unequal attention to different stakeholder groups a problem, it should explain with respect to which norm this is a problem, and how this norm applies to the problem situation.

To be applicable to a problem situation, norms often have to be operationalized, for example by choosing values of variables as indicators. We call an operationalized norm a *criterion*.

There is usually a *set* of norms, related to different stakeholders, and a problem description should identify these relationships. Which stakeholders require which norms? And why does a stakeholder require adherence to a norm because the norm is intrinsically valuable, or because adherence to the norm leads to adherence to another norm? For example, is the norm of fairness intrinsic, or is it instrumental for another norm, namely avoidance of implementation failure?

During problem-solving, new norms may be identified that were not apparent earlier, or that were not relevant earlier. For example, suppose a business process delivers results too late too many times, and this is diagnosed as a problem with coordination between two departments. We may improve the coordination between two departments by means of (1) weekly meetings, or by (2) the use of a workflow system. If we choose solution option (2), the additional norm that this system runs on currently available platforms becomes relevant. This norm is not relevant in case option (1) is chosen, and so far it has not been identified in the problem investigation. So some relevant norms in a problem situation may only become relevant after a solution option is chosen. And some may become relevant only after a solution is implemented. For example, interoperability may become an issue after a workflow system is implemented. Generalizing from this example, some norms define the problem (the number of late results must be below a certain level), whereas others are constraints that are only relevant for some solutions and not for others.

In addition to describing phenomena, norms, and stakeholders, there is another concern that the engineer may have to describe. Usually, the problem of interest is part of a complex bundle of problems, involving many phenomena, many norms, and many stakeholders. For example, in addition to being unfair to some stakeholder groups, ERP requirements negotiation may lack a true insight into the cost of implementing various requirements, no renegotiation takes place when new cost data become available, etc. If an RE paper proposes a solution for "the" ERP requirements negotiation problem, which of these problems does it solve? Are the different problems distinguished, are priorities given to them, and does the author make a choice which problem to solve?

To sum up, our first list of criteria for evaluation papers reporting on a solution to a world problem is this:

- 1.1 Phenomena
- 1.2 Norms
- 1.3 Relation between norms and phenomena
- 1.4 Stakeholders
- 1.5 Problem-solver's priorities

How was the problem solved? These criteria follow the engineering cycle, which we discussed at length in the preceding section. In a description of the outcome of an engineering process in which several alternative solutions were designed and validated, and one was chosen,

implemented, and evaluated, one should find the following elements:

- 2.1 Problem investigation
 - 2.1.1 Causal relationships between phenomena
- 2.2 Solution design
 - 2.2.1 Source of the solution. Is it borrowed or adapted from some published source, or is it totally new?
 - 2.2.2 Solution specification
- 2.3 Design validation
 - 2.3.1 Solution properties
 - 2.3.2 Solution evaluation against the criteria identified in problem investigation.
- 2.4 Choice of a solution
 - 2.4.1 Alternatives considered
- 2.5 Implementation description
- 2.6 Implementation evaluation
 - 2.6.1 Observations
 - 2.6.2 Relating observations to the criteria relevant for this implementation.

Solution validity, which is the sufficient proof of that the solution works for this problem, is often called “soundness” of the solution in the RE literature.

What is the relevance of the solution? A minimal relevance criterion is novelty. The proposed and implemented solution must not have been proposed and implemented before. However, relevance also means that this solution not only works for this particular problem, but for a whole class of problems. For example, in a paper describing a new ERP requirements negotiation technique, is this technique applicable to all ERP implementations? To other implementation processes? In addition, since the paper described solution properties and describes properties of an implementation, does the paper also contain a significant addition to our knowledge?

This leads to the following three relevance criteria.

- 3.1 Novelty of the solution
- 3.2 Relevance for classes of world problems
- 3.3 Relevance for theory

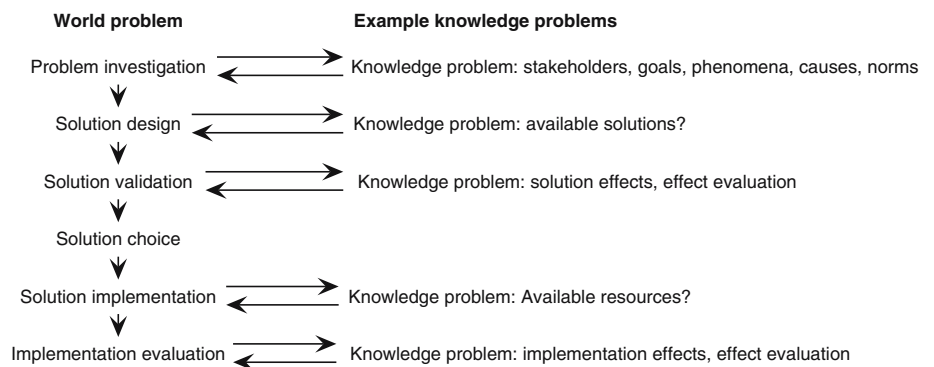
Discussion We do not mean that each paper about world problems should satisfy all these criteria. It may take a Ph.D. project to analyze a problem, propose a solution technique, and do a proof-of-concept validation. It may take two more years to do a proper validation, and an additional 10 years to implement it in the real world and evaluate its use. Journal papers may present a solution specification and a validation, and conference papers may only specify a solution and illustrate it, without validating it. Instead of requiring every engineering paper to satisfy all criteria, we use the engineering cycle to classify a paper, and then choose appropriate criteria to evaluate it.

2.4 The research cycle

We observed that there are several knowledge problems in the engineering cycle. Figure 2 summarizes them. An answer to a knowledge problem may be found by various means such as looking it up in the literature, consulting other people who solved similar problems, or by using one’s own previous experience with similar problems. Here we are interested in the case that original research must be done to find an answer. Hence, it is relevant for us to describe the research cycle.

To solve a knowledge problem by doing research, we have to figure out what we want to know, design the research by which we want to acquire this knowledge, validate this design, do the research as designed, and finally evaluate the outcome of this. This *research cycle* has the same structure as the engineering cycle, and for a good reason both have the structure of rational action, in which we plan an action before we act, and in which we evaluate the outcome of an action after the act. More in detail, the research cycle has the following structure.

Fig. 2 Knowledge problems in the engineering cycle



1. *Problem investigation* Before we design the research, we must investigate the nature of the problem we need to solve. Examples of questions to be answered at this stage are the following:
 - What do we want to know?
 - Why do we want to know this? This, the research goal, relates the research problem to a world problem. Which world problem provides the context for this knowledge problem? Is this part of the problem investigation, or of a design validation, or of an implementation evaluation?
 - Who wants to know this?
 - What do we already know about this problem? Is a simple literature study sufficient, or must we design a more elaborate research?
 - Is it an empirical or a conceptual problem? In this paper, we restrict ourselves to empirical problems.

For example, if we are investigating a world problem, then we want to know what the problematic phenomena are, who are the stakeholders and which norms they have with respect to which these phenomena are problematic. We want to know this because we need to know which actions can help solve this world problem. We might have solved this type of world problem before, in which case no research needs to be done because we already understand the problem; or the problem may be hitherto unsolved, in which case we need to do research in order to understand the problem.

2. *Research design* If we have decided that we must do research to acquire the desired knowledge, then we need to design the research. Can we do laboratory experiments, or must we do field research? Are there cases available that we can study, or should we do simulations or build prototypes to study? Or should we do action research to try out our techniques in reality? In general, a research design should state what we observe, how we observe it, and how we analyze the observations. More information about research design can be found elsewhere [3, 11, 17, 25].
3. *Design validation* The research design must be validated in several ways. First, any design for empirical research uses a conceptual framework within which to describe the relevant phenomena. The design must be checked for *construct validity*, which means that the relevant concepts defined in this framework have been operationalized adequately into observable variables. Second, empirical research results in claims about causal relationships among these variables. For example, a paper may claim that insufficient attention to user groups in ERP implementation significantly adds to the implementation costs. This is a causal claim. Other possible claims are about statistical relationships or are just descriptive. Whatever the knowledge claim, a research design is *internally valid* if the claim indeed follows from the observations.

Internal validity is a property of a research design. Third, even when constructs have been properly operationalized and knowledge claims are internally valid, we must still check whether the result generalizes to other cases, beyond the one observed in the research. This is the question of *external validity*. See Cooper and Schindler [3] for more about the various forms of validity.

4. *Research* Having designed the research and validated the design, we must do it. It is important that the research be performed in a repeatable and intersubjective manner. If we would do it again, we should get the same results, and if others would do it, they should also get the same results.
5. *Evaluation of outcomes* The knowledge acquired by the research should be related to the original research questions and research context. If a world problem has been analyzed, we should now check if we understand enough of the problem to start thinking about a solution. If a design of a solution to a world problem has been validated, we should now check whether we know enough about the design to trust that it solves the original world problem. And if an implementation has been evaluated, we should now check whether we know enough about the implementation to assess it with respect to the relevant criteria.

See Babbie [2] for more on the structure of knowledge problems.

The research cycle identifies six research tasks that may not always be performed in sequence. During a research project, there will be many jumps back and forth through the research cycle. However, to justify the results of research, the preceding tasks have to be performed and they must have the described relationships to one another. To justify the research procedure, the research design must be described; to justify the design, the relationship to the problem to be investigated must be indicated, and it must be justified that with respect to these questions, the design is valid. In other words, the research cycle is the structure of justification and not necessarily the structure of discovery [12].

2.5 Criteria for the presentation of a solution to a knowledge problem

When a paper reports about research, it should answer the following questions:

- Which knowledge problem is solved?
- How was the problem solved?
- What is the relevance of this outcome?

These are the same questions we asked about world problems earlier. We discuss each of these questions in more detail. Appendix 2 summarizes the criteria that come out of this discussion.

Which knowledge problem is solved? As indicated in Fig. 2, engineering research consists of investigating world problems, validating solution designs, or evaluating implementations. Whatever the context, to indicate which knowledge problem is solved, we must indicate which *phenomena* are of interest, and what properties of these phenomena, called *variables*, we are interested in. The knowledge problem may consist of finding values of the variables of interest in the observed situation, or it may consist of finding causal or statistical relationships among the variables. The problem is usually operationalized in the form of a list of *research questions*. If there are several knowledge problems, are they distinguished and *prioritized*? Is it clear which one the paper is proposing to solve? This leads us to the following criteria.

- 1.1 Phenomena
- 1.2 Variables
- 1.3 Relationships among variables
- 1.4 Research questions
- 1.5 Priorities

How was the problem solved? This involves a *design* of the research by which the answer to the research questions will be found. To describe a research design we must describe the *population* studied, the *measurement procedure* followed and the method by which the results were *analyzed*. Measurement is the assignment of numbers to phenomena according to some procedure [10, page 177]. In the conceptualization of the knowledge problem, the phenomena have been described by means of variables. Measurement is then the determination of values of these variables according to some procedure such as direct observation, probes, questionnaires, interviews, or some other method. Analysis of the results can be qualitative (conceptual analysis) or quantitative (using some statistical method).

Following the research cycle, we also require a report on the *construct validity*, *internal validity* and *external validity* of the research.

When the research is then actually done, we perform *measurements* that must be reported.

Finally, we need to know what conclusions the authors drew from the measurements. What is the *answer* to the research questions that follows from these measurements? Are there *theoretical explanations* of these answers? And is there any way in which these explanations could be wrong? Could there be *fallacies* in the research design, in the measurements, or in the analysis that could invalidate the results?

Summarizing, we have the following evaluation criteria.

- 2.1 Research design
 - 2.1.1 Population
 - 2.1.2 Measurement procedure
 - 2.1.3 Analysis method

- 2.2 Validity
 - 2.2.1 Construct validity
 - 2.2.2 Internal validity
 - 2.2.3 External validity

2.3 Measurements

2.4 Analysis

- 2.4.1 Answers to research questions
- 2.4.2 Theoretical explanations
- 2.4.3 Possible fallacies

What is the relevance of this solution? There are two kinds of relevance that we must check. First, is the knowledge acquired an interesting addition to the current body of knowledge? And second, what is the implication for the engineering context from which we started? So the two criteria here are the following.

- 3.1 Relevance for theory
- 3.2 Relevance for engineering practice

Discussion Papers that report about research should be evaluated according to all of these criteria. Progress reports may restrict themselves to a part of the research cycle, but to report about the entire research, all these criteria must be satisfied. This is different from a report about solving a world problem, where we saw that a paper can restrict itself to part of the engineering cycle, such as problem investigation or solution design.

3 The problem: the methodological structure of RE papers

Having set out our conceptual framework for solving knowledge problems and world problems, we now return to our original question what the methodological structure of RE papers is. We will describe our research using our checklist of evaluation criteria for knowledge problems, summarized in Appendix 2.

The *phenomena* of our problem are RE papers written by researchers and submitted to a scientific audience. This includes papers that have been published in scientific RE conferences and journals as well as those that have been rejected by them. This reflects our concern with the methodological soundness of all submissions to scientific RE conferences and journals. Given this set of phenomena, our original *research question* is split into two:

- Q1 What is the methodological structure of submitted RE papers?
- Q2 Is the methodological structure of accepted papers significantly different from that of rejected papers?

Note that our phenomena are papers that *have been* submitted; it does not contain papers that *will be* submitted.

The *variables* used to describe the phenomena have been defined in our conceptual framework. The checklists of Appendices 1 and 2 is in fact a list of variables whose value we would like to determine. The question, what the methodological structure of an RE paper is?, is thus operationalized into the set of questions what the values of our variables are for the observed phenomena.

Since we do not have a problem bundle, there is no need to set *priorities* among a set of problems.

4 Research design

4.1 Population

Since the set of all submitted RE papers is very large and mostly inaccessible, we have to draw a sample. To answer research question Q1, we selected the first 37 submissions to the RE'03 conference [19]. We call this set *sample 1*. The reason for choosing the submissions to the RE'03 conference is that this set of papers was accessible to us. There is no particular reason for selecting the first 37 submissions; we could also have selected the last 37, or any other subset. There is no reason to assume that the first 37 submissions are systematically different from any other set of 37 submissions.

To answer our second question, we extended sample 1 with the set of all accepted papers. This yielded a set, called *sample 2*, consisting of 30 rejected papers and all the 25 accepted papers (Seven papers in sample 1 were accepted papers.)

4.2 Measurement procedure

We scored both sample on the checklist of Appendix 2. One of us coded the material. After that, the other repeated the coding. Inconsistencies were resolved together.

4.3 Analysis method

Our analysis method contains nothing more complicated than counting the scores of papers on the different variables and summarizing the results.

5 Validity

Construct validity is the extent to which we really measured what we claim to be measuring. There might be interpretation errors in our data. Where one observer reads a description of the phenomena, norms, and stakeholders of a world problem in a paper, another observer might not read the same description in this way at all. However, this did not occur in our observations.

The three or four inconsistencies in coding that we encountered were easily resolvable, and in the vast majority of cases where one of us considered a paper to satisfy a criterion, the other did so too.

A different issue is that papers may satisfy criteria only to a certain extent. Whether or not a paper about a world problem describes the relevant norms, phenomena, and stakeholders is a matter of degree: if the description consists of one sentence and some hand-waiving, is it present or absent? Most criteria are of the kind that can be satisfied to a greater or lesser extent. In almost all cases, when one of us thought that a paper satisfied a criterion only to a small extent, the other also thought so. We therefore introduced a category “satisfies the criterion, but insufficiently”.

Internal validity is the extent to which knowledge claims inferred from the measurements are indeed valid. Our selection of cases and our coding procedure ensure a sufficient level of internal validity.

External validity is the generalizability of our results to the entire population. This is not possible from our two samples. They are too small to draw conclusions about the set of all submitted RE papers. Nevertheless, we will argue later that we can draw some conclusions from our results that are relevant for authors of future RE papers.

6 Measurements

We summarize our measurements in Figs. 3 and 4. We discuss each of the samples in turn.

Papers about world problems There are 37 papers in sample 1, of which 27 are classified as papers about world problems, 6 as papers about knowledge problems and 4 as papers about both a world problem and a knowledge problem. For example, a paper that extracts a method from a number of empirical case studies is classified both as being about a world problem (a method is proposed that is claimed to improve something) and about a knowledge problem (empirical case studies are performed). Clearly, sample 1 has a bias towards world problems.

Few papers give an adequate problem definition. In particular the relationship between norms and phenomena is described clearly by only a few papers. Almost all papers consider a problem in isolation, and do not give priorities. About half of the papers diagnoses the problem.

All papers specify their solution adequately. Validation of the solution receives attention in only a small number of papers. Alternatives are rarely considered. Only a few papers describe an implementation.

All papers in sample 1 claim novelty. However, this was not always substantiated. More than half of the papers claim relevance for practice. Relevance for theory is not a point of attention.

Fig. 3 Score of the two samples on the checklist for world problems (Appendix 1). The *numbers* under “+” indicate how many papers score clearly positive on the criterion, and the *numbers* under “+–” indicate how many papers score weakly on that criterion

	Sample 1		Sample 2			
	N = 31		Rejected = 26		Accepted = 20	
	+	+–	+	+–	+	+–
1. Problem definition						
1.1 Phenomena	7	13	6	11	4	9
1.2 Norms	8	14	8	9	3	14
1.3 Norms/phenomena	2	8	2	8	2	4
1.4 Stakeholders	9	12	7	8	2	12
1.5 Priorities	3	0	0	0	0	0
2. How solved						
2.1 Problem investigation						
2.1.1 Causal	10	7	8	7	8	4
2.2 Solution design						
2.2.1 Source	3	7	3	5	1	6
2.2.2 Specification	31	0	26	0	20	0
2.3 Validation	4	1	4	1	6	1
2.4 Choice						
2.4.1 Alternatives	3	2	3	2	1	3
2.5, 2.6 Implementation	2	1	1	1	1	0
3. Relevance						
3.1 Novelty	31	0	26	0	20	0
3.2 Relevance for practice	11	10	8	9	8	6
3.3 Relevance for theory	1	1	1	1	0	0

Fig. 4 Score of the two samples on the checklist for knowledge problems (Appendix 2). The *numbers* under “+” indicate how many papers score clearly positive on the criterion, and the *numbers* under “+–” indicate how many papers score weakly on that criterion

	Sample 1		Sample 2			
	N = 10		Rejected = 8		Accepted = 6	
	+	+–	+	+–	+	+–
1. Problem definition						
1.1 Phenomena	10	0	8	0	6	0
1.2 Variables	6	3	6	2	4	2
1.3 Questions	1	0	1	0	0	1
1.4 Priorities	1	1	1	1	0	0
2. How solved?						
2.1 Research design						
2.1.1 Population	4	0	3	0	4	1
2.1.2 Measurement procedure	3	1	2	1	4	0
2.1.3 Analysis method	3	2	1	2	4	1
2.2 Validity	2	1	0	1	3	1
2.3 Measurements	3	2	2	2	4	1
2.5 Analysis	0	1	0	1	4	1
3. Relevance						
3.1 For theory	2	3	2	0	4	0
3.2 For practice	4	2	1	1	0	0

Sample 2 displays a bias towards world problems too. The quality of problem definition is slightly worse for accepted papers than it is for rejected papers in sample 2.

Diagnosis of the problem occurs just as little in accepted as in rejected papers. Validation of the solution receives attention in relatively more accepted papers than

rejected papers. Alternatives are hardly considered in accepted papers. Relevance for practice is claimed by relatively more accepted than rejected papers, but the difference is small.

Papers about knowledge problems Turning to Fig. 4, 10 papers in sample 1 and 16 in sample 2 are research papers. That is about 25% of each sample. We find the same percentage in the accepted and rejected part of sample 2: about 25% of each part is research papers.

We see that in both samples 1 and 2, almost all research papers define their variables, but that research questions are hardly stated and priorities not set. Less than half of the papers describe their research design, and validity is not an issue for most research papers. More accepted than rejected papers (relatively speaking) present their measurements. More accepted than rejected papers (relatively speaking) claim relevance for theory, but accepted papers do not claim relevance of their results for practice.

7 Analysis

7.1 Answers to research questions

To summarize, the pattern that emerges from these data is that in our samples, papers about world problems present solutions to isolated problems. Relevant norms and stakeholders are not always indicated, and diagnosis of the causes of the problem is omitted in almost half of the papers in each sample. Solutions are specified clearly, but the sources for them are usually not indicated and validation is almost always omitted. Very few papers describe an implementation. More than 60% of the papers in each sample claim relevance for practice. Relevance for theory is hardly an issue for the papers in our samples.

These observations are also made of the sets of accepted and rejected papers in sample 2. An important difference between accepted and rejected papers in sample 2 is that more accepted than rejected papers devote space to validation.

Most research papers in our samples do not formulate research questions, and about half do not describe their research design and almost none discuss validity. This means that generalizability of results is at issue. If a research design is not described, others cannot repeat the research. And if validity is not discussed, generalizability to the entire population cannot be claimed.

7.2 Explanations

The provision of explanations is not a part of our research goal. Nevertheless, in this section we speculate about possible explanations of our results. These explanations are hypothetical and need further research to confirm or refute them.

The first hypothetical explanation for these data that we offer is that authors of RE papers are designers: They design new techniques and then publish them, illustrating them with an example. If it is not the problem but the solution that is of interest, then this would explain the absence of problem diagnosis, and of setting priorities in (real-life) problem bundles. It does not explain the lack of validation though, because validation is the investigation of solution techniques.

The lack of validation can be explained by the fact that RE papers are about RE processes performed by people. To validate that a technique improves RE practice, the researcher could do experiments or use other empirical methods to discover properties of the technique, and relate these properties to the situations in which the technique can be used. This is hard work that usually takes long to accomplish, and moreover they are investigations into what happens when *people* use the new technique. This requires research skills similar to those of social scientists, and the designers who invented the techniques may not be attracted by this kind of research. They find it less fun to do the research than to invent the techniques. This is the hypothetical explanation that we offer for the lack of validation in RE'03 papers. The lack of attention to research design in the research papers in our samples is also explained by it.

The absence of any discussion of implementation in papers about world problems follows from the same hypothesis. RE techniques are to be used by people. Implementing those techniques means transferring them to industrial practice. This is a slow process that can take anywhere from 5 to 20 years. It is unrealistic to expect discussion of implementations soon after a technique is invented. However, given that the field of RE is about 15 years of age, it does not explain why there are only one or two discussions of implementation in our samples.

7.3 Soundness

This brings us back to the validity of our results. We do not know whether our samples are representative for the entire population of hitherto submitted RE papers, and so we cannot claim external validity. But neither can we find a reason why our samples would be different from all other RE papers. The call for papers of RE'03 differed from previous calls in that slightly more attention was spent on the evaluation criteria for the different paper categories. Validation was mentioned explicitly. If there is any difference between our samples and earlier RE submissions, then we would expect our samples to spend more attention on validation than earlier submissions.

Is there any possibility that our results are wrong? Is there a flaw in our reasoning that invalidates our observations? The coding of the papers in terms of our checklists involves a qualitative interpretation, as our distinction between “+” and “+−” shows. This is a

fuzziness shared with a lot of social science research conducted by questionnaires. Subjects who try to answer a question may feel that the question is slightly beside the point that under different conditions, not mentioned in the question, different answers are possible, or that there is a definite answer that is not in the list. As a result, their answers may not represent what they really think. Many such errors of interpretation may occur if a questionnaire has not been tried out before being used.

We think we have included sufficient measures to exclude interpretation errors in the data. The criteria of Appendices 1 and 2 have clear definitions and, as indicated earlier, are in agreement with the accepted views of research methodologists. If the concepts would have been fuzzily defined, then many coding inconsistencies should have occurred. Since there were hardly any coding inconsistencies, the remaining fuzziness is part of the data itself and not of the concepts introduced by us. And since we agreed on almost all our qualitative judgments about the extent to which a paper satisfied a criterion, we think the data in Figs. 3 and 4 are reliable.

8 Discussion

Returning to our motivation to do this research, the majority of papers in our samples discusses solutions to world problems, but few of them do problem investigation or solution validation. We claim that this inhibits the transfer of research results to practice. To facilitate such a transfer, it should be clear to practitioners for which problems a technique has shown to be successful. Therefore, researchers should select relevant problems to solve, and they should deliver techniques that have been validated for particular problem classes.

This is true for both problem- and technique-driven engineering research. In *problem-driven* research, engineers study real-world problems and try to design and validate solutions for them. In *technique-driven* research, engineers study the properties of techniques and try to find real-world problems that can be solved by them. In either case, problems and solutions are matched by investigating both problem properties and solution properties. If we do neither problem investigation nor solution investigation, then chances are that the techniques we come up with are irrelevant for practice.

The relevance of our research for practice consists of the normative conceptual framework we defined to do this research. If our results—lack of problem investigation and solution validation—would be falsified by all future RE papers, then this would be a good thing, for then two barriers to transfer of research results to practice would have been removed. But even in that case, we think our two case studies show that it is useful to publish a conceptual framework for engineering research such as the one we describe, for at least in our samples there are too many papers that do not follow the precepts of this framework.

We have simply applied principles known from research methodology, and we do not claim any contribution to that field. However, we do claim that we have clarified in Sect. 2 the relationship between research and engineering, and to have defined two convenient checklists of criteria not published before. Just as our engineering checklist identifies two obstacles for transfer of techniques to industry, namely lack of problem investigation and lack of validation, so our research checklist identifies two obstacles towards creating an accumulation of knowledge, namely the lack of research designs and the lack of discussions of validity of results. This prevents researchers to reuse and build upon research results published by others.

Acknowledgements This paper benefited from discussions with Klaas van den Berg and the participants of the CERE04 workshop and from the comments by anonymous reviewers.

Appendix 1. Checklist for papers about solutions to world problems

An engineering paper does not need to satisfy all criteria listed below. Which criteria are relevant depends on the task in the engineering cycle that the paper reports about.

1. Which world problem is solved?
 - 1.1 Phenomena
 - 1.2 Norms
 - 1.3 Relation between norms and phenomena
 - 1.4 Stakeholders
 - 1.5 Problem-solver's priorities
2. How was the problem solved?
 - 2.1 Problem investigation
 - 2.1.1 Causal relationships between phenomena
 - 2.2 Solution design
 - 2.2.1 Source of the solution. Is it borrowed or adapted from some published source, or is it totally new?
 - 2.2.2 Solution specification
 - 2.3 Design validation
 - 2.3.1 Solution properties
 - 2.3.2 Solution evaluation against the criteria identified in problem investigation.
 - 2.4 Choice of a solution
 - 2.4.1 Alternatives considered
 - 2.5 Implementation description
 - 2.6 Implementation evaluation

- 2.6.1 Observations
- 2.6.2 Relating observations to the criteria relevant for this implementation.

3. Is the solution relevant?

- 3.1 Novelty of the solution
- 3.2 Relevance for classes of world problems
- 3.3 Relevance for theory

Appendix 2. Checklist for papers about solutions to knowledge problems

A research paper must satisfy all criteria.

1. Which knowledge problem is solved?

- 1.1 Phenomena
- 1.2 Variables
- 1.3 Relationships among the variables
- 1.4 Research questions
- 1.5 Priorities

2. How was the problem solved?

- 2.1 Research design
 - 2.1.1 Population
 - 2.1.2 Measurement procedure
 - 2.1.3 Analysis method
- 2.2 Validity
 - 2.2.1 Construct validity
 - 2.2.2 Internal validity
 - 2.2.3 External validity
- 2.3 Measurements
- 2.4 Analysis
 - 2.4.1 Answers to research questions
 - 2.4.2 Theoretical explanations
 - 2.4.3 Possible fallacies

3. What is the relevance of this solution?

- 3.1 Relevance for theory
- 3.2 Relevance for engineering practice

References

1. Archer B (1969) The structure of the design process. In: Broadbent G, Ward A (eds) *Design methods in Architecture*. Lund Humphries, pp 76–102
2. Babbie E (2001) *The practice of social research*, 9th edn. Wadsworth, Belmont
3. Cooper D, Schindler P (2003) *Business research methods*, 8th edn. Irwin/McGraw-Hill
4. Cross N (1994) *Engineering design methods: strategies for product design*, 2nd edn. Wiley, New York
5. Cross N (2001) Design cognition: results from protocol and other empirical studies of design activity. In: Eastman C, McCracken M, Newstetter W (eds) *Design knowing and learning: cognition in design education*. Elsevier, Amsterdam, pp 79–103
6. Davis A, Hickey A (2004) A new paradigm for planning and evaluating requirements engineering research. In: 2nd international workshop on comparative evaluation in requirements engineering, pp 7–16
7. Gause D, Weinberg G (1989) *Exploring requirements: quality before design*. Dorset House Publishing
8. Glass R, Ramesh V, Vessey I (2004) An analysis of research in the computing disciplines. *Commun ACM* 47(6):89–94
9. Hicks M (1999) *Problem solving in business and management: hard, soft and creative approaches*. International Thomson Business Press
10. Kaplan A (1998) *The conduct of inquiry. Methodology for behavioral science*. Transaction Publishers, 1998. First edition 1964 by Chandler Publishers
11. Kitchenham B, Pfleeger S, Hoaglin D, Emam K, Rosenberg J (2002) Preliminary guidelines for empirical research in software engineering. *IEEE Trans Softw Eng* 28(8):721–733
12. Kuhn T (1970) Logic of discovery or psychology of research? In: Lakatos I, Musgrave A (eds) *Criticism and the growth of knowledge*. Cambridge University Press, Cambridge, pp 1–23
13. Lakatos I (1976) Proofs and refutations. In: Worall J, Zahar E (eds) *Cambridge University Press, Cambridge*
14. March J (1994) *A primer on decision-making*. Free Press
15. Pahl G, Beitz W (1986) *Konstruktionslehre. Handbuch für Studium und Praxis*. Springer, Berlin Heidelberg New York
16. Parnas D, Clements P (1986) A rational design process: how and why to fake it. *IEEE Trans Softw Eng SE-12*:251–257
17. Pfleeger S (1995) Experimental design and analysis in software engineering. *Ann Softw Eng* 1:219–253
18. Roozenburg N, Eekels J (1995) *Product design: fundamentals and Methods*. Wiley, New York
19. Sikkil N, Wieringa R (eds) (2003) In: *Proceedings of the 11th IEEE international requirements engineering conference*. IEEE Computer Science Press
20. Suchman L (1983) Office procedures as practical action: models of work and system design. *ACM Trans Office Inf Syst* 1:320–328
21. Suchman L, Wynn E (1984) Procedures and problems in the office. *Office Technol People* 2:135–154
22. Tichy W, Lukowicz P, Prechelt L, Heinz E (1997) Experimental evaluation in computer science: a quantitative study. *J Syst Softw* 28:9–18
23. Wieringa R (1996) *Requirements engineering: frameworks for understanding*. Wiley, New York
24. Witte E (1972) Field research on complex-decision-making processes—the phase theorem. *Int Stud Manage Organ* 2:156–182
25. Wohlin C, Runeson P, Höst M, Ohlsson MC, Regnell B, Weslén A (2002) *Experimentation in software engineering: an introduction*. Kluwer, Dordrecht
26. Zelkowitz M, Wallace D (1997) Experimental validation in software engineering. *Inform Softw Technol* 39:735–743