Validating the R-CMM

Sarah Beecham, Tracy Hall and Austen Rainer

Technical Report No 373

Department of Computer Science
Faculty of Engineering and Information Sciences

April 2003

1: Introduction

We have created a framework that aims to represent the essence of good requirements practices. This Requirements Capability Maturity Model (R-CMM[1]) prioritises the requirements process and provides a pathway for software practitioners to follow in their requirements process improvement activities. The R-CMM is based on the established software improvement model developed by the Software Engineering Institute (SEI). We believe that the SEI framework has considerable strengths and since its release in 1991, the Software Capability Maturity Model (SW CMM) has become increasingly popular as a method for controlling and improving software practices (Paulk *et al.*, 1995). However, there has been little formal validation of the SEI model and we therefore look to the literature for a baseline or standard to guide our model evaluation activities, e.g. (Lindland *et al.*, 1994; Berry and Jeffery, 2000; Dybå, 2000; Berry and Vandenbroek, 2001; Kitchenham *et al.*, 2002b). In this report we present our validation methodology and report the findings of a detailed validation questionnaire that involved a group of experts in the fields of Requirements and Software Process Improvement (SPI).

The main motivator for creating this specialisation of the SW CMM is our empirical study into the problems companies are having with their process improvement activities (Hall *et al.*, 2002) and (Beecham *et al.*, 2003). The 12 companies in our study represent four levels of maturity were all using the SW CMM as their model for improving their processes. Although technical problems eased off as companies matured, the organisational problems are not supported by increased maturity. Our results showed that companies in our study had a general lack of control over their requirements processes. Requirements best practices, requirements process models and the SW CMM do not appear to be reaching these companies. This may be due to Laueson's finding that 25% of requirements defects cannot be prevented by any technique Laueson (2001). Yet, a high maturity company did report that shifting resources to the requirements process resulted in less defects in the software, e.g. (Krasner *et al.*, 1994). The R-CMM therefore endeavours to introduce the benefits of a strong requirements process to less mature companies.

This report represents the final stage of the first cycle of model development. This validation stage however, does not mark the end of development, as findings from this study will be fed back into the model. The four main stages to our model development comprise creating a framework; populating the framework; applying the practices in an assessment; and a validation of the model. These four strands of work are reported separately as shown in Figure 1.

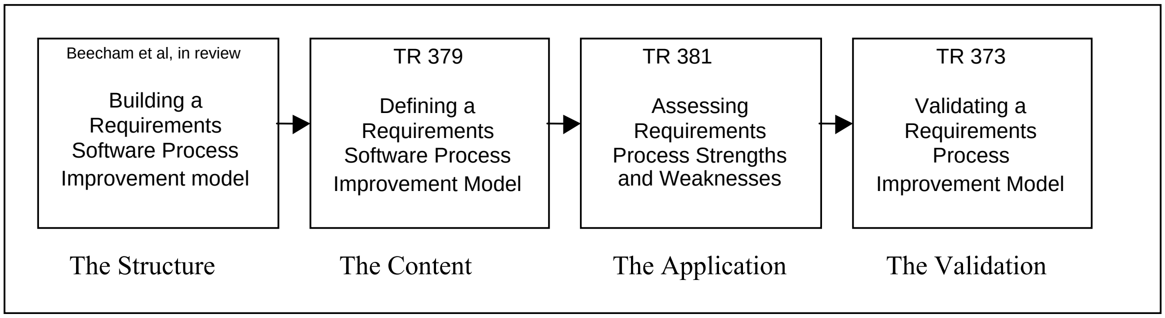

Figure 1: 4 dimensions of model development

It is in this report that we discuss how the model is validated and give some of the key results of the validation process. The main processes involved in validating the R-CMM are:

- Consider the criteria for success devised during the initial stages of model development that includes requirements and objectives of the model;
- create a questionnaire to query/test all the success criteria;
- select a panel of experts to test the model through participating in the questionnaire;
- analyse the expert panel responses to the questionnaire;
- report the findings of the questionnaire.

We refer to the literature to guide us in all these processes and emulate best practices from previous studies that validate similar models. Future work will involve considering how to implement the questionnaire findings to create an improved model.

2. Expert Panel

We emulated previous methods used in the validation of improvement models and measurement ‘instruments’ by inviting a panel of experts to complete a detailed questionnaire (e.g. (El Emam and Madhavji, 1996) and (Dybå, 2000)). A targeted group of experts were asked whether they would be able and willing to participate in a tick box questionnaire that the pilot study showed took 1 hour to complete. The experts were given a brief background to the study.

We assessed that the expertise needed to validate the R-CMM falls into 4 categories as given in Figure 2.

	CMM	Requirements
Practitioners	✓	✓
Academics	✓	✓

Figure 2: Four areas of expertise represented by the R-CMM validation panel

Although Figure 2 shows 4 distinct areas of expertise, in practice there is an overlap in the experience of panel members. For example many academics also have some practical experience of requirements, and some quality managers implementing the CMM also have a strong awareness of the requirements process. These subjects were selected to be representative of the population and allow us to draw conclusions from the results ("Prelim guidelines', Kitchenham et al, 2002). For example, the needs of the practitioner and the knowledge of researcher are incorporated into the R-CMM analysis. This is in line with Kitchenham et als (2002)(Framework) suggestion that different audience groups should have been used to assess their model. In our case, practitioners needed to be represented as they are the primary motivator for the R-CMM framework and contents (See Hall et al, 2003). The views of academia are also needed as the R-CMM incorporates some best practices suggested by this group. The SEI recommendations for experts participating in a process assessment support this cross-section of knowledge as shown in Figure 3.

SEI suggest the following team take part in process maturity assessment (Paulk *et al.*, 1995):	We suggest the following team for validating the CMM Requirements Model:
EXPERIENCED PEOPLE	
	practitioners and academics
→ knowledgeable in the process	→ knowledgeable in the CMM method (researcher)
→ knowledge in the technology (software development: coding, design etc.…)	→ knoweldgeable/practiced in requirements (elicitation, specification, validation: traceablily, modelling etc.) (researcher and practitioner)
→ knowledge of the application area	→ Participated in SPI: process assessment; modelling; measuring (practical experience – practitioner)

Figure 3: Adapting SEI assessment team recommendations to requirement validation team attributes

Each expert was sent two booklets as given in Appendices A and B. The first booklet is the questionnaire and the second booklet is the accompanying documentation to which the questionnaire refers. The documentation is a collection of models showing how the R-CMM grows from the SW CMM framework into a specialised model that views requirements through a goal question metric paradigm (Basili and Romach, 1988). The documentation booklet also includes an example of how the R-CMM assesses the strength of a requirements process. Feedback from the pilot study indicated that providing any further documents and detail would require an unacceptable level of input from the panel member, yet many of the terms and processes represented in the model need further clarification. Therefore, to limit the amount of documentation, we placed detailed definitions of all the terms and processes listed in the model on a web page for our panel to access.

The questionnaire review process consisted of several activities:

- establishing the expert's view on the established SW CMM as a process improvement framework;
- noting how strongly the expert felt about the need for requirements process support;
- measuring how well the model complied with the success criteria listed in Appendix C.

The two initial questions were important indicators; as if (1) the expert did not believe the SW CMM was a useful model they were unlikely to view an offshoot of this model positively. Also if the expert's response to the second assertion, that requirements were in need of further support, was weak, they are unlikely to see value in a detailed model of requirements. Questions relating to the success criteria (3) were designed to show the strengths and weaknesses of the model.

2.1 Expert Panel Demographics

Twenty-eight experts were approached and we received a positive response from 23. We sent the questionnaire (Appendix A) together with accompanying model documentation (Appendix B) to the 23 participants, of whom 20 completed the questionnaire. The 20 experts represent the roles given in Figure 4:

	Practitioner	Academic	Practitioner & academic	Total
SPI/CMM	1	0	0	1
Requirements	6	4	1	11
CMM and Requirements	3	3	2	8

Figure 4: Distribution of expertise in the R-CMM expert panel

Figure 4 shows that all our experts, with the exception of one, have a good or expert knowledge of requirements (19/20), whereas only 9/20 have a good or expert knowledge of the CMM. We therefore expect some of the CMM related questions to be given a 'no opinion/ don't know' response, whereas the requirements questions should elicit replies from most participants.

Experts were targeted in academia for having published work on requirements and/or software process improvement, in industry the candidates were selected for their 'hands on' experience in the field of requirements and/or implementing improvement programs. In order to confirm our assessment, we included a section for participants to assess their own level of expertise at the beginning of the questionnaire. Only in two cases did we alter the participants' responses to be consistent with their associated details that revealed a different profile. For example, a participant who was personally involved in CMM assessments or participated in CMM activities was assessed to have a 'good' knowledge of the CMM rather than a 'fair' knowledge of the CMM.

2.2 Reliability of expert panel responses - knowledge

Validation of the R-CMM is dependent on the expert panel's ability to give accurate feedback. (Lauesen and Vinter, 2001) recommend using experts with different backgrounds to counter this problem, and we therefore targeted experts from different backgrounds. For example the requirements research community tend not to support the SW CMM as a means of controlling, managing and improving the requirements process as is seen in the work of Sommerville and Sawyer with their alternative Requirements Good Practice guide (Sommerville and Sawyer, 1997). And in a similar way to Dyba (2000), we use several questions to measure each area under analysis. Combining multiple items to test each success criteria averages out errors and gives a more accurate and consistent (reliable) measurement.

2.3 Reliability of expert panel responses – scope

Naturally, the larger the sample group and the broader the spread of knowledge the more confident we will be in the results. However Laueson and Vinter (2001) used only 3 experts and therefore were able to insist on them reaching a consensus rather than taking an average. The experiment found that the ability of the experts to predict techniques to prevent requirements defects were very high when put into practice (Lauesen and Vinter, 2001). Kitchenham et al (Kitchenham *et al.*, 2002a) analysed the accuracy of several methods of estimating project effort. Although results of this empirical study were based on projects in a single company, the rigourous statistical analysis showed that human centred estimating processes incorporating expert opinion can substantially outperform simple function point models. Although the context in which we have used experts differs substantially to Kitchenham et al's, it is perhaps relevant to note that human involvement in the form of expert mediation can substantially outperform simple algorithmic models.

Dyba used 11 experts to conduct his review process emphasising the importance of each expert's depth of knowledge and hands-on experience. We therefore conclude that using 20 experts, with a proven record of experience, from diverse backgrounds is a valid way to validate the R-CMM. The depth and scope of the R-CMM expert panel meet the requirements set by previous work in the area of expert panel validation.

When measuring responses to quality issues we are not dealing with interval data, and El Emam and Madhavji discuss this *construct* when analysing questionnaire data:

> "A basic concept for comprehending the reliability of measurement is that of a *construct*. A construct refers to a meaningful conceptual object. A construct is neither directly measurable nor observable. However, the quantity or value of a construct is presumed to cause a set of observations to take on a certain value. An observation can be considered as a question in a maturity questionnaire (this is also referred to as an *item*). Thus the construct can be indirectly measured by considering the values of those items." (El Emam and Madhavji, 1995)

2.2 Questionnaire Design

The questionnaire is a tool used to elicit information from the expert panel in order to establish how useful the model is, how it can be improved, and whether it meets the success criteria set out at the beginning of the project. A blank questionnaire is given in Appendix A.

The reliability of the results are dependent on the questionnaire asking the right questions to the right people. Berry and Jeffrey (2000) ran a skirmish test to assess the respondents' perceptions of each question in their development of an instrument for assessing software measurements programs. Using a seven point ordinal scale, they asked respondents to rank the following:

1. How confident are you that you understand this question?
2. To what extend to you have the knowledge to answer this question?
3. How difficult was it to respond in terms of the alternatives given?
4. How relevant is this question to the subject of [in our case: Requirements Process Improvement]?

This line of questioning is only useful if we can use the information to change the questionnaire. As resources (time and people) did not allow for this, we dealt with the 4 points above as follows:

1. The question of understandability was covered in five ways;

Method	Result
Examining responses to the pilot test. (However we did not make a special point of asking participants to note any ambiguity in the question)	The 4/5 researchers who piloted the questionnaire said that they found some of the questions difficult to answer as they did not feel 'qualified' to answer them, not that the

	questions were ambiguous.
Highlighting to the experts in our covering letter that they were not expected to be able to answer (understand) all questions as the questionnaire dealt with two specific areas.	20 completed questionnaires were returned from the 24 experts who agreed to participate. Experts used the 'no opinion/don't know" when unable to answer the question.
Including a "further comments" blank page at the end of the questionnaire.	Three participants used this to explain problems they had in understanding some of the questions. And one participant wrote a separate note as our demographics section was not suitable for listing his qualifications
Creating an online support page with definitions and further model related documentation.	We are uncertain how many participants referred to this page, but from some feedback it is clear that it was used "Your explanation and the materials [on the web] help, thanks".
Providing the participants with a contact name and e-mail address for any queries they might have.	One participant did ask for clarification on one question.

2. The question of knowledge is not so relevant to this assessment as we selected people who we knew had the required knowledge to answer a cross section of the questions.

3. The question of how appropriate the response categories are is difficult to assess without direct reference to each question. However one participant did create a further 'middle/neutral' scale point not given in the original response categories. Our design does not include a neutral point in the middle of a bi-polar scale as we wanted to encourage a positive or a negative response to the question. The neutral scale point was off set to the side of the scale so participants had to opportunity to use this if they did not understand the question, had no opinion or were neutral. In 99.9% of cases this design proved workable, although we cannot comment on the reasons the 4 questionnaires were not returned. We emulated previous questionnaire scales that model developers used with panel of experts to validate their models. Dyba (Dybå, 2000) considers the relative merits of different measurement scales and concludes that a 5 point scale was the most reliable measure. Figure 4 shows how we adapt a Likert 5 point bi-polar scale (Likert, 1932).

Figure 4: The bipolar 5 point scale used to show agreement/disagreement with questionnaire statements:

Strongly disagree (1)	Disagree (2)	Agree (3)	Strongly agree (4)	No opinion /neutral (5)

When using a 2-tailed form of questioning we employ another form of bi-polar scale as shown in Figure 5.

Too few		Correct number		Too many	
		←	→		
(1)	(2)	(3)	(4)	(5)	no opinion
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Figure 5: A two tailed response scale

These bi-polar categories, however, do not suit our second form of 'open' questioning where we want to elicit ordinal measurement details from the participants. For example to find out "How consistent is the level of detail given within the Requirements CMM?" we need an ordinal measure scale as given in Figure 5. We interpret this scale as 1 = 1% (i.e. not at all consistent); 2 = 35% (i.e. below average/unsatisfactory consistency); 3 = 65% (i.e. above average/satisfactory consistency); 4 = 99% (i.e. consistency is very high).

Not at all (1)	(2)	(3)	Very (4)	no opinion/ neutral (5)

Figure 5: The ordinal scale gives a measure of response to open questions

No opinion responses

We attribute no 'value' to the no opinion response, but include the number of responses in our measurement, i.e. N = 5 which has implications when reporting individual results as a percentage of the responses.

(However when questions relate specifically to an area that cannot be answered by some of the participants and we give a second result as follows-

We do not include the number of 'no opinion' responses in the result, so that N= 4. When reporting the percentage of findings for each category this is likely to be a more accurate result. This is because the 'no opinion' category is used by the experts who do not have experience of the area under analysis. We argue that if the participant is not an expert in this area their input should not be included? Or maybe we can't make this assumption?? If we just report the raw/observed figures and don't give % this gets around this problem. DON'T DO THIS?)

5 point scales used

Dyba's in depth study on how many points to use on a rating scale considered that too few scale points results in a coarse answer scale and much information is lost "because the scale does not capture the discriminatory powers that respondents are capable of making." Conversely, using too many scale points the scale can become graded so finely that it is beyond the respondents' powers of discrimination. Research has found that the 5-point scale is more reliable that other scales. This finding is also supported by Likert 1932 (Likert, 1932).

Also I have taken some of the questions El Emam & Jung used (JSS 2001) in validating the ISO/IEC 15504 assessment model (Software Process Improvement Capability dEtermination (SPICE)) trials.(El Emam and Birk, 2000) **(Search out paper again)**

3. Criteria for Success

Appendix D gives a breakdown of the criteria for success against which we test the R-CMM. This criteria was considered and set at the start of model development. The success conditions state that the model must adhere to CMM characteristics; be limited in scope; be consistent; understandable, easy to use; tailorable and verifiable.

The questionnaire was designed specifically to test how closely the R-CMM meets these objectives.

3.1 Questionnaire Results: Testing the Criteria for Success through an analysis of Expert Panel responses

This section gives presents the results of the questionnaire and analyses the significance of these results. To reduce the likelihood of random errors in single items each success criteria is measured by a combination of items. We use the term 'item' in this context to refer to each question and its associated 5-point scale response.

Success Criteria 1: Adherence to SW CMM characteristics

(i) The panel was asked question 15(a) at the end of the questionnaire when they had answered all the questions that relate to the R-CMM in the 'accompanying documentation' booklet. The question comes in the section headed 'Overall impression'.

Q15(a): How well does the new requirements framework retain the SW CMM concept?

Not at all (1)	(2)	(3)	Very well (4)	no opinion (5)
0		7	7	6

70% of participants gave a positive response. 30% had no opinion.

(ii) Question 6(b) comes from the section in the questionnaire that looks at the design of the framework with special reference to the structure and context.

Q 6(a): How appropriate is it to adapt the SW CMM level characteristics to create maturity goals for the Requirements CMM?

Not at all (1)	(2)	(3)	Very (4)	no opinion
0	2	12	5	1

85% of participants gave a positive response to this question

(iii) Question 8(e) examines how well the R-CMM conforms to the SW CMM concept by questioning whether processes in each requirements phase are repeatable

Q8(e): How well do the questions (based on 5 requirements phases) relate to the Level 2 goal?

Not at all (1)	(2)	(3)	Very (4)	no opinion
	2	7	10	1

Question 8(g) asks whether the 5 phases adequately cover all the activities involved in the requirements stage.

Q8(g): How well do questions (Q1 - Q5) cover all the key activities involved in the requirement stage of software development?

Not at all (1)	(2)	(3)	Very (4)	no opinion
1	1	12	6	1

(iv) When looking at the statement in the question 10(g), the participants are advised that processes at Level 2 maturity refer to best practice at a project level and not at an organisation-wide level. (Q10g refers to the list of 20 processes given in question 9.)

Q10(g): Each process relates to Maturity Level 2 (baseline processes)

Strongly disagree	Disagree	agree	Strongly agree	No opinion
0	3	8	2	7

25% disagreed, 50% thought that the processes related to Level 2 characteristics. 35% had no opinion.

(v)

Q10(h): All Key processes are represented (for a baseline process level – the SW CMM claim that their repeatable level contains all key processes needed to establish repeatable processes in software)

Strongly disagree	Disagree	agree	Strongly agree	No opinion	Missing
2	5	9	1	2	1

53% (discounting 'missing' from calculation) gave a positive response. 37% gave a negative response.

(vi) Question 10(a) reflects the CMM criteria that all processes should work together to achieve a level of process maturity.

10(a) All processes work together to achieve requirements process improvement at a repeatable level (all processes are repeatable at level 2 in the SW CMM)

Strongly disagree	Disagree	agree	Strongly agree	No opinion	Missing
	5	11	2	1	1

25% disagreement; 65% agreement

(vii) Questions 10(c) and 10(d) were designed to test participant response to the idea that benefits can be gained by implementing one process without the other processes at one level of maturity. The statement also implies that there is no need to prioritise requirements process implementation. The more negative the response the closer to the CMM concept that processes are not independent of the others.

10 (c) The Requirements process can be improved by implementing the individual processes in any order

Strongly disagree	Disagree	agree	Strongly agree	No opinion	Missing
4	13	0	0	3	0

10(d) Processes can be implemented gradually (one-by-one) as each process is independent of the others

Strongly disagree	Disagree	agree	Strongly agree	No opinion	Missing
4	10	6	0	0	0

No participant strongly agreed with this statement, while 70% disagreed with statement.

STILL TO DO….
(viii)
We extend the adherence of CMM characteristics to ask more globally relevant questions that relate to whether adhering to CMM characteristics is in fact a suitable way to present requirements processes. To assess this we asked the experts what they thought about the current CMM.

Q5 – give a breakdown of results.

a) The CMM clearly defines software process activities
b) The CMM's 5 stage framework helps companies to prioritise process implementation
c) The CMM framework can be tailored to suit a company's specific needs
d) The guidelines in the CMM represent current best practice in software development

(ix)
Q14
(f) This assessment method retains the CMM level concept

(Suitability of adherence of CMM characteristics test?)

3.2 Questionnaire Results Analysis: Testing the Criteria for Success through an analysis of Expert Panel responses

This section analyses the significance of the results given in 3.1.

Success Criteria 1: Adherence to SW CMM characteristics

(i) How well does the new requirements framework retain the SW CMM concept?

11 out of the 20 participants had very little previous knowledge of the SW CMM. Therefore when asked to give an 'overall impression' of the R-CMM, 30% were unable to give an opinion on this conceptual question. This question was given at the end of the questionnaire when participants had been through all the levels of detail presented in the R-CMM. The participants who were able to give an opinion were unanimous in confirming that the new requirements framework retained the SW CMM concept.

(ii) How appropriate is it to adapt the SW CMM level characteristics to create maturity goals for the Requirements CMM?

When answering this question (b), participants were directed to look at Document 1 in the accompanying documentation booklet. It was easier for them to interpret this question than question (a) as they had something concrete to work against. Therefore most had an opinion despite 55% having little or no previous knowledge of the CMM. No-one thought is was totally inappropriate to adapt SW CMM characteristics to create maturity goals for the new model. None of the participants thought it was totally inappropriate to adapt SW CMM level characteristics to create maturity goals for the Requirements CMM. However 10% thought it was only marginally appropriate, with 5% having no opinion.

(iii) There is very strong agreement that the 5 phases relate to the level 2 goal and that the phases represent all the key activities required at a baseline level.

(iv) Each process relates to Maturity Level 2 (baseline processes)

50% agreed that the processes listed in the model related to Level 2 characteristics. 35% had no opinion - this high percentage is likely to be due to many of the participants being unfamiliar with the CMM. The 15% who disagreed with the statement are therefore likely to have a good knowledge of the CMM.

NB: Tracy… Need to work out what an acceptable level is, 50% is probably too low to allow us to confidently state that the processes relate to maturity level 2. For all these analyses we need to test for compliance/acceptance of the statement. I.e. Is there a statistics test that will support levels of confidence? We could do a frequency table of all the answers and then see if there is any polarisation in a corresondence analysis? Here we want to see a low inertia with people agreeing to the statements.

(v) All Key requirements processes are represented (for a baseline process level – the SW CMM claim that their repeatable level contains all key processes needed to establish repeatable processes in software)

The generic quality of the model makes this statement difficult to interpret. The statement produced a wide variety of responses, and although the majority agreed with the statement, with 2 disagreeing strongly and 5 disagreeing, we need to investigate further which processes are believed to be missing, or which processes need to be more defined. Some experts did add some suggestions as to additional processes they would like included at a baseline level, e.g. risk assessment.

This question can be compared with the answers given in question iii above – question iii is a higher level view of the process (looking at the phases), while this question looks at the processes that populate each phase.

(vi) All processes work together to achieve requirements process improvement at a repeatable level (all processes are repeatable at level 2 in the SW CMM)

Processes need to complement each other to achieve a strong basis for process improvement. With 65% agreement with this statement the model is showing some compliance with this need. However, the 25% disagreement shows that there is not a concensus.

(vii)

The strong disagreement with the statements in the two questions covered in this section, suggest that the experts believe that there should be a strict order and control over how the processes are implemented. It would not benefit the process to implement processes in any given order, and there is also a suggestion the processes will not work independently as they are dependent on other processes to work effectively. This result is very much in line with CMM concepts, and we take the disagreement as affirmation of the need for a structured requirements process. For example, had the experts said that requirements processes can be implemented in any order (agree with statement) there would be little need of a model to guide them. Also, if processes could be implemented independently of each other, there would be little need for the co-ordination of effort offered by the R CMM – where achieving a maturity level depends on implementing all key processes.

References

Basili, V. R. and Romach, H. D. 1988. The Tame Project: Towards Improvement-oriented software environments. *IEEE Transactions on Software Engineering*, 14 (6): 758-773.

Beecham, S., Hall, T. and Rainer, A. 2003. Software Process Improvement Problems in 12 Software Companies: An Empirical Analysis. *Empirical Software Engineering*, 8 (1): 7-42.

Berry, M. and Jeffery, R. 2000. An Instrument for Assessing Software Measurement Programs. *Empirical Software Engineering*, 5: 183-200.

Berry, M. and Vandenbroek, M. 2001. A Targeted Assessment of the Software Measurement Process. Seventh International Software Metrics Symposium, April 04 - 06, London, England. IEEE Computer Society

Los Alamos, California. 222-235.

Dybå, T. 2000. An Instrument For Measuring The Key Factors Of Success in Software Process Improvement. *Empirical Software Engineering*, 5 (4): 357-390.

El Emam, K. and Birk, A. 2000. Validating the ISO/IEC 15504 Measure of Software Requirements Analysis Process Capability. *IEEE Transactions on Software Engineering*, 26 (6): 541-566.

El Emam, K. and Madhavji, N. H. 1995. *The Reliability of Measuring Organizational Maturity*: John Wiley & Sons.

El Emam, K. and Madhavji, N. H. 1996. An Instrument for Measuring the Success of the Requirements Engineering Process in Information Systems Development. *Empirical Software Engineering*, 1: 201-240.

Eriksson, D. M. 2003. A framework for the constitution of modelling processes: A Proposition. European Journal of Operational Research, 145: 202-215.

Hall, T., Beecham, S. and Rainer, A. 2002. Requirements Problems in Twelve Companies: An Empirical Analysis. IEE Proceedings for Software, October, 149: No.5: 153-160.

Kitchenham, B., Pfleeger, S. L., McColl, B. and Eagan, S. 2002a. An empirical study of maintenance and development estimation accuracy. *Journal of Systems and Software*, 64 (1): 57-77.

Kitchenham, B. A., Pickard, L., Linkman, S. G. and Jones, P. 2001. A Preliminary Risk-based Software Bidding Model. Keele University, Keele. Technical Report TR0103.

Kitchenham, B. A., Pickard, L., Linkman, S. G. and Jones, P. 2002b. A Framework for Evaluating a Software Bidding Model. Proceedings of the Conference on Empirical Assessment in Software Engineering, 8th - 10th April 2002, Keele University.

Krasner, H. K. C., Pyles, J. S. and Wohlwend, H. S. 1994. A Case History of the Space Shuttle Onboard Systems Project. SEMATECH Technology Transfer, Texas. Technical Report Technology Transfer # 94093551A-TR.

Lauesen, S. and Vinter, O. 2001. Preventing Requirement Defects: An Experiment in Process Improvement. *Requirements Engineering Journal*, 6 (1): 37-50.

Likert, R. 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology*, 22 (140).

Lindland, O. V., Sindre, G. and Solvberg, A. 1994. Understanding Quality in Conceptual Modeling. *IEEE Software*, 11 (2): 42-49.

Paulk, M. C., Weber, C. V., Curtis, B. and Chrissis, M. B. 1995. *The Capability Maturity Model: Guidelines for Improving the Software Process*. Reading, Massachusetts: Addison Wesley Longman, Inc.

Sommerville, I. and Sawyer, P. 1997. *Requirements Engineering A good practice guide*. Chichester: John Wiley & Sons Ltd.

Appendix A: The R-CMM expert panel validation questionnaire

Appendix B: Models used in Validation of the R-CMM : Questionnaires "accompanying documentation"

Appendix C: The Expert Panel

Name of Participant	Current/most recent company	Position/relevant experience
Bangert, A		ITS Consultant
Childs, P	B C Electrical Techniques Ltd	IT Consultant
Fox, D	Clerical Medical Investment Group	IS Project Manager
Homan, D	NORTEL (ex)	Quality Manager
Hough, A	Moneyfacts Group Plc	IT Director
Kujala, Sari (PhD)	Helsinki University of Technology	Senior Researcher (involved in assessment of the REAIMS model)
Kutar, M (PhD)	University of Hertfordshire	Lecturer (expert in Requirements methods)
Maiden, N (PhD)	City University, London	Head of Research Centre
McBride, T (PhD?)	University of Technology, Sydney	Lecturer/trained SPICE assessor, trained ISO 9001 auditor/ ex chairman of NSW Software Quality Assoc, on ISO sub-committee to develop software engineering standards
Nuseibeh, Bashar (PhD)	Open University, Computing Dept	Professor
Anonymous	Insurance Company	IT Business Analyst
Robinson, J (PhD)	Rand, USA	Senior Information Scientist (many years experience as software requirements Engineer)
Sawyer, Pete (PhD)	Lancaster University	Head of Computing Dept, co-author of Text Book on Requirements Engineering
Smith, R	CSE International Ltd	Consultant
Spooner, A	Norwich Union	Web Development Manager/Project manager
Steele, J	BAe Systems	Head of Hardware Engineering
Stephens, M		Senior Information Analyst
Sutcliffe, A (PhD)	UMIST Dept of Computation	Professor
Wilkinson, V	SEMA (ex)	Analyst/Programmer
Wilson, D (PhD)	University of Technology, Sydney	Associate Dean (involved in PPP project)

Appendix D: Criteria for Success

Objective	Purpose	Rule
Adherence to CMM characteristics	The new model should be recognisable as a CMM offshoot By tapping into the SW CMM the requirements model takes the strengths of a proven improvement structure and becomes more accessible and compatible, avoiding redundant activities.	– CMM maturity model levels must be implemented – Each level should have a theme consistent with CMM – Key Requirements processes must be integrated – The model should be recognisable as a CMM offshoot
Limited Scope	By defining the boundaries we set out clearly what the model must include. It is equally important to define what the model excludes to create a useful model that is a simplification of a complex system.	– Activities relating to technical and organisational requirements processes will be included – Processes will be included on a priority basis. – Only processes directly relevant to the R-CMM process areas will be included – Processes will be generic and abstract to allow for individual adaptation
Consistency	SW CMM presentation is used to limit misinterpretation and limit the need to familiarise the user with the new model. The R-CMM has two forms of internal consistency to help guide the user: Depth: Consistent language is used to link high level processes to detailed process guidelines Breadth: Consistent structure is used across levels to ensure that there is a clear progression from one maturity level to the next.	– Language will be consistent with CMM – There will be a consistency in language (with clear links) between models of differing granularity (detail) that are at the same level of maturity. – There will be a consistency in structure between models at the same level of granularity that are modelling different maturity levels.
Understandable	All users of the model should have a common understanding of the model in order to gain a shared understanding of the requirements process and where it needs improvement. There should be no ambiguity in interpretation, especially when goals are set for improvement	– All terms should be clearly defined (i.e. have only one meaning). – All relationships between processes and model architecture should be unambiguous and functional.
Ease of use	Over-complex models are unlikely to be adopted as they require extra resources, and are often too challenging for the user to interpret and follow without extensive training. The model will have differing levels of decomposition starting with the most high level in order to gradually lead the user through from a descriptive model towards a more prescriptive solution	– The model should be decomposed to a level that is simple to understand – The model should be simple yet retain meaning – The chunks of information should clearly relate as they develop into more complex structures – The model should require little or no training to be used
Tailorable	The model must be structured so that it can be extended and tailored to particular development environments	– The structure must be flexible – The structure must be modular – The structure must be transparent
Verifiable	To assess whether a process is useful, well implemented the criteria needs to be verifiable Objectives should not be based on a wish list that cannot be verified. An external validation of the model will help to improve the model and add confidence in its representation.	– The objectives set at the outset of the model development must all be verifiable, i.e. we must be able to ask whether our model has met the objectives set out in this table. – Seek external validation

APPENDIX E: Results of Expert Panel Validation of R-CMM

We compare responses to two questions (8b and 10f) as the contingency tables suggested that they were significantly different.

QUESTION REFERENCE No.	Strongly disagree	Disagree	agree	Strongly agree	No opinion/ neutral
8b) How appropriate is it to include organisational processes (e.g. skills audit) and technical processes (e.g. techniques to trace requirements) in one model?	1	1	5	13	0
10f) Each process relates to requirements engineering activities	3	6	8	3	0

We ran a chi square test to see if there is a significance in the way the two questions are treated.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.514[a]	3	.009
Likelihood Ratio	12.446	3	.006
Linear-by-Linear Association	9.031	1	.003
N of Valid Cases	40		

a. 4 cells (50.0%) have expected count less than 5. The minimum expected count is 2.00.

With χ^2 value of 11.51, df 3 p = .009, we reject the null hypothesis that the two questions are being answered in a similar way (i.e. there is a significant difference in the way the two questions are being answered). However, I am not convinced that a chi square test is what we need to show the difference in how each participant responded to this question, as it takes into account the row differences as well as the column differences. I therefore looked back at the data to pair the responses (as given in Table below) in order to run a paired t-test. (Again, not sure if this is in fact the correct test as it is meant to compare two conditions to show difference). However Kitchenham (JSS 2002) uses the paired t-test of the difference between the 'absolute residuals' for two models to test whether one model was significantly better than the other.

Participant	q8 (b)	q10 (f)	Participant	q8 (b)	q10 (f)
1	3	4	11	4	3
2	4	3	12	4	2
3	2	1	13	4	2
4	4	4	14	4	3
5	4	4	15	4	3
6	1	1	16	3	3
7	4	2	17	3	3
8	4	3	18	4	2
9	4	1	19	3	2
10	4	3	20	3	2

The paired t-test result also confirms that the responses to Question 8b) are significantly different from the responses given to question 10f.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	QUES8B	3.50	20	.827	.185
	QUES10F	2.45	20	1.050	.235

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	QUES8B & QUES10F	20	.333	.151

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	QUES8B - QUES10F	1.05	1.099	.246	.54	1.56	4.273	19	.000

So, can I report the results as follows:

A related 2 tailed test showed the difference in means to be significant, t(19)= 4.27, α = ? Confidence level 95%. The response to question on requirements processes was significantly lower to the response to the question on including the organisational and technical processes in one model.

We interpret this result as showing that a model that combines both organisational and technical processes is appropriate … but that some of these processes are not considered/perceived to be requirements engineering activities.

Appendix F: Model Validation Frameworks

Barbara Kitchenham's paper (Kitchenham *et al.*, 2002b) identifies a set of five quality dimensions that can be used to assess the quality of a model: syntactic quality, semantic quality, pragmatic quality, test quality and value. .Kitchenham equates the methods for achieving quality goals with evaluation actions. Kitchenham's framework is particularly helpful as a guide to how to assess whether the goals have been met.

We have similar goals:

A comparison of evaluation goals and methods:

Stage	Bidding Model Goal	Bidding Model Methods	R-CMM Methods
1	Evaluating how well the model meets its requirements	Comparison of requirements against model components	Questionnaire used to compare model Objectives/success factors against model components
2	Evaluating whether model users accept the model as a reasonable representation of the bidding problem (requirements problem)	Survey of potential users' attitudes to the model. Running the model on example bidding scenarios	Future work: Survey of potential users' attitudes to the model Running the model (tool) on example requirements scenarios
3	Evaluating whether model users find the model improves their understanding of the bidding process and the implication of their estimates and assumptions	Survey of model users experiences	Future work: Survey of model users experiences Survey to include: are model users able to measure whether their understanding of the requirements process is improved, and if so, does this knowledge help improve their overall requirements process? Does an improved requirements process create a better requirements specification? How can we measure successful requirements?

Similar to Kitchenham's first report (Kitchenham *et al.*, 2001) we will attempt to answer only stage 1 of the evaluation process. However we shall learn from the problems they had with their initial ideas that our evaluation methods must be objective.

In the same way as the Bidding Model, our model is also generic and would need specialising before practical application, making 2 and 3 difficult to address in the model's current state. As Kitchenham et al (2002) explain, "since our model is only semi-generic we would not expect an evaluation of the model's value to its users".

(Boehm in Kitchenham, 2002) model evaluation include:

Boehm's (?) model evaluation	(Eriksson, 2003) method	R-CMM	Questionnaire items that test these qualities
Definition: Has the model clearly defined the requirements processes and their relevant maturity levels?	Definition of problems and possibilities within a specified area of interest	List of Requirements problems and key processes within the area of requirements engineering process improvement are defined.	

Fidelity - the model should give consistent results, i.e. a level 2 requirements process should equate to a level 2 Software CMM criteria.			
Constructiveness: Can a user tell why the model relates different processes to different levels of maturity?	Generation of propositions for management of defined problems in the form of a framework of ideas and a methodology	All processes are placed in a structured framework defined by the SW CMM. The framework suggests an order of implementation and assessment of individual requirements processes.	
Detail: Does the model allow for sub processes to be developed from processes: e.g. guidelines?			
Stability: Can the model cope with different user needs, i.e. inputs? – is only superficially tested through expert panel who don't physically use model	Application of the framework to the area of interest using the methodology, and reflection upon that application	Model sections are sent to a group of experts in the area of interest. Examining the model, the panel of experts complete a questionnaire to assess its suitability for the purpose	
Scope: Does the model cover the class of software projects whose 'requirements' you need to assess?			
Ease of Use: Are the model inputs and options easy to understand and apply? I.e. the requirements goals?			
Parsimony: Does the model avoid highly redundant factors or factors which make no contribution to improved requirement processes?			

Kitchenham 2002, uses Lindlands framework (Lindland *et al.*, 1994) to model quality goals and how to achieve them. Since Lindland based his framework on Requirements models, we have studied his framework before adapting Kitchenham's version.

Kitchenham (2002) links Lindland and Boehms evaluation framework - "we have attempted to retain the conceptual simplicity of Lindland et al's framework, while incorporating additional elements that are required for the purpose of evaluating our bidding model. She states that "Although the revised model evaluation framework was developed with reference to our bidding model, most elements of the framework are relevant to any decision and risk management models represented as influence diagrams where the model elements and input values are based on expert opinion".

In our evaluation of the model we also consider the syntactic quality, semantic quality, pragmatic undertanadability and potential value of the model as shown in the table below:
A different perspective on quality aspects to test for than that taken by errikson and boehm. There is clearly some overlap in areas such as fidely, definitions
Kitchenham's Revised model evaluation Framework (Kitchenham *et al.*, 2002b)

Quality aspect	Goal	Model Properties	Means
Syntactic quality	Syntactic Correctness	Defined syntax	Manual checking of the model – the diagrams must include only factors and links between factors and there must be no circular links. **There should be a clear path from goal through to process**
Semantic quality	Feasible validity Feasible completeness	Traceability to domain	Inspect the model to check that the model includes: Definitions Detail information Scope information Sensitivity analysis – look to see if there are any unnecessary features Consistency checking – aimed at ensuring the model is internally consistent
Pragmatic quality	Feasible Comprehension		Means to enable comprehension – this is achieved through linked information option (i.e. through referencing supporting data)
	Feasible Understandability	Structuredness Expressive economy	Documentation guidelines and standards covering format and content. Contents standards should define elements that must be included in model documentation such as The knowledge the user is expected to have The purpose of the model Explanation of the input required for assessment Explanation of the output of the assessment
Test quality	Feasible test coverage	Executability	Requirements process fidelity can be tested by a company that has already achieved level 2 process maturity. (test for consistency) Test for different scenarios, different application areas, different company sizes, different maturity levels for sensitivity and stability of model.
Value	Practical Utility		Means to enable model use including appropriate user interface design, user manuals and training. Means to evaluate model value for example empirical study of model users view of using the model (experiment, interviews or self administered questionnaire)

1. Syntactic quality:

Model development is necessarily the construction of knowledge and the Requirements Capability Maturity Model (R-CMM) attempts to guide actions, while taking account of modelling approaches that have been empirically validated (Eriksson, 2003). In Figure XX, we show how we have used Eriksson's generic modelling framework to explain the purpose, structure and behaviour of the R-CMM. We incorporate this with Boehm's model evaluation criteria.

Erikson's framework gives an order to model development by focussing on the purpose, behaviour and structure of the model. Development of the R-CMM includes activities 1, 2 and part of 3 and 4. 3 and 4 are only partially completed as the model has not been fully tested by companies producing software. However the experts have looked at the model and answered questions that address the problem area, and in this report we compare the results (as in 4) with our original defined success criteria.

Our methodology is similar to that of Eriksson (2003), as shown in table below

Activities	Erikson method	Our methodology – applied
1	Definition of problems and possibilities within a specified area of interest	List of Requirements problems and key processes within the area of requirements engineering process improvement are defined. (The Purpose of the Model is tested)
2	Generation of propositions for management of defined problems in the form of a framework of ideas and a methodology	All processes are placed in a structured framework defined by the SW CMM. The framework suggests an order of implementation and assessment of individual requirements processes. (The Structure of the model is tested)
3	Application of the framework to the area of interest using the methodology, and reflection upon that application	Model sections are sent to a group of experts in the area of interest. Examining the model, the panel of experts complete a questionnaire to assess its suitability for the purpose (behaviour is tested)
4	Compilation of the results thereby attained and comparison with the problems and possibilities initially defined	The results from the validation stage (3) are evaluated and findings are fed back into the original model design for refinement (results are analysed)

Table 1: …The approach, structure and behaviour of modelling is validated

Interpretation

"Those occupied with practical model construction are likely to ask themselves the following question: *How can we make sense of contributions to the field of model construction so as to ensure they are properly understood and successfully used in practical contexts*?"(Eriksson, 2003) A comment made by an expert who validated the requirements R-CMM concurs "The biggest problem with any of these models is interpretation, if the model can be interpreted differently it will be… The SW CMM is very concise but interpretation still plays a big part".