A METHODOLOGY FOR MEASURING THE RISK ASSOCIATED WITH A SOFTWARE REQUIREMENTS SPECIFICATION

Trevor T. Moores and Ralph E. M. Champion
Department of Information Systems
City University of Hong Kong
Kowloon Tong, HONG KONG.

ABSTRACT

This paper presents a six-step metrics-based methodology for assessing the risks associated with and hence the resources required to implement the requirements contained within a software requirements specification (SRS). The method seeks to eliminate the use of subjective probability assessments in models of risk exposure (RE) and risk reduction leverage (RRL). Measurements are taken of the number of requirements and the class of risk, the number of change requests and their date of issue, and the cost of each requirement change. The class of requirements risk is tailored to a given organisation using the Delphi method. The information collected is stored as an historical database for use in the analysis of subsequent SRSs.

INTRODUCTION

Identifying and controlling the risk associated with the development of an information system is seen as critical for the management of a software development project (Alter, 1979; McFarlan & McKenney, 1983; Charette, 1989; Chittister & Haimes, 1993). "Risk" is typically defined as 'the possibility of incurring misfortune or loss'. In software development terms, this usually implies the potential for some disruptive event to impinge upon the project. The key problem with risk is that it cannot be avoided, it can only be anticipated and its effects reduced. For a programmer this may result in the need to redo work. For a project manager this may result in a failure to deliver on-time and on-budget. For an organisation this may result in a failure to carry out its business (Tate, 1988; Haimes, 1991).

While a number of studies have been carried out to identify the range of factors which represent risk to a software development project (e.g., Anderson & Narasimhan, 1979; Zmud, 1980; Death, 1983; Charette, 1990), the traditional use of probability or uncertainty as a measure of risk has meant that a quantitative assessment of risk exposure (RE) and risk reduction leverage (RRL) relies on a subjective estimate of the associated probability based on experience. It is well known, however, that humans subjects are notoriously bad at reasoning using probabilities (Kahneman & Tversky, 1982).

This paper will develop a six-step metrics-based methodology which allows the subjective probability ratings used in risk assessment models to be replaced by objective (empirical) measurements. The focus here will be on the verification of IS development (assessing the evolution of the system) rather than on validation (assessing the IS solution). Since the cost of reworking software requirements is more than 100 times cheaper than the cost of making changes at the operations stage (Boehm, 1981), the most cost-effective point at which to apply the risk management methodology is at the requirements stage. Hence, the methodology will focus on the applicability of risk analysis at this stage. The next section will discuss the RE and RRL models used to assess the potential impact of risk on a software development project. The applicability of software metrics to a software requirements document will then be discussed before providing a detailed description of the techniques required to implement such a methodology within an organisation. Areas of further research will then be identified, in particular, the possibility of developing a knowledge-based front-end to the methodology.

ASSESSING AND MANAGING RISK

Given the definition of risk as the potential for an event to occur, the control of risk is a two-step process involving: the assessment of which risk factors are likely to be important; and, the management of risk by measuring and prioritising the probability of the event and its impact on the project. According to Kaplan & Garrick (1981), risk assessment asks three questions for each phase of the development process:

1. What can go wrong?
2. What is the likelihood that it will go wrong?
3. What would the consequences be?

These studies have typically been carried out by asking project managers or users to rate (an existing list of factors) or specify (where no prior model is in place) the importance of a list of potential risks. Factor analysis allows the subsequent list to be condensed and grouped under general headings. For instance, Boehm (1989, 1991) identifies four major headings of systems failure which set out what can go wrong, namely: hardware problems, software problems, organisational problems, and human problems.
Boehm suggests that the full list of problems (see Table 1) can be used as a checklist by a project manager to identify the most likely risks. Boehm defines risk exposure (RE) and risk reduction leverage (RRL) as the key measures, where:

\[
RE = \text{Prob(UO)} \times \text{Loss(UO)}
\]

\[
\text{RRL} = \frac{(\text{RE}_{before} - \text{RE}_{after})}{\text{Risk Reduction Cost}}
\]

### Table 1. The 10 most important risk factors by category

(Boehm, 1989, 1991)

<table>
<thead>
<tr>
<th>Category</th>
<th>Risk factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>1. Personnel shortfalls</td>
</tr>
<tr>
<td></td>
<td>2. Unrealistic schedule and budgets</td>
</tr>
<tr>
<td>Resources</td>
<td>3. Developing the wrong software functions</td>
</tr>
<tr>
<td></td>
<td>4. Developing the wrong user interface</td>
</tr>
<tr>
<td></td>
<td>5. Gold-plating</td>
</tr>
<tr>
<td></td>
<td>6. Continuing stream of requirements changes</td>
</tr>
<tr>
<td>Requirements</td>
<td>7. Shortfalls in externally developed software</td>
</tr>
<tr>
<td></td>
<td>8. Shortfalls in external personnel</td>
</tr>
<tr>
<td>External</td>
<td>9. Shortfalls in real-time performance</td>
</tr>
<tr>
<td>Technology</td>
<td>10. Straining at the bounds</td>
</tr>
</tbody>
</table>

**Figure 1. Risk exposure (RE) contours**

Risk exposure is the product of the probability of some unsatisfactory outcome, Prob(UO), and the loss due to this outcome incurred by the project, Loss(UO). The project manager would need to calculate RE for each identified risk and decide a level of acceptable risk exposure. For factors which exceed the acceptable level of RE, the project manager would then have to either reduce the probability of the event occurring, or reduce the loss associated with it (see Figure 1).

For instance (from Boehm, 1989, p8), if the loss associated with having a particular type of interface error is estimated at US$1 million and from experience we estimate that the probability of such a fault - on a scale of 0 to 1.00 - is 0.30, then the risk exposure is given as:

\[
RE = (0.30) \times \text{US$1000K}
\]

\[
= \text{US$300K}
\]
An approach to dealing with this problem could entail buying a requirements and design interface checker for US$20K which would reduce the Prob(UO) to 0.1, or, carrying out a detailed interface test for an additional US$150K of manpower which would reduce the Prob(UO) to 0.05. The RRL of these two approaches (denoted as RRL1 and RRL2, respectively) can be calculated as follows:

\[
RRL_1 = \frac{([0.30]\times US$1000K - [0.10]\times US$1000K)}{US$20K} = 10
\]

\[
RRL_2 = \frac{([0.30]\times US$1000K - [0.05]\times US$1000K)}{US$150K} = 1.67
\]

Therefore, the interface checker approach can be seen to deliver the highest risk reduction. Although Boehm’s approach may seem reasonable, care must be taken in interpreting these figures. Firstly, although the models of RE and RRL appear to be objective they are not, since although the financial impact of the risk reduction may be accurately calculated, the crucial assessment of the probability of risk is subjectively derived. In other words, the inputs to these models are best guesses by the project manager and so, the risk associated with the guess may in itself account for the difference in result between RRL1 and RRL2. Furthermore, cheap alternatives to the above two approaches which do little to reduce the risk can appear to provide better RRL scores. For instance, employing a US$500-a-day consultant for one day may not reduce the risk a great deal, say the Prob(UO) becomes 0.29, but the RRL score, RRL3, calculates as follows:

\[
RRL_3 = \frac{([0.30]\times US$1000K - [0.29]\times US$1000K)}{US$0.50K} = 20
\]

Neither can it be assumed that the consultant can reduce the risk exposure by 0.01 every day, since most errors are typically found early in the testing phase. Given that RRL is simply a ratio, it is unlikely that the project manager will be happy with option 3, since the risk to the project remains at 0.29 and it is only the cheapness of the risk reduction cost that suggests option 3 is a better solution than options 1 or 2. The problem with RE and RRL is the use of “probability” in the model. Although risk is sometimes defined in terms of “uncertainty” rather than probabilities (e.g., Barki et al, 1993), the fact remains that subjective ratings are inadequate measures of risk. Principally, the calculation of a probability is based on an assessment of the number of choices available within an event. The difference between probabilistic and empirical ratings can be illustrated with the following example: Suppose a coin is flipped ten times and it came up heads nine times and tails once. According to probability, the chance of the next flip producing a head, \( P(h)_{\text{prob}} \), remains 0.5 since it must be one of two equally likely outcomes; however, the empirical result suggests that \( P(h)_{\text{emp}} = 0.9 \) since a head has been observed on nine of ten trials. Over a long period of time, it would be expected that \( P(h)_{\text{prob}} = P(h)_{\text{emp}} \) but this could only be derived from empirical observation. In other words, \( P(h)_{\text{emp}} \) is a surrogate for \( P(h)_{\text{prob}} \) when the theoretical basis for calculating \( P(h)_{\text{prob}} \) is not available.

Software development is sufficiently complex for it to be assumed that \( P(h)_{\text{prob}} \) is not available and any models making use of probabilities - such as RE and RRL - must be revised. Rather than the use of probabilities, therefore, there is a need to apply objective measurements of the risks involved.

**METRICS-BASED RISK ASSESSMENT**

A software metric is a measurement of some attribute of the software process or products that relates directly or indirectly to some property of the system that needs to be assessed or predicted. For instance, metrics have been developed to measure the attribute “size” of a software system in order to assess properties such as the effort required to implement the system (e.g., Halstead, 1977; Albrecht & Gaffney, 1983), and predict properties such as the cost of the project as a whole (e.g., Putnam, 1978; Freiman & Park, 1979; Boehm, 1981). The general methodology for collecting metrics in order to quantify an attribute of the software development process is a six-step process, as follows:

1. Identify which attribute needs to be assessed (or predicted).
2. Identify those properties which are to be used as a measure of the attribute identified in ‘1’.
3. Collect the measures defined in ‘2’ until a database of sufficient size has been created.
4. Quantify the relationship between the properties ‘1’ and ‘2’.
5. Apply the model on the metrics database to test its accuracy.
6. If the model works sufficiently well, use the model to estimate '1' on a live project; otherwise, start again at step ‘2’.

This sequence of steps sets up a project-metrics-model-tool cycle (see Figure 2), where the tool is evaluated against real world projects, and the data collected (now including the estimate) is used to update and refine the models.

**Figure 2. The project→metrics→model→tool cycle**

---

### Defining a set of measurements

Steps ‘1’ and ‘2’ are the key, since only with a sensible definition of the attribute to be assessed and the measurements required can useful metrics models be derived. Given that software requirements specification (SRSs) are more structured than the initial user-defined requirements document, the methodology will be applied to the SRS rather than the user document. The application of software metrics requires both a purpose and an intuitive link between the attribute to be measured and the metric used to measure the effect of the attribute (Basili & Rombach, 1988; Fenton, 1994). In this case, the attribute to be measured is risk; the purpose is to quantify models of RE and RRL; while, the property which is meant to have an intuitive link to risk is the changeability of the requirements and the cost of change. While the model of RRL remains the same, the model of RE becomes:

\[
RE = \text{Risk}_\text{change} \times \text{Cost}_\text{change}
\]

In other words, the probabilistic Prob(UO) factor is replaced by a quantifiable measure of the risk associated with a change to a requirement, Riskchange, and the cost needed to accommodate this change, Costchange. Both Riskchange and Costchange can be quantified if the changeability of any given requirement and the cost of the change can be measured.
Figure 3. A representation of the requirements change process

The measurements defined here are based on the model given in Figure 3, which puts a boundary around the process under study. As can be seen, the implementation of a requirement is represented as a transformation from its written statement in the software requirements specification (SRS), to its implementation in the operational system. Two parties are involved in this process: the users who decide on the requirements; and, the developers who implement each of the required functions. A change to a requirement is initiated by a change request from the user and in response, the developers make some change to its implementation. It should be noted that the change may include cancellation, and the cost of the change in this case would be the effort already expended. In order to quantify this process, the measurements required are as follows:

- the number of requirements (actual number and category type);
- the number of change requests (actual number with date of issue);
- the cost of each change request (in the relevant currency, e.g., US$).

Counting the actual number of requirements depends principally on the way in which the requirements document is written. On the assumption that "simplest is best", the suggestion here is to take high-level counts of the number of requirements, such as the number of functions defined within a form-based approach (Heninger, 1980).

The ability to derive useful categories of requirements risk is clearly crucial. Boehm (1989) derived his five categories (see again Table 1) based simply on experience of developing large-scale software systems. The approach here, however, is to derive the categories more rigorously using the Delphi method (Delbecq et al., 1975). The Delphi method is often used to generate agreement on such things as MIS issues seen as being important for IS executives (e.g., see Niederman et al., 1991; Wang, 1994). The objective of the technique is to discuss drafts of a list of issues in order to surface new issues and move participants towards a consensus. Starting with a base-line questionnaire, participants are asked to rate the issues, typically on a scale of 1 to 10 with 1 being the lowest and 10 being the highest. Succeeding questionnaires summarise group responses and the participants are asked to re-evaluate their opinions based on the new evidence. A three-round Delphi method is common and stops after three rounds of questionnaire, although it is possible for the procedure to continue until a reasonable level of consensus is reached.
As an example, in a survey of IS managers in large UK corporations and computing companies (Moores, 1993), one of the questions asked was: "What are the major reasons for projects failing to meet user requirements?" A ranked list of responses is given in Table 2. Clearly, this particular list would need further analysis since it contains both symptoms and causes of requirements risk. The symptoms listed include a general lack of understanding, requirements being too changeable and too speculative, while the potential causes listed include poor/conflicting user participation, insufficient time for analysis, uncontrolled changes and inadequate testing. This list can be used as the baseline Delphi questionnaire, however. Subsequent questionnaires are then phrased so as to focus participants towards specific causes rather than symptoms of requirements risk.

The list can be tailored for a particular organisation by asking project managers to rate each of the reasons given in the baseline questionnaire or suggest their own additional reasons. Once consensus has been reached, factor analysis based on the rating scores can help to identify the underlying concepts being used. The final list is then the range of categories which are to be used to classify requirements risks. For instance, data would be collected on the number of requirements changes that were identified as being due to *Insufficient time for analysis*. The advantage of this approach over other studies is that it takes into account the idiosyncrasies of the particular organisation, a factor seen as important in developing some types of local metrics models (Kitchenham, 1992).

Counting the number of change requests and the cost of each change depends on their being an adequate change management process in place. The count of change requests may be taken to be the number of Change Request Forms (CRFs) which were actioned, that is, actions were initiated in response to the request. CRFs tend to be simple forms (e.g., see Sommerville, 1992, pp555-8) documenting problems within the system. By only counting those CRFs which were actioned, CRFs that report the same or trivial problems can be ignored. Before a CRF is actioned, the change manager needs to analyse the cost of the change and it's impact on the system (Pressman, 1992). Within this process the cost of the change is immediately available as the cost of the rework scheduled to deal with the change request. Since change requests are possible at any point during the development process, and as Boehm (1981) points out the cost of any change increases down the life-cycle, both Riskchange and Costchange are likely to have different values depending on when the change occurs. By recording the date at which the change request is received, this issue can be taken into account.

**Building the risk models**

The final steps of the metrics-based methodology (steps '3' to '6') are carried out by using the measures defined above to quantify the following relationships:

\[
\begin{align*}
\text{Req}_X & = \text{The total number of requirements assigned to risk category } X \\
\text{Change}_X & = \text{The total number of Req}_X \text{ that required a change} \\
\text{Cost}_X & = \text{The total cost of Change}_X \\
\text{Risk}_{\text{change}} & = \text{Change}_X/\text{Req}_X \\
\text{Cost}_{\text{change}} & = \text{Cost}_X/\text{Change}_X
\end{align*}
\]
Figure 4. Calculating Risk\textsubscript{change} for four categories of requirements risk

<table>
<thead>
<tr>
<th>Project No.</th>
<th>Requirement risk category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
</tr>
</tbody>
</table>

\( X = \text{Requirement assigned to risk category X} \)

With measurements taken across a number of projects, the risk exposure can be calculated for each category of requirements risk. For example, consider that measurements are taken across three projects using four broad categories of requirements risk (see Figure 4). With five instances of requirements placed in category 1 and only one instance of a change, Risk\textsubscript{change} is calculated as 0.20. Similarly, Risk\textsubscript{change} for category 3 is calculated as 0.66, while categories 2 and 4 have no recorded changes and so Risk\textsubscript{change}=0, Cost\textsubscript{change}=0 and RE=0. In other words, the perceived risk has not been found to entail any actual cost. Cost\textsubscript{change} for categories 1 and 3 is the average value of Change\textsubscript{XY} and this completes the RE calculation. Each category of risk would be derived using the method described above.

The accuracy of a metrics model (step ‘5’) is typically assessed as a function of both the general predictive ability of the model, PRED, and the size of error attributable to each estimate, MMRE (Conte et al, 1986). A metrics model is said to be accurate if the estimate is within 25% of the actual 75% of the time, given as PRED(0.25)>0.75, and the mean magnitude of the relative error is less than 0.25, given as MMRE<0.25, where,

\[
\text{MMRE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|\text{Actual}_i - \text{Estimate}_i|}{\text{Actual}_i} \right)
\]

and,

\[
\text{Actual}_i = \text{actual value for the } i^{th} \text{ data point}
\]

\[
\text{Estimate}_i = \text{estimated value for the } i^{th} \text{ data point}
\]

This definition has been taken on board and widely used although the value of ±25% is clearly an arbitrary threshold. For instance, without formally defining PRED as a relationship, Boehm (1981) took it that a model could be deemed to be accurate if it produced estimates within 20% of the actual 70% of the time, that is, PRED(0.20)>0.70, while other studies have required a higher level of PRED at PRED(0.20)>0.80 (Moore & Edwards, 1992). Given the accuracy of risk models have not been properly assessed to-date a weaker level of accuracy is set initially, using PRED(0.25) ≥0.70 and MMRE≤0.25. As more data becomes available and the model is fine-tuned, the accuracy of the models will increase over time.
Having completed the above steps, the risk models derived using this methodology can now be employed on live projects. The person responsible for reviewing the SRS first decides which requirements are subject to some level of risk. For those requirements where a risk category applies, the risk exposure is calculated. For instance, if a requirement dealing with checking valid data entry is known to have been specified without much discussion with the user, then the requirement might bring to the project a risk under the heading of Poor/conflicting user participation. If this category of risk is recorded as effecting changes in 1-in-4 projects with an average cost of US$10K, then the RE can be calculated as:

\[ RE = \text{Riskchange} \times \text{Costchange} \]
\[ = 0.25 \times \text{US$10K} \]
\[ = \text{US$2.5K} \]

An SRS with twenty similar requirements would result in \( RE=\text{US$50K} \). It is at this point that either the requirements would have to be investigated further so as to eliminate the risk, or instigate strategies of risk management (using RRL) in order to reduce the level of RE to one acceptable to the project manager.

CONCLUSION

This paper has sought to replace the probabilistic models of risk exposure and risk reduction leverage (RE and RRL, respectively) with a 6-step metrics-based methodology which quantifies the risk of change and cost of change to requirements (Riskchange and Costchange, respectively). The metrics required to quantify risk and the techniques required for their collection have been described in detail. The key to the methodology is that subjective ratings of the probability of a particular risk impacting on a project is replaced by empirically derived weights. These weights accurately reflect the number of times a class of risk has actually affected a project. The classes of requirements risk are derived for a given organisation using the Delphi method. A further advantage of the metrics database is that data would be provided on the most common risks. Such information would enable organisations to investigate reasons for their occurrence and remedies for the problem. As with most research, a number of issues require further investigation. Firstly, the methodology described above suggests the RE for a project is the sum of the RE for each risk. Clearly, there will be a certain amount of interdependence between risk factors. For example, when considering general project risk, Boehm (1989) has noted that Personnel shortfalls and Unrealistic schedule and budgets are two of the most important risk factors (see again Table 1). However, any calculation of schedules and budgets would take into account the efficiency of the development team, and it is only by knowing that they are of below average quality that the schedules and budgets can be judged to be “unrealistic.” Care must be exercised in using the Delphi method and factor analysis to derive a meaningful set of risk categories, therefore, since it is this list which drives the collection of data and allows the appropriate Riskchange and Costchange to be assigned.

Finally, although the metrics-based methodology has removed one potential flaw in the use of risk models - the use of subjective risk probability assessments - there is still an element of skill involved in risk assessment and management. Primarily, it is still up to the project manager to recognise that a requirement contains a certain class of risk. Furthermore, even if the RE is established there will almost certainly be a major difference in how changes to requirements are dealt with. One project manager may postpone delivery of some functionality in order to meet deadlines, whilst another may negotiate extra funding and deliver the system the customer needs. Clearly, skill in terms of negotiation has an impact here.
The use of skill opens up the possibility of developing a knowledge-based front-end to the methodology described in this paper. Such a front-end would help to further remove subjectivity from the assessment or risk by making operational the heuristics in use. The metrics database would provide the information upon which to test the veracity of the heuristics, while the heuristics would help to clarify the data needed to assess and manage risk. The authors are currently attempting to implement such a research programme.

REFERENCES


