This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

# Factor Reduction and Clustering For Operational Risk In Software Development

## ABSTRACT

Software development failures frequently emerge as a result of the failure to understand and identify risks. The aim of this paper is to identify the most salient risk factors during a software development project lifecycle, in terms of occurrence likelihood and impacts on cost overrun. A questionnaire survey was circulated to 2000 software development companies, IT consultancy and management companies, and web development companies in the UK, USA, Europe, India, China, Japan, Canada, Australia and Asian countries. This asked respondents to evaluate a number of risk factors. However, many factors were closely related and so we apply a factor reduction and clustering process to allow a smaller number of crucial risk factors to be identified. The three main clusters of risk factors identified in this study are 'feasibility study', 'project team management', and 'technology requirements'. While 'feasibility study' may be unlikely to occur it can have significant impact on outcomes; 'project team management' is likely to occur but has relatively little impact on outcomes in comparison to 'technology requirements'. Professionals will need to carefully check and balance these factors and generate a risk mitigation plan to reduce the severity of the project failures. These results allow them to connect the probability of occurrence and overall impact to focus their scarce resources on reducing the most pertinent risks in their project.

**Keywords:** factor reduction, factor clustering, software development risk, risk occurrence likelihood, cost overrun

## 1.0 INTRODUCTION

Software projects are dynamic and tend to have volatile requirements making them difficult to manage and control as the project scope can change frequently. Failure to understand and to identify risks leads to software development failures (Stahl & Bosch, 2014; Reyes et al., 2011). Literature on how to manage a development project often refers to cost, time, and quality as the key project success criteria; however, there are also many other different, broad, and overlapping definitions of project success and failures (Portillo-Rodríguez et al., 2014; Jorgensen, 2010; Baccarini et al., 2004; Linberg, 1999; Ropponen, 2000). Various researchers observe that in practice it may be very difficult to claim the project was really successful or a failure (Martín & Yelmo, 2014). A single project can be considered successful by one stakeholder and failure by another (Naquin & Tynan, 2003; Thüm et al., 2014).

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

The probability of occurrence and impact of software risks on project performance (Han & Huang, 2007) has been examined to understand how different methods influence software practitioners' risk perception and decision making processes. Specific context have been studied, including risks in e-commerce development (Ngai & Wat, 2005); software project performance to establish risk dimensions and developed an exploratory model (Wallace et al., 2004). An authoritative list of common risk factors in general software projects was developed by Schmidt et al. (2001); Keil et al. (2002) used a Delphi study to reconcile users' and project managers' perceptions on IT project risks.

IT literature has produced a number of conceptual frameworks to explain different types of software development risk, risk management strategies and measures of software project performance (Dingsoyr et al., 2012; Nidumolu, 1996; Wallace & Keil, 2004). Many studies suggest that failure to manage risks causes common problems such as cost overruns, unsuitability for intended task, unmet user requirements and schedule overruns (Appari & Benaroch, 2010; Capilla et al., 2014; Magdaleno et al., 2012).

Clearly, software development projects can present serious risks to the well-being of an organization (Marina et al., 2014). Various risk checklists and frameworks have been proposed but the underlying dimensions of the risks in each stages of the development life cycle of a software project, their likelihood occurrence, and the impact on the cost overrun are poorly understood. This is particularly true when examined from the perspective of professionals within development teams. However, managing risks effectively requires a small subset of risks to be identified so resources can be appropriately focused. This research seeks to improve risk management, by examining a range of risks and attempt to present a list of underlying factors that influence risk, based on the perspectives of professionals. Therefore, we seek empirically grounded factors that can progress research in this area. From this, we connect the identified risk factors to the software development project lifecycle, in terms of the likelihood of occurrence and impacts on cost overrun. This can help professionals to identify and manage the most important risks at different points in the project. To do this we first discuss the concepts associated with factor reduction before we discuss how we administered the questionnaire. Then, we discuss the analysis process. Finally, we connect the identified clusters to the software development lifecycle.

## 2.0 FACTOR REDUCTION AND CLUSTERING

Factor reduction and clustering is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of variables (Morgan et al., 2004; Punch, 2005). The presence of many variables often makes it difficult to be systematically understand data and patterns. The factor reduction and clustering technique can reduce the number of variables without losing too much of the information the original variables provide (Field, 2005).

Multicollinearity can be a problem in multiple regressions with a lot of variables; factor reduction and clustering could solve this problem by combining variables that are collinear (Field, 2005). Data reduction is achieved by looking for variables that correlate highly with a group of other variables, but do not correlate with variables outside that group. This technique examines variables clustered together in a meaningful way. When two or more variables are correlated, an underlying common factor can be proposed which these variables share and which explains the correlations between them (Punch, 2005).

Factor reduction and clustering can be used either in hypothesis testing or in searching for constructs within a group of variables for a more easily understood framework. The process begins by finding a linear combination of variables that accounts for as much variation in the original variables (Morgan, et al., 2004). It then finds another component that accounts for as much of the remaining variation as possible and it is uncorrelated with the previous component. The process cycles and 'rotation' continues in this way until there are as many components as original variables. Usually, a few components will account for most of the variation, and these components can be used to replace the original variables (Punch, 2005). By reducing a data set from a group of interrelated variables into a smaller set of factors, data clustering achieves parsimony by explaining the maximum amount of common variance in a correlation matrix using the smallest number of explanatory concepts (Field, 2005).

## 3.0 RESEARCH PROCEDURES AND PROCESS

A questionnaire survey was employed among software practitioners. There were 2000 questionnaires circulated electronically to software development companies, IT consultancy and management companies, and web development companies in the UK, USA, Europe, India, China, Japan, Canada, Australia and Asian countries. Respondents were approached across a variety of IT job functions and management hierarchy with varying amount of experience. The sample size of this questionnaire survey is 2000 and the respondents were 324 (16.2%). The percentage of overrun stated in the scale was to quantify the magnitude of the scale. Most respondents had more than 10 years experience in software development practice with the mean value at 11.8 years and a standard deviation at 5.29. The average number of software projects undertaken was 9.2 with a standard deviation at 5.31. Statistical Package for the Social Science (SPSS) and Microsoft Excel were employed for data analysis of the 104 risk factors involved in this survey.

### 3.1 Questionnaire design

Questionnaires are an effective technique for statistical data and opinion collection (Parasumraman, 1991; Burns, 2000) and enable data collection over different geographical regions. The draft questionnaire was created based on risk factors identified and sourced from the relevant literatures. Other risk factors were identified based on experience and conversations with other researchers. This draft was shared with academic staff, professional colleagues, fellow researchers with a request for suggestions in terms of wordings, sequence of questions, layout of the questionnaire and contents. These suggestions led to changes in wording, sequencing of questions, and overall layout. Next, the questionnaire was piloted by sending the questionnaires to experienced academics and researchers. This further feedback led to adjustments to improve the comprehensibility of some elements.

The questionnaire consists of three parts. Part 1 collects the respondents' background and demographic data. Part 2 lists the selected risk factors and the respondents need to assess and rate the degree of significance of each risk factor using the 5-point scale. Part 3 asks for respondents' opinions in relation to the effective risk mitigation strategies. We capture the respondents' opinions relating to project failures rather than objective measurements of project failures as many respondents would be unwilling to candidly share objective measures of failure with us.

Our respondents were requested to rate each risk factor using a scales in a rating approach. This likert-type scale enabled us to capture the respondents' opinions and experience relating to certain statements. This approach is consistent with other research conducted in this area of study (for example: Jiang & Klien, 1999; Ropponen & Lyytinen, 2000; Cule et al, 2000; Schmidt et al, 2001; Wat & Ngai, 2005; Procaccino et al, 2005; Perera et al,2006; and Eun Hee Kim, 2006; *inter alia*).

In Part I of the questionnaire, respondents' demographic data and general information were collected and summarized in Table 1.

In Part II, respondents were provided with the list of 104 risk factors relating to software projects, categorised into the six stages of project management development lifecycle. The respondents need to assess the likelihood of the occurrence of each risk factors on a scale of 1 to 5 (1-No occurrence; 2-Unlikely; 3-Likely; 4-Highly likely; and, 5-Very highly likely). The respondents were asked to rate the likely impact of each of the 104 risk factors on cost overruns as a percentage of the original estimate, also on a scale of 1-5 (1-Very low (1-10% overrun); 2-Low (11-20% overrun); 3-Moderate (21-30% overrun); 4-High (31-40% overrun); and, 5-Very high (>40% overrun)).

In Part III, we examined the practitioners' perceptions of the effectiveness of 30 risk mitigation strategies in reducing the risk generally, without specifying their relationship to the risk likelihood or impact on cost overrun. This enables correlation analysis of factors scores and extraction of components from risk likelihood and impact on cost overrun to these mitigating strategies. Respondents were asked to rate the effectiveness of the risk strategies in response to the risk factors using the scale of 0-6 (0 - don't know; 1 – not effective strategy; 2 – very slightly effective strategy; 3 – generally effective strategy; 4 – highly effective strategy; 5 – very highly effective strategy; and, 6 – exceptionally effective strategy)

### 3.2 Sampling

Three main category of businesses related to IT/IS were chosen: Software development companies, IT consultancy and management companies, and Web development companies. Due to the diversity of services that other IT/IS related companies offer, we opted to include these general IT/IS related companies as well.

Project managers and users tend to identify and rank highly risks that are perceived to be outside their own control (Bannerman, 2008). That is, they tend to identify risks in the responsibility domains of others, rather than point to factors as risks within their own areas of responsibility (March and Shapira, 1987; Schmidt et al., 2001). It is crucial that the views of all key software practitioners groups are taken into account in the risk identification and management process. Therefore, we sought respondents from within the development team itself (e.g., Project managers, Software Developers/programmers, and IT technical support staff) and also the Managing Directors/Board of Directors. The Managing Directors/ Board of Directors perceptions was perceived to be relevant as the risk factors within the project could have significant impacts on wider business operations. IT technical support perceptions were important as the staff have a direct involvement with the users during implementation and providing technical support when the software is up and running.

Past research shows questionnaire responses dominated by firms from the USA and the UK. We aimed to replicate these results and undertook an electronic search of appropriate firms before we collected their contact and email addresses from the internet. The questionnaire was circulated electronically to software development companies, IT consultancy and management companies and web development companies in the UK, USA, Europe, India, China, Japan, Canada, Australia, and some other Asian.. 2000 questionnaires were circulated and a total of 324 valid questionnaires were returned which constitutes a response rate of 32.4%. Demographic data are summarised in Table 1. We consider this an adequate sample size that is likely to ensure component estimate stability [see Guadagnoli and Velicer (1988)].

Table 1: Demographic data for respondents

| Participant characteristic | Responses | Percentage (%) |
|---|---|---|
| **Company type** | | |
| Software development company | 122 | 37.7 |
| IT consultancy & management | 104 | 32.1 |
| Web development | 98 | 30.2 |
| **Years of experience in software projects** | | |

| | | |
|---|---|---|
| Less than 3 | 23 | 7.1 |
| 3 – 6 | 47 | 14.5 |
| 7 – 10 | 55 | 17.0 |
| 11 – 14 | 81 | 25.0 |
| 15 – 18 | 86 | 26.5 |
| More than 18 | 32 | 9.9 |
| | | |
| **Number of software projects** | | |
| Less than 3 | 27 | 8.3 |
| 3 – 6 | 103 | 31.8 |
| 7 – 10 | 83 | 25.6 |
| 11 – 14 | 48 | 14.8 |
| 15 – 18 | 40 | 12.4 |
| More than 18 | 23 | 7.1 |
| | | |
| **Geographic distribution** | | |
| USA | 131 | 40.4 |
| UK | 81 | 25.0 |
| India | 28 | 8.6 |
| Canada | 18 | 5.6 |
| Others | 174 | 20.4 |

The geographical spread of respondents is weighted towards USA- and UK-based respondents, similarly to previous research samples, although our geographic base is slightly wider than past studies. Thus, the survey respondents reflect geographical sources of other studies but with slightly more geographical representation. No other country accounts for more than 10% of the sample and hence we do not expect any geographical bias in the results that differs from extant studies focussed on the USA and the UK. As a consequence, it is felt that the respondent sample reported here is broadly comparable with other research in this area and the sample composition is similar to that used in other researcher in the field and therefore it is comparable with earlier research.

### 3.4 Analysis of responses

Analysis of the responses to Parts I and II relied on principal components analysis (PCA) to determine which of the risk factors cluster into statistically meaningful groupings as has been done with previous research (see Conger, Loch, & Helft, 1995; Wallace, Keil, & Rai 2004; Jiang and Klein, 1999, 2000; *inter alia*). Using the Kaiser (1960) criterion, eigenvalues over 1.0 were retained. This reduces the candidate risk factor list to the most influential and gives us a list of extracted risk factors (these were termed 'risk factors' to distinguish them as confirmed from the risk factor candidate list used in the survey questions). These risk factors were then grouped as explanatory factor loadings into risk components. The aggregation of the remaining risk factors into risk components is a clustering process that is a PCA-determined method based on the degree of collinearity between risk factors.

The empirically derived risk components relating to questions Part I and II are then interpreted in terms which are meaningful in relation to the life-cycle of the project. This reversion to the research risk construct is an important element of the research contribution, on the basis of theory and further empirical work, as it addresses why the risk components observed in practice are likely to be important. Thus, it enabled us to interpret the risk components observed in relation to project life cycle which typically involves reconfiguring the observed components to those that meaningfully relate to the risk construct, using methods that involve both appeal to theory and further statistical investigation.

## 4.0 DATA INTERPRETATION AND RESULTS

### 4.1 KMO & Bartlett test

Before factor reduction and clustering, the KMO & Bartlett test was performed to check the possible presence of multicollinearity and correlation among the risk variables. Kaiser-Meyer-Olkin (KMO) measures sampling adequacy and the Bartlett Test of Sphericity presents correlations, and the results are shown in Table 2. Kaiser (1974) recommends accepting values greater than 0.5; values between 0.5 and 0.7 are good, values between 0.7 and 0.8 are great, and values above 0.8 are very good. Table 2 shows the sampling is adequate to conduct factor reduction and clustering. Bartlett's test examines whether the population correlation matrix resembles an identity matrix. If the population correlation matrix resembles an identity matrix then every variable correlates with all other variables (i.e., all correlations coefficients are close to zero). Bartlett's test examines the null hypothesis ($H_0 > 0.05$) that the original correlation matrix is an identity matrix. We find significant p-values for both likelihood occurrence and impact on cost overruns, indicating the correlation matrix is not an identity matrix. Therefore, factor reduction and clustering is appropriate for these variables in this survey as there are similarities between the variables.

Table 2: KMO & Bartlett test

| **Stages** | **Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy** | | **Bartlett's Test of Sphericity (Significance value)** | |
|---|---|---|---|---|
| | Likelihood occurrence | Impact on cost overrun | Likelihood occurrence | Impact on cost overrun |
| Feasibility study | 0.598 | 0.605 | 0.000 | 0.000 |
| Project planning | 0.704 | 0.750 | 0.000 | 0.000 |
| Requirement | 0.607 | 0.584 | 0.000 | 0.000 |
| Development | 0.724 | 0.742 | 0.000 | 0.000 |
| Implementation | 0.850 | 0.687 | 0.000 | 0.000 |
| Operation maintenance | 0.506 | 0.505 | 0.000 | 0.000 |

The results indicate that there is a basis of interpretability that provides sufficient distinctness between project stages within which risk components may be identified and analysed. Moreover, the large sample size used in this study exceeds that normally considered to be adequate for research of this nature and at a point test parameters become stable irrespective of the participant to variable ratio (Kass & Tinsley, 1979). The Cronbach's Alpha values of our emerging risks and their impact on cost overrun questions are 0.963 and 0.968 respectively, suggesting that the risk components have significantly high internal consistency.

### 4.2 Likelihood of occurrence

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

Each component was set according to a series of correlations between different risk factors as shown in Table 3. The first column is labelled as initial Eigenvalues related to Eigenvalue of the correlation matrix. To carry out the factor reduction and clustering, only components with Eigenvalues more than 1 were selected as suggested by Punch (2005).

The initial and rotated Eigenvalues were used to confirm the variation explained by each extracted risk component. Lower values indicate that the contribution to the explanation of the variances in the set of risk attributes is minimal. Nine components carry Eigenvalues at 1 and above, and represent 88.45% of the total variance; therefore, these nine components are deemed as representative of the overall 104 risk factors. This means less than 12% of the existing information is compromised. A scree plot is shown in Figure 1, whose purpose is to provide a graphical picture of the Eigenvalue for each component extracted in SPSS. The slope of screen is decreasing, while moving towards components with Eigenvalue less than 1. The point of interest is defined between components nine and ten, where the curve connects to the points, starting to flatten out and horizontal. In a screen plot, the place where a sharp change in angle occurs is considered as the exact point that Eigenvalues of less than 1 are placed according to Morgan et al. (2004). On the sharp slope of curve, the Eigenvalues bigger than 1 are located, while in the flatten part of the curve, the Eigenvalues smaller than 1 are plotted.

Table 3: Likelihood occurrence

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 31.998 | 30.767 | 30.767 | 31.998 | 30.767 | 30.767 | 29.255 | 28.030 | 28.030 |
| 2 | 27.799 | 26.729 | 57.496 | 27.799 | 26.729 | 57.496 | 24.813 | 23.759 | 51.789 |
| 3 | 15.440 | 14.631 | 72.127 | 15.440 | 14.631 | 72.127 | 14.152 | 13.607 | 65.396 |
| 4 | 12.249 | 11.951 | 84.078 | 12.249 | 11.951 | 84.078 | 11.778 | 11.391 | 76.787 |
| 5 | 7.476 | 7.088 | 91.166 | 7.476 | 7.088 | 91.166 | 6.282 | 6.040 | 82.827 |
| 6 | 1.993 | 1.907 | 93.073 | 1.993 | 1.907 | 93.073 | 1.834 | 1.816 | 84.643 |
| 7 | 1.817 | 1.647 | 94.720 | 1.817 | 1.647 | 94.720 | 1.548 | 1.489 | 86.132 |
| 8 | 1.639 | 1.476 | 96.196 | 1.639 | 1.476 | 96.196 | 1.295 | 1.245 | 87.377 |
| 9 | 1.426 | 1.271 | 97.467 | 1.426 | 1.271 | 97.467 | 1.114 | 1.071 | 88.448 |
| 10 | 0.805 | 0.374 | 97.841 | | | | | | |
| 11 | 0.672 | 0.246 | 98.087 | | | | | | |
| 102 | -1.287E-14 | -1.237E-14 | 100.000 | | | | | | |
| 103 | -1.540E-14 | -1.481E-14 | 100.000 | | | | | | |
| 104 | -1.634E-14 | -1.571E-14 | 100.000 | | | | | | |

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

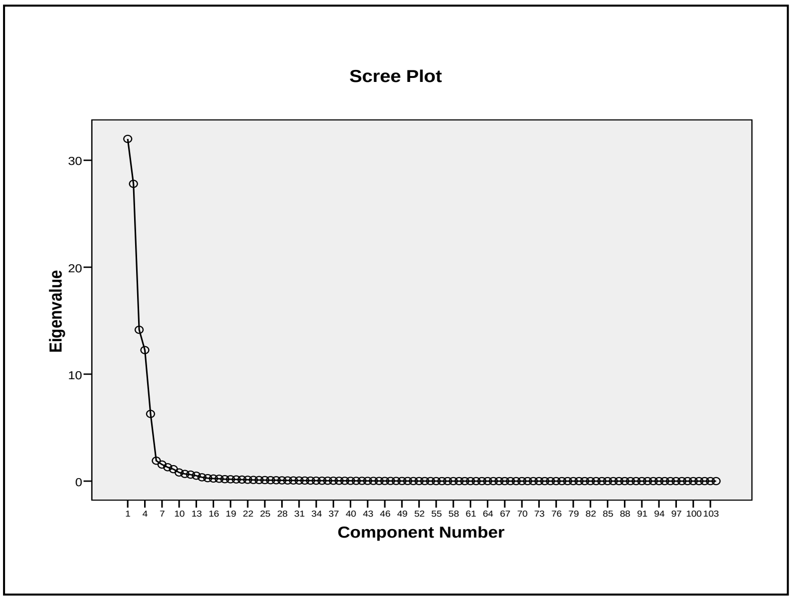



Figure 1: Scree plot of risk factors for likelihood occurrence

In the principal component analysis, the nine components with Eigenvalues greater than 1 were selected for the next phase. This extraction of rotated component matrix was used to identify which risk factors contributed the highest level of influence on the software project. The Matrix loading score presented in Table 4 shows the degree of influence of each risk factor in the whole survey, and the risk factors with the highest rate of influence were distinguished. This factor loading tells the relative contribution that a variable makes to a factor. Most variables have higher loadings on more important factors. It normally interprets factor loadings with an absolute value greater than 0.4 (ignoring the +ve or –ve sign) around 16% of the variance in the variable (Stevens, 1995; Maccallum et al., 1999; Morgan et al., 2004; Field, 2005); thus, only loading scores with the degree of influence greater 0.4 are shown in Table 4. For example, the risk factor (F1; 0.656) has greater influence on component 7 compared to other components. Whereas, the risk factor (P5; 0.509) has more influence on component 6 in relation to other components, and (R3; 0.526) has more influence on component 4 in relation to other components. For risk factors D18-D20 and D24-D35, the loading scores were below 0.4 thus not selected.

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

Table 5: Matrix loading score for likelihood occurrence

| Factor | Component | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| F1 | | 0.175 | -0.156 | | | 0.17 | **0.656** | -0.166 | |
| F2 | | **0.576** | | | -0.165 | | | 0.145 | 0.237 |
| F3 | -0.274 | 0.208 | 0.117 | | | **0.637** | | -0.131 | |
| F4 | | | -0.184 | 0.296 | 0.288 | -0.182 | | | **0.495** |
| F5 | | **0.561** | 0.178 | -0.238 | | 0.257 | 0.185 | | 0.126 |
| P1 | 0.146 | | -0.141 | 0.178 | -0.157 | **0.457** | | | -0.211 |
| P2 | | 0.212 | 0.137 | | 0.194 | | -0.221 | **0.532** | 0.234 |
| P3 | | | | | | | 0.278 | **0.722** | 0.213 |
| P4 | | -0.223 | | | -0.269 | 0.12 | | **0.78** | |
| P5 | 0.234 | | 0.171 | | 0.105 | **0.509** | 0.291 | -0.165 | 0.245 |
| P6 | 0.197 | | 0.181 | -0.217 | 0.268 | **0.418** | 0.279 | 0.189 | -0.138 |
| P7 | -0.149 | -0.265 | 0.152 | 0.181 | 0.209 | **0.443** | | 0.248 | 0.154 |
| P8 | | | | | | | 0.137 | **0.875** | -0.137 |
| P9 | | 0.286 | 0.123 | 0.192 | **0.62** | 0.274 | 0.285 | 0.133 | -0.241 |
| P10 | | 0.143 | **0.807** | | 0.15 | | -0.145 | -0.114 | 0.102 |
| P11 | 0.123 | -0.156 | -0.144 | | | **0.848** | | 0.276 | 0.102 |
| P12 | | | | | **0.59** | 0.21 | | 0.2 | |
| P13 | | | -0.175 | | **0.892** | | -0.259 | 0.15 | 0.142 |
| P14 | | | -0.258 | 0.121 | | **0.676** | | | -0.158 |
| R1 | | | | **0.803** | | | | | |
| R2 | | | | **0.553** | | | | | |
| R3 | | | | **0.526** | | | | | |
| R4 | | | | **0.456** | | | | | |
| R5 | | | | **0.656** | | | | | |
| R6 | | | | **0.572** | | | | | |
| R7 | **0.551** | | | | | | | | |
| D1 | | **0.827** | | | | | | | |
| D2 | | **0.493** | | | | | | | |
| D3 | | **0.708** | | | | | | | |
| D4 | | **0.856** | | | | | | | |
| D5 | | **0.688** | | | | | | | |
| D6 | | **0.568** | | | | | | | |
| D7 | | **0.691** | | | | | | | |
| D8 | | **0.474** | | | | | | | |
| D9 | | **0.611** | | | | | | | |
| D10 | | **0.835** | | | | | | | |
| D11 | **0.55** | | | | | | | | |
| D12 | | **0.427** | | | | | | | |
| D13 | | | | | **0.808** | | | | |
| D14 | | | | | | | | **0.477** | |
| D15 | **0.448** | | | | | | | | |
| D16 | 0.097 | -0.314 | 0.337 | -0.305 | 0.002 | 0.027 | 0.204 | 0.004 | -0.224 |
| D17 | **0.414** | | | | | | | | |
| D18 | 0.358 | -0.059 | 0.113 | 0.073 | 0.043 | 0.328 | -0.305 | 0.179 | 0.2 |
| D19 | -0.061 | 0.385 | 0.207 | 0.295 | 0.082 | -0.123 | 0.016 | 0.239 | -0.066 |
| D20 | 0.081 | -0.396 | 0.3 | -0.31 | 0.191 | 0.095 | 0.033 | -0.343 | 0.063 |
| D21 | | | **0.551** | | | | | | |
| D22 | | | **0.592** | | | | | | |
| D23 | | | **0.692** | | | | | | |
| D24 | -0.194 | 0.054 | 0.272 | -0.264 | 0.037 | 0.214 | 0.001 | -0.187 | 0.099 |
| D25 | 0.094 | 0.002 | -0.242 | 0.163 | 0.071 | -0.05 | 0.211 | 0.073 | -0.139 |
| D26 | 0.031 | -0.132 | 0.225 | 0.104 | -0.002 | 0.164 | 0.286 | 0.131 | 0.008 |
| D27 | -0.058 | 0.089 | 0.207 | -0.309 | 0.256 | 0.09 | -0.182 | -0.05 | 0.161 |
| D28 | 0.351 | 0.045 | 0.084 | 0.31 | 0.229 | 0.336 | 0.171 | 0.202 | -0.075 |
| D29 | -0.112 | 0.176 | 0.092 | 0.188 | 0.002 | 0.227 | 0.151 | 0.068 | 0.075 |
| D30 | 0.134 | 0.306 | -0.211 | 0.347 | 0.079 | -0.065 | 0.12 | -0.189 | 0.16 |
| D31 | 0.084 | -0.103 | 0.329 | -0.067 | 0.069 | 0.062 | 0.103 | 0.02 | -0.161 |
| D32 | 0.106 | 0.156 | 0.163 | 0.049 | 0.243 | -0.16 | -0.058 | -0.174 | 0.094 |
| D33 | 0.067 | -0.058 | 0.037 | -0.246 | 0.145 | 0.109 | 0.363 | 0.022 | -0.137 |
| D34 | -0.137 | 0.072 | 0.124 | 0.012 | 0.192 | 0.203 | 0.161 | 0.117 | 0.039 |
| D35 | 0.045 | -0.281 | 0.026 | 0.199 | -0.294 | 0.119 | 0.215 | -0.103 | 0.034 |

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

The most influential risk factors of each component were extracted to form a reduced list, thus the 104 risk factors were reduced to nine components as shown in Table 5. The percentages of variance for each component in Table 5 are extracted from Table 3, yet common themes of the components were identified and each component was given new terms for reference.

Table 5: Components for likelihood occurrence

| Risk component | Extracted eigenvalue | Extracted sum of squared loadings: variance % | Rotation sum of squared loadings: variance % | Risk factors aggregated to component following rotation | |
|---|---|---|---|---|---|
| | | | | **Factor loading score** | |
| **Component 1**<br>Project user engagement | 31.99 | 30.76 | 28.03 | 0.551<br>0.550<br>0.448<br>0.414 | Lack of users involvement in requirement stage: R7<br>Failure of user acceptance test: D11<br>Lack of users involvement and commitment: D15<br>Ineffective communication within development team members: D17 |
| **Component 2**<br>Technology failure | 27.79 | 26.73 | 23.76 | 0.576<br>0.561<br>0.827<br>0.493<br>0.708<br>0.856<br>0.688<br>0.568<br>0.691<br>0.474<br>0.611<br>0.835<br>0.427 | Too narrow focus on the technical IT issues: F2<br>Inappropriate technology chosen from the feasibility study: F5<br>Improper handover from the requirement team: D1<br>Inappropriate development methodology used: D2<br>Unsuitable working model and prototype: D3<br>Programming language and CASE tool selected not adequate: D4<br>High level of technical complexities: D5<br>Project involves the use of new technology: D6<br>Difficulty in defining the input and output of system: D7<br>Immature technology: D8<br>Technological advancements and changes: D9<br>Failures and inconsistencies of unit/modules test results: D10<br>Time consuming for testing: D12 |
| **Component 3**<br>Project personnel | 15.44 | 14.63 | 14.15 | 0.807<br>0.551<br>0.592<br>0.692 | Project management & development team not properly set up: P10<br>Inexperienced team members: D21<br>Lack of commitment to project among team members: D22<br>Ineffective and inexperienced project manager: D23 |
| **Component 4**<br>Technology and system requirements | 12.25 | 11.95 | 11.39 | 0.803<br>0.553<br>0.526<br>0.456<br>0.656<br>0.572 | Unclear and inadequate identification of systems requirements: R1<br>Incorrect systems requirements: R2<br>Misinterpretations of the systems requirements: R3<br>Conflicting system requirements: R4<br>Gold plating or unnecessary functions and requirements: R5<br>Inadequate validation of the requirements: R6 |
| **Component 5**<br>Project implementation | 7.48 | 7.09 | 6.28 | 0.620<br>0.590<br>0.892<br>0.808 | Critical and non-critical activities of project not identified: P9<br>Lack of contingency plan/back up: P12<br>System conversion method not well planned: P13<br>Resources shifted from project during development: D13 |
| **Component 6**<br>Project planning | 1.99 | 1.91 | 1.82 | 0.637<br>0.457<br>0.509<br>0.418<br>0.443<br>0.848<br>0.676 | Overlooked the management and business impact issues: F3<br>Unclear project scope and objectives: P1<br>Improper change management planning: P5<br>Inaccurate estimate of resources: P6<br>Unrealistic project schedule: P7<br>Unclear line of decision making authority throughout the project: P11<br>Improper planning of timeframe for project reviews and updating: P14 |
| **Component 7**<br>Feasibility study | 1.82 | 1.65 | 1.49 | 0.656 | Wrong justifications of cost benefit analysis from feasibility study: F1 |
| **Component 8**<br>Project process | 1.64 | 1.48 | 1.24 | 0.532<br>0.722<br>0.780<br>0.875<br>0.477 | Undefined project success criteria: P2<br>Lack of quality control procedure and mechanism: P3<br>Project milestones for stages not well established: P4<br>Inadequate detail work breakdown structure: P8<br>Changes in management of organisation during development: D14 |
| **Component 9**<br>Feasibility study decision | 1.43 | 1.27 | 1.07 | 0.495 | Wrong justification of investment alternatives and opportunity cost: F4 |

Table 5 reports both the variance explained by these retained factors from the total variance of all 104 factors as well as the factor loadings (and their variances) following varimax rotation (an orthogonal rotation method) in which the variance of each factor is maximised. This facilitates the interpretability of the resulting factors. The retained risk factors (45 in total) and their grouping into risk components are shown in the right column. The initial identifiers (R, D, F and P) in the right column – followed by a number – refer to the project stages

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

and their initial number within each stage. The stages are: F- Feasibility Study; D – Development; R – Requirement; and P – Project Planning.

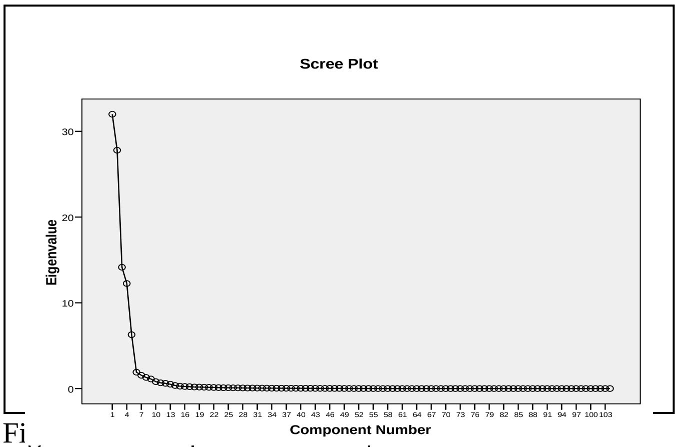
### 4.3 Risk impact on cost overrun

For risk impact on cost overrun, the Eigenvalue of the first factor in Table 6 is 32.202 and the proportion of the total test variance accounted for by the first factor is 30.963%. Only seven components carry Eigenvalues greater than 1, account for 87.92% of variance for risk impact on cost overrun (and less than 13% of original data were compromised).

Table 6: Total variance explained – risk impact on Cost overrun

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 32.202 | 30.963 | 30.963 | 32.202 | 30.963 | 30.963 | 28.333 | 27.143 | 27.143 |
| 2 | 21.481 | 20.578 | 51.541 | 21.481 | 20.578 | 51.541 | 19.625 | 18.809 | 49.952 |
| 3 | 16.762 | 16.079 | 67.620 | 16.762 | 16.079 | 67.620 | 15.727 | 15.022 | 60.974 |
| 4 | 14.385 | 13.694 | 81.314 | 14.385 | 13.694 | 81.314 | 12.662 | 12.175 | 73.149 |
| 5 | 12.290 | 11.602 | 92.916 | 12.290 | 11.602 | 92.916 | 10.793 | 9.493 | 82.642 |
| 6 | 3.817 | 3.661 | 96.577 | 3.817 | 3.661 | 96.577 | 3.609 | 3.559 | 86.201 |
| 7 | 2.319 | 2.129 | 98.706 | 2.319 | 2.129 | 98.706 | 1.791 | 1.722 | 87.923 |
| 8 | .217 | .181 | 98.887 | | | | | | |
| 9 | .156 | .131 | 99.018 | | | | | | |
| 102 | -1.287E-14 | -1.237E-14 | 100.000 | | | | | | |
| 103 | -1.540E-14 | -1.481E-14 | 100.000 | | | | | | |
| 104 | -1.634E-14 | -1.571E-14 | 100.000 | | | | | | |

The point of interest in the Scree plot in Figure 2 is defined between components 7 and 8, where the curve connects to the points, starting to flatten out and horizontal. Table 7 presents the degree of influence of each risk factor on cost overrun, where the risk factors with the highest rate of influence and high loading matrix were distinguished.




Fi

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk, 9*(3), 53-88.

Table 7: Matrix loading score for risk impact on cost overrun

| Factor | Component | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| F1 | | | | 0.19 | **0.553** | | 0.169 |
| F2 | | **0.65** | | -0.128 | | 0.176 | 0.298 |
| F3 | **0.535** | 0.167 | -0.109 | 0.154 | 0.174 | | |
| F4 | | -0.257 | | | **0.619** | -0.176 | 0.185 |
| F5 | -0.125 | **0.623** | -0.248 | | | | -0.12 |
| P1 | **0.557** | | | | 0.127 | -0.122 | |
| P2 | 0.298 | | 0.149 | | | | **0.843** |
| P3 | -0.19 | -0.115 | | 0.282 | -0.191 | | **0.827** |
| P4 | 0.178 | | | | 0.144 | | **0.813** |
| P5 | -0.189 | | | -0.112 | | **0.407** | |
| P6 | **0.577** | -0.156 | 0.102 | 0.3 | 0.118 | -0.227 | |
| P7 | | 0.191 | 0.243 | | | **0.637** | -0.146 |
| P8 | | 0.111 | | 0.103 | -0.191 | 0.262 | **0.842** |
| P9 | | | | | 0.125 | -0.232 | **0.587** |
| P10 | -0.236 | | 0.223 | -0.264 | | 0.162 | **0.84** |
| P11 | | | | 0.254 | -0.149 | **0.654** | -0.103 |
| P12 | 0.235 | | -0.278 | -0.249 | | **0.448** | 0.172 |
| P13 | | | | **0.773** | 0.159 | | |
| P14 | **0.482** | -0.151 | 0.222 | 0.269 | | | -0.131 |
| R1 | | | **0.529** | | | 0.178 | |
| R2 | -0.258 | 0.297 | **0.593** | | | | |
| R3 | | | **0.539** | -0.26 | | -0.112 | 0.275 |
| R4 | | 0.136 | **0.757** | | | 0.189 | |
| R5 | | 0.101 | **0.836** | 0.257 | 0.203 | 0.156 | 0.174 |
| R6 | | | **0.56** | | 0.24 | | 0.275 |
| R7 | | **0.546** | **0.782** | -0.173 | 0.203 | -0.123 | |
| D1 | | | | | | **0.502** | |
| D2 | | | | **0.467** | | -0.176 | 0.25 |
| D3 | | -0.235 | | **0.729** | 0.176 | | |
| D4 | | | 0.114 | **0.717** | -0.263 | | -0.167 |
| D5 | | **0.727** | -0.171 | | | | |
| D6 | | **0.662** | 0.129 | | -0.122 | 0.209 | |
| D7 | | -0.107 | | **0.577** | 0.228 | | |
| D8 | -0.173 | **0.747** | | 0.254 | -0.221 | 0.208 | 0.27 |
| D9 | -0.134 | **0.82** | 0.149 | -0.258 | | 0.127 | |
| D10 | | -0.235 | | **0.729** | -0.176 | | |
| D11 | | | | **0.72** | 0.171 | | -0.167 |
| D12 | | -0.121 | | **0.897** | | 0.16 | |
| D13 | -0.245 | | | | -0.166 | **0.499** | 0.196 |
| D14 | 0.285 | | | 0.28 | -0.214 | **0.475** | |
| D15 | **0.561** | | | | | | 0.25 |
| D16 | 0.047 | 0.339 | -0.068 | -0.043 | 0.05 | 0.017 | 0.155 |
| D17 | 0.24 | 0.157 | | 0.259 | 0.284 | **0.666** | 0.149 |
| D18 | -0.138 | 0.176 | 0.094 | 0.002 | 0.297 | -0.335 | 0.138 |
| D19 | 0.182 | | -0.164 | -0.16 | | 0.147 | **0.774** |
| D20 | -0.025 | -0.354 | 0.163 | 0.355 | 0.15 | -0.014 | 0.386 |
| D21 | 0.136 | -0.303 | 0.373 | 0.168 | -0.105 | 0.173 | -0.035 |
| D22 | **0.436** | 0.14 | -0.282 | | | -0.29 | -0.265 |
| D23 | **0.776** | 0.24 | -0.295 | 0.19 | 0.068 | -0.077 | 0.244 |
| D24 | 0.078 | 0.085 | 0.084 | 0.121 | 0.159 | 0.256 | -0.131 |
| D25 | -0.04 | 0.278 | -0.276 | -0.128 | -0.296 | 0.131 | 0.202 |
| D26 | 0.079 | -0.183 | 0.205 | 0.324 | 0.351 | -0.088 | 0.318 |
| D27 | 0.052 | 0.12 | -0.038 | -0.022 | -0.18 | 0.185 | -0.242 |
| D28 | -0.309 | 0.097 | 0.183 | 0.092 | 0.224 | -0.113 | 0.303 |
| D29 | 0.029 | -0.095 | -0.111 | -0.009 | 0.182 | 0.034 | -0.023 |
| D30 | -0.082 | 0.332 | 0.272 | 0.139 | -0.117 | -0.096 | 0.27 |
| D31 | 0.096 | -0.254 | -0.061 | -0.21 | 0.089 | 0.076 | -0.384 |
| D32 | -0.379 | -0.123 | 0.194 | 0.128 | 0.201 | 0.124 | 0.041 |
| D33 | 0.081 | 0.333 | 0.1 | -0.14 | 0.12 | -0.269 | -0.271 |
| D34 | -0.3 | 0.183 | -0.003 | 0.25 | -0.001 | 0.084 | 0.069 |
| D35 | 0.31 | -0.164 | 0.203 | 0.216 | -0.049 | -0.344 | 0.12 |

The most influential risk factors of each component were extracted in Table 8 to form a reduced list, which is manageable without losing a large amount of data. By applying factor reduction and clustering and data reduction in this survey, the 104 risk factors are reduced to seven components. The percentages of variance for each component in Table 8 are extracted from Table 6.

This is a copy of the "Post-print" (i.e., final authors' copy, post-refereeing). Published as:
Mohd-Rahim, F. A., Wang, C., Boussabaine, H., Abdul-Rahman, H., & Wood, L. C. (2014). Factor reduction and clustering for operational risk in software development. *Journal of Operational Risk*, *9*(3), 53-88.

Table 8: Components for risk impact on cost overrun

| Risk component | Extracted eigenvalue | Extracted sum of squared loadings: variance % | Rotation sum of squared loadings: variance % | Risk factors aggregated to component following rotation: Loading factor | Risk factors aggregated to component following rotation |
|---|---|---|---|---|---|
| **Component 1**<br>Project team planning | 32.202 | 30.963 | 27.143 | 0.535<br>0.557<br>0.577<br>0.482<br>0.561<br>0.436<br>0.776 | Overlooked the management and business impact issues: F3<br>Unclear project scope + objectives: P1<br>Inaccurate estimate of resources: P6<br>Improper planning of timeframe for project reviews and updating: P14<br>Lack of users involvement and commitment: D15<br>Lack of commitment among development team members: D22<br>Ineffective and inexperienced project manager: D23 |
| **Component 2**<br>Technology appropriateness | 21.481 | 20.578 | 18.809 | 0.650<br>0.623<br>0.727<br>0.662<br>0.747<br>0.820 | Too narrow focus on the technical IT issues: F2<br>Inappropriate technology chosen from the feasibility study: F5<br>High level of technical complexities: D5<br>Project involves the use of new technology: D6<br>Immature technology: D8<br>Technological advancements and changes: D9 |
| **Component 3**<br>Technology specification | 16.762 | 16.079 | 15.022 | 0.529<br>0.593<br>0.539<br>0.757<br>0.836<br>0.560<br>0.782 | Unclear and inadequate identification of systems requirements: R1<br>Incorrect systems requirements: R2<br>Misinterpretations of the systems requirements: R3<br>Conflicting system requirements: R4<br>Gold platting or unnecessary functions and requirements: R5<br>Inadequate validation of the requirements: R6<br>Lack of users involvement in requirement stage: R7 |
| **Component 4**<br>Technology and implementation | 14.385 | 13.694 | 12.175 | 0.773<br>0.467<br>0.729<br>0.717<br>0.577<br>0.729<br>0.720<br>0.897 | System conversion method not well planned: P13<br>Inappropriate development methodology used: D2<br>Unsuitable working model and prototype: D3<br>Programming language and CASE tool selected not adequate: D4<br>Difficulty in defining the input and output of system: D7<br>Failures and inconsistencies of unit/modules test results: D10<br>Failure of user acceptance test: D11<br>Time consuming for testing: D12 |
| **Component 5**<br>Feasibility study | 12.290 | 11.602 | 9.493 | 0.553<br>0.619 | Wrong justification of cost benefit analysis from feasibility study: F1<br>Wrong justification of investment alternatives and opportunity cost: F4 |
| **Component 6**<br>Project team management | 3.817 | 3.661 | 3.559 | 0.407<br>0.637<br>0.654<br>0.448<br>0.502<br>0.499<br>0.475<br>0.666 | Improper change management planning: P5<br>Unrealistic project schedule: P7<br>Unclear line of decision making authority throughout the project: P11<br>Lack of contingency plan/back up: P12<br>Improper handover from the requirement team: D1<br>Resources shifted from project during development: D13<br>Change in management during development: D14<br>Ineffective communications within development team members: D17 |
| **Component 7**<br>Project team activities | 2.319 | 2.129 | 1.722 | 0.843<br>0.827<br>0.813<br>0.842<br>0.587<br>0.774 | Undefined project success criteria: P2<br>Lack of quality control procedure and mechanism: P3<br>Project milestones for stages not well establish: P4<br>Inadequate detail breakdown structure: P8<br>Critical and non-critical activities of project not identified: P9<br>Inadequately trained development team members: D19 |

## 5.0 DISCUSSIONS

### 5.1 Clustering of risk likelihood occurrence

This initial clustering into nine components and the relationship to individual risk factors is entirely empirically determined, in common with previous research adopting this approach [see, for example, Ropponen and Lyytinen (2000)]. The factor loadings following rotation indicate a shift in importance of individual factors to the risk components and redistribution in the overall explanation with the total variance accounted for is marginally reduced to 88.45%. Varimax rotation was used to maintain the orthogonality of the individual factors and this enhances their interpretability. The major components (those with the largest variances and with variances reported in parenthesis) are: 1 (28.03%), 2 (23.76%), 3 (14.15%) and 4 (11.39%). The identification of risk factors in Table 4 provides a guide to the interpretation of the risk components and articulating the findings, to present a view on clustering in the context of the whole life cycle project, determining risks in relation to meaningful project stages.

Consequently, the nine risk components interpreted what clusters could be formed and may be placed into a whole life cycle context. This is similar to the approach taken by Wallace and Keil (2004), who employ socio-technical systems theory to help establish the dimensionality of risks they observe in their own survey (n=507). Our approach allows factor reduction and clustering to establish initial dimensionality (the nine risk components of Table 4) and to interpret the results in terms of the whole project life cycle. This latter element is akin to that employed by Barki et al (1993). In examining the make-up of the risk components (that is, from the risk factors in the Table 4), a number of themes were observed to be consistent with categorisations from a generalised project plan over the whole-life cycle. This is supported by the examination of the factor loadings of the risk factors which are reported in Table 5, which reports the loading factors extracted from the rotated component matrix of the risk data sample. This is the main basis for the component interpretation used.

From the analysis, only 45 risk factors out of the 104 were selected which account for 88.45% of the variance that could be explained. The 45 risk factors were selected based on the *'eigen-one'*, Kaiser (1974) criterion cut-off, which used only risk factors that have factor loading of 0.400 and above. The nine components extracted from factor reduction and clustering were then clustered together to form a few clusters that have some common themes. For each cluster, the likelihood occurrence of risk is calculated based on percentage of variance of each component derived from Table 3. The analysis shows three main and internally consistent clusters emerge (Table 9): a) Cluster 1: Feasibility study; b) Cluster 2: Project and team management; and c) Cluster 3: Technology requirement.

Table 9: New clustering for likelihood occurrence of risk

| Cluster 1 | Component | % variance | Main risk factors | Total % variance |
|---|---|---|---|---|
| **Cluster 1 Feasibility study** | Component 7 | 1.49 | F1: Wrong justifications of cost benefit analysis from feasibility study. | **2.56 %** |
| | Component 9 | 1.07 | F4 : Wrong justification of investment alternatives and opportunity cost | |
| **Cluster 2 Project and team management** | Component 1 | 28.03 | R7: Lack of users involvement in requirement stage<br>D11: Failure of user acceptance test<br>D15: Lack of users involvement and commitment<br>D17: Ineffective communication within development team members | **51.52 %** |
| | Component 3 | 14.15 | P10: Project management & development team not properly set up<br>D21: Inexperienced team members<br>D22: Lack of commitment to project among development team members<br>D23: Ineffective and inexperienced project manager | |
| | Component 5 | 6.28 | P9: Critical and non-critical activities of project not identified<br>P12: Lack of contingency plan/back up<br>P13: System conversion method not well planned<br>D13: Resources shifted from project during development | |
| | Component 6 | 1.82 | F3: Overlooked the management and business impact issues<br>P1: Unclear project scope and objectives<br>P5: Improper change management planning<br>P6: Inaccurate estimate of resources<br>P7: Unrealistic project schedule<br>P11: Unclear line of decision making authority throughout the project:<br>P14: Improper planning of timeframe for project reviews and updating | |

| | | | | |
|---|---|---|---|---|
| | Component 8 | 1.24 | P2: Undefined project success criteria<br>P3: Lack of quality control procedure and mechanism<br>P4: Project milestones for stages not well established<br>P8: Inadequate detail work breakdown structure<br>D14: Changes in management of organisation during development | |
| **Cluster 3**<br><br>**Technology requirement** | Component 2 | 23.76 | F2: Too narrow focus on the technical IT issues<br>F5: Inappropriate technology chosen from the feasibility study<br>D1: Improper handover from the requirement team<br>D2: Inappropriate development methodology used<br>D3: Unsuitable working model and prototype<br>D4: Programming language and CASE tool selected not adequate<br>D5: High level of technical complexities<br>D6: Project involves the use of new technologies<br>D7: Difficulty in defining the input and output of system<br>D8: Immature technology<br>D9: Technological advancements and changes<br>D10: Failures and inconsistencies of unit/modules test results<br>D12: Time consuming for testing | **35.15 %** |
| | Component 4 | 11.39 | R1: Unclear and inadequate identification of systems requirements<br>R2: Incorrect systems requirements<br>R3: Misinterpretations of the systems requirements<br>R4: Conflicting system requirements<br>R5: Gold plating or unnecessary functions and requirements<br>R6: Inadequate validation of the requirements | |

**Cluster 1** comprises components 7 and 9 and represents 2.56% of the total variance explained. Only two risk factors make-up this cluster and they relate to cost benefit analysis and an analysis of opportunity costs in the initial evaluation. It is not surprising that these factors could be related in this manner; a full cost and benefits analysis would seek to avoid the risks associated with both incorrect conclusions from the analysis undertaken and a subsequent failure to incorporate all relevant factors. From the perspective of IT professionals, it could be argued that any failure in assessment at the project feasibility stage might manifest as project problems later-on. This would consequently have an impact on project success. The expectation is that IT professionals would prioritise this issue, if only to avoid dealing with the consequences of a situation not of their making later in the project life cycle.

However, the issue could become problematic if they are not involved in the project at this stage. Furthermore, the identification of opportunity costs requires a deep knowledge of the organisational context in order to be able to allocate them successfully. Invariably, this will need an early and accurate projection of cash flows including the opportunity costs of capital (Ballantine & Stray, 1998). The fact that IT professionals highlight this as a significant weight in the survey response might be reflective of their lack of oversight and detailed knowledge of organisational context which would provide an appropriate framework to judge what may or may not be appropriate feasibility risks to identify and evaluate.

What is clear from this cluster is the potential for organisations to generate problems in the project due to poor interaction and communication between managers within the organisation and those external consultants who will be charged with managing the project later in the timeline. Ultimately, this issue reflects the assumptions held by the various parties involved in the decision making and the relative boundaries of consideration they hold relative to the potential for embedding risk in the project.

**Cluster 2** comprises components 1, 3, 5, 6 and 8 and represents 51.52% of the total variance explained. This has been labelled Project and Team Management. The basis of this, in looking at the associated risk factors in Table 3, is indicated by a range of factors concerning the interaction of the project with non-technical software areas, and particularly regarding interaction with staff and other firm resources. Again, the issues of communication and information sharing could be seen to play an important role in shaping this problem. Whilst there is one factor (*lack of user involvement in the requirement stage*) that may overlap with the Feasibility study cluster, the remaining factors relate to the immediacy of project implementation.

Components 1 and 3 relate to project interaction with resources available and Components 5, 6 and 8 relate to the consequences of failure at the planning stage. Thus, on the resource side the observed risk factors were ranging from lack of human input, to inexperience, lack of commitment, mid-project resource-shifting, inadequate line management, along with overlooked business impact issues where the resources interaction with the wider organisational context is evident. With respect to the consequences of planning failure, a full spectrum of risks were observed that were inadequately anticipated but specifically contextualised in terms of project performance. The risk factors are often placed in the past tense and responses are provided with the benefit of hindsight and with the project running or even complete. It might be right to question why these *ex post* factors do not appear as planning issues specifically. As such, they could be regarded as risk factors that naturally pre-suppose a limit on the ability of actors to forecast risk at the feasibility stage which arguably validates our division of risk components along project cycle criteria.

It could be argued that Cluster 2 is really incorporates experiences of risks relating to earlier stages and also of new risks uniquely related to project implementation. On an aspect noted earlier, it was mentioned that no risk factors relating to *Implementation* and *Operation and Maintenance* survived the cut-off of the extracted factors. This now seems not to be an oversight on the part of IT professionals, but a perspective on risks that they could not agree on since they did not emerge as factors exhibiting sufficient correlations to factorise. It is possible that of those risk factors identified in Cluster 2, the risk factors observed became evident during project build and completion and are therefore the consequences of the risks relating to *Implementation* and *Operation and Maintenance*. In this respect, the wider context of risks relating to 'Project and Team Management', as was labelled Cluster 2, more accurately reflect the risk perspective of IT professionals of risk factors identified at some distance from the organisational detail and context.

**Cluster 3** is defined as 'Technology requirement' and this is comprised of components 2 and 4 and represents 35.15% of the total variance explained. The range of risks in this cluster anticipate a wide range of problems but, as with Cluster 2, they appear to involve both the crystallisation of risks not earlier anticipated in Cluster 1 and of new risks emerging that relate to the inadequacies of various aspects of technology as they first become operational. With respect to failure of planning, it was noted that the inappropriate choice of technology at the feasibility stage was identified in this cluster along with unclear and inadequate identification of systems requirements and even incorrect systems requirements. Of the risks that are likely to appear only once the project is partly implemented, the underperformance of technology is a dominant feature.

This manifests in terms of factors including the narrow focus; inadequacy of development methodologies, programming languages, and working models, and programme or module failure; the use of new, immature, highly complex or even outdated technology; misinterpretation of systems requirements and conflicting systems requirements; and the risks relating to project testing, specifically that of extended time periods and inadequate validation. As before, it was noted that there is a build-up of risks from failures at earlier stages combined with risks that are unique to technology that could only become apparent at a point when some part of the project has been implemented. This cluster also is seen as the technology context for *Implementation* and *Operation and Maintenance* risk factors that did not earlier survive the cut-off. This follows naturally, given the perspective of IT professionals.

**5.2 Clustering of Risk impact on cost overrun**

Table 10 details the results relating to the factor reduction and clustering of risks relating to their impact cost overrun. Again, only factors with eigenvalues of more than 1 are retained and seven risk components are identified. Using the same concept of clustering of the Likelihood occurrence of risk previously explained, only 45 risk factors of the impact on cost overrun out of the 104 initially surveyed was selected which account for 87.92% of the variance that could be explained. The 45 risk factors were selected based on the '*eigen-one*', Kaiser (1974) criterion cut-off, which used only risk factors that have factor loading of 0.400 and above. The major components (those with the largest variances and with variances reported in parenthesis) are: 1 (27.14%), 2 (18.81%), 3 (15.02%) and 4 (12.18%).

Given the discussion in the literature concerning the role of cost in project IT risk analysis, it is no surprise that the interpretability and mapping of risk onto a cost view of the project life cycle extracts similar risk factors, although their clustering is somewhat different. This can be considered to be an early validation of the cost perspective as a relevant and distinct partition of risk. In looking for themes with which to analyse the components and their factor content in the context of the project life cycle, project clusters were formed from the identified components and their rotated risk factors. Three clusters are identified and, as before, labelled them *Cluster 1*: Feasibility study; *Cluster 2*: Project and Team Management; and *Cluster 3:* Technology Requirement.

**Cluster 1** comprises component 5 only and represents 9.49% of the explained variance. The risk factors are identified with the likelihood of occurrence and with an improved level of explanation (variance of 2.56%, previously reported for the likelihood of occurrence). Component 5 loads positively on risk factors F1 and F4. One view, which the research tentatively offer, is that risk relating to an adequate feasibility study is likely to be interpretable fully in terms of a cost outcome. Moreover, given the proportion of the variance explained (from the rotated factors), it is clear that IT professionals judge this to be a crucial stage of the project life cycle. This suggests that it is possible to take a view of the success of a project, and its risk factors, from an early stage in respect of both their likelihood of occurrence and of their impact on the potential for cost overrun. Our result here appears to be similar to other results. In particular, to the analysis of risk by Han and Huang (2007) on components and project performance which reveals that the composite impact of the planning and systems requirement risk dimensions showed a higher impact on the cost performance of the project. Na et al. (2007) reported that functional systems requirements risks were positively correlated with the cost overrun of software projects.

Table 10: New clustering for risk impact on cost overrun

| Cluster 1 | Component | % variance | Main risk factors | Total % variance |
|---|---|---|---|---|
| **Cluster 1 Feasibility study** | Component 5 | 9.493 | F1: Wrong justification of cost benefits analysis from feasibility study.<br>F4: Wrong justification of investment alternatives and opportunity cost. | 9.493 |
| **Cluster 2 Project team management** | Component 1 | 27.143 | Overlooked the management and business impact issues: F3<br>Unclear project scope + objectives: P1<br>Inaccurate estimate of resources: P6<br>Improper planning of timeframe for project reviews and updating: P14<br>Lack of users involvement and commitment: D15<br>Lack of commitment among development team members: D22<br>Ineffective and inexperienced project manager: D23 | 32.424 |
| | Component 6 | 3.559 | Improper change management planning: P5<br>Unrealistic project schedule: P7<br>Unclear line of decision making authority throughout the project: P11<br>Lack of contingency plan/back up: P12<br>Improper handover from the requirement team: D1<br>Resources shifted from project during development: D13<br>Change in management during development: D14<br>Ineffective communications within development team members: D17 | |
| | Component 7 | 1.722 | Undefined project success criteria: P2<br>Lack of quality control procedure and mechanism: P3<br>Project milestones for stages not well establish: P4<br>Inadequate detail breakdown structure: P8<br>Critical and non-critical activities of project not identified: P9<br>Inadequately trained development team members: D19 | |
| **Cluster 3**<br><br>**Technology requirement** | Component 2 | 18.809 | Too narrow focus on the technical IT issues: F2<br>Inappropriate technology chosen from the feasibility study: F5<br>High level of technical complexities: D5<br>Project involves the use of new technology: D6<br>Immature technology: D8<br>Technological advancements and changes: D9 | 46.006 |
| | Component 3 | 15.022 | Unclear and inadequate identification of systems requirements: R1<br>Incorrect systems requirements: R2<br>Misinterpretations of the systems requirements: R3<br>Conflicting system requirements: R4<br>Gold platting or unnecessary functions and requirements: R5<br>Inadequate validation of the requirements: R6<br>Lack of users involvement in requirement stage: R7 | |
| | Component 4 | 12.175 | System conversion method not well planned: P13<br>Inappropriate development methodology used: D2<br>Unsuitable working model and prototype: D3<br>Programming language and CASE tool selected not adequate: D4<br>Difficulty in defining the input and output of system: D7<br>Failures and inconsistencies of unit/modules test results: D10<br>Failure of user acceptance test: D11<br>Time consuming for testing: D12 | |

One potentially significant difference between the cost overrun analysis of this section and of the risk occurrence analysis of the previous section is that risk factors comprising this cluster are identified in a single component in the cost overrun analysis whereas they were identified as two separate single risk factor components earlier on. This suggests that the cost overrun view of risk provides a stronger message in terms of interpretability of risk as it relates to IT software projects than dealing with risk occurrence more generally. This point is elaborated below following further evidence relating to the remaining clusters.

**Cluster 2** comprises components 1, 6, and 7 and accounts for over 30% of the variance. This cluster is mainly composed of project development risks all of which have very high loadings and which are positively correlated. In common with the earlier analysis of risk occurrence, a partition of risk factors was observed between those that are unique to the stage in the project cycle which the cluster is mostly closely associated – that is, post-feasibility study – and of risk factors emerging as a consequence of risk emerging from inadequate planning and foresight at the feasibility stage or in Cluster 1. Thus, it can be seen that the same risk factors emerging but with some re-organisation into fewer components which we take to be indicative of a stronger and clearer message concerning risk identification and impact. For example, there are now fewer components to interpret as a single cluster compared with the components relating to risk impact.

In Cluster 2, the reduction from five to three components as the research moves from risk occurrence to cost overrun. In relation to risk occurrence, components 1 and 3 were separately identified components within Cluster 2. Both contain risk factors relating to user involvement. In Cluster 2, the cost overrun analysis these two components largely merged into a single component. As with the merging of components identified in relation to the feasibility study in the risk occurrence results, there is a similar effect for cost overruns. Again, it appears that taking a cost overrun view of risk perhaps clarifies the commonalities underlying risk that are determined by cost impact. The research argued previously that it was possible to justify the separation of components observed in relation to risk occurrence on the grounds of updating of risk and that, once a project team had become involved, it was possible to discern aspects of risk that would emerge following implementation of some aspect of the project which was a view not available for those involved in the feasibility study specifically. From a cost overrun basis, it does not appear to matter that this division or sequencing of risk factors is relevant. This is also observed with the other components. Thus, component 6 of the cost overrun Cluster 2 has much in common with components 5 and 6 of the risk occurrence analysis. Component 7, the final component of Cluster 2 for cost overrun, appears to map fairly directly with its counterpart component 8 of the risk analysis section.

**Cluster 3** is comprised of components 2, 3, and 4 and accounts for 46.01% of variance. The fact that there were only two components identified for the risk occurrence analysis but three for the cost overrun appears to contradict the research arguments concerning the consolidating effects of a cost overrun view of risk. Whilst each component for both sections is comprised of risk factors satisfying the extraction cut-off tests, it should be noted that the Cluster 3 of both risk occurrence and cost overrun indicates that risk factors relating to a too narrow focus on technical issues, choice of inappropriate technology, high complexity, immature technology, and out-of-date technology are key drivers of risk as they relate to technology requirements. Thus, in terms of both risk occurrence and cost overrun, there does not appear to be a divergence of opinion in terms of the general thrust of which factors are loading on the respective components.

Both groups of clusters have total percentage of variance below 90% for the likelihood of occurrence and risk impact on cost overrun. In fact, through the factor reduction and clustering process, less than 13% of information is compromised (Table 11)

Table 11: Percentage of variance for clusters

| **Cluster** | **% variance of likelihood occurrence** | **% variance of risk impact on cost overrun** |
|---|---|---|
| Feasibility study | 2.56% | 9.49 % |
| Project team management | 51.52% | 32.42 % |
| Technology requirement | 35.15% | 46.01 % |

Significant contrasting views were found in the user related risk factors such as failure of user acceptance test, users resistance to change, and failure to manage the users expectations. Although these factors rarely happen, they could make a drastic impact on cost overrun once they occur. Jiang et al. (2002) and Barki et al. (1993) suggested user related risks as the extent to which prospective software users participate in software project development, their readiness to accept the proposed software system, their attitude toward the software and their experience in software project development. These issues make it difficult to understand and to predict the users' expectations and requirements, and thus the completion of the final project within the timeframe and the budget allocated. Jiang & Klein (2000) stated that poor communications among development team members does not allow for the coordination necessary to conduct the individual tasks required to complete the project in an orderly fashion. Much of the time might be spent on duplication of efforts and progress will be towards individual's goal rather than the project goal. A research carried out by Proccaccino et al. (2005) also highlighted the importance of actively nurturing effective communication that improves interpersonal relationships of their team members. As pointed out by Linberg (1999) and Glass (1999), particular attention should also be given to internal intrinsic items, as they relate to motivation and productivity. Other factors such as inappropriate development methodology used, inexperienced project managers, high level of technical complexities, time consuming for testing, and lack of regular reviews were rated high in terms of the likelihood of occurrence but low in their impact on the cost overrun. Project management literature also suggested the importance of the communications between the development team and the users in defining the project scope and controlling the project changes. Lapses in these tasks could lead to increased uncertainty throughout the development cycle and could contribute to project overruns. The project managers and the development team need to build, create and maintain good relationship and trust with the users, to avoid being caught in a situation where supports and commitments for the project suddenly evaporates.

## 6.0 CONCLUSION AND FUTURE RESEARCH

The research sets out to explore what IT professionals thought of IT software development risks and took the project life cycle as the research risk construct, and identified 3 main clusters of risk for both the likelihood occurrence and its impact on cost overrun. We find that the likelihood of occurrence of each factor is not necessarily related to the impact on cost overruns. Specifically, though the likelihood of difficulties in the feasibility study may be low, they can lead to significant overruns. Meanwhile, problems with project team management may be most common, the overall impact on cost overruns is relatively modest. Based on this research, professionals should spend most time and resources evaluating and managing risks relating to the technology requirements, as these represent the most significant cost overruns, even though they are only the second most likely to occur.

Future studies should address risk mitigation plan development processes and build on this research to establish a stronger connection between the factors identified and actual cost overruns. This would address the key limitation of this study: we have identified factors grounded in the opinions of the managers, we have not connected these to actual outcomes. Therefore, future research must use these factors and establish whether they have explanatory power in the management of risk by using objective project performance measures such as the actual magnitude of cost overruns, and the measured risks associated with these factors. Therefore, this type of check will point the way to managers balancing these phenomena to prioritize risks and to develop an appropriate risk mitigation plan which would either emphasize reducing the likelihood of a risk or reducing the impact on the cost overrun. Our research design has a self-selection bias that could not be adequately corrected given the current design, yet it is something that should be accounted for in future studies, perhaps by measuring correlation between the experience with or frequency of such overruns and the magnitude of such.

## REFERENCES

Appari, A., and Benaroch, M. (2010), "Monetary pricing of software development risks: A method and empirical illustration", *Journal of Systems and Software*, 83(11), 2098-2107.

Baccarini, D., Salm, G., Surrey, P. (2004), "Management of risks in Information Technology projects", *Journal of Industrial Management & Data Systems*, 104(4), 286-295.

Ballantine, J & Stray, S. (1998), "Financial appraisal and the IS/IT investment decision making process", *Journal of Information Technology*, 13 (1), 3-14.

Barki, H., Rivard, S., and Talbot, J. (1993), "Toward as assessment of software development risk", *Journal of Management Information Systems*, 10(2), 203-223.

Capilla, R., Bosch, J., Trinidad, P., Ruiz-Cortés, A., Hinchey, M. (2014), "An overview of Dynamic Software Product Line architectures and techniques: Observations from research and industry", *Journal of Systems and Software*, 91, 3-23.

Dingsoyr, T., Nerur, S., Balijepally, V., Moe, N.B. (2012), "A decade of agile methodologies: Towards explaining agile software development", *Journal of Systems and Software*, 85(6), 1213-1221.

Glass, R.L. (1999), "Evolving a new theory of project success", *Communications of the ACM*, 42(17), 17-19.

Han, W.M., and Huang, S.J. (2007), "An empirical analysis of risk components and performance on software projects", *Journal of Systems & Software*, 80(1), 42-50.

Jiang, J., and Klein, G. (2000), "Software development risks to project effectiveness", *Journal of Systems & Software*, 52, 3-10.

Jiang, J., Klein G., Chen, H.G., and Lin, L. (2002), "Reducing user-related risks during and prior to system development", *International Journal of Project Management*, 20(7), 507-515.

Jorgensen, M. (2010), "Identification of more risks can lead to increased over-optimism of and over-confidence in software development effort estimates", *Information and Software Technology*, 52(5), 506-516.

Kaiser, H.F. (1974), "An index of factorial simplicity", *Psychometrika*, 39, 31-36.

Kass, R.A. & Tinsley, H.E.A. (1979), "Factor Analysis", *Journal of Leisure Research*, 11, 120-138.

Keil, M., Tiwana, A., Bush, A. (2002), "Reconciling user and project manager perceptions of IT project risk: a Delphi Study", *Information Systems Journal*, 12(2), 103-119.

Linberg, K.R. (1999), "Software developer perceptions about software project failure: a case study", *Journal of System & Software*, 49, 177-192.

Maccallum, R.C., Widaman, K.F., Zhang, S., Hong, S. (1999), "Sample size in factor analysis", *Psychological Methods*, 4(1), 84-99.

Magdaleno, A.M., Werner, C.M.L., de Araujo, R.M. (2012), "Reconciling software development models: A quasi-systematic review", *Journal of Systems and Software*, 85(2), 351-369.

Marina, P., Simmonds, J., Astudillo, H. (2014), "Semi-automated Tool Recommender for Software Development Processes", *Electronic Notes in Theoretical Computer Science*, 302(25), 95-109.

Martín, Y., Yelmo, J.C. (2014), "Guidance for the Development of Accessibility Evaluation Tools Following the Unified Software Development Process", *Procedia Computer Science*, 27, 302-311.

Morgan, G.A., Leech, N.L., Gloeckner, G.W., Barret, K.C. (2004), *SPSS for introductory statistics: Use and interpretation, 2nd edition*. Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey.

Na, K.S., Simpson, J., Li, X.T., Singh, T., Kim, K.Y. (2007), "Software development risk", *Journal of Systems & Software*, 80(4), 596-605.

Naquin, C.E., and Tynan, R.O. (2003), "The team halo effect: why teams are not blamed for their failures", *Journal of Applied Psychology*, 88, 332-340.

Nidumolu, S. (1996), "A comparison of the structural contingency and risk-based perspectives on coordination in software development projects", *Journal of Management Information Systems*, 13(2), 77-113.

Portillo-Rodríguez, J., Vizcaíno, A., Piattini, M., Beecham, S. (2014), "Using agents to manage Socio-Technical Congruence in a Global Software Engineering project", *Information Sciences*, 264(20), 230-259.

Procaccino, J.D., Verner, J.M., Shelfer, K.M., and Gefen, D. (2005), "What do software practitioner's really think about project success: an exploratory study", *Journal of Systems & Software*, 78(2), 194-203.

Punch, K.F. (2005), "Introduction to Social Research: Quantitative & qualitative approaches. 2nd edition", SAGE Publications, London.

Reyes, F., Cerpa, N., Candia-Véjar, A., Bardeen, M. (2011), "The optimization of success probability for software projects using genetic algorithms", *Journal of Systems and Software*, 84(5), 775-785.

Ropponen, J., Lyytinen, K., (2000), "Components of software development risk: How to address them? A project manager survey", *IEEE Transactions on Software Engineering*, 26 (2), 98-112.

Schmidt, R., Lyytinen, K., Keil, M., and Cule, P. (2001), "Identifying software project risks: An international Delphi Study", *Journal of Management Information Systems*, 17(4), 5-36.

Stahl, D., Bosch, J. (2014), "Modeling continuous integration practice differences in industry software development", *Journal of Systems and Software*, 87, 48-59.

Stevens, J.P. (1995), *Applied multivariate statistics for the social sciences, 2nd edition*, Erlbaum Publications. Hillsdale, NJ.

Thüm, T., Kastner, C., Benduhn, F., Meinicke, J., Saake, G., Leich, T. (2014), "FeatureIDE: An extensible framework for feature-oriented software development", *Science of Computer Programming*, 79(1), 70-85.

Wallace, L., and Keil, M. (2004), "Software project risks and their effect on outcomes", *Communications of the ACM*, 47(4), 68-73.

Wallace, L., Keil, M., Rai, A. (2004), "How Software Project Risk Affects Project Performance: an investigation of the dimensions of risk and exploratory model", *Decision Sciences*, 35(2), 289-321.