Chapter 3
Research Problem and Methodology

Examination of the existing literature confirms that the organizations are embracing distributed agile development for creating low cost and flexible solutions. Further, combining distributed development with agile approach involves significant risks which need to be addressed. Although, risk management is important for software development, there has been dearth of some comprehensive work for managing risks in distributed agile development (DAD). This study aims at investigating the most significant risk factors which impact distributed agile development projects and suggests the most frequently used methods to control them. Hence, we define our research problem in the following research question.

3.1 Research Questions and Objectives

Research Question 1: What are the most significant risk factors in DAD projects as perceived by the practitioners?

Software project risk points to an aspect of development task, process or environment, which if ignored tends to adversely affect the outcome of a software project (Lyytinen et al., 1993). A risk factor can be defined as a condition that can present a serious threat to the successful completion of a software development project (Schmidt et al., 2001). We decided to conduct an extensive exploratory qualitative study for unearthing the risks which affect the successful delivery of DAD projects by conducting in-depth interviews and analysis of DAD project work documents. We believed that identifying the most important risk factors from the set of factors which would be generated by exploration was necessary. Hence, we further decided to use quantitative descriptive study to detect those risk factors which are perceived to be the most important for DAD projects.

Risk management is concerned with a phased and systematic approach to analyze and control risks occurring in a specific context (Charette, 1989). Hence, along with identifying the risk
factors which may hinder the success of DAD projects, we also need to analyze the risk management approaches which can be used to lessen the effect of those risks.

**Research Question 2:** What are the risk management techniques frequently used in practice for managing the identified risk factors?

The objective of this work is to make the risks impacting DAD projects visible to the practitioners so that they can predict the risks which may impact their project and take measure accordingly to control them. This is possible when the logically related risk factors are grouped together to create a risk classification. This classification of risks will help the practitioners managing DAD projects to analyze the risks and their causes in a more economical and effective way (Hoodat & Rashidi, 2009; Pandian, 2007). Hence, we formulated Research Question 3 as stated below.

**Research Question 3:** How are the risk factors categorized to facilitate real world adoption?

Project involves unique or risky undertakings which have to be completed by a certain date, for a certain amount of money and within some expected level of performance. The triple criteria of project success require meeting cost, schedule and performance targets (Barki et. al., 2001; Nidumolu, 1995). Hence, it becomes necessary to study the constraints which are being impacted by those risks in DAD projects. The following question addresses the concern related to analysis of project constraints on the occurrence of a risk factor in DAD project.

**Research Question 4:** Which of the project constraints of Time, Cost and Quality are impacted most significantly by specific risk factors identified in the exploratory study from the perspective of the practitioners?

The research questions lead to the following research objectives for this study:
1. To identify the conditions (risk factors) which impact the success of DAD projects.
2. To identify the dimensions of those risk factors which include, description, cause of the risk, software development methodology which is the source of the risk and the corresponding risk management methods which can be used to reduce their impact.

3. To create a categorization of risk factors impacting DAD project by grouping related risk factors together.

4. To segregate the most important risk factors from the identified set of factors which impact the DAD projects.

5. To provide the frequently used risk management approach for the risk factors so that the practitioners can control the risk more effectively.

6. To identify the risk factors which may lead to time overrun, those risk factors which may lead to cost overrun and those risk factors which may lead to low quality solutions.

7. To develop a risk management framework that can easily be adopted by the organizations and signifies the important risk factors, corresponding risk management techniques in practice and the impact of risk factors on the project constraints of time, cost and quality.

3.2 Research Methodology

The existing literature was studied, which helped us to understand the type of risks and challenges frequently encountered by the distributed agile teams. These studies also gave insights on the risk management approaches which are used to control the risks. Literature survey gave us a limited view of risks in DAD methods, and hence we decided to conduct an extensive exploratory study involving in-depth interviews and project work document analysis.

In order to strengthen the finding of the exploratory study, we planned to conduct a quantitative descriptive study too. The details of the two studies are given in this section.

3.2.1 Methodology for Qualitative Exploratory Study

Exploratory approach is useful when the area of investigation is vague and the variables to be studied are unknown or not thoroughly defined (Cooper & Schindler, 2006). We decided to use exploration because there was scarcity of work in this area and an extensive view of the risks and challenges faced by DAD teams was not available. Exploration relies more on qualitative techniques like individual in-depth interviews, participant observations, films and photographs case studies, document analysis etc. We decided to use ‘in-depth interviews’ of practitioners in
order to identifying the risks in DAD projects. We planned to conduct in-depth interviews in the form of an experience survey, in which we will look for the ideas or views of the practitioners (Cooper & Schindler, 2006), who have experience in managing DAD projects, on the risks they have encountered and the methods they used to overcome them. Besides, the information on risk factors and the corresponding management approaches, we also planned to collect data pertaining to the cause of the risk factor and the source of the risk factor in terms of the software development approaches being used. The risk factors which are caused due to blending of distributed development and agile will be referred to as ‘DAD Risk Factors’, while other risk factors will be termed as ‘Non DAD Risk Factors’. Non DAD risk factors will be the ones that are caused due to other methods of software development like agile development, distributed development and traditional development. We decided to pilot test the questionnaire for exploratory data collection in order to improve its quality. It will help us to obtain relevant data which can provide us good insights in the problems faced by practitioners while managing DAD projects.

Qualitative methods usually rely on combining the data from multiple methods and sources of information. This principle is known as triangulation (Denzin & Lincoln, 1970), and it helps in reducing the possible bias in the findings of the research study, which may be introduced due to the limitations of a specific method. It helps in validating the results obtained from one method of data collection and further leads to development of a generalized solution (Maxwell, 1992). Hence, we decided to supplement our findings from exploratory in-depth interviews with that of ‘Work Document Analysis’ of DAD projects.

The in-depth interviews and project work documents would provide us with the qualitative data describing risk factors in DAD projects and the corresponding risk management approaches as used by the practitioners. It was required to use qualitative methods for analysis of this information in order to create an initial risk categorization for DAD projects. We found ‘Constant Comparison Method’ suitable for the analysis of the qualitative data (Glaser & Strauss, 1967).
Constant comparison method is concerned with generating and plausibly suggesting (not provisionally testing) many properties and hypotheses about a general phenomenon (Glaser, 1965), which in our case, is the Risk Management in Distributed Agile Development (DAD) projects. This method involves four stages (Glaser, 1965):

1. **Comparing incidents applicable to each category**
   At this stage, the analyst starts coding each incident in his data in as many categories as possible. While coding an incident for a category, it is compared to the previous incidents coded in the same category. This process leads to the generation of theoretical properties of the category and relationship of these categories with other categories. After coding three to four times, the analyst needs to stop coding and record a memo on ideas. At this point there is a need to reflect back over the theoretical notions to determine alternate ways of coding. This will help him to bring clarity in his ideas, with which he can further go back to the data for more coding and constant comparison.

2. **Integrating categories and their properties**
   As the coding continues, the units of constant comparison method changes from comparison of incidents to incidents with properties of the category which resulted from the initial comparison of the incidents. Hence, accumulated knowledge of properties of categories which comes due to constant comparison starts integrating. Hence, categories start getting integrated with other categories and their properties, which further leads to theory development.

3. **Delimiting theory**
   As the theory develops, various delimiting at the level of (1) theory and (2) original list of categories proposed for coding takes place. As the next incidents are compared, major modifications becomes fewer and fewer. These later modifications are mainly on the order of logical clarity, pairing off non-relevant properties, integrating elaborating details of the properties into major outline of interrelated properties and then reduction. Reduction helps us to create higher level concepts, which are smaller in number. This reduction in terminology leads to a theory which can be generalized. Delimiting theory also results in a delimiting of the original list of proposed categories for coding. The categories then become theoretically saturated. After one has coded incidents for the same category a number of times, it can be easily seen if the next incident points to a new
aspect of the category. If yes, then the incident is coded and compared. If no, the incident is not coded, since it is not adding anything new to the theory. Hence, the delimitation of the universe of data economizes the research resources as it forces the analyst to spend his time and effort on data which is relevant only to his categories.

4. **Writing theory**

At the end of the process the analyst has coded the data, a series of memos are created and theory has been developed. The analyst now needs to collate all the memos on each category, which forms the major themes as written in books or papers.

We will be using exploratory study to gather qualitative data corresponding to risk factors in DAD projects, which will be converted into a risk categorization. Constant comparison method would help us to group the related risk factors together to form a ‘risk area’, which would further be used to create broad risk categories. In our study, the information on risk factors, which we will obtain from the respondents will correspond to the ‘incidents’ given in the explanation of constant comparison method (Glaser, 1965). These incidents will be compared and the properties of the risk factors will emerge. Based on the properties of the risk factors, they will be grouped to form risk areas. Integration of related areas will further lead to the formation of risk categories. The risk factors at one level, risk areas at second level and risk categories at higher level will lead to the formation of an initial risk categorization.

This will be followed by the analysis of the data that will be obtained from project work documents of DAD projects in order to strengthen the risk categorization created. Iterative use of constant comparison method will help us to consolidate all the qualitative information into concrete risk categories containing risk areas and further risk factors.

The risk categories that will emerge will then be mapped to certain theoretical themes. We chose ‘Leavitt’s Model of Organizational Change’ since, relationship between the components of Leavitt’s model and various aspects of software development was very apparent.

As discussed in the previous chapter, according to Leavitt’s model (1964) organizations form multivariate systems consisting of four interacting components —task, structure, actor, and
technology. Leavitt’s’ proposes that these component can be translated into well known elements of software development ‘actors’ cover all stakeholders including users, managers, and designers; ‘structure’ denotes project organization and other prevailing institutional arrangements; ‘technology’ means development tools, methods, and hardware and software platforms; and ‘task’ signifies expected outcomes in terms of goals and deliverables.

Leavitt’s model has been related to the concept of software risk as follows: a change in any Levittian component in a system development process can create disturbances which in worst case may lead to software failure. Such disturbances are perceived as risky incidents which make the estimation of performance of the development project difficult (Nidumolu, 1995). Risky incidents are events or states in the real world which have a potential to cause loss and thus make the development project to fail. Risk factors (risk items) are derived from postulated causal dependencies between the risky incidents and losses (Lyytinen et al., 1998).

According to Leavitt, these risky incidents always fall into one or several Levittian elements and their connections (Lyytinen et al., 1998). Hence, we anticipated that we will be able to relate all the risk factors (risk items) which we plan to identify through exploratory study, to the components for Leavitt’s model. This will help us to take into account all the aspects of software development from the organization’s perspective from where the risks may arise and become a threat to the success of a DAD project.

3.2.2 Methodology for Quantitative Descriptive Study
We will be creating a risk factor categorization for DAD risks by using the data collected through the exploratory study, which will be containing number of risk factors which impact DAD projects. In order to identify those risk factors which are most significant in DAD projects from the view of the practitioners, we decided to conduct a quantitative descriptive study of risk factors. This will help the practitioners to concentrate on those risks which will pose threat to the DAD projects to a larger extent and hence needs more attention.

We further planned to obtain quantitative information for risk management approaches corresponding to all the risk factors based on the ‘frequency with which they are being used in
the industry’. This will provide the practitioners, with a set of ‘most frequently used methods’ for controlling the risks, which will in turn help them to perform more effective risk management for DAD projects.

Along with obtaining the data on importance of risk factors and risk management methods, we also decided to gather data corresponding to the project constraints which are being impacted by those risk factors. It was decided to use the triple constraints, time, cost and quality for this purpose. Since agile methods consider changing the scope instead of compromising on quality, schedule or resources (Cohn, 2010), we decided to not to consider the impact of risk factors on the ‘scope’.

3.2.2.1 Proposition and Hypothesis

Hence, based on our research problem, we propose that

1. The identified risk factors for distributed agile development projects are perceived by practitioners to have an impact on the success of the projects.
2. The proposed risk management approaches for the risk factors are frequently used to reduce the impact of the risk in distributed agile development projects.
3. The identified risks occurring in DAD projects also have impact on the project constraints, namely, time, cost and quality.

Our proposition is operationalized by the following set of hypothesis:

3.2.2.1.1 Hypothesis for Risk Factors in DAD Projects

1. $H_{0RC}$: The identified Risk Categories do not have an impact on the success of DAD project.
   $H_{1RC}$: The identified Risk Categories have an impact on the success of DAD project.

2. $H_{0RA}$: The identified Risk Areas do not have an impact on the success of DAD project.
   $H_{1RA}$: The identified Risk Areas have an impact on the success of DAD project.

3. $H_{0RF}$: The identified Risk Factors do not have an impact on the success of DAD project.
   $H_{1RF}$: The identified Risk Factors have an impact on the success of DAD project.
3.2.2.1.2 Hypothesis for Risk Management Methods for DAD Projects
1. $H_{0RM}$: The suggested Risk Management Approaches are not used frequently for reducing the impact of the risk factors in DAD projects.
   $H_{1RM}$: The suggested Risk Management Approaches are used frequently for reducing the impact of the risk factors in DAD projects.

3.2.2.1.3 Hypothesis for Impact of Risk Factors on Project Constraints in DAD projects
1. $H_{0T}$: The identified Risk Factors do not have an impact on the Time of the DAD Project.
   $H_{1T}$: The identified Risk Factors have an impact on the Time of the DAD Project.

2. $H_{0C}$: The identified Risk Factors do not have an impact on the Cost of the DAD Project.
   $H_{1C}$: The identified Risk Factors have an impact on the Cost of the DAD Project.

3. $H_{0Q}$: The identified Risk Factors do not have an impact on the Quality of the DAD Project.
   $H_{1Q}$: The identified Risk Factors have an impact on the Quality of the DAD Project.

We plan to test the above listed hypothesis by conducting the quantitative descriptive study. We describe the planned methodology for quantitative study and the statistical methods which were deemed to be appropriate for analyzing the quantitative data.

3.2.2.2 Analysis Methods to be used for Quantitative Data
1. Ranking Data of Risk Categories, Risk Areas and Risk Factors: It was decided to use the Kendall’s Test of Concordance on the rank order data in order to find out the extent to which the respondents agree on the ranking. This test would be applied for risk categories, risk areas and risk factors separately since we will be obtaining the ranks for risk categories at level 1, ranks for risk areas under each risk category at level 2 and ranks for risk factors under each risk areas for each of the risk category at level 3.

2. Rating Data of Risk Management Approaches: This data would represent the rating given to the risk management approaches for each risk factor. In order to find out the
statistically significant ‘frequency of usage’ of each risk management method to control a particular risk, we will use Chi-square test of Independence.

3. **Project Constraint data for Risk Factors**: We will be obtaining the project constraint which is being most significantly impacted when a particular risk occurs in DAD project. We decided to use chi-square test in order to identify the risk factors which are statistically significantly impacting one of the project constraints. In the section below, we present how application of various tests on the quantitative data of risk factors and risk management methods.

3.2.2.2.1 Methodology for Risk Factors Rank Order Data Analysis

The risk categorization will provide us with set risk factors which will be grouped in risk areas and further broad level risk categories. We want to identify the most important risk categories, the most important risk areas under each risk category and the most important risk factor under each risk areas. Hence, we decided to obtain ranking data corresponding to risk categories, risk areas and risk factors at three different levels. A self-administered questionnaire will be used to obtain the ranks for risk categories, risk areas and risk factors.

The ranking of risk factors will be done based on the ‘severity of impact on the DAD projects’. A ranking task requires the respondent to rank order a small number of items on the basis of overall preference or some characteristics of the stimulus (Cooper and Schindler, 2006). In our study the ‘characteristic of stimulus’ corresponds to the severity with which the risk impacts the success of a DAD project. The scale to be used for ranking the risk factors is defined as follows: 5: very strong impact, 4: strong impact, 3: some impact, 2: negligible impact, 1: no impact. The ‘Don’t Know Condition’ will also be considered while collecting the ranking data for risk factors. We will allow the respondents to give a rank ‘0’ if they are not aware of a particular risk factor.

At the first level, the respondents will be required to provide the categories which have relatively stronger impact on the DAD projects than others. The second level, we will obtain the relative ranks of the risk areas based on their impact on the success of DAD projects. The risk areas which are considered to have comparatively stronger impact on DAD projects will be given
higher ranks. Finally, at the third level, we will obtain the ranking of risk factors under each risk areas. As described above, the same scale of 5 to 1 will be used to obtain ranked data of risk areas and risk categories as well.

We planned to perform a pilot test for this study also by sending the questionnaire to experts and obtaining their views on the content and quality of the questionnaire.

**Kendall’s Test of Concordance for Risk Factor Ranking**

In order to measure the level of agreement of the respondents on the ranking of the risk factors it will be required to use Kendall’s test of Concordance (W) on the rank data of risk factors (Kvam & Vidakovic, 2007). The value of Kendall’s coefficient ‘W’ ranges from 0 to 1, where 0 reveals perfect disagreement and 1 reveals perfect agreement in the ranking of the respondents (Gibbons, 1993). The same applies to ranking of risk areas and risk categories.

Kendall and Smith (1939) provided a descriptive measure of agreement or concordance for data comprised of M sets of ranks, where M > 2. We consider an artificial rank structure as presented by Table 3-1. We have two variables, X and Y. Variable X consist of values $x_i$, $i = 1... N$ with N being the number of ranks in each set of ranks. Variable Y consists of values $y_j$, $j = 1... M$ with M being the number of sets of ranks. Each value of variable X, $x_i$, has a rank $r_{ji}$, assigned by the value of the ranking variable Y, $y_j$. $R_i$ is the rank total for value $x_i$ of the variable X.

If perfect agreement were observed between the j values of the ranking variable, one value of the variable X would be assigned a ‘1’ by all j values of the ranking variable, and the rank total would be M. Another value of the variable X would be assigned a 2 by all j values of the ranking variable, and the rank total would be 2M. Therefore, when perfect agreement exists among ranks assigned by M values of the ranking variable, the rank totals are M, 2M, 3M, ...... NM. The total sum of ranks for M values of the ranking variable is $MN (N + 1) / 2$, and the mean rank sum is $M (N + 1) / 2$.

The degree of agreement between the values of the ranking variable reflects itself in the variation in the rank totals (Ferguson, 1966, pp. 225–226). When all the values of the ranking variable are
in agreement, this variation is at a maximum. Disagreement between the values of the ranking variable reflects itself in a reduction in the variation of rank totals. For maximum disagreement, the rank totals tend to be equal.

Since \( R_i \) is the rank total for value \( x_i \) of the variable \( X \), the sum of squared deviations of rank totals from the average rank total for \( N \) values of variable \( X \) is:

\[
S = \sum_{j=1}^{M} (R_j - \bar{R})^2
\]  

where \( \bar{R} = \frac{1}{N} \sum_{j=1}^{M} R_j \)  

The maximum value of this expression occurs when perfect agreement exists between the values of the ranking variable. This value is given by the expression:

\[
\frac{1}{12} M^2 N (N^2 - 1)
\]  

The coefficient of concordance, \( W \), is defined as the ratio of sum of squared deviations of rank totals from the average rank total to the maximum possible value of the sum of squared deviations of rank totals from the average rank total (Kendall and Smith, 1939; cf. Kendall, 1970).

\[
W = \frac{12S}{M^2N (N^2 - 1)}
\]  

The measure of concordance is given by expression (3). When perfect agreement exists between the values of the ranking variable, \( W = 1 \). When maximum disagreement exists, \( W = 0 \). Kendall’s coefficient of concordance (\( W \)) does not take negative values and is thus bounded on the interval \( 0 \leq W \leq 1 \).
Table 3-1: Ranks assigned to N values of variables X by M values of variables Y (Kendall’s Test of Concordance)

<table>
<thead>
<tr>
<th>Variable y</th>
<th>x₁</th>
<th>........</th>
<th>xᵢ</th>
<th>........</th>
<th>xₙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>y₁</td>
<td>r₁₁</td>
<td>.......</td>
<td>rᵢⱼ</td>
<td>.......</td>
<td>rᵢₙ</td>
</tr>
<tr>
<td>...</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
<td></td>
</tr>
<tr>
<td>yᵢ</td>
<td>rᵢ₁</td>
<td>rᵢᵢ</td>
<td>rᵢⱼ</td>
<td>rᵢₙ</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
<td>.......</td>
<td></td>
</tr>
<tr>
<td>yₘ</td>
<td>rₘ₁</td>
<td>rₘᵢ</td>
<td>......</td>
<td>rₘₙ</td>
<td></td>
</tr>
<tr>
<td>Rᵢ</td>
<td>R₁</td>
<td>Rᵢ</td>
<td>......</td>
<td>Rₙ</td>
<td></td>
</tr>
</tbody>
</table>

When this test will be applied to rank order data of risk factors, three quantities will be generated. It will give Kendall’s coefficient of Concordance (W), value of observed significance level (p-value) and the mean rank of risk factors within a particular risk area. The same applies, when this test is will be applied to a set of risk areas under a particular risk category and similarly to the set of five major risk categories. We will be doing this test for significant level of 0.05. The observed p-value will be compared with the significance level (0.05). If observed p-value is less than 0.05, it will indicate that the results will be statistically significant and hence can be generalized for the whole population.

3.2.2.2.2 Methodology for Risk Management Rating Data Analysis

Corresponding to the risk factors in DAD projects, the suitable risk management methods to control them will also be collected during exploration. Since, there would be multiple methods which can be used to reduce the impact of a particular risk, it becomes necessary to find out those methods are used ‘most frequently’ by the practitioners. Hence, it was decided to obtain the rating data for the risk management methods which will be suggested by the findings of exploratory study.

We planned to obtain the rating of risk management method based on the ‘frequency of use’ of the risk management methods to control a particular risk. Rating asks the respondents to estimate the magnitude of a characteristic or quality that an object possesses. It is used when a participant score an object or indicant without making a direct comparison to another object or attitude (Cooper & Schindler, 2006). In our study, corresponding to each risk factor, we would be asking the respondent to provide us the rating for each of the risk management methods in a given set of suggested methods, based on their frequency of use in practice. Amongst the
suggested methods for controlling a particular risk, the methods which are used more frequently in the industry were rated higher than the others. A scale of 5 to 1 (0 for don’t know condition) was used for rating data which is defined as: ‘5’ for methods which are always used, ‘4’ for methods which are often used, ‘3’ for methods which are sometimes used, ‘1’ for those methods which are never used and ‘0’ if the respondents is not aware of the usage of risk management method. Analysis of rating data risk management techniques helped us to identify the methods which are most frequently used by agile practitioners for controlling the risks in DAD projects. Since, we wanted to find out the statistically significant ‘frequency of usage of the risk management methods’, for each risk factor in DAD projects, it was decided to use ‘Chi-Square Test of Independence’.

**Chi-Square Test of Independence**

Chi-Square Test of Independence is used to analyze the frequencies of two variables with multiple categories to determine whether the two variables are independent. For example: Financial investors might want to determine whether type of preferred stock investment is independent of the region where the investor resides. Chi-square test is used to measure any level of data measurement, but it is useful in analyzing nominal data (Black, 2009).

Chi-square test of Independence is used to analyze if the two variables are independent (not related). The null hypothesis of a chi-square test of independence is that the two variables are independent. The alternate hypothesis is that the variables are not independent. In general, if two variables are independent, the expected frequency values of each cell can be determined by

$$e_{ij} = \frac{(n_i \cdot n_j)}{N}$$

where:

i = the row

j = the column

$n_i$ = the total of row i

$n_j$ = the total of column j

N = the total of all frequencies.
Chi-square test of independence uses the *expected frequencies*, which are the theoretical frequencies of the categories from the population distribution and the *observed frequencies*, which refers to the actual frequencies from the distribution, for the calculation of the chi-square test statistic.

A contingency table is created before we apply the chi-square test of independence for analyzing if two variables are associated as shown in Figure 3.1:

![Figure 3.1: Contingency Table used by Chi-Square Test of Independence](image)

The Chi-square test statistic is calculated as follows:

\[
\chi^2 = \sum \sum \frac{(f_o - f_e)^2}{f_e}
\]

where:  \( df = (r - 1)(c - 1) \)

- \( r \) = number of rows
- \( c \) = number of columns

In our study, we intend to use a 5 to 1 point scale which measure the ‘frequency with which the risk management methods is used’ for controlling a particular risk factor. The scale will be defined as follows: 5: Always Used, 4: Often Used, 3: Sometimes Used, 2: Seldom Used, 1: Never Used. We will also allow the respondents to give the rating ‘0’ to the risk management methods if they are unaware of the risk management methods or the risk factor. This will be referred to as the ‘Don’t Know’ condition.
By using chi-square test of independence, we will be able to separate out those risk management methods from the ones identified during exploratory study, that are being used more frequently than the ones that are less frequently used, in the industry. A two-way contingency table will be created, where the ‘rows’ of the table, will contain the risk management methods for each of the risk factor, while the ‘columns’ will represent the ratings which will be given to the risk management methods on a 5-point scale. Each cell of the contingency table will represent the ‘frequency of respondents’ who would give this rating on a 5-point scale to a particular risk management method. This ‘frequency of respondents’ represents the rating given by the set of respondents which are maximum in number.

This test aims to identify the ‘statistically significant frequency’ with which the risk management methods are used, as provided by the set of respondents. For each method, the rating value that has maximum number of corresponding respondents will be considered. The methods which attain the rating ‘5’ or ‘4’ (as given by maximum number of respondents) will be considered as ‘More Frequently Used’ methods, while those which attain the rating ‘3’, ‘2’ or ‘1’ (as given by maximum number of respondents) will be considered as ‘Less Frequently Used’ methods.

Chi-square test on the rating data will generate three values (i) Chi-square test statistic (ii) Observed value of Significance Level (p-value) (iii) corresponding to each of the risk management method in the set of methods that can be used to control a particular risk factor, we will get the ‘frequency of respondents who opted for a particular rating for that method’.

We will be using significance level of 0.05 and will be comparing the p-value with 0.05. Obtaining p-value less than 0.05 will indicate that the rating of the risk management methods is statistically significant. This will imply that the rating obtained by each of the risk management method, corresponding to each risk factor for DAD project, can be generalized.

Hence, we will be able to identify the risk management methods which are statistically significantly more frequently used methods for controlling a particular risk in DAD projects. The results may not represent the ideal methods to be used for the risks, instead, the ones which are
more popular amongst the agile practitioners. The practitioners can select more frequently used methods for managing the risks and hence perform effective risk management for DAD projects.

### 3.2.2.2.3 Methodology for Project Constraint Data Analysis

Along with performing a study on risks in DAD projects and the corresponding risk management methods, we planned to study the project constraints which are being impacted when the risks occur in DAD projects. We planned to consider three project constraints, namely, time, cost and quality, for the project constraints analysis. It was decided to obtain the project constraint which is being ‘impacted most significantly’ when a particular risk occurs. We planned to use ‘Chi-Square test of Independence’ for analyzing the association between the risk factors and the project constraints in DAD project.

As explained in the above section, chi-square test of independence helps us to find out if two variables are independent or not. We will be creating a contingency table containing the risk factors under a particular risk area as one dimension (row) and the project constraints as the other dimension (column). The cell of this contingency table will contain the number of respondents who relate a particular risk factor with a particular project constraint (Time or Cost or Quality). In other words, for each row representing a risk factor, each cell in that row will have the ‘count of the respondents’ who believes that the risk factor is impacting a particular project constraint (time or cost or quality) the most.

The chi-square test on this data will provide us with three sets of data (i) Chi-square test statistic (ii) observed significance level (p-value) (iii) corresponding to each risk factor, the frequency of respondents who selected time as the project constrained being impacted most significantly by that risk factor, frequency of respondents who selected cost and frequency of respondents who selected cost as the project constraint. We will set the significance level at 0.05, and will compare the observed p-value with the significance level (0.05).

If the p-value obtained by applying Chi-Square test is less than the significance level (0.05), it will indicate that each of the risk factor corresponding to a particular risk area is statistically significantly related to one of the project constraint considered for analysis. For each of the risk
factor, we will consider the project constraint (Time or Cost or Quality), which is selected by maximum number of respondents from the given set of respondents, as the one which is being impacted by the risk factor most significantly.

This will help us to segregate all the risk factors which are impacting the ‘time of the project’, those which impact the ‘cost of the project’ and those which impact the ‘quality of the project’. The managers executing DAD projects can use these results to concentrate only those risk factors which are impacting the ‘project constraint’ which is most important for a particular project.

3.2.3 Methodology for Validation

The findings of the qualitative descriptive study and quantitative descriptive study will be consolidated to create the risk management framework for DAD projects. We plan to evaluate the applicability and performance of the risk management framework by getting it validated from the companies that execute DAD projects.

Companies will be provided with the partial framework due to the paucity of time. According to the relevance of risk category to the present status of the project, one or two of the risk categories will be implemented. The risk management framework will be suggesting the risk factors in the order of their importance and the corresponding risk management methods which are used most frequently to control the risk in DAD projects.

We will be collecting quantitative data corresponding to the impact and the probability of occurrence of the risk factors for validating the framework, through two different self-administered questionnaires.

The risk exposure (magnitude of risk) will be measured based on the perception of the project team before and after implementing the framework. We will be obtaining the ‘Estimated Magnitude of Risk’ or the ‘Estimated Risk Exposure’ for the risk factors, before implementing the framework using one questionnaire. A second questionnaire will be used to obtain the ‘Actual Magnitude of Risk’ or the ‘Actual Risk Exposure’ for the risk factors, after implementing the partial framework for a limited period of time. The benefit of implementing
the framework will be obtained by the difference between the risk magnitude before and after implementation of the framework.

Difference in Risk Exposure = Estimated Risk Exposure – Actual Risk Exposure

This will help us to identify the following:

1. The number of risk factors which will be totally eliminated due to the framework implementation.
2. The number of risk factors whose risk exposure (magnitude of risk) reduced: Those risk factors for which the difference in risk exposure will be a ‘non-zero positive value’.
3. The number of risk factors for which the magnitude of risk remained the same: Those risk factors for which the difference in risk exposure will be ‘zero’.
4. The number of risk factors for which the magnitude of risk inflated: Those risk factors for which the difference in risk exposure will be ‘a negative value’.

The percentage of risk factors whose risk magnitude has reduced and the percentage or risk factors which were altogether eliminated, will tell us about the benefit of the framework.

The risk management framework will be suggesting the most appropriate approaches which can be used to control the risk in DAD projects. We will also ask the respondents to specify the risk management methods which they used for controlling the risk which they will consider. This will help us to evaluate the usage and applicability of the framework.