Causal Analysis of Factors Governing Collaboration in Global Software Development Teams

Prateeti Mohapatra  
ABB Corporate Research  
Industrial Software Systems  
ABB GISL  
Bangalore, India  
prateeti.mohapatra@in.abb.com

Petra Björndal  
ABB Corporate Research  
Industrial Software Systems  
ABB AB  
Västerås, Sweden  
petra.bjorndal@se.abb.com

Karen Smiley  
ABB Corporate Research  
Industrial Software Systems  
ABB Inc.  
Raleigh, NC, United States  
karen.smiley@us.abb.com

Abstract—Globally distributed software development (GSD) is increasing in popularity in industry. However, as it is coupled with its challenges of distance, time, and culture, it increases the importance of identifying and understanding the specific factors that enable and hinder GSD teams. This paper presents the approach and preliminary findings from an exploratory study of the enabling and inhibiting factors that affected several globally distributed projects in a large commercial organization. Our quantitative analysis includes grouping these factors to reduce the dimensional complexity, studying their underlying causal relationships, and identifying the most influential factors using factor analysis and structural equation modeling. The paper concludes by presenting preliminary findings, limitations, and directions for future work.

Keywords - Global Software Development, Coordination, Causal Relationship, Factor Analysis, Structural Equation Modeling

I. INTRODUCTION

Developing software using teams that are distributed across countries is no longer a rare phenomenon. At ABB (www.abb.com), a leading global industrial company specialized in power and automation technologies, we are exploring the factors that influence collaboration and team performance during globally-distributed software development (GSD). The benefits of GSD are many: search for specialized talents, expansion through acquisition, reduction in development costs, reduction in time to market, and proximity to customers have all propelled the growth of globally distributed software teams [1-2].

However, development of software using globally-distributed teams introduces collaboration difficulties, with the dispersed team members having to interact across the barriers of distance, time zones, language, and culture [3]. The increasing prevalence of GSD, coupled with these challenges, has elevated the importance of identifying and understanding the factors that enable and hinder the quality and development schedule of globally developed software. This paper describes our recent steps in this direction.

II. PRIOR AND RELATED WORK

Past studies on collaboration in GSD have identified a number of factors – mainly personnel and team related – that govern collaboration in globally dispersed teams. Kiel [4] found that it is hard to deploy, execute, and control projects in GSD environments because of non-technical factors, i.e. social, cultural, behavioral, and political. Based on an extensive review of literature, Raffo and Setamanit [5] selected a number of factors influencing GSD. In a comprehensive survey of the literature on virtual teams, Powell et al. [6] categorized the published articles along different issues in virtual teams. With Snipes et al. [7], we used the Goal-Question-Metric method to define a set of twenty-five metrics for collaboration in GSD projects. Casey and Richardson [8] identified ten key factors to establish and facilitate the operation of globally distributed virtual teams. At the core of all these factors and challenges lies the issue of Coordination [9][10][11][12]. Kraut and Streeter [13] have emphasized the problem of coordination among software developers while developing large software systems as a major contributor to the “software crisis”, and Herbsleb [18] specifically identified coordination over distance as a key GSD issue. Based on this, we decided to examine the contribution impact of factors on Coordination Effectiveness.

Many studies have captured relationships between various factors. Prikладници et al. [14] presented relationships of GSD factors using a conceptual map. Cataldo et al. [15] presented a qualitative analysis and concluded that coordination problems in GSD projects depend on relationships of several factors. Setamanit et al. [16] developed an interaction effect model of factors based on literature regarding GSD. These results were hypothetical in nature, but were validated with information from the literature and industry standard data.

To the best of our knowledge, no prior work has reported to what extent the findings of a study/survey precisely represent causal relationships in industrial practice among various factors. The factors considered in the previous works also addressed a large number of issues pertaining to GSD, such as product architecture, development strategy, etc. However, these studies did not clearly distinguish team-related factors and product or development-related factors. The dimensional complexity of the factors complicates the analyses.

In consideration of the above-made points, we identified two pertinent research questions:

1) What categories of factors significantly influence collaboration in global software engineering?
2) What are the causal relationships among these factors and categories?

To address these questions, we undertook an exploratory study as described in the following sections. Based on the data we gathered, we identified important factors, and grouped them into broad categories using exploratory factor analysis based on the different perceptions of the respondents [17]. We then used the confirmatory structural equation modeling approach [17] to find causal relationships among the constructs (broad categories of related factors).

III. STRUCTURED SURVEY QUESTIONNAIRE (SSQ)

To understand more fully how the different GSD-related factors interact in practice, we designed the SSQ based on the factors gathered from past studies and from our own past work in which we had defined a set of GSD metrics [7]; selected a diverse set of globally distributed software development projects; and conducted structured interviews guided by the questionnaire.

Since the focus of our study is limited to only the issues of team collaboration, we designed the SSQ to collect project participants' opinions on only those factors which are relevant for inter-site collaboration. A total of twenty-one factors were considered in this study: Attrition Issues, Commitment to the Shared Goals, Communication Media Richness, Communication, Synchronization, Cultural Differences, Encouraging Innovative Solutions, Expressing New Ideas, Face-to-Face Interaction, Frequency of Planned Meetings, Language Barriers, Informal Spontaneous Meetings, Meeting before Project Startup, Planning of Communication Strategy, Prior Work Together, Reason for Involvement of Different Sites, Sharing New Ideas, Sharing Personal Information, Team Building Activities, Time-Zone Issues, Training on Communication Processes and Tools, Trust.

We interviewed twenty-nine members of software development teams who were involved in six projects and were located in Scandinavian countries (Finland, Norway, Sweden), China, India and USA. Each interview session lasted for about one and a half hours. Interview modes used were face-to-face, videoconferencing, and telephone.

![Experience of Respondents](image)

Figure 1. Experience of Respondents

The interviewees included six Project Managers, five Project Leaders, seventeen Team Members, and one Technical Support person. The total experience and GSD-related experience of the twenty-nine interviewees are given in Figure 1. GSD experience varied from a minimum of 0 years to a maximum of 19 years, with mean of 6.5 years. Their industry experience varied between 2 years and 20 years, with mean of 8.7 years.

IV. ANALYSIS OF RESPONSES

With factor analysis, several high-leverage factors (identified from their standardized regression weights) having similar implications are grouped under one category, or construct. This approach considerably reduces the dimensional complexity of the factors and makes them easily comprehensible. To apply this approach to our case study, we first examined the measured sampling adequacy (MSA) for each factor, and excluded factors having low MSA values. In our study, Reason for Involvement of Different Sites and Planning of Communication Strategy had low MSA values and were excluded from further consideration.

A factor analysis was then performed on the responses with the help of the MATLAB software package. A principal component analysis was performed using varimax factor rotation to group the factors and thus determine the appropriate categories of factors (constructs).

Seven constructs were formed, shown in Table I with their constituent factors. We decided to eliminate the factors Sharing Personal Information and Attrition Issues that had factor loadings with absolute magnitude less than 0.6, as suggested by Hair et al. [17].

TABLE I. CONSTRUCTS DERIVED FROM THE RESPONSES

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Factor Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coordination Effectiveness</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>0.810</td>
</tr>
<tr>
<td>Communication</td>
<td>0.778</td>
</tr>
<tr>
<td>Cultural Differences</td>
<td>-0.664</td>
</tr>
<tr>
<td>Synchronization</td>
<td>0.679</td>
</tr>
<tr>
<td>Frequency of Planned Meetings</td>
<td>0.783</td>
</tr>
<tr>
<td>2. Teamness</td>
<td></td>
</tr>
<tr>
<td>Commitment to the Shared Goals</td>
<td>0.639</td>
</tr>
<tr>
<td>Sharing New Ideas</td>
<td>0.871</td>
</tr>
<tr>
<td>Expressing New Ideas</td>
<td>0.805</td>
</tr>
<tr>
<td>Encouraging Innovative Solutions</td>
<td>0.716</td>
</tr>
<tr>
<td>3. Project Start-up Activities</td>
<td></td>
</tr>
<tr>
<td>Face-to-Face Interaction</td>
<td>0.781</td>
</tr>
<tr>
<td>Team Building Activities</td>
<td>0.618</td>
</tr>
<tr>
<td>4. Communication Mechanisms</td>
<td></td>
</tr>
<tr>
<td>Language Barriers</td>
<td>-0.602</td>
</tr>
<tr>
<td>Communication Media Richness</td>
<td>0.784</td>
</tr>
<tr>
<td>5. Acquaintance before Project Start-up</td>
<td></td>
</tr>
<tr>
<td>Prior Work Together</td>
<td>0.785</td>
</tr>
<tr>
<td>Meeting before Project Startup</td>
<td>0.821</td>
</tr>
<tr>
<td>6. Time-Zone Issues</td>
<td>-0.902</td>
</tr>
<tr>
<td>7. Miscellaneous Factors</td>
<td></td>
</tr>
<tr>
<td>Training on Comm.Tools/Processes</td>
<td>0.820</td>
</tr>
<tr>
<td>Informal Spontaneous Meetings</td>
<td>0.714</td>
</tr>
</tbody>
</table>

As can be seen from Table I, Cultural Differences, Language Barriers, and Time-Zone Issues have negative signs. This means that these factors have a negative effect on their respective constructs.
After reducing the complexity of the factors by grouping them into seven constructs, we explored the relationships among the constructs. We first made the most appropriate hypotheses based upon logic and prior work, and then tested them with the help of structural equation modeling.

A. Hypotheses on Causal Relationships among Constructs

To determine the causal relationships among the six constructs, we first proposed a set of hypotheses, each describing a direct relationship between two constructs. In what follows, we give the findings of past works and our logical reasoning to formulate different hypotheses.

Raffo and Setamanit [5] identified the loss of teamness as one of the major factors hindering GSD projects, leading us to hypothesis H1:

H1: Teamness has a positive effect on Coordination Effectiveness.

Cataldo et al. [15] suggested that face-to-face interactions during the initial project formulation stage are successful mechanisms for collaboration in GSD projects. On the basis of this observation, we formulated hypothesis H2:

H2: Project Startup Activities have a positive effect on Teamness.

According to Bird et al. [21], it is helpful to know each other on a personal basis for effective communication. This conclusion leads to hypothesis H3:

H3: Acquaintance before Project Startup has a positive effect on Teamness.

According to Casey and Richardson [8], the availability of adequate communication media is important for GSD projects. This led us to hypothesis H4:

H4: Communication Mechanisms have a positive effect on Coordination Effectiveness.

Casey and Richardson [8] suggest that training in technical, communications, and cultural skills are specific to the needs of a GSD environment. Communication in GSD projects also suffer due to a lack of informal meetings [19]. These factors are grouped as Miscellaneous in Table I. The conclusions reached in these past works led us to hypothesis H5:

H5: Miscellaneous Factors have a positive effect on Coordination Effectiveness.

Time-zone differences have been noted to hinder and limit opportunities for GSD collaboration [14][21]. Coordination, communication, and cooperation are all identified as negatively impacting temporal distance [8]. Therefore, our hypothesis H6 is that time-zone issues will have a negative effect on Coordination Effectiveness:

H6: Time-Zone Issues have a negative effect on Coordination Effectiveness.

To test these hypotheses, a path analysis was done using the structural equation modeling approach.

B. The Structural Equation Modeling Approach

Structural Equation Modeling (SEM) is recognized as a comprehensive and flexible approach for data analysis. SEM provides a straightforward method of dealing with multiple relationships simultaneously, has the ability to assess the relationships comprehensively, and provides a transition from exploratory to confirmatory analysis [17]. It aggregates a series of hypothesized cause-effect relationships between constructs into a composite hypothesis [20]. We use the SEM approach to confirm or reject the proposed hypothesis.

The most obvious difference between SEM and other multivariate techniques is its ability to accommodate multiple interrelated dependence relations in a single model [17]. The approach becomes a useful tool when one dependent variable needs to be treated as an independent variable in a subsequent analysis. For instance, in our case, Teamness is initially treated as a dependent variable and later treated as an independent variable. The inputs to SEM are either raw data or variance/covariance or correlation matrix, and a model to be evaluated [19]. The focus of SEM is not on individual responses, but on patterns of relationships across the samples. In addition, SEM is a comprehensive statistical approach to test hypotheses about relations among the constructs shown in Table I [17]. Every construct is linked to at least one factor that is measurable, thus making measurement of the construct feasible.

In this study, the SEM approach was used to test the set of seven hypotheses made earlier. The AMOS 4.0 software package [22] was used for the purpose. Figure 2 is a path diagram showing the hypothesized relationships among the constructs. The arrows indicate the direction of influence, and the numbers appearing on the arrows indicate the associated standardized regression weights obtained. The weights are significant, at p < 0.01 each.

Following Hair et al. [17], goodness of fit index, normalized fit index, root mean square approximation, chi-square statistics, and degree of freedom estimates were computed to validate the model. Table II gives the computed values of these measures. Columns 1, 2 and 3 give the values of chi-square ($\chi^2$), degrees of freedom (DF), and their ratio respectively. Columns 4, 5 and 6 give the value of goodness of fit index (GFI), normalized fit index (NFI), and root mean square error approximation (RMSEA) of the model. For social science data, acceptable values of GFI and NFI are the values greater than 0.5 [23]. The values in Table II indicate that the model is reliable, and that we can use this model to predict the relationships among the constructs.

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>DF</th>
<th>$\chi^2$/DF</th>
<th>GFI</th>
<th>NFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>167.8</td>
<td>13</td>
<td>1.271</td>
<td>0.734</td>
<td>0.533</td>
<td>0.079</td>
</tr>
</tbody>
</table>

In the path diagram, the standardized regression weights indicated on the branches were used to accept or reject a hypothesis. According to Hair et al. [17], hypotheses corresponding to standardized regression weights less than 0.1 should be refuted. All six hypotheses tested were supported because their standardized regression weights were more than 0.1 (Table III).
TABLE III. RESULTS OF THE HYPOTHESIS TEST

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Description</th>
<th>Standardized regression weight</th>
<th>Inference drawn</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Teamness has a positive effect on Coordination Effectiveness.</td>
<td>0.15</td>
<td>Supported</td>
</tr>
<tr>
<td>H2</td>
<td>Project Startup Activities have a positive effect on Teamness.</td>
<td>0.12</td>
<td>Supported</td>
</tr>
<tr>
<td>H3</td>
<td>Acquaintance before Project Startup has a positive effect on Teamness.</td>
<td>0.16</td>
<td>Supported</td>
</tr>
<tr>
<td>H4</td>
<td>Communication Mechanisms have a positive effect on Coordination Effectiveness.</td>
<td>0.96</td>
<td>Supported</td>
</tr>
<tr>
<td>H5</td>
<td>Miscellaneous Factors have a positive effect on Coordination Effectiveness.</td>
<td>0.48</td>
<td>Supported</td>
</tr>
<tr>
<td>H6</td>
<td>Time-Zone Issues have a negative effect on Coordination Effectiveness.</td>
<td>-0.14</td>
<td>Supported</td>
</tr>
</tbody>
</table>

V. INTERPRETATION OF FINDINGS

As evident from the values of the standardized regression weights in Table III, Communication Mechanisms is the most influential construct affecting coordination effectiveness, with the next most important construct being Miscellaneous Factors, followed by Teamness.

We constructed the Path Model in Figure 2 to determine the hypothesized relationships among the constructs, and to test the hypotheses made above. The figure shows the two factors (Language Barriers and Communication Media Richness) which influence Communication Mechanisms. The high standardized regression weight of Language Barriers (-0.44) identifies it as the most important contributing factor for teams we studied. The calculated importance of Language Barriers is consistent with our interview data, which showed a high frequency of interviewees reporting language barriers as the main reason hindering coordination effectiveness on their project. The next most important contributing factor to this construct is Communication Media Richness; finding it significant in these industrial projects corroborates identification of this factor by many past researchers [2][5][6][8] as having relevance in distributed teams.

Within the construct of Miscellaneous Factors, Training on Communication Tools and Processes was not found to have a significant influence on coordination effectiveness for the teams we studied. However, Informal Spontaneous Meetings were found to have a significant influence on coordination effectiveness for these teams.

While Teamness was the third most significant construct affecting Coordination Effectiveness, it surprisingly did not have as strong an effect as Communication Mechanisms and Miscellaneous Factors. Software developers may be more affected by media richness, language barriers, and informal meetings, and less by other team-formation factors, e.g. free flow of new ideas.

We note the following limitations of the study:
- Due to the small number of teams and interviewees, the captured user perceptions and patterns may not truly reflect factors which impact collaboration on other GSD teams, or in a general way.
- Our data is based upon project participants’ reported perceptions of the studied factors, e.g. trust. These data values may not reflect the ‘true’ values (e.g. the actual levels of trust could be lower or higher than reported), or may not accurately predict the true causative impact on coordination effectiveness.
- For some respondents the number of years of GSD-related experience was low, and the average number of globally distributed projects in which a respondent worked was low. We did not directly assess prior GSD project experience (with different teammates or at other companies), or other possible context variables, as contributing factors.
- Certain questions in the SSQ were open-ended. Those responses were not analyzed to assess their consistency with the findings of this analysis.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a quantitative analysis of factors that are considered important in one large commercial organization. Based on past works, we used questionnaire-guided interviews to explore 21 factors that were likely to affect effective collaboration in globally distributed software development projects. We grouped the factor data into seven constructs using exploratory factor analysis. Confirmatory analysis was performed using structural equation modeling to assess our hypotheses regarding the relationships among the constructs. All six hypotheses were supported by the data.

We visualize a number of areas in which the work can be extended. Additional hypotheses on relationships among the constructs could be assessed. Data gathering for additional GSD teams and project should be pursued, and scope could be extended to investigate the issues surrounding teamness in the organization. One could also explore estimation of impact on task and project duration in a GSD environment for various levels of the influencing factors. Further studies on global collaboration could consider the social networks of professionals.
REFERENCES