Causal Analysis of Factors Governing Collaboration in GSD Teams

ICGSE 2010 – Presentation Outline
Causal Analysis of Factors Governing Collaboration in Globally Distributed Software Development Teams

- Walking the Talk: Our Globally Distributed Team
- Motivation and Context for This Study
- Approach for This Study
  - Objectives and Research Questions
  - Why We Chose SEM (Structural Equation Modeling)
  - Steps of Study Method
- Study Execution
  - Data Collection
  - Factor Analysis
  - Definition of Hypotheses
  - Hypothesis Testing
  - Conclusions
- Summary / Q&A
Motivation and Context for This Study
Background in ABB

- The ABB group of companies (www.abb.com) is a world leader in power and automation technologies
- Operates in >100 countries, sells products all over the world
- Increasing number of products with software components
- Increasing number of globally-distributed software development (GSD) projects:
  - Specialists
  - Acquisitions
  - Reduction in development cost
  - Reduction in time to market
  - Proximity to customers (global market)

⇒ High-quality GSD performance is a key concern for ABB
Motivation and Context for This Study

Key Challenges of Global Software Development

Distributed development:
- Communication
- Trust
- ‘Coordination over distance’

Globally distributed development:
- Temporal separation (time zones)
- Culture
- Language

GSD Collaboration Research Plan

- Survey state of the art in measuring GSD collaboration
- Define a set of measures for GSD collaborations, using G-Q-M to address:
  - 3 sub-goals: efficiency, quality, innovation
  - Collaboration factors and effectiveness
  - Project outcomes
- Develop data gathering instruments:
  - SSQ: Structured Survey Questionnaire, to guide project interviews and metrics collection
  - DMQ: Online Distributed Meeting Questionnaire, for assessing GSD team meetings and tool usage
- Conduct benchmarking studies with completed and ongoing GSD projects
- Identify tools and tactics for specific development tasks and GSD configurations that yield measurable gains in project performance
Approach for This Study
Objectives and Research Questions

Objectives:
- To explore the enabling and inhibiting factors that affect globally distributed projects at ABB
- To examine the contribution impact of these factors on coordination effectiveness

Research Questions:
- What categories of factors significantly influence coordination in global software engineering at ABB?
- What are the causal relationships among these factors?

Why We Chose Structural Equation Modeling

- Structural Equation Modeling looks at how well a proposed set of causal relations explains the pattern of covariance among a set of variables.
- Provides the ability to have multiple interrelated dependence relationships in a single model
- Helps in modeling complex relationships:
  - Estimating many equations at once
  - A dependent variable in one equation can be an independent variable in other equation(s)
- Path Analysis:
  - Consists of path diagrams (a pictorial representation of the relationships)
  - Calculates the strength of each relationship (path) depicted in the path diagram
Approach for This Study
Steps of Study Method

<table>
<thead>
<tr>
<th>Step</th>
<th>Name</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Collection</td>
<td>• Gather data in exploratory studies, using SSQ-guided interviews and DMQ + project artifacts</td>
</tr>
<tr>
<td>2</td>
<td>Factor Analysis</td>
<td>• Identify factors to be considered in the data analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Conduct exploratory analysis to group factors into constructs</td>
</tr>
<tr>
<td>3</td>
<td>Definition of Hypotheses</td>
<td>• Propose a set of hypotheses to examine possible causal relationships among constructs</td>
</tr>
<tr>
<td>4</td>
<td>Hypothesis Testing</td>
<td>• Perform Path Analysis using Structural Equation Modeling to test the hypotheses</td>
</tr>
<tr>
<td>5</td>
<td>Conclusions</td>
<td>• Assess results of analyses</td>
</tr>
</tbody>
</table>

Study Execution – 1. Data Collection

"Global Software Project Management: A Case Study", SEAFOD 2010

Exploratory studies with 29 GSD team members and managers in 6 projects (3 partially agile; internal distributed development)

- Product development projects distributed across Scandinavia, Central Europe, China, India, USA
  - Max time zone differences: 4.5-11.5 hrs across 2-4 sites
- Conducted SSQ-guided interviews (mostly F2F, a few V/C)
- Gathered DMQ data on various distributed meetings
Study Execution – 2. Factor Analysis
Overview of Approach

Method of data reduction:
- Analyze the structure of the correlations among a large number of variables *(in our case, questionnaire responses)* by defining a set of common underlying dimensions
- Group correlated variables together, and separate them from other variables with low or no correlation

General Steps in Factor Analysis:
- Step 1: Selecting and Measuring a set of variables
- Step 2: Data screening to prepare the correlation matrix
- Step 3: Factor Extraction
- Step 4: Factor Rotation to increase interpretability
- Step 5: Interpretation
- Further Steps: Validation and assessment of reliability of the measures

Study Execution – 2. Factor Analysis
Initial Examinations of Response Data

Constituent factors to include in constructs were selected from the 22 candidates, following standard guidelines:
- Examined ‘Measured Sampling Adequacy’ (MSA) of the response data for each factor.
  - Two factors having low MSA values (< 0.5) were excluded
- Performed Principal Component Analysis using *varimax factor rotation* to group the factors, and thus determine the appropriate categories of factors (constructs).
  - Two factors having absolute magnitude < 0.6 for factor loading were excluded
- High-leverage factors having similar implications were then grouped into categories or *constructs*.
  - Seven constructs were formed from 18 factors
### Study Execution – 2. Factor Analysis

#### 22 Candidate ‘Factors’ (4 Eliminated)

<table>
<thead>
<tr>
<th>Candidate Factor</th>
<th>Loading</th>
<th>Candidate Factor</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attrition Issues</td>
<td>(&lt;0.6)</td>
<td>Meeting Before Project Startup</td>
<td>0.821</td>
</tr>
<tr>
<td>Commitment to the Shared Goals</td>
<td>0.639</td>
<td>Planning of Communication Strategy (N/A – MSA low)</td>
<td>0.785</td>
</tr>
<tr>
<td>Communication Media Richness</td>
<td>0.784</td>
<td>Prior Work Together</td>
<td>0.785</td>
</tr>
<tr>
<td>Communication</td>
<td>0.778</td>
<td>Reason for involvement of Different Sites (N/A – MSA low)</td>
<td>0.810</td>
</tr>
<tr>
<td>Cultural Differences</td>
<td>-0.664</td>
<td>Sharing New Ideas</td>
<td>0.871</td>
</tr>
<tr>
<td>Encouraging Innovative Solutions</td>
<td>0.716</td>
<td>Sharing Personal Information (&lt; 0.6)</td>
<td></td>
</tr>
<tr>
<td>Expressing New Ideas</td>
<td>0.805</td>
<td>Synchronization</td>
<td>0.679</td>
</tr>
<tr>
<td>Face-to-Face Interaction</td>
<td>0.781</td>
<td>Team Building Activities</td>
<td>0.618</td>
</tr>
<tr>
<td>Frequency of Planned Meetings</td>
<td>0.783</td>
<td>Time-Zone Issues</td>
<td>-0.902</td>
</tr>
<tr>
<td>Informal Spontaneous Meetings</td>
<td>0.714</td>
<td>Training on Communication Processes and Tools</td>
<td>0.820</td>
</tr>
<tr>
<td>Language Barriers</td>
<td>-0.602</td>
<td>Trust</td>
<td>0.810</td>
</tr>
</tbody>
</table>

### Study Execution – 2. Factor Analysis

#### 7 Constructs Defined, Covering 18 Factors

1. **Coordination Effectiveness**
   - Trust
   - Communication
   - Cultural Differences
   - Synchronization
   - Frequency of Planned Meetings

2. **Teamness**
   - Commitment to Shared Goals
   - Sharing New Ideas
   - Expressing New Ideas
   - Encouraging Innovative Solutions

3. **Project Start-Up Activities**
   - Face-to-Face Interaction
   - Team Building Activities

4. **Communication Mechanisms**
   - Language Barriers
   - Communication Media Richness

5. **Acquaintance Before Project Start-Up**
   - Prior Work Together
   - Meeting Before Project Start-Up

6. **Time-Zone Issues**

7. **Miscellaneous Factors**
   - Training on Communication Tools/Processes
   - Informal Spontaneous Meetings
Study Execution – 3. Definition of Hypotheses

<table>
<thead>
<tr>
<th>ID</th>
<th>Hypotheses on Causal Relationships Among Seven Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Teamness has a positive effect on Coordination Effectiveness</td>
</tr>
<tr>
<td>H2</td>
<td>Project Startup Activities have a positive effect on Teamness</td>
</tr>
<tr>
<td>H3</td>
<td>Acquaintance before Project Startup has a positive effect on Teamness</td>
</tr>
<tr>
<td>H4</td>
<td>Communication Mechanisms have a positive effect on Coordination Effectiveness</td>
</tr>
<tr>
<td>H5</td>
<td>Miscellaneous Factors have a positive effect on Coordination Effectiveness</td>
</tr>
<tr>
<td>H6</td>
<td>Time-Zone Issues have a negative effect on Coordination Effectiveness</td>
</tr>
</tbody>
</table>

Study Execution – 4. Hypothesis Testing

SEM Path Diagram

Used Structural Equation Modeling (Path Analysis) to measure the constructs
Study Execution – 4. Hypothesis Testing
Are These SEM Results Reliable?

- **Computations performed to validate the model:**
  - GFI - Goodness of fit index
  - NFI - Normalized fit index
  - RMSEA - root mean square approximation
  - $\chi^2$ - chi-square statistics
  - DF - degree of freedom estimates

<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>

- Values of GFI and NFI > 0.5 indicate that the model is reliable, and that we can use this model to predict the relationships among the constructs.

---

Study Execution – 5. Conclusions
Constructs Impacting Coordination Effectiveness

<table>
<thead>
<tr>
<th>ID</th>
<th>Hypotheses on Causal Relationships Among Seven Constructs</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>

Criterion for refuting hypotheses: $|\text{weight}| < 0.1$
Study Execution – 5. Conclusions
Factors Impacting Coordination Effectiveness

<table>
<thead>
<tr>
<th>Constructs and Their Constituent Factors</th>
<th>Standardized Regression Weights</th>
<th>Constructs and Their Constituent Factors</th>
<th>Standardized Regression Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Coordination Effectiveness</strong></td>
<td></td>
<td><strong>3 Project Startup Activities</strong></td>
<td>0.12 (T)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.82</td>
<td>Face-to-Face Interaction</td>
<td>0.66</td>
</tr>
<tr>
<td>Communication</td>
<td>0.83</td>
<td>Team Building Activities</td>
<td>0.46</td>
</tr>
<tr>
<td>Cultural Differences</td>
<td>-0.86</td>
<td><strong>4 Communication Mechanisms</strong></td>
<td>0.96</td>
</tr>
<tr>
<td>Synchronization</td>
<td>0.70</td>
<td>Language Barriers</td>
<td>-0.44</td>
</tr>
<tr>
<td>Frequency of Planned Meetings</td>
<td>0.55</td>
<td>Communication Media Richness</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>2 Teamness (T)</strong></td>
<td>0.15</td>
<td><strong>5 Acquaintance Before Project Start-Up</strong></td>
<td>0.16 (T)</td>
</tr>
<tr>
<td>Commitment to the Shared Goals</td>
<td>0.57</td>
<td>Prior Work Together</td>
<td>0.85</td>
</tr>
<tr>
<td>Sharing New Ideas</td>
<td>0.79</td>
<td>Meeting Before Project Startup</td>
<td>0.61</td>
</tr>
<tr>
<td>Expressing New Ideas</td>
<td>0.73</td>
<td><strong>6 Time-Zone Issues</strong></td>
<td>-0.14</td>
</tr>
<tr>
<td>Encouraging Innovative Solutions</td>
<td>0.66</td>
<td><strong>7 Miscellaneous Factors</strong></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Informal Spontaneous Meetings</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Training on Communication Processes and Tools</td>
<td>X -0.06</td>
</tr>
</tbody>
</table>

Criterion for refuting significance of factor: |weight| < 0.1

---

Summary
Preliminary Findings From This Study

- Quantitatively analyzed GSD factors that are considered important in one large commercial organization.
- Used questionnaire-guided interviews and online surveys to examine 22 factors that were likely to affect coordination effectiveness in global software development collaboration.
- Applied exploratory factor analysis to stratify and group the factor data into seven constructs to reduce complexity.
- Performed confirmatory analysis using Structural Equation Modeling to assess six hypotheses regarding the relationships among the constructs. All six were supported.

*Insights gained from the SEM approach described can be used to support future decisions on GSD projects.*
Summary
Limitations Of The Study

- Small number of teams and interviewees
- Data values *reported* by interviewees for the studied factors may not reflect the ‘true’ values
  - May not accurately predict the true causative impact on coordination effectiveness, but
  - Reported values may still be useful indicators or proxies for guiding decision-making
- Other context variables, e.g. prior GSD project experience (with different teammates or at other companies), were not directly assessed as possible contributing factors
- Interviewees’ responses to open-ended questions were not *formally* analyzed to assess their consistency with the findings of this analysis

Summary
Scope For Future Work

Additional areas which could be explored include:

- Data gathering for additional GSD teams and projects, with scope extended to investigate the issues surrounding teamness in the organization *(underway)*
- Sensitivity analyses on using alternate groupings of factors to form the constructs *(e.g. inclusion of Trust factor with Teamness)*
- Assessment of additional hypotheses on relationships among the constructs
- Estimation of impact on task and project duration in a GSD environment for various levels of the influencing factors
- Further studies considering the social networks of professionals, and examining metadata available in artifact management tools
Selected References


For full references, see the paper in the ICGSE 2010 proceedings.

Contact Information / Q&A

Authors: Prateeti Mohapatra
         Petra Björndal
         Karen Smiley (today's presenter)

The authors may be reached collectively after ICGSE 2010 via
ICGSE2010@gsd.agileteams.com
or through their Global Software Development blog:
http://gsd.agileteams.com
Authors' papers are available via http://publications.agileteams.com

Please join the “GSD Collaboration” LinkedIn group!

QUESTIONS?
Power and productivity for a better world™