Globally Distributed Software Development Project Performance: An Empirical Analysis

Narayan Ramasubbu
Rajesh Krishna Balan

Singapore Management University

(by Mohammad Seyed Alavi)
Distributed Software Development

Is dispersing software development tasks as beneficial as we think at first glance?
Distributed Software Development

... Do you think about these things:

- Extend work beyond the regular office hours at a single site
- Less cost of development
- New technology, Easier collaboration
- New workforces with high capabilities are emerging, go to use them
Distributed Software Development

Let's see the effects of dispersing software development tasks on productivity and quality ...

To start the study, we should model globally distributed software development
Work Dispersion

Quality management approaches
- Prevention-based
- Appraisal-based
- Failure-based

Project Performance
- Development Productivity
- Conformance Quality

Control Variables
- Team size, Code size, Reuse, Upfront investment, Design Rework
Work Dispersion

Work dispersion =

$100^2 - (\% \text{ effort at first development center})^2$

$- (\% \text{ effort at second development center})^2$
Quality Management Approaches

• Prevention-based
  – Programming training
  – Business domain training
  – Process training
  – Configuration management
  – Task Planning and Scheduling

(effort spent on training, project planning and configuration management activities)
Quality Management Approaches

• Appraisal-based
  – Requirement, Specification and Design reviews
  – Code inspection
  – Status reviews

(effort spent on peer reviews of requirement, design, status reviews and code inspection)
Quality Management Approaches

- Failure-based
  - Unit testing
  - Module testing
  - Integration testing
  - System testing
  - Error tracking and correction

\(\textit{effort spent on unit tests, module integration tests and system tests}\)
Project Performance

• Development Productivity:

\[
\frac{\text{software code size in function points}}{\text{total development effort in person-hours}}
\]

• Conformance Quality:

\[
\frac{1}{(\text{defects} + 1)}
\]
Control Variables

• Productivity Variables
  – Team Size
  – Reuse

• Quality Variables
  – Code Size
  – Upfront Investment

• Common Variables
  – Design Rework
Empirical Equation

Now...

Formulate empirical equations for Development productivity and Conformance quality...
Empirical Equation

- Development productivity =
  \[ f \left( \text{conformance quality, work dispersion, prevention-based approach, appraisal-based approach, failure-based approach, reuse, team size, design rework} \right) \]

- Conformance quality =
  \[ f \left( \text{development productivity, work dispersion, prevention-based approach, appraisal-based approach, failure-based approach, code size, upfront investment, design rework} \right) \]
Empirical Equation

Effects of size and effort on quality and productivity are not linear

therefore...

Effects of dispersion and quality management practices on conformance quality and development productivity are not linear
Empirical Equation

\[ \ln(\text{Development productivity}) = \alpha_0 + \alpha_1 \ln(\text{conformance quality}) + \alpha_2 \ln(\text{work dispersion}) + \alpha_3 \ln(\text{prevention}) + \alpha_4 \ln(\text{appraisal}) + \alpha_5 \ln(\text{failure}) + \alpha_6 \ln(\text{reuse}) + \alpha_7 \ln(\text{design rework}) + \alpha_8 \ln(\text{team size}) + \varepsilon_1 \]

\[ \ln(\text{Conformance quality}) = \beta_0 + \beta_1 \ln(\text{development productivity}) + \beta_2 \ln(\text{work dispersion}) + \beta_3 \ln(\text{prevention}) + \beta_4 \ln(\text{appraisal}) + \beta_5 \ln(\text{failure}) + \beta_6 \ln(\text{design rework}) + \beta_7 \ln(\text{code size}) + \beta_8 \ln(\text{upfront investment}) + \varepsilon_2 \]
Data Collection

A leading software service company:

19,000 employees
17 countries
annual revenue, more than one billion dollars
CMMI level 5 (research site under study)
Data Collection

**42** completed projects in **2 years**
*(development of commercial business applications)*

Ethnographic observation

Structured interview with:
- 2 senior business development manager
- 4 project manager
- 10 randomly selected project team member
Dispersion of Teams

2 development center: India, United States

Human Resource allocation was primarily sourced from Indian center
Finding coefficients

Regression method used for fitting data...

... and coefficients were found
<table>
<thead>
<tr>
<th>Variables</th>
<th>Development Productivity</th>
<th>Conformance Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Productivity</td>
<td>NA</td>
<td>0.893   β1</td>
</tr>
<tr>
<td>Conformance Quality</td>
<td>-0.308 α1</td>
<td>NA</td>
</tr>
<tr>
<td>Dispersion</td>
<td>-1.018 α2</td>
<td>-0.621 β2</td>
</tr>
<tr>
<td>Prevention-based</td>
<td>0.059 α3</td>
<td>0.470 β3</td>
</tr>
<tr>
<td>Appraisal-based</td>
<td>0.573 α4</td>
<td>0.363 β4</td>
</tr>
<tr>
<td>Failure-based</td>
<td>0.324 α5</td>
<td>0.656 β5</td>
</tr>
<tr>
<td>Reuse</td>
<td>0.845 α6</td>
<td>NA</td>
</tr>
<tr>
<td>Design Rework</td>
<td>-0.106 α7</td>
<td>0.233 β6</td>
</tr>
<tr>
<td>Team Size</td>
<td>-0.182 α8</td>
<td>NA</td>
</tr>
<tr>
<td>Code Size</td>
<td>NA</td>
<td>-1.067 β7</td>
</tr>
<tr>
<td>Upfront Investment</td>
<td>NA</td>
<td>0.375 β8</td>
</tr>
<tr>
<td>Constant</td>
<td>2.530 α0</td>
<td>8.050 β0</td>
</tr>
</tbody>
</table>
Effect of Dispersion on Project Performance

![Graph showing the relationship between productivity (Func. Pts / Person Hrs) and dispersion (% of work at second site). The productivity decreases as the dispersion increases.]
Effect of Dispersion on Project Performance

The graph illustrates the relationship between productivity (in terms of function points per person hour) and quality (expressed as the reciprocal of defects plus one). Three lines are shown, representing different dispersion levels:

- Lower Dispersion
- Mean Dispersion
- Higher Dispersion

As the quality decreases (moving to the right on the x-axis), productivity decreases for all dispersion levels. The graph demonstrates that higher dispersion is associated with lower productivity at any given quality level.
Reducing Dispersion Effects through QMA
Reducing Dispersion Effects through QMA
Relative Effects of Different QMAs

![Graph showing relative effects of different QMAs](image)
Relative Effects of Different QMAs
Summary

• Dispersion has a significant effect on productivity and a harder-to-capture secondary effect on quality
• The effect of dispersion can be significantly mitigated through the use of structured software engineering processes
• Different QMAs have significantly different impacts on different dimensions of project performance
Questions?
Suggestion

- Uncertainty
  - Information asymmetry between remote teams
  - Ambiguous authority

- Solution
  - Learning-before-doing: preventive and appraisal-based
  - Learning-by-doing: failure-based
Robustness of Analysis

- Cook's Distance statistics
Robustness of Analysis

- Cook's Distance statistics
Correctness of data

Internal company software engineering process database (*because of CMM-level 5 compliance*)

Routinely auditing the reports

audited by independent external assessment auditor for CMM & ISO 9001
Regression

Endogeneity between conformance quality and development productivity (Durbin-Wu-Hausman)

Two Stage Least Squares instead of Ordinary Least Square