Follow-the-Sun Software Development: A Controlled Experiment to Evaluate the Benefits of Adaptive and Prescriptive Approaches

Josiane Kroll, Alan R. Santos, Rafael Prikladnicki, Estevão R. Hess, Rafael A. Glanzner, Afonso Sales, Jorge L. N. Audy, Paulo H. L. Fernandes
Computer Science School
Pontificia Universidade Católica do Rio Grande do Sul (PUCRS)
Porto Alegre, Brazil
{josiane.kroll, alan.ricardo, estevao.hess, rafael.audy}@acac.pucrs.br, {rafaelp, afonso.sales, paulo.fernandes, audy}@pucrs.br

Abstract—Follow-the-Sun (FTS) has been used in the context of global software development projects in order to take the advantages of temporal differences between several productions sites located in different time zones. However, the lack of FTS experience in the software industry is observed as the main barrier for its adoption. Recent studies suggest that FTS is more suitable for adaptive approaches (e.g. agile methodologies). For this reason, we designed and executed a controlled experiment to investigate the benefits of adaptive and prescriptive approaches for FTS. We used fictional maps with teams distributed in two sites. Each site had two teams representing software designers and developers. Our results indicate that the use of adaptive approaches increases the speed, but not the accuracy and the quality of the work.

Keywords—Global software development; follow-the-sun; adaptive approach; prescriptive approach; software engineering.

I. INTRODUCTION

Many software development companies are looking for costs reduction and the increase of productivity [1]. Follow-the-Sun (FTS) is a software development strategy where several productions sites are located in different time zones in a way that work can be done in a twenty-four hour basis per day, seven days per week. The main goal of FTS is to reduce the time-to-market, or increase the development speed [2].

FTS is not easy and many challenges have been reported in the literature, such as communication difficulties, process coordination, team management and cultural diversity [3]. Carmel et al. say that few FTS success cases are reported in the literature [4], and several authors discuss that the use of adaptive approaches for software development is a promising way to make FTS work [2] [5] [6].

Adaptive approaches or adaptive methods (e.g. agile methods and iterative development) are aimed to rapidly adapt to the changing reality. An adaptive approach emphasizes communication and collaboration in an iterative process [7, 16]. On the other hand, prescriptive approaches or methods (e.g. waterfall lifecycle) aimed at future plans in detail. The team works in linear-sequential planned activities, and usually struggle to deal with change [2]. For this reason, we conducted a research in order to investigate adaptive and prescriptive approaches for FTS. We performed a controlled experiment with students from a postgraduate Computer Science program PUCRS, a private University in the south of Brazil. We used fictional maps and simulated distributed teams in two sites. Each site had two teams representing software designers and developers. We collected and compared data about speed, accuracy and quality of the work developed by the teams in each approach.

This paper is structured as follows: in Section 2 we present some background about Follow-the-Sun (FTS). In Section 3, our research questions and hypothesis are presented. In Section 4, we present the experimental design and explain the experiment execution. In Section 5, we present the results obtained in this experiment. In Section 6, we discuss the results and Section 7 concludes the paper.

II. FOLLOW-THE-SUN SOFTWARE DEVELOPMENT

One of the main purposes of FTS is the reduction of software development cycle duration [2]. FTS software development teams are globally distributed across different time zones and production locations [8]. When a team finishes its own regular hours of work, another team located in another site and another time zone will receive the tasks to start its workday. Daily production handoffs are performed by teams following to the next production site that will be in a different time zone and will continue the work of the previous teams [2] [9].

In FTS, tasks are handed off from one team to another in the end of a working day in a specific site [4] [8]. A handoff is then the transition between teams and can be defined as a check-in of a work unit to the next production site [4]. In the context of FTS, handoffs are conducted on a daily basis, at the end of each site shift. However, daily handoffs coordination is very difficult and requires different practices to reduce coordination costs [2].
FTS is an important research area, but so far, it is relatively understudied within Software Engineering. The success cases of FTS usage in industry are still small. Carmel, Espinosa and Dubinsky (2009) claim that there are few documented success cases in industry and many difficulties in applying FTS practices [4]. There are several challenges related to communications, process coordination, team’s management, cultural differences and geographic differences [10].

III. THE EXPERIMENT

Our goal in this study is to investigate the use of the adaptive and prescriptive approaches in the context of FTS. We have defined the following research questions (RQ):

RQ1: Are teams using adaptive approaches faster than teams using prescriptive approaches for software development in the context of FTS?

RQ2: Do teams using adaptive approaches deliver more accurate work than teams using prescriptive approaches for software development in the context of FTS?

RQ3: Do teams using adaptive approaches deliver more quality than teams using prescriptive approaches for software development in the context of FTS?

A. The Study Design

Our experimental design was inspired in the laboratory study conducted by Espinosa, Nan and Carmel [12]. The authors executed an experiment to check the impact of time zone overlap on speed and accuracy. The teams used simple fictional map that consist of a background image, with various objects and colored arrows styles. The maps production was done in using PowerPoint® slides. Participants worked in pairs and the objective was to complete the maps according to the defined requirements. We borrowed from the authors the tasks performed by the teams during the development of our experiment. The maps design task was also used in another study [8]. In this study, the goal was to check the impact of increasing the number of sites in the FTS cycle on the quality of the work that was delivered at the end.

To conduct the experiment with teams working on FTS, we chose the context of a university with students from a computer science postgraduate program at PUCRS. In this environment, the experiment was simulated with the definition of distributed teams. The experiment dimensions are defined as:

- Process: in vitro, with the participants located in a controlled environment. This experiment was executed by the development of experiment activities as part of a post graduation class in two isolated labs.

- Participants: the experiment was executed with postgraduate students.

- Scenario: the studied problem was an academic task, and it was represented by the execution of tasks in a fictional task map [12].

- Generality: the experiment was specific and it is valid only in the scope of this study.

Selection of participants: We invited by convenience 12 students from Computer Science postgraduate program at PUCRS.

Type of experiment and experimental units: The experiment was performed using one factor and two treatments. We used the following notation:

- vAdp: Maps developed using the adaptive approach.
- vPre: Maps developed using the prescriptive approach.

In this experiment, the factor was represented by developing maps and the treatments were prescriptive and adaptive approaches (Table 1).

<table>
<thead>
<tr>
<th>#Participant</th>
<th>vAdp</th>
<th>vPre</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S6</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S7</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S8</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S9</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S10</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S11</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>S12</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

The participants’ distribution to setup the teams was randomly defined between the factors adaptive and prescriptive. It was defined two designers and two makers for adaptive group and four designers and four makers for prescriptive group because the prescriptive group required two designers or two makers per shift.

The experiment was planned to have five shifts of 15 minutes each (representing one workday). For each approach (adaptive and prescriptive) the participants were organized in two sites where each site had two teams. We used FTS with no time-zone overlap between the distributed teams.

The members of each team could play the following roles: Map Designer (D) or/and Map Maker (M). As proposed by
Espinosa, Nan, and Carmel [12] these roles are similar to that of a software designer who needs to communicate the design specifications to a programmer. Each map designer (D) had a set of 13 maps while each map maker had a set of 13 blank PowerPoint© slides.

B. Instrumentation, Training & Execution

The experiment was executed in two isolated labs to avoid outside interference. Table II presents the instruments used to perform this experiment.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>When</th>
<th>How</th>
<th>When</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Prescriptive</td>
<td>Experiment</td>
<td>Maps Development</td>
<td>Execution</td>
</tr>
<tr>
<td>Unit</td>
<td>Adaptive</td>
<td>Experiment</td>
<td>Maps Development</td>
<td>Execution</td>
</tr>
<tr>
<td>Document</td>
<td>Experiment</td>
<td>Experiment</td>
<td>Experiment preparation</td>
<td>Preparation</td>
</tr>
<tr>
<td>Guide</td>
<td>Training</td>
<td>Pre-experiment</td>
<td>To present the experiment context and motivation</td>
<td>Preparation</td>
</tr>
<tr>
<td>Metric</td>
<td>Report</td>
<td>Post-experiment</td>
<td>Data was gathered using a manual report</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

To prepare the experiment execution, the following aspects were observed [11]:

- **Experiment consensus**: Participants were provided the required knowledge about the experiment, and clarification about the experiment goals;
- **Sensitive results**: Participant’s names are anonymous throughout the experiment description.

All variables and resources were carefully defined before the experiment execution. We applied one treatment for each group (Prescriptive and Adaptive), contextualizing the objectives, motivation and technical procedure for the experiment.

Fig. 2 illustrates how the participants were distributed in the isolated labs. Map designers (D) were in Lab1 and map makers (M) were in Lab2. A map designer’s task was to provide instructions to the map makers about how to replicate the maps. Each map was composed by one background picture, five arrows and two extra icons. The joint of the arrows composing the path and icons in the background picture was executed by the map makers.

Teams that were working within the same shift (adaptive approach) had the possibility to use a communication tool. For this experiment we used GTalk [18]. Participants could “chat” with their teammates whenever they wanted. For the prescriptive approach the participants could only add instructions and comments in text log files from one shift to the next. Those text files were named as handoff files. GTalk logs helped us to rule out potential confounding effects of media richness [12]. The experiment execution had five shifts of 15 minutes each. Each team had 15 minutes to perform their tasks.

<table>
<thead>
<tr>
<th>Adaptive</th>
<th>Prescriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAB1</td>
<td>LAB1</td>
</tr>
<tr>
<td>LAB2</td>
<td>LAB2</td>
</tr>
</tbody>
</table>

Fig. 3 presents the map designer perspective. In the designer perspective, the idea of the task execution is to describe to makers a path composed by five joint arrows and the addition of two extra icons in a background picture. The designer must define the color (blue, green or red) of the arrows, the start and the end position of each arrow, as well as if these arrows are solid or dashed. Regarding the extra icons, the designer must precisely describe where these icons must be placed in the background picture. In the maker perspective, the success of a task depends on the accurate description done by the designer, and the correct interpretation done by the maker. The maker receives a set of steps that describe where the extra icons must be placed and the type and color of arrows, as well as where each arrow must be set in the figure in order to compose the path of a rows. At the end, it was possible to create an experiment structure that could simulate adaptive and prescriptive software development using activities with fictional maps. During the experiment execution, researchers were available for clarification of questions from participants.
C. Dependent Variables and Measures

RQ1 checks the speed obtained by one team using an adaptive approach and another team using a prescriptive approach. The metric associated to RQ1 is calculated for each team dividing the amount of maps delivered (md) by the total number of maps (tm).

**Speed = md / tm**

RQ2 checks the accuracy of the maps delivered in each approach. The metric is calculated by dividing the amount of points from correct elements in each map (ce) by the number of total points (tp).

**Accuracy = ce / tp**

RQ3 verifies the quality the project delivered in each approach. The metric for quality is calculated by the sum of total accuracy points (ap) divided by the total number of maps delivered (md) by each team.

**Quality = ap / md**

To conduct the experiment we have formulated some hypotheses. We consider null and alternative hypothesis as following:

1. **The speed obtained in the adaptive and prescriptive approaches are the same.**

   **Null Hypothesis, H₀:** Speed obtained by the team using the adaptive approach is the same as the speed obtained by a team using the prescriptive approach.

   - **Metrics:** We calculated the speed as following:
     
     - **SP** – Represents the speed obtained by the team using a prescriptive approach.
     - **SA** – Represents the speed obtained by the team using an adaptive approach.

   **H₀:** SP = SA

2. **Alternative hypothesis, H₁:** Teams using the adaptive approach are faster than teams using the prescriptive approach.

   **H₁:** SP < SA

3. **Alternative hypothesis, H₂:** Teams using the prescriptive approach are faster than teams using the adaptive approach.

   **H₂:** SP > SA

We have also collected information and tested similar hypotheses for accuracy and quality of the work delivered by the teams in each one of the approaches.

IV. RESULTS

After five shifts, adaptive teams produced six out of thirteen maps and prescriptive teams produced four out of thirteen maps. Adaptive teams obtained 16% more speed than prescriptive teams. This result was obtained using the metric associated to RQ1:

Adaptive: 6/13 = 0.46

Prescriptive: 4/13 = 0.30

The following conditions were applied in order to check the accuracy obtained: Each map was composed by one background image plus two icons plus five arrows. It was the same composition used by Solingen and Valkema [8]. We measured the accuracy using the metric associated to RQ2. In addition, we also considered if map elements (arrows and icons) were in the right position. The maximum points per map are 12. Each accuracy criteria is described in details next:

- Background image = 1 point;
- Arrow color = 1 point;
- Arrow style = 1 point;
- First icon = 1 point;
- Second icon = 1 point;
- Position of each arrow = 1 point if the position match and 0.5 point when within the limit of 1 inch;
- Position of each icon = 1 point if the position match and 0.5 point when within the limit of 1 inch.

The results are described in Table 3. Seven maps were not produced (N/P) by adaptive teams and nine maps were not produced by prescriptive teams.

TABLE III. MAP ACCURACY IN DETAIL

<table>
<thead>
<tr>
<th></th>
<th>Adaptive</th>
<th>Prescriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map 1</td>
<td>0.75</td>
<td>0.58</td>
</tr>
<tr>
<td>Map 2</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>Map 3</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Map 4</td>
<td>0.71</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 5</td>
<td>0.54</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 6</td>
<td>0.58</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 7</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 8</td>
<td>N/E</td>
<td>0.92</td>
</tr>
<tr>
<td>Map 9</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 10</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 11</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 12</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Map 13</td>
<td>N/E</td>
<td>N/E</td>
</tr>
<tr>
<td>Total</td>
<td>4.17</td>
<td>3.38</td>
</tr>
</tbody>
</table>

N/E – Task not executed.

In Fig. 4, we present a comparison of the results presented in Table III regarding map accuracy.

The results indicate that teams using an adaptive approach produced more maps than teams using a prescriptive approach.
However, when we observed the total points of correct elements for each map, prescriptive teams have more accuracy. We also computed the quality metric, associated to RQ3:

- **Adaptive**: 4.17/6 = 70%
- **Prescriptive**: 3.38/4 = 85%

Our findings indicate that teams using adaptive approach had 15% less quality; in contrast, these teams had 16% more speed. This means that there is a trend showing that teams using adaptive approaches might have less quality but more speed. However, we did not have enough data points to support statistical analysis.

V. DISCUSSION

**FTS is a research area with many important aspects to investigate** [3] [4] [13]. In theory, the use of FTS can significantly reduce the duration of the software development lifecycle [14]. In this study, we investigated the difference and benefits of using prescriptive and adaptive approaches in the context of FTS.

The speed obtained by teams using an adaptive approach is higher than the speed obtained by teams using a prescriptive approach. The percentage obtained by adaptive teams was 30% higher than the speed obtained by prescriptive teams. Phalnikar, Deshpande and Joshi [15] argue that agile teams are able to produce software faster, and Smite, Moe and Agerfalk [16] state that adaptive approaches such as agile methodologies aim to increase productivity of software teams [16]. Thus, adaptive approaches could benefit FTS by reducing the project cycle time. Moreover, FTS aims to reduce time-to-market [4], which it seems that can be obtained using adaptive approaches.

The accuracy obtained by teams in the delivered tasks was better in two tasks performed by prescriptive teams when we compare the first three tasks. On the other hand, when comparing the average accuracy of delivered tasks, adaptive teams are better than prescriptive teams. Teams using adaptive approaches are able to implement more requirements than teams using prescriptive approaches. This result shows that adaptive approaches may accelerate communication between teams and thus increasing the productivity, but we observed a lack of formal protocols to check the accuracy of the work.

The number of sites interacting in a development cycle may cause a small impact on the average accuracy [8]. We observed that the perception in relation to a task delivered is also a factor that affects the accuracy of the work. Our findings showed that students had less perception on task accuracy using adaptive approaches.

Quality is also an important factor for project success [5]. We observed this factor verifying the total accuracy points and the total number of maps delivered by each team. Results obtained showed that adaptive teams had 15% less quality than prescriptive teams. Carmel, Espinosa and Dubinsky [4] obtained similar result in their quasi-experiment. The decreasing of quality in maps delivered by teams using adaptive approaches could be caused by the lack of participant’s perception in relation to the quality criteria required for each task.

Based on the results found, our experiment suggests that adaptive approaches could perform better than prescriptive approaches in the context of FTS. This is also claimed by Carmel, Espinosa, and Dubinsky [4] and Gupta [17], and should be deeply investigated in the future, with experiments also executed in real software development settings.

VI. THREATS TO VALIDITY

One of the key issues in experimentation is evaluating the validity of the results [11]. In this section we discuss the potential threats that are relevant for our study and how they are addressed.

A. **Construct Validity**

Construct validity concerns the degree of accuracy to which the variables defined in the study measure the constructs of interests. In this study, we have followed the constructs defined in the two original FTS experiments [8] [12].

B. **Internal Validity**

Threats to internal validity influence the conclusions about a possible causal relationship between the treatment and the outcome of a study. We identified a couple of such kind of threats.

- **Skill to perform the task**: selected students had different skills and could potentially influence tasks performance. We randomly select students for each approach to minimize this threat.
- **Experiment unit**: we used an experimental condition different from a real scenario of software development. Students’ iteration included additional tasks, which are not usually done in a typical “real” iteration.
- **Measures**: our measures can be imperfect, since we simulate a full day work with only few minutes. We acknowledge that, but observe that any experiment would need to drastically reduce effort to represent a full day in an experimental setting.
- **The type of task**: participants manipulate maps and do not carry out real programming tasks. In this case, we refer to the original experiments where the authors indicate that they “decided in favor of a fictial map task instead to eliminate possible confounds due to differences in the software programming abilities of the participants” [8] [12].

C. **External Validity**

External validity describes the study representativeness and the ability to generalize the results outside the scope of the study. For any academic laboratory experiment the ability to generalize results to industry practice is restricted by the usage of students as study participants. Although the students may not be representative of the entire population of software professionals, it has been shown that the differences between students and real developers may not be as large as assumed by previous research [19].

555
D. Conclusion Validity

Conclusion validity is concerned with the relationship between the treatment and the outcome. We acknowledge that the small number of data points is not ideal from the statistical point of view. Small sample sizes, especially when the key experimental unit is at the team level, are a known problem difficult to overcome.

VII. Conclusion and Future Work

In this paper, we executed an experiment in order to investigate both the adaptive and prescriptive approaches in the context of FTS software development. We found that the usage of adaptive approaches increases the speed, but they do not always enhance accuracy and quality of the work done by distributed sites. We believe that this experiment has important findings that contribute to the literature on global software engineering and follow the sun software development.

For future work, we suggest new studies in order to replicate this experiment with more participants and different shifts scenarios (e.g. some time-zone overlap). We also suggest the replication of this experiment with industry participants.

ACKNOWLEDGMENT

The authors are funded by the PDTI program, funded by Dell Computers of Brazil Ltd. (Law 8.248/91). The third author is also funded by CNPq (projects 483125/2010-5 and 550130/2011-0). The authors thank all the students who participated in the experiment and Alberto Espinosa, Ning Nan and Erran Carmel for providing us the material used in their experiment.

REFERENCES