# Language Matters

Yi Wang
IBM Almaden Research Center
650 Harry Road, San Jose, CA 95120, USA
wangyi@us.ibm.com

***Abstract*—*Context*: In order to gain competitive advantages in the global market, many software outsourcing organizations in non-English speaking countries adopted the use of English as the lingua franca mandate in their organizations.**
***Goal*: We sought to assess the influence of the English lingua franca mandate on the teamwork in software outsourcing vendors empirically.**
***Approach*: We conducted a field experiment performed in a Chinese software outsourcing company. We measured the teamwork related constructs repeatedly with the progress of field experiment process.**
***Results*: This study reveals enforcing the English lingua franca may lead to: (1) lower satisfaction at work, (2) the decrease of teamwork quality, and (3) the lack of coordination in terms of social-technical congruence.**
***Conclusion*: Decision makers may need to be careful before launching the English lingua franca mandate. We also conceptually discusses how the English lingua franca mandate's influence on software quality. In a broader sense, the research calls attentions to the "language" issues in global software engineering practices, which is well-known but has not yet received enough attentions in software engineering research.**

> *...for good or ill the language of software development is English - albeit the US variant.*
> *- A. Bryant [1], 2000.*

> *I think it is more or less ridiculous. All of us are Chinese, but we need to pretend we are foreigners and speak English in office. It is so unnatural, so I decide to keep silent.*
> *- A senior software engineer at a Chinese software outsourcing company experimented English only policy.*

## I. Introduction

Language is a challenge for global software engineering [2]. As Bryant [1] claimed: "the language of software development is English." For software engineers who are not native English speakers, English is often considered to be an important communication skill [3] [4] in their workplace. Global software development organizations, especially those offshore outsourcing vendors in non-English speaking countries, language is often a barrier for them to expand their business. Decision makers in these organizations are often interested in setting English as the lingua franca in their organizations. They believe this would help them develop their human resource, improve communication with oversea partners, improve the product development process, and eventually, bring competitive advantages [5].

However, using English as the lingua franca does not necessarily always bring favorable outcomes. It may lead to some unfavorable side effects, for instance, avoidance of communication among non-native speakers [6], status struggles [7], power relationship [8], subgrouping [9], and so on. However, current research related to English as the lingua franca is mostly done by management scholars whose main interests are psychological and political aspects of organizational behaviors rather than teamwork. Software engineering researchers also mainly focus on techniques that support multilingual communications [10], [11]. Hence, we know little about how English-only policy influence team collaboration and coordination, especially in the setting of global software engineering context. Since the English-only policy may impact the communication within the team, it may further influence the teamwork. We have not yet built a clear understanding on this issue.

These issues around language are critical, particularly for outsourcing vendors in non-English speaking countries. Let's take China as an example, to overcome the poor English competency[1], Chinese outsourcing vendors are much more interested to establish English only policy in their organizations. Their eagerness of improving English competency may lead them to overlook the risks arising with the English-only policy, particularly, the negative influences on team collaboration and coordination.

While the use of English as the lingua franca receives increasingly scholarly attention in recent years, software engineering research has not yet investigates its influence to software development in non-English speaking software outsourcing organizations. The present study aims to contribute to develop a first-of-a-kind empirical inquiry focusing on the influences of English as the lingua franca on teamwork in software organizations, particularly, on individual's satisfaction [12], teamwork quality, and coordination. We designed and performed a field experiment at Chosco (a pseudonym), a Chinese outsourcing company, with the assistance from its management team.

In particular, the study provided the opportunity to explore several related research questions:

RQ1: *How does the use of English as the lingua franca influence team member's satisfaction at work?*

[1]The Future of China's Outsourcing Industry: http://goo.gl/qdAJzH, an industrial survey conducted by Shanghai Jiao Tong University and Fuqua Business School in Duke University.

RQ2: *How does the use of English as the lingua franca influence the teamwork quality?*

RQ3: *How does the use of English as the lingua franca influence the coordination among team members?*

This study reveals enforcing the English lingua franca may lead to: (1) lower satisfaction at work, (2) the decrease of teamwork quality, and (3) the lack of coordination in terms of social-technical congruence. These findings indicate the policy makers in software outsourcing vendors may need to be careful when making decisions about using English as lingua franca in their organization.

The reminder of this article proceeds as follows. Section III presents the field experiment design. Section IV discusses the measures for three main constructs. Section V introduces how we analyzed collected data. Section VI presents the results and findings. Section VII summarizes the main findings, discusses the implications, limitations, and future work. In this section, we link this research with the empirical software engineering literature to develop a conceptual model which related the English lingua franca mandate to software quality and productivity. Section VIII concludes the paper.

## II. Research Approach

The collaboration with Chosco gave us the opportunity to address the three research questions with a field experiment in real world settings. This section presents the context, study design, and execution. The main method employed in this study is **field experiment**, which refers examining an experimental intervention in the real world (i.e., the natural settings) rather than in the laboratory. The study design, analysis, and report[2] followed the standard field experiment design [13] and the best experiences of empirical software engineering experience [14] [15] [16] [17] [18].

The three research questions indicates three major constructs: *satisfaction at work*, *teamwork quality*, and *coordination*. The measures for them will be introduced in next section. In this section, we simply suppose they are measurable, and, of course, they are.

### A. Research Context: Chosco

Chosco is an entrepreneurial software development company located in Shanghai. Its has three major software development divisions for clients from different regions. These divisions are European and American, Domestic, and Japanese business divisions. The European and American business division had 137 full-time employees (07/01/2014) and a pool of around 150 part-time employees[3]. Chosco maintains a tight relationship with a highly prestigious university in Shanghai. The majority of its part-time employees are senior undergraduate and graduate students in that university. Almost all staff members in the division have college or above education. They have acceptable business English proficiency. There were 39 on-going projects within the division when the English-only policy started. 36 of them spanned the whole experiment period (07/01/2014-09/03/2014).

*Lingua Franca at Chosco*

The senior management of Chosco is ambitious to expand their outsourcing business. They view the lack of English competency as one of the major barriers in bidding foreign contracts. The assumed the use of English as the lingua franca might be a good choice. Compared with other alternatives (e.g., hiring profession English training agencies), this method was supposed more cost-effective.

Initially, Chosco's CEO contacted the author to seek personal suggestions on the English lingua franca mandate. Due to the lack of theory and empirical evidences on this topic, the author suggested to conduct an experiment in his company to examine the policy, and decided whether or not to make the English lingua franca mandate to be a long term requirement for developers. Then, the experiment plan was approved. The management team decided to launch the English lingua franca mandate temporarily and reserved the right of revoking or continuing the policy according to the preliminary results of the experiment. In this sense, the field experiment is more or less an instance of *action research* which actually influenced the organizational policy.

The launching of the English lingua franca mandate was in 07/01/2014. The policy has three major requirements to the influence developers:

- Only English is allowed to speak in the workspace.
- All working email, chatting, and meeting note should be in English.
- All documents, including those only for internal use, must be in English.

### B. Field Experiment Design & Process

*Projects and Subjects*

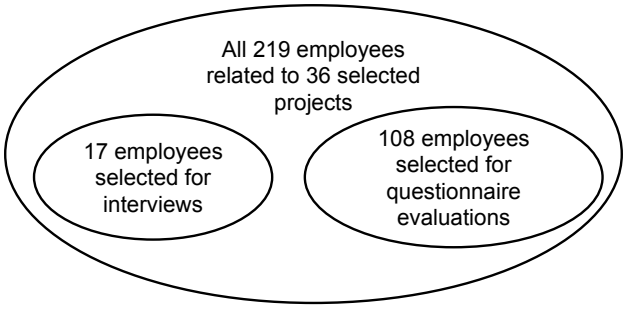


Fig. 1: The relationship between two selected samples.

A total of 36 software development teams from Chosco participated in this research. These teams were in Chosco's European and American business division, so they were the

[2]A non-academic/technical report was also prepared to Chosco for its internal use.

[3]Most of part-time employees are students in local universities. Hiring temporary staff is mostly based on operating cost considerations. The study was conducted in summer, so most of part-time employees can work full-time for summer break in their university.

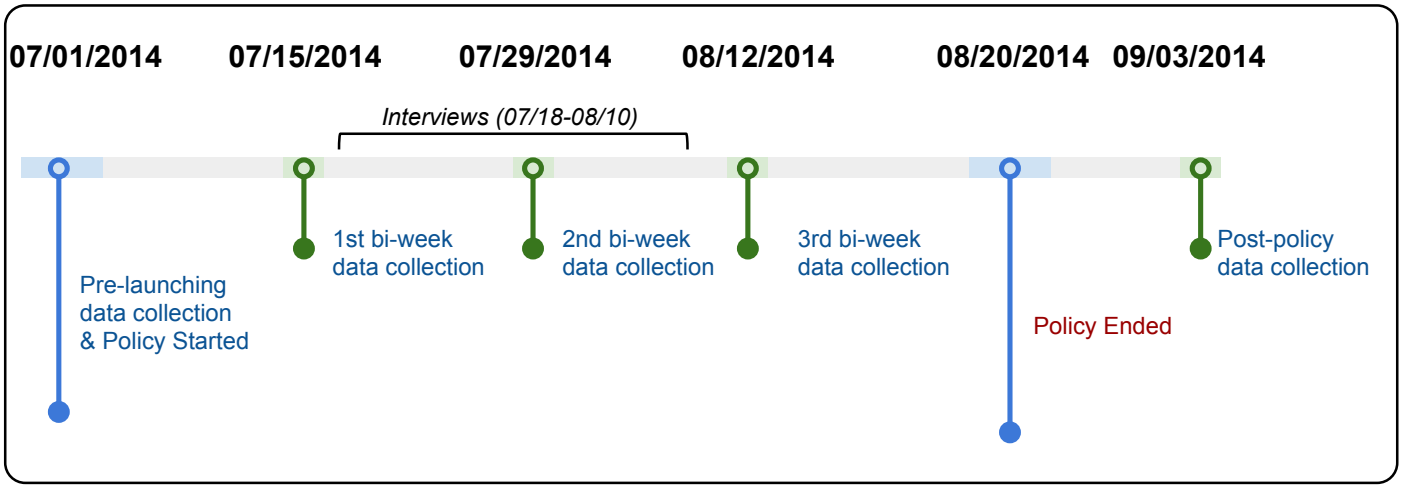


Fig. 2: The time-line describes the progress of the field experiment process. We marked the major events and data collection points in the time-line.

teams impacted by the implementation of lingua franca policy. All these teams maintained active status during the whole period of the experiment. There are 219 technical members (including full-time and part-time employees) in these 36 teams. Chosco's management provided a list of projects including names and contact information of team members to the author, while informing the team members that a series of division-wide evaluations was to be conducted. We did not directly tell them the purpose of the experiment, but used an excuse that the company was working with an external consultant to improve the software process. This intervention helped to avoid conscious and unconscious bias introduced by knowing the goal of research.

For each team, we pick the manager and two random selected to provided their evaluations through pre-designed questionnaires (see APPENDIX) focusing on the first two research questions.Another 17 employees were picked for short interviews. Figure 1 shows the relationship between two selected samples. The reason for selecting two separate samples of subjects is that we wanted to exclude the potential mutual influences between interviews and questionnaire evaluations.

*Experiment Process*

The author and a web engineer in Chosco created an internal website to host the questionnaire. The administration of this questionnaire distribution and collection was performed by two internal human resource associates in Chosco. We leveraged Chosco's Human Resource management system to distribute questionnaire links and to track responses, which ensured all selected subjects responded the questionnaire for they didn't want to leave "uncooperative" records in the system. Doing so also enabled us to track who answered the questionnaire but still kept their response anonymous.

The time-line in figure 2 show the process of the experiment. The main tasks in each step are as follows.

Step 0 06/23/2014: Informed the launching of English lingua franca policy, and sent the data collection notices to the 108 selected employees (this step is not shown in figure 2.

Step 1 07/01/2014: Performed the pre-launching data collection (overall, the 1st data collection). The English lingua franca mandate started after we collected all data.

Step 2 07/15/2014: Performed the 1st bi-weekly data collection (overall, the 2nd data collection). Interviews started in 07/18.

Step 3 07/29/2014: Performed the 2nd bi-weekly data collection (overall, the 3rd data collection). Interview ended in 08/10.

Step 4 08/12/2014: Performed the 3rd bi-weekly data collection (overall, the 4th data collection).

Step 5 08/20/2014: The English lingua franca mandate ended. In 08/17/2014, a report of preliminary results was delivered to Chosco's management team. Chosco's CEO decided to end the policy according to the report.

Step 6 09/03/2014: Performed the post-policy data collection (overall, the 5th and the last data collection).

After all steps and finishing data analyses, a note was sent to all participants to debrief the study (including the real purpose and the high-level overview of findings) and to acknowledge their participation.

*Data Collection*

The quantitative data for first two constructs was collected through the five-wave questionnaires (Step1-4, and Step6). Meanwhile, we computed the measure of *coordination* (the third construct) for five times accordingly.

We collected two types of qualitative data. The first type of data comes from the communication records of the studied projects. We also conducted semi-structured interviews with team members to understand their perception and experience

with English lingua franca mandate during the "experimenting" period. We interviewed 17 individuals with an identical interview protocol. The 17 individuals were selected from those who were NOT selected to do questionnaire evaluations (see figure 1). Each interview is around 20 minutes[4]. Interviewees were allowed to talk about anything they want to share. The same researcher conducted all interview sessions while the interviews were administrated in Chinese.

## III. Measures

In this section, we introduces the measures for the the three main constructs: *job satisfaction*, *teamwork quality*, and *coordination*.

First two constructs (satisfaction at work and teamwork quality), and are measured by subjective, self-reported measures and adapted from standard psychological tests. Using standard measurements brings benefits such as better reliability, and also facilitates the replication of the empirical software engineering research. Both constructs considered in this study use the team as the unit of analysis. Therefore, all measures were evaluated on the team level although we collected the raw data from the individual respondents. Thus, we asked respondents to evaluate the characteristics and behaviors of their team as a whole. The individual evaluations from the same team were aggregated together. Both measurements demonstrated strong reliability (see the Cronbach's alpha coefficients[5] in the APPENDIX). Before the aggregation, we calculated the interrater reliability (IRR) with the method in [19]. In general, the results indicate quite strong agreement of ratings within the same team. The average scores of this test across all teams are between 0.79 and 0.96.

The measure of the third constructs (*coordination*) is developed by software engineering researchers [20]. It was extracted from the artifacts and did not require collecting self-report data. We will introduce its definition and calculation in next subsection.

### A. Details of the Measures

*Satisfaction at Work*

We adapted the 3-items questionnaires developed in [21] to measure the "Satisfaction at Work". In [21], Mathew et al. used it to measure the job attitude of software engineers working for Indian outsource vendors. Therefore, using it to measure people with similar job role in similar organizations should be no problem. To increase the validity of using it, we also tested the reliability with collected data.

*Teamwork Quality: TWQ*

To measure the teamwork quality, we adapted the well-developed, and widely-adopted teamwork quality questionnaire (TWQ) [22]. TWQ consists of 37 items that belongs to one of the 6 facets (i.e., *Communication, Coordination, Balance of Member Contributions, Mutual Support, Effort, Cohesion*).

The TWQ is a high order construct of the six facets. To calculate the TWQ, we first performed a factor analysis (principle component) on the team level using aggregated team member responses. The results show the correctness of latent construct assumption concerning the TWQ construct. Therefore, simply using the average of 6 facets' score may be not valid. Therefore, we followed the procedures suggested in Hogel & Gemuenden [22] to compute the TWQ as a one-factor measure, and tested its reliability. This method is well-adopted by researchers in extracting one-factor measures [23].

*Coordination: Social-technical Congruence*

*1) Definition of STC:*
In addition to measuring subject self-reported coordination as a part of team work quality, we also measured coordination using social-technical congruence (STC) developed in [20]. STC measures the differences between *coordination needs (CN)* and *actual coordinations (AC)*. Coordination needs are dyadic relationships that indicate that two people should be coordinating. It can be identified through studying the technical dependencies between the work items and the assignments of these work items to each developer. Actual coordination reflects how people in the organization do coordinate with each other. Both coordination needs and actual coordinations are represented by two $n \times n$ matrix, where $n$ is the number of developers in a team.

Suppose there are $n$ developers and $m$ work items, we use a $n \times m$ matrix $A$ to denote the assignments of $m$ work item to the $n$ developers, and a $n \times n$ matrix $D$ to denote the dependencies between $n$ work items. The coordination needs $(CN)$ can be written as:

$$CN = A \times D \times A^t \tag{1}$$

$CN_{ij} \neq 0$ means developers $i$ and $j$ should coordinate. The actual coordination matrix $AC$ can be built through mining the communication records between team members. $AC_{ij} \neq 0$ means developer $i$ and $j$ have actually coordinated. STC can be calculated by comparing $CN$ to $AC$. If a non-zero element $AC_{ij}$ exists in the actual coordination matrix and the corresponding non-zero $CN_{ij}$ exists in the coordination needs matrix, then there is congruence between developers $i$ and $j$ (the coordination need between $i$ and $j$ is satisfied). We define two measures in equation (1) and (2):

$$DIFF(CN, AC) = card\{diff_{ij} | CN_{ij} \neq 0 \& AC_{ij} \neq 0\} \tag{2}$$

$$|CN| = card\{CN_{ij} \neq 0\} \tag{3}$$

Then, we can calculate the STC as:

$$STC = \frac{DIFF(CN, AC)}{|CN|}$$

*2) Extracting STC:*
The assignment matrix $A$ was directly extracted from the project lifecycle management platform (Most teams were using Atlassian JIRA[6]). For some organizational restrictions,

[4]We kept the interview concise to avoid excessive interruptions to interviewees' work.

[5]The Cronbach's alpha coefficients were calculated at individual level. We reported the averages of all five times of measurements during the experiment process.

[6]https://www.atlassian.com/project-management-software

we cannot access the source code to extract the actual code dependencies. The dependency between work items (matrix $D$ in equation (1)) were provided by project architects (or project mangers server architect duties) of the target projects. The matrix $AC$ was extracted from the lifecycle management platform (including issue tracking) and the email archival. Please note that, we did not only consider the new work items but also incorporate all work items that have someone actively worked on in the bi-week time frame. Doing so avoided the problem of missing work items whose life cycle crossed the data collection time points.

## IV. Data Analysis

We took mixed methods [24] to analyze the quantitative and qualitative data collected from the experiment. We gave higher priority to quantitative analysis. Doing so assumes a postpositivist stance, as well as the simultaneous belief that the inclusion of qualitative data and analysis is likely to increase understanding of the underlying phenomenon [25].

*Quantitative Data Analysis*

Obviously, the three constructs were repeatedly measured (see figure 1). To assess the influences of the use of English as the lingua franca, we performed Repeated Measures ANOVA. We performed post-hoc analyses to examine the pairwise comparison results, to assess the sphericity, and necessary statistical corrections. All statistical analyses were performed with `R 3.1.2` statistical package [26].

*Qualitative Data Analysis*

We performed light-weight qualitative analysis [27] on the interview transcripts, and the selected company communication materials using ATLAS.ti[7], a qualitative data coding software [28]. The qualitative data analysis aimed to identify the underlying reasons behind the quantitative findings. It also could be viewed as the cross-examinations of quantitative findings. We did not perform a complete qualitative analysis since it is subordinate to the quantitative analysis. We drove the qualitative data analysis with the quantitative findings. Thus, the qualitative data was not open coded but analyzed with quantitative findings as "a priori codes".

## V. Results

In this section, we reported the results and findings emerge from the data analysis. The report of statistical analyses adheres to the recommendations in the APA publication manual [29].

### A. RQ1: Team Members' Perceptions

For software engineer's *stratification at work*, a repeated measures ANOVA with a Greenhouse-Geisser correction[8] determined that the *satisfaction at work* differed statistically significantly between five data collection time points (see figure 1). The statistics is: $F(3.033, 106.151) = 14.116, P < 0.001$. Post hoc tests using the Bonferroni correction revealed a set of pair observations are significantly different. They are: (1) Pre-experiment vs. 1st data collection ($P < 0.001$); (2) Pre-experiment vs. 2nd data collection ($P < 0.001$); (3) Pre-experiment vs. 3rd data collection ($P < 0.001$); (4) 2nd data collection vs. Post-experiment ($P < 0.001$); (5) 3nd data collection vs. Post-experiment ($P = 0.041$). Figure 3 shows the distribution of *satisfaction at work* over the five data collection points. It shows that the *satisfaction at work* is relative low compared to pre and post experiment data collection.

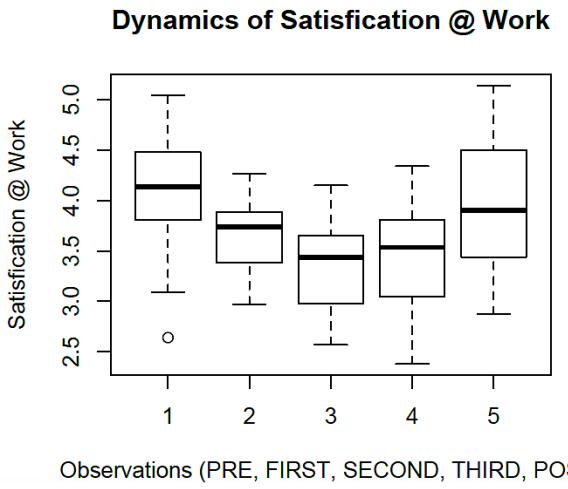


Fig. 3: The distribution of *satisfaction at work* over the five data collection points.

Based on the statistical analyses, we can conclude:

> **Finding 1:** *Introducing English lingua franca mandate elicits a statistically significant decrease in satisfaction at work. After the policy ended, satisfaction at work gets some quick increase.*

The qualitative data also indicates the general negative attitudes towards the English lingua francs mandate. As one interviewee commented:

> *I understand why they want us to speak English here. But the problem is, is it useful? At least for many of us, it is useless. I am a programmer and have no opportunities to speak with clients. I think our CEO may want the client feel happy when they visit our office. But I doubt the effect of it. Seeing a crowd of people speaking poor* Chinglish *in the office would not make the clients happy. (P2)*

### B. RQ2: Teamwork Quality

For project's *teamwork quality*, a repeated measures ANOVA with the assumption of sphericity determined that the TWQ differed statistically significantly between data collection time points (see figure 1). The statistics is: $F(4, 140) = 31.273, P < 0.001$. Post hoc tests using the Bonferroni

[7] http://atlasti.com/

[8] If data violated the assumption of sphericity, a Greenhouse-Geisser correction is necessary.

correction revealed a set of pair observations are significantly different. They are: (1) Pre-experiment vs. 1st data collection ($P < 0.001$); (2) Pre-experiment vs. 2nd data collection ($P < 0.001$); (3) Pre-experiment vs. 3rd data collection ($P < 0.001$); (4) Pre-experiment vs. Post-experiment ($P < 0.001$); (5) 1st data collection vs. Post-experiment ($P < 0.011$); (5) 2nd data collection vs. Post-experiment ($P = 0.033$). Figure 4 shows the distribution of *TWQ* over the five data collection points. It shows that the *TWQ* is relative low compared to pre- and post-experiment data collection. Different from the *satisfaction at work*, it seems the recovery of *teamwork quality* is relative slow (no significant differences with 3rd data collection).

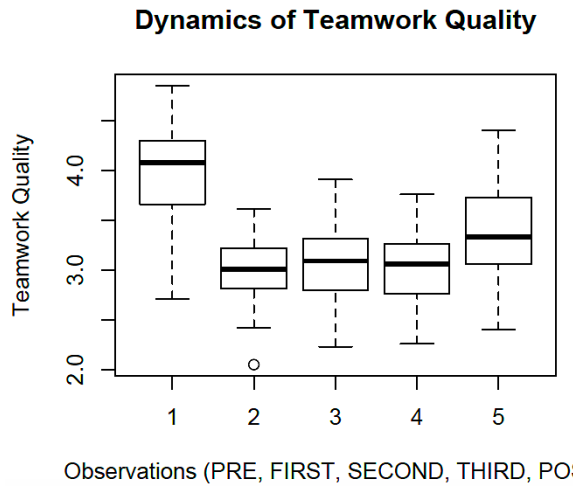


Fig. 4: The distribution of *teamwork quality* over the five data collection points.

Five out six dimensions of TWQ were lower in the samples of 1st, 2nd, and 3rd data collections. Surprisingly, *cohesion* as one of the six dimensions of TWQ, exhibited a slight increase in the 1st data collection points (Average: 3.81 vs. 3.97). The improved cohesion may result from the "sympathy" emerged from sharing sufferings, complaints, and negative attitudes towards the management team. Some literature (e.g., [30]) has shown sympathy (or even anger) may potentially lead to high cohesion. As one interviewee said:

> *I think all my colleagues don't like it. I can hear people's complaint... complaints in English! This makes me feel a little bit better. At least, I am not the alone. (P6)*

The qualitative data indicates the low TWQ may be caused by three reasons. The foremost one is *communication avoidance*. There are many reasons for reduce the communications with other team members. The most mentioned reason is the fear of losing face ("mianzi" in Chinese [31]). For instance, an interviewee said: "*Even myself feel embarrassed or funny when listening my Szechwan*[9] *accent English, let alone others.*" The second reason is the incapability of using English to describe complex problems. The last reason is a "non-cooperative" attitude towards non-sense organizational policy.

Based on above analyses, we can conclude:

> **Finding 2:** *Introducing English lingua franca mandate elicits a statistically significant decrease in teamwork quality. After the policy ended, there is some increase in teamwork quality but the increase is not fast. However, cohesion may be slightly improved.*

### C. RQ3: Coordination (STC)

For *social-technical congruence*, a repeated measures ANOVA with the assumption of sphericity determined that the STC differed statistically significantly between data collection time points (see figure 1). The statistics is: $F(4, 140) = 8.820, P < 0.001$. Post hoc tests using the Bonferroni correction revealed a set of pair observations are significantly different. They are: (1) Pre-experiment vs. 1st data collection ($P < 0.001$); (2) Pre-experiment vs. 2nd data collection ($P < 0.001$); (3) Pre-experiment vs. 3rd data collection ($P < 0.001$). We found that the the difference is marginal significant in pairwise comparison of Pre-experiment vs. Post-experiment ($P = 0.087$).

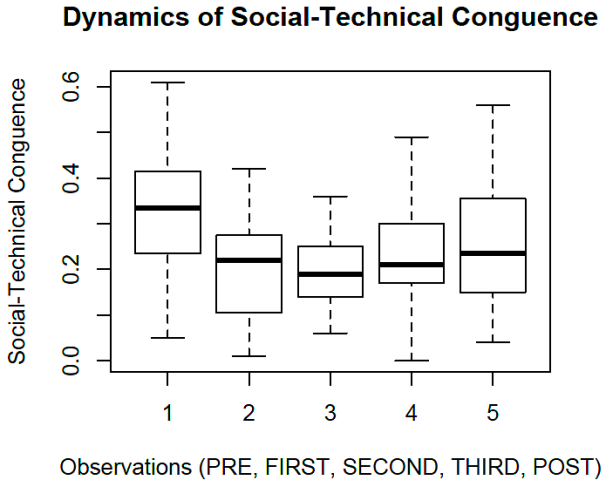


Fig. 5: The distribution of *social-technical congruence* over the five data collection points.

Figure 5 shows the distribution of *STC* over the five data collection points. Figure 5 indicates the variances become larger in Post-experiment condition although the change of average STC is not significant. This may indirectly indicate some teams are more sensitive to the policy change and have better execution. An interviewee was pretty proud for his team's agility in facing organizational policy changes:

> *None of us like it [the English lingua franca]. But I think it is not a big problem for us. We can deal with it. Our team is good at dealing with tough things.*

[9]Szechuan is a province in southwestern China.

His team actually did well. Their STCs in the five data collection points are: 0.52, 0.36, 0.34, 0.38, 0.45, which is much better than the majority.

Based on above analysis, we can conclude:

**Finding 3:** *Introducing English lingua franca mandate elicits a statistically significant decrease in coordination (measured by social-technical congruence). Coordination may be restored faster in some teams while the recovery is slow in general.*

## VI. Discussion

### A. Contributions to Literature

First of all, the research calls attention to the *language* issue, especially the use of English as the lingua franca in global software engineering teams. Language is important in global collaboration [2]. However, it is fair to say there is not much serious literature on this topic. This research calls attentions on the research related to "language" GSE. We hope there would be more future work on this topic.

Although we did not directly evaluated the influence of the use of English as the lingua franca on software artifacts, we still can hypothesize that it may lead to some quality loss. Software engineering researchers have built solid knowledge on the positive influence of good "teamwork" (e.g., [32], [22]) and "coordination" (e.g., [20], [33]) on software quality. The construct "satisfaction at work" is not so directly linked to software quality. However, literature such as [34] [35] has established the relationship between satisfaction and developer's motivation, and associated highly motivated developers with improved software productivity and quality.

Combining with these results, we developed a conceptual model. Figure 6 depicts the model. The left part is developed in this paper while the right part is drawn from the literature. This model indicates the English lingua franca may lead to software quality loss through lowering the *satisfaction at work*, *teamwork quality*, and *coordination*. This model can be examined by future empirical studies, and used as the workbench to incorporating more constructs.

The way we develop this research reflects its interdisciplinary nature. We borrowed measures from management literature, introducing it to software engineering community. Particularly, GSE researchers may need to look at the literature in areas such as international business, organizational behavior, etc. Results from software engineering, e.g., social-technical congruence, may be valuable for management researchers.

Last but not least, this paper demonstrates field experiment can be a useful method in empirical software engineering. It can be used in testing existing theories, methods, techniques, and to inspire reflection on of how to adapt them to make sense in real world environment.

### B. Contributions to Practice

The research has several implications to practices. First of all, software outsourcing vendors in non-English speaking countries may need to be more careful when introduce the English lingua franca mandate. It is necessary to evaluate whether the benefit over the loss. They may adopt the experimenting methods developed in this paper in their evaluation to minimize the potential side effect. From the client's perspective, it might be not necessary to force the vendors to adopt the English lingua franca mandate.The client should careful evaluate the technical qualification of different vendors. Simply associating "establishing English lingua franca" with "high competency" could be very risky and encourage software outsourcing vendors to cut corners.

### C. Research and Industrial Practices: Rethink Researchers' Role

The research demonstrates close collaboration with industrial partners will not only yield insightful research findings but also will make real world impacts. Researchers may need to rethink the role of empirical software engineering as a way to improve the industrial practices, as well as their potential role as advisors or consultants to industry. In the history of software engineering research, many important results the researchers' deep involvements in industry practices. For instance, *The Mythical of Man-Month* resulted from Brooks' experiences at IBM when managing the development of OS/360 [36], Barry Boehm's COCOMO [37] is based on his experience in TRW Aerospace. Given the nature of software engineering as an Engineering discipline, keeping our tradition on research highly relevant to practices may be very important, and crucial to the success of our discipline.

### D. Limitations and Potential Extensions

Our study has several limitations that may also serve as extensions for future research. One limitation arises from the use of a single company to collect data. Conducting field experiment in natural setting is difficult. The opportunity of conducting this study is almost by accident but not through seeking. For example, the part-time developers were working full-time due to the summer break in Chinese universities, which enabled us to reach most of them. However, we can conduct well-designed laboratory experiment in future with the minimal sacrifice of reality. Lab experiment will also provide better controls for developing more generalizable results. Other researchers who have strong industry connections may try to replicate this study. To facilitate the replications, we enclosed all questionnaire items in the APPENDIX.

Second, we could not control the progress of the projects studied in this research. In different phase of a project's life cycle, team members may experience a different level of pressure, hence may have different levels of satisfaction. Therefore, it may partially contribute to the correlation between the English-only policy and low satisfaction at work. In a broader sense, different teams or projects may have different characteristics, for instance, some team may have better execution (see the indirect evidence in section V-C). Investigating the variations among teams and projects would have potential to develop a rich future research agenda.

Third, management literature has explored a set of constructs that may be influenced by the lingua franca mandate

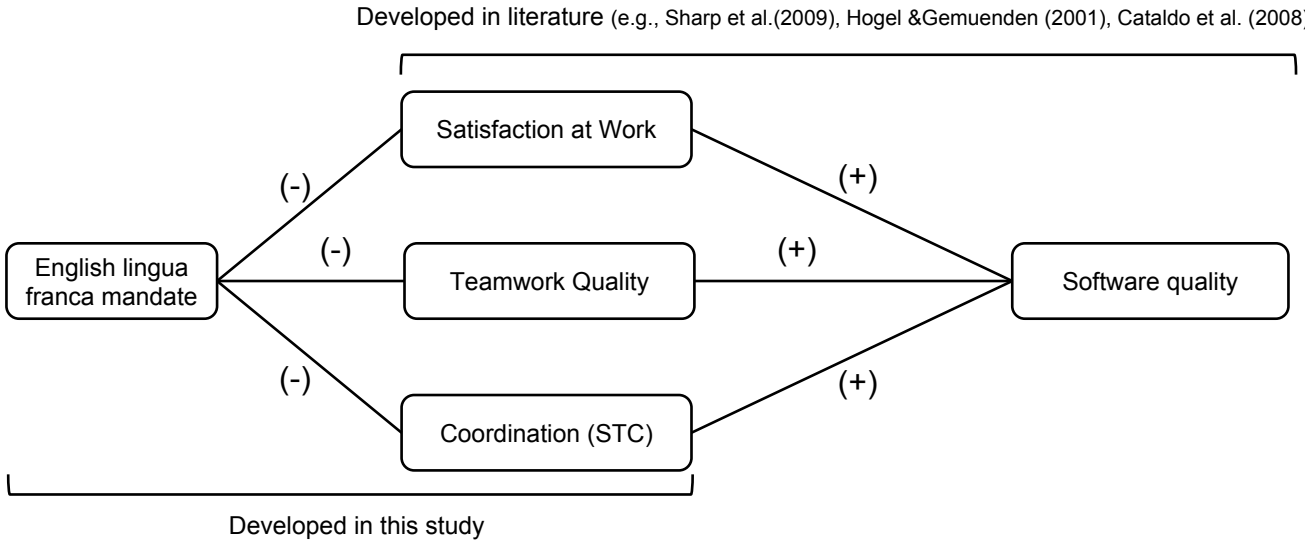


Fig. 6: The conceptual model linking the English lingua franca mandate with software quality based on this study and prior literature.

(e.g., status, powers, see section I). It might be valuable to leverage and integrate them with software engineering research through the way what we outlined in figure 6, and then, to evaluate them empirically. This study opens up opportunities for future study that cuts across the traditional disciplinary boundaries of software engineering and management.

## VII. Conclusion

In this paper, we describe a field experiment that empirically assesses the influence of the English lingua franca mandate on the teamwork in non-English speaking software outsourcing vendors. The results demonstrate the English lingua franca mandate may arise detriment to individual's satisfaction at work, teamwork quality and coordination among developers. The findings suggest the decision of the English lingua franca mandate need to be carefully evaluated. In another place, clients may not need to require the vendors to adopt the English lingua franca mandate. The research calls attentions to the "language" issues in global software engineering practices, which is well-known but has not yet received enough attentions in research. It also demonstrates the feasibility of interdisciplinary study that combines software engineering and management.

## Acknowledgment

The author would like to thank Chosco management team, particularly its CEO James Guo and all participants of this study for their support to this study.

## References

[1] A. Bryant, It's engineering Jim? but not as we know it: software engineering solution to the software crisis, or part of the problem?. *Proceedings of the 22nd international conference on Software engineering (ICSE'00)*, 78-87, 2000.

[2] J. D. Herbsleb, Global software engineering: The future of socio-technical coordination. *Proceedings of the 2007 Future of Software Engineering (FOSE'07)*, 188-198, 2007.

[3] P. Bourque, R. Dupuis, A. Abran, J. W. Moore, and L. Tripp. The guide to the software engineering body of knowledge. *IEEE software*, Vol.16, No. 6, 35-44, 1999.

[4] M. J. Riemer, English and communication skills for the global engineer. *Global Journal of Engineering Education*, Vol. 6, No. 1, 91-100, 2002.

[5] D. Kapur, and R. Ramamurti, India's emerging competitive advantage in services. *The Academy of Management Executive* Vol. 15, No. 2, 20-32, 2001.

[6] J. Lauring, and A. Klitmøller, Corporate language-based communication avoidance in MNCs: A multi-sited ethnography approach. *Journal of World Business*, Vol. 50, No. 1, 46-55, 2015.

[7] T. B. Neeley, Language matters: Status loss and achieved status distinctions in global organizations. *Organization Science*, Vol. 24, No. 2, 476-497, 2013.

[8] M. Boussebaa, S. Sinha, and Y. Gabriel, Englishization in offshore call centers: A postcolonial perspective. *Journal of International Business Studies*, Vol. 45, No. 9, 1152-1169, 2014.

[9] P. J. Hinds, T. B. Neeley, and C. D. Cramton, Language as a lightning rod: Power contests, emotion regulation, and subgroup dynamics in global teams. *Journal of International Business Studies*, Vol. 45, No. 5, 536-561, 2013.

[10] F. Calefato, F. Lanubile, R. Prikladnicki, and J. H. S. Pinto, An empirical simulation-based study of real-time speech translation for multilingual global project teams. *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '14)* 56, ACM.

[11] F. Calefato, F. Lanubile, T. Conte, and R. Prikladnicki, Assessing the impact of real-time machine translation on multilingual meetings in global software projects. *Empirical Software Engineering*, 1-33, 2015.

[12] H. Sharp, N. Baddoo, S. Beecham, T. Hall, and H. Robinson, Models of motivation in software engineering. *Information and Software Technology*, Vol. 51, No. 1, 219-233, 2009.

[13] A. S. Gerber and D. P. Green, *Field Experiments: Design, Analysis, and Interpretation*. WW Norton, 2012.

[14] T. Dybå, V. B. Kampenes, and D. I. Sjøberg, A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, Vol. 48, No. 8, 745-755, 2006.

[15] S. Easterbrook, J. Singer, M. A. Storey, and D. Damian, Selecting empirical methods for software engineering research. *Guide to Advanced Empirical Software Engineering*,285-311, Springer London, 2008.

[16] J. E. Hannay, J. E., D. I. Sjøberg, and T. Dybå, A systematic review of theory use in software engineering experiments. *IEEE Transactions on Software Engineering*, Vol.33, No. 2, 87-107, 2007.

[17] A. Jedlitschka, M. Ciolkowski, and D. Pfahl, Reporting experiments in software engineering. *Guide to Advanced Empirical Software Sngineering*, 201-228, Springer London, 2008.

[18] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Björn, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[19] L. James, R. Demaree, and G. Wolf, Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, Vol. 69, No. 1, 85-98, 1984.

[20] M. Cataldo, J. D. Herbsleb, and K. M. Carley, Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '08)*, 2-11, 2008.

[21] J. Mathew, E. Ogbonna, and L. C. Harris, Culture, employee work outcomes and performance: An empirical analysis of Indian software firms. *Journal of World Business*, Vol. 47, No. 2, 194-203, 2012.

[22] M. Hoegl, and H. G. Gemuenden, Teamwork quality and the success of innovative projects. *Organization Science*, Vol. 12, No. 4, 435-449, 2001.

[23] B. Thompson, *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association, 2004.

[24] S. C. Petter, and M. J. Gallivan, Toward a framework for classifying and guiding mixed method research in information systems. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*. 2004.

[25] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner, Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, Vol. 1, 112-133, 2007.

[26] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/, 2013.

[27] M. Q. Patton, *Qualitative Evaluation and Research Methods*. SAGE Publications, 1990.

[28] S. Friese, *Qualitative Data Analysis with ATLAS.ti*. SAGE Publications, 2011.

[29] American Psychological Association, *Publication Manual of the American Psychological Association, 6th edition*. American Psychological Association (APA), 2009.

[30] J. Cragan, D. W. Wright, and C. Kasch, *Communication in Small Groups: Theory, Process, and Skills*. Cengage Learning, 2008.

[31] C. L. H. Chang, The dilemma of renqing in ISD processes: interpretations from the perspectives of face, renqing and guanxi of Chinese cultural society. *Behaviour & Information Technology*, Vol. 31, No. 5, 481-493, 2012.

[32] I. Evans, *Achieving software quality through teamwork*. Artech House, 2004.

[33] I. Kwan, A. Schroter, and D. Damian, Does socio-technical congruence have an effect on software build success? A study of coordination in a software project. *IEEE Transactions on Software Engineering*, Vol. 37, No. 3, 307-324, 2011.

[34] C. França, H. Sharp, and F. Q. da Silva, Motivated software engineers are engaged and focused, while satisfied ones are happy. *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, No. 32 .2014.

[35] Y. Wang, What motivate software engineers to refactor source code? evidences from professional developers. *Proceedings of IEEE International Conference on Software Maintenance (ICSM'09)*, 413-416, 2009.

[36] F. P. Brooks Jr, *The Mythical Man-Month: Essays on Software Engineering, Anniversary Edition, 2/E*. Pearson, 1995.

[37] B. Boehm. *Software Engineering Economics*. Englewood Cliffs, NJ:Prentice-Hall, 1981

## APPENDIX

### *APPENDIX A: Satisfaction at work (3 items)*

*Cronbach's alpha = 0.76*
1. I like the people I work with.
2. There is too much infighting in this organization (r).
3. My organization is unfair to me (r).

### *APPENDIX B: TWQ (6 factors, 37 items)*

#### *1. Communication*

*Cronbach's alpha = 0.92*
1-1. There was frequent communication within the team.
1-2. The team members communicated often in spontaneous meetings, phone conversations, internet chat, etc.
1-3. The team members communicated mostly directly and personally with each other.
1-4. There were mediators through whom much communication was conducted (r).
1-5. Project-relevant information was shared openly by all team members.
1-6. Important information was kept away from other team members in certain situations (r).
1-7. In our team there were conflicts regarding the openness of the information flow (r).
1-8. The team members were happy with the timeliness in which they received information from other team members.
1-9. The team members were happy with the precision of the information received from other team members.
1-10. The team members were happy with the usefulness of the information received from other team members.

#### *2. Coordination*

*Cronbach's alpha = 0.79*
2-1. The work done on subtasks within the project was closely harmonized.
2-2. There were clear and fully comprehended goals for subtasks within our team.
2-3. The goals for subtasks were accepted by all team members.
2-4. There were conflicting interests in our team regarding subtasks/subgoals (r).

#### *3. Balance of Member Contributions*

*Cronbach's alpha = 0.73*
3-1. The team recognized the specific potentials (strengths and weaknesses) of individual team members.
3-2. The team members were contributing to the achievement of the team's goals in accordance with their specific potential.
3-3. Imbalance of member contributions caused conflicts in our team (r).

#### *4. Mutual Support*

*Cronbach's alpha = 0.87*
4-1. The team members helped and supported each other as best they could.

4-2. If conflicts came up, they were easily and quickly resolved.
4-3. Discussions and controversies were conducted constructively.
4-4. Suggestions and contributions of team members were respected.
4-5. Suggestions and contributions of team members were discussed and further developed.
4-6. Our team was able to reach consensus regarding important issues.

*5. Effort*

*Cronbach's alpha = 0.86*
5-1. Every team member fully pushed the project.
5-2. Every team member made the project their highest priority.
5-3. Our team put much effort into the project.
5-4. There were conflicts regarding the effort that team members put into the project (r).

*6. Cohesion*

*Cronbach's alpha = 0.81*
6-1. It was important to the members of our team to be part of this project.
6-2. The team did not see anything special in this project (r).
6-3. The team members were strongly attached to this project.
6-4. The project was important to our team.
6-5. All members were fully integrated in our team.
6-6. There were many personal conflicts in our team (r).
6-7. There was personal attraction between the members of our team.
6-8. Our team was sticking together (r).
6-9. The members of our team felt proud to be part of the team.
6-10. Every team member felt responsible for maintaining and protecting the team.

*APPENDIX C: Notes*

*1. (r) denotes the item was reversed coded.*
*2. A 5-point Likert scale is used for items of multi-item constructs.*
*3. At the end of each questionnaire, we asked a question at the end of the questionnaire. That is: "What do you think is the purpose of this study?". This question was used to check whether a participant figured out the real purpose. If he or she found the real purpose, the data case should be excluded in the data analysis.*