

How does working from home affect developer productivity? — A case study of Baidu during the COVID-19 pandemic

Lingfeng BAO¹, Tao LI², Xin XIA^{3*}, Kaiyu ZHU², Hui LI² & Xiaohu YANG¹¹College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China;²Baidu Inc., Beijing 100085, China;³Faculty of Information Technology, Monash University, Melbourne 3800, Australia

Received 18 December 2020/Revised 10 March 2021/Accepted 19 May 2021/Published online 14 March 2022

Abstract Nowadays, working from home (WFH) has become a popular work arrangement due to its many potential benefits for both companies and employees (e.g., increasing job satisfaction and retention of employees). Many previous studies have investigated the impact of WFH on the productivity of employees. However, most of these studies usually use a qualitative analysis method such as surveys and interviews, and the studied participants do not work from home for a long continuing time. Due to the outbreak of coronavirus disease 2019 (COVID-19), a large number of companies asked their employees to work from home, which provides us an opportunity to investigate whether WFH affects their productivity. In this study, to investigate the difference in developer productivity between WFH and working onsite, we conduct a quantitative analysis based on a dataset of developers' daily activities from Baidu Inc., one of the largest IT companies in China. In total, we collected approximately four thousand records of 139 developers' activities of 138 working days. Out of these records, 1103 records are submitted when developers work from home due to the COVID-19 pandemic. We find that WFH has both positive and negative impacts on developer productivity in terms of different metrics, e.g., the number of builds/commits/code reviews. We also notice that WFH has different impacts on projects with different characteristics including programming language, project type/age/size. For example, WFH has a negative impact on developer productivity for large projects. Additionally, we find that productivity varies for different developers. Based on these findings, we get some feedback from developers of Baidu and understand some reasons why WFH has different impacts on developer productivity. We also conclude several implications for both companies and developers.

Keywords WFH, developer productivity**Citation** Bao L F, Li T, Xia X, et al. How does working from home affect developer productivity? — A case study of Baidu during the COVID-19 pandemic. *Sci China Inf Sci*, 2022, 65(4): 142102, <https://doi.org/10.1007/s11432-020-3278-4>

1 Introduction

Working from home (WFH) is a work arrangement in which employees do not need to work at a central place (e.g., office building, warehouse, or store). WFH has various names, such as remote work, teleworking, or telecommuting. These terms are used differently and interchangeably from study to study [1–3]. Nowadays, since WFH is facilitated by many tools such as virtual private networks, cloud computing, and online meeting software, more and more companies allow their employees to work from home. A survey in 2018 from OWL labs shows that 52% of employees work from home at least once a week and 56% of companies allow remote work¹⁾. WFH can offer some benefits to both companies and employees, for instance, when employees can work from home, they feel more trusted and are better able to balance work and life responsibilities, which can increase employee retention and make them happier and more productive.

* Corresponding author (email: Xin.Xia@monash.edu)

1) <https://www.owllabs.com/state-of-remote-work/2018>.

WFH is also adopted by many IT companies, for instance, a recent report mentioned that Twitter announced staff can continue WFH permanently²). Developers can perform their daily tasks (e.g., writing code, debugging, building projects, and code review) as usual by remotely accessing resources of companies when working from home. WFH might have different impacts on productivity, which is a big concern of software developer organizations [4]. Understanding the difference of developer productivity when working from home and the reasons behind it can help improve the management of companies and projects, increase the job satisfaction of developers, and make developers more productive.

The survey of OWL labs reported that employees who work remotely at least once a month are 24% more likely to feel productive in their roles than those who do not or cannot work remotely. On the contrary, WFH might have a negative impact on productivity. For example, it would decrease the efficiency of developer communication, which plays an important role in software development [5]. Many studies have investigated the impacts of WFH on productivity [6–10]. However, most of these studies use a qualitative approach (e.g., survey or interview) based on the feedback from general workers (not only developers). Additionally, the studied participants usually do not work from home for a long continuing time. In this study, we aim to investigate the impacts of developer productivity when working from home for a long time in a quantitative way.

Due to the outbreak of coronavirus disease 2019 (COVID-19)³, which is an infectious disease caused by a newly discovered coronavirus, a large number of IT companies ask their employees to work from home, which provides us an opportunity to investigate how their productivity is affected when working from home for a long continuing time.

In this study, we collect the data from Baidu, Inc., China, which contains the development activities of 107 developers in 70 working days. There is a part of records in this dataset on which developers work from home due to the COVID-19 pandemic. We compare developer productivity when working from home with working onsite in terms of multiple aggregated values such as mean, median of several metrics (e.g., the number of builds, commits, and code reviews). We summarize our findings and contributions as follows:

- To the best of our knowledge, we are the first to investigate the impacts of WFH on developer productivity based on developers' daily activities. We find that WFH has both positive and negative impacts on developer productivity in terms of different metrics, such as the number of builds/commits/code reviews.
- We investigate the impacts of WFH on projects with different characteristics including program language and project type/age/size and find that WFH has different impacts on different kinds of projects. For example, WFH has a negative impact on developer productivity for large projects. We also find that productivity varies for different developers.
- We conclude the reasons why developers have different productivity when working from home and provide implications based on our findings and the feedback from Baidu.

Paper structure. The remainder of the paper is structured as follows. Section 2 describes the dataset and research questions in this study. Section 3 presents the results of the analysis for the six research questions. Section 4 discusses implications and threats to validity. Section 5 briefly reviews related work. Section 6 concludes the paper and discusses future directions.

2 Case study setup

In this section, we first present the dataset from Baidu. Then, we describe the research questions and their corresponding motivations.

2.1 Dataset

We collected a dataset of developers' daily activities from Baidu, Inc., which is the world's largest Chinese language Internet search provider⁴, the largest knowledge and information centered Internet platform company in China, and a world-leading artificial intelligence (AI) company.

2) <https://www.bbc.com/news/technology-52628119> (accessed at August 7, 2020).

3) <https://www.who.int/health-topics/coronavirus>.

4) <https://www.baidu.com/>. As the world's largest Chinese language Internet search provider, Baidu responds to a huge amount of search queries from more than 100 countries and regions every day, serving as the most important way for netizens to access Chinese language information. With its mission to "make the complicated world simpler through technology", Baidu promotes constant technological innovation and is committed to being a world-leading technology company that understands its users and helps them grow.

Table 1 The overview of the dataset

	Date range	#Working days	#Records
DATA_ONSITE	2019/12/23–2020/02/02	42	1325
DATA_WFH	2020/02/03–2020/03/01	28	1103
Total		70	2428

Table 2 The metrics of each record of developers' daily activities provided by Baidu

Feature	Description
date_partition	The date of a record.
username_e	The encrypted user name of a developer.
repo_name_e	The encrypted repository name.
commit_count	The number of commits submitted by a developer.
line_inserted	The number of lines of code inserted by a developer.
line_deleted	The number of lines of code deleted by a developer.
review_count	The number of code reviews performed by a developer.
job_status_build	The status of build performed by a developer.
build_count	The number of builds performed by a developer; A build refers to the process of continuous integration, including compilation, test, and deploy.
job_status_release	The status of release performed by a developer.
release_count	The number of releases performed by a developer.
compile_count	The number of compilations performed by a developer.

The dataset we get from Baidu contains 107 developers' daily activities from eight projects in 70 working days. Table 1 presents the overview of the dataset. As shown in this table, the time of WFH is from 2020/02/03 to 2020/03/01 because Baidu asked all its employees to work from home after the outbreak of the COVID-19 pandemic in China⁵). We refer to this part of records as to DATA_WFH. On the other hand, we also got a list of records during which developers work onsite, i.e., from 2019/12/23 to 2020/02/02 (referred to as DATA_ONSITE). The development activities in the dataset DATA_WFH and DATA_ONSITE are similar since they happened very closely. This dataset has a total of 2428 records. Among these records, there are 1325 and 1103 records that belong to DATA_ONSITE and DATA_WFH, respectively.

Each record in the dataset has several metrics that represent the activities of a developer in one day. Table 2 presents the fields of a record. Each record has a date (date_partition) on which a developer's activities are reported. The developers' names and their projects in the record are encrypted into unique IDs due to the security and privacy policy of Baidu so that we can still track records over time. Each record has the following numeric metrics: commit_count, line_inserted, line_deleted, review_count, build_count, release_count, compile_count. Although these numeric metrics are dependent on various factors such as developer experience, programming languages, and styles [11], many previous studies have used similar quantitative metrics such as lines of code to measure developer productivity [12–14]. Therefore, we believe these metrics can potentially indicate developer productivity.

Owing to the security policy of Baidu, the numeric metrics are standardized by the following formula: $z = \frac{X - \mu}{\sigma}$, where z is the standardized value, X is the real value of a metric in a record, μ is the mean of a metric in the dataset, and σ is the standard deviation of a metric in the dataset. Thus, the standardized values do not affect the distribution of a metric and the findings in this study since the findings are based on the comparison between the values of metrics when developers work from home and work onsite. Moreover, the standardized values can be positive or negative. Additionally, there are two other fields, i.e., job_status_build and job_status_release, which represent the status of builds/releases performed by a developer. There are four possible values for the status of builds/releases, including success, failed, canceled, or NULL. Compared to the other three statuses, a successful build or release means that a developer is more productive on that day.

Table 3 presents the information about the eight projects in the dataset. These projects are created in two different years, i.e., 2017 and 2018. There are three types of projects, i.e., APP (application software), SERVER (server software such as web services, API libraries), and SDK (software development kits). Among these projects, four projects are written in C++ while the other four projects are written in

⁵) Wuhan is the first city in China to be lockdown since 2020/01/23, and the whole country started lockdown from 2020/01/27, which is during the Spring Festival (2020/01/25) — the most important holiday in China. After the holiday, Baidu asked its employees to work from home.

Table 3 The project information

	Created year	Project type	Language	#Developer
P1	2017	APP	C++	13
P2	2017	APP	Java	13
P3	2018	SERVER	Java	4
P4	2017	SERVER	Java	25
P5	2018	SERVER	C++	53
P6	2018	SDK	C++	10
P7	2018	SERVER	Java	7
P8	2017	SERVER	C++	14

Java. Additionally, we count the number of developers who have records of development activities in the dataset for each project. We also report that these projects have different numbers of developers, for example, project P5 has the most number of developers (i.e., 53) and P3 has the least number of developers (only 4 developers).

2.2 Research questions

In this section, we present six research questions we address in our study.

RQ1. Are there any significant differences in developer productivity between working from home and working onsite?

Motivation. In this RQ, we want to investigate whether WFH can affect developers' productivity compared to working onsite. Given the dataset from Baidu, we measure the overall productivity of all developers by combining their activities together, then compare the overall productivity when working from home and working onsite.

RQ2. Do different programming languages affect developer productivity when working from home?

Motivation. Previous studies show that programming languages have an important impact on developers' activities, such as programming comprehension [15] and being a long-time contributor of open source projects [16, 17]. WFH might have different impacts on developers using different programming languages. For example, since C++ projects in Baidu are usually larger and require more computing resources than Java projects, developers often need to build and debug these C++ projects on a powerful machine remotely. Meanwhile, for most Java projects, developers can write code and debug it in their own computers at home. Thus, in this RQ, we want to investigate whether developers using different programming languages have different productivities when working from home.

RQ3. Do different project types affect developer productivity when working from home?

Motivation. As shown in Table 3, eight projects in the dataset have three different types, i.e., APP, SERVER, and SDK. The projects with different project types could have different project management methods and styles of schedules, different development, and communication tools, which might have a potential impact on developers' productivity [18, 19]. WFH changes the way of project management and development, which have different impacts on developer productivity for projects with different types. For example, many APP projects develop mobile apps, which usually rely on some specific framework and have predefined code styles and specifications. While the software developed by SERVER projects is usually applied in much more complicated scenarios and depends on many different frameworks and programming languages. Thus, it might be more different for developers when working from home to build, test, and debug a SERVER project than an APP project. Thus, we want to investigate whether different project types have an impact on developer productivity when working from home.

RQ4. Do different project ages affect developer productivity when working from home?

Motivation. Different project ages might affect developers' activities. For example, our previous study found that developers in older projects spend more time on program comprehension activities than those of projects in the beginning stage [15]. WFH might expand such effects caused by project age, for example, for an older project, developers need to read the source code and documents more frequently since such a project usually has more maintenance tasks; but they cannot access these resources and communicate with colleagues easily when working from home, which might lower their productivity. Thus, we want to investigate whether different project phases have an impact on the productivity of developers when working from home in this RQ.

RQ5. Do different project sizes affect developer productivity when working from home?

Motivation. Different project sizes (measured by the number of developers of a project in this study) might have an impact on developers' productivity [20]. Zhou et al. [21] found that size as a factor has been always considered as a confounding effect in different approaches in software engineering. Due to different project sizes, projects might have different ways of project management and software development, which would be affected by WFH differently. For example, it might be more difficult to communicate with each other in a large project when working from home, which might decrease the productivity of the project.

RQ6. Do individual developers have different productivity when working from home?

Motivation. Compared with working onsite, developers might have different productivity when working from home due to some personal factors, e.g., experience, personality, habit, and skills. For example, developers are much easier to be interrupted by some other stuff when working from home. In RQ1 we use the aggregated activity data of developers while in this RQ we use the daily activities of individual developers to investigate their productivity. Identifying developers who have different performances when working from home and the potential reasons behind it can help project leaders manage the projects. Thus, in this RQ we want to investigate whether individual developers have different productivity when working from home.

3 Case study results

In this section, we present the results of these six research questions one by one.

3.1 RQ1. The overall developer productivity when working from home

Methodology. To compare developer productivity when working from home with that when working onsite, we first group the records in the dataset by day, and compute several aggregate values including mean, median, sum, max, and min for each numeric metric, as shown in Table 3. Thus, we can know a more accurate distribution of developers' productivity in terms of each metric. For example, although the mean of the submitted commits (commit_count) in one day indicates the average workload of developers in one day, sometimes a few experienced/core developers usually contribute more commits than junior/periphery developers, the mean of commits might still be very high. For all the metrics except for sum, we calculate these aggregate values of each metric of all developers for each day. For sum, we only consider the developers whose records are both in DATA_WFH and DATA_ONSITE, i.e., the sum of DATA_WFH and DATA_ONSITE includes the same number of developers. For the two non-numeric features job_status_build and job_status_release, we compute the success rate of build/release in one day, that is, the ratio of the number of times that a build/release is successful.

For each aggregation function of a metric, we have two groups, i.e., the days on which developers work from home and work onsite, respectively. The number of data points of a group is equal to the number of days of the corresponding group. Then, we apply the Wilcoxon rank-sum test [22] to investigate whether the difference is statistically significant in terms of one kind of aggregate values of each metric. We also compute Cliff's delta [23]⁶⁾ to quantify the amount of difference between the two groups. Consequently, we can compare developer productivity when working from home with working onsite in terms of different metrics.

Results. Since we only have standardized values for these metrics in the dataset, we use violin plots to show the distributions of these metrics when working from home and working onsite, as shown in Figure 1. Notice that the values of these metrics have been standardized (see Subsection 2.1), their values can be negative. From this figure, these metrics have fewer wide distributions when developers work from home than those when developers work onsite except for the number of lines deleted every day. For example, the range of the mean of commit_count by day when working onsite is approximately from -0.4 to 0.6 , while the range of the mean of build_count by day when WFH is approximately from -0.5 to 1.8 . This indicates that developers might have more stable productivity when working from home than when working onsite.

Table 4 presents p-values and Cliff's deltas (δ) for DATA_WFH compared with DATA_ONSITE in terms of different aggregate values of each metric in one day. There are some cases whose values are “-” in these tables, which is caused by the fact that two compared lists are completely the same. We find

⁶⁾ Cliff defines a delta of less than 0.147, between 0.147 and 0.33, between 0.33 and 0.474, and above 0.474 as negligible, small, medium, and large effect size, respectively.

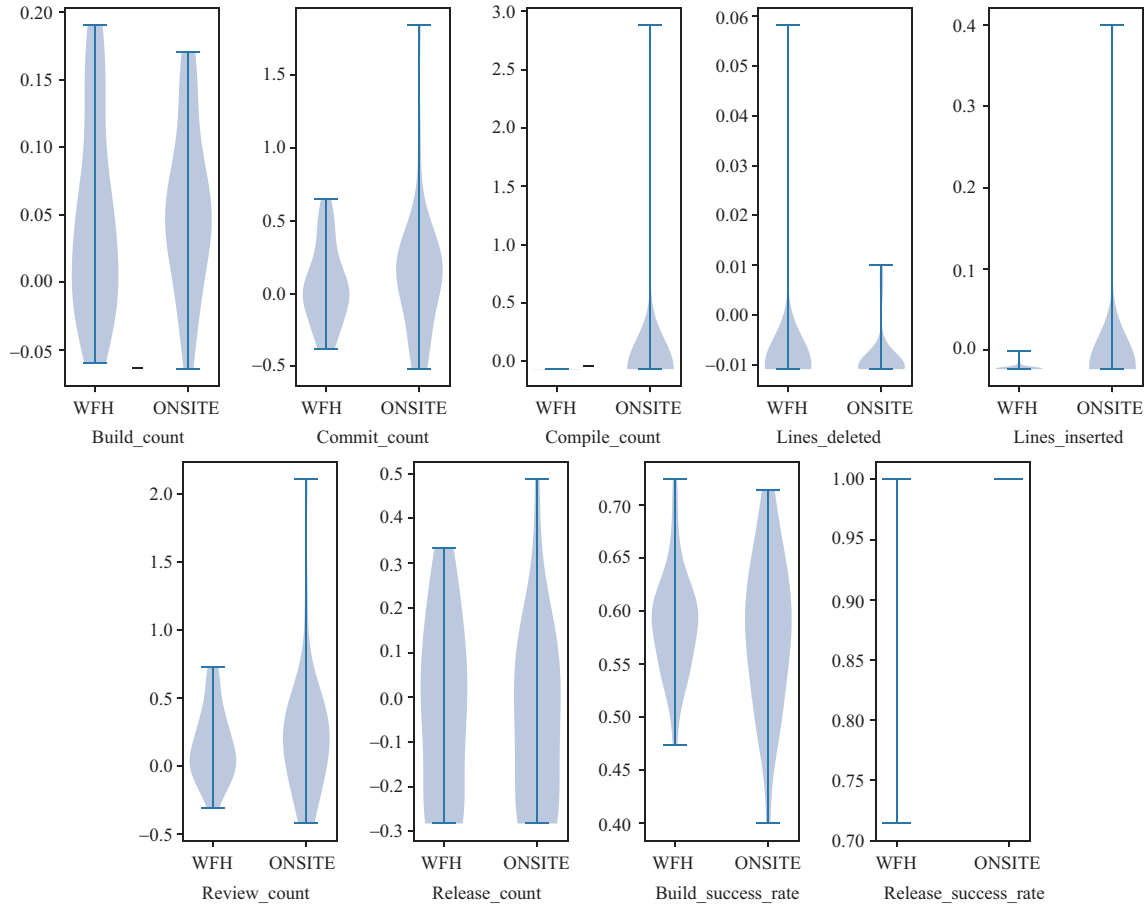


Figure 1 (Color online) Violin plots for DATA_WFH and DATA_ONSITE in terms of the mean of numeric metrics grouped by day and build/release successful rate in one day.

Table 4 p-value and Cliff’s delta (δ) for DATA_WFH compared with DATA_ONSITE in terms of different aggregate values of each metric in one day

Metric	Mean		Median		Sum		Max		Min	
	p-value	δ	p-value	δ	p-value	δ	p-value	δ	p-value	δ
build_count	0.112	0.173	0.001	-0.419 (medium)	0.153	0.146	0.012	0.320 (small)	0.002	-0.262 (small)
commit_count	0.220	0.111	0.034	0.245 (small)	0.227	0.107	2.05E-04	0.500 (large)	-	-
compile_count	0.036	-0.225 (small)	0.216	-0.024	0.027	-0.273 (small)	0.216	-0.024	-	-
lines_deleted	0.005	-0.371 (medium)	0.139	0.152	0.006	-0.359 (medium)	0.378	0.045	-	-
lines_inserted	0.006	-0.357 (medium)	0.175	0.132	0.007	-0.350 (medium)	0.172	0.135	-	-
release_count	0.131	0.160	-	-	0.119	0.168	0.027	0.259 (small)	-	-
review_count	0.144	0.151	0.075	0.182	0.161	0.141	0.015	0.308 (small)	-	-

that there are some cases in which developer productivity when WFH is significantly different from that when working onsite. To ease inspection, we highlight those Cliff’s delta values when the p-values are significant in different levels of the effect size. For example, the sum of compile_count, line_deleted and line_inserted when WFH is significantly less than those when working onsite, and all the effect sizes are not at the negligible level. This might indicate WFH has a negative effect on developer productivity in terms of the number of compilations, deleted lines, and inserted lines every day. We also compute p-values and Cliff’s deltas for the success rate of build/release for DATA_WFH compared with DATA_ONSITE. We find that the success rate of the build when WFH is not significantly different from when working onsite (p – value = 0.207), while the success rate of release when WFH is significantly less from when working onsite (p – value = 0.002 and δ = -0.842). This might be because a project release requires more collaboration and communication that are affected by WFH.

Moreover, we find that WFH has different impacts on developer productivity in terms of different metrics. For example, in terms of build_count, its median values for developers when WFH are less than these when working onsite. On the contrary, the maximum value of build_count for developers when WFH is larger than these when working onsite. This might indicate that a small number of developers

Table 5 The cases in which WFH values for C++ and Java projects are significantly different from ONSITE values

Language	Metric	Positive	Negative
C++	build_count	Sum, max	Min
	commit_count	Mean, max	-
	compile_count	-	Mean, sum
	lines_deleted	-	Sum
	lines_inserted	-	Sum
	release_count	Mean	-
	review_count	Mean, max	-
Java	build_count	-	Mean, median, sum, min
	commit_count	-	Sum
	compile_count	-	-
	lines_deleted	-	-
	lines_inserted	-	-
	release_count	-	-
	review_count	-	Mean

perform more builds when working from home, while most of the developers perform fewer builds. As shown in the results of RQ6 (see Subsection 3.6), this might be because some developers can be more productive when working from home.

Overall, WFH has different impacts on developer productivity in terms of different metrics.

3.2 RQ2. The impact of programming language

Methodology. For both the records in DATA_WFH and DATA_ONSITE, we divide them into two groups: those from the projects using C++ and Java. For each group, we use the same method in the RQ1, then investigate whether the difference in developer productivity between WFH and working onsite is statistically significant in terms of the aggregate values of each metric.

Results. Table 5 presents the cases in which the WFH values are significantly different from those of working on site, and the effect sizes of WFH are not negligible compared with these of working onsite for C++ and Java projects. The column ‘Positive’/‘Negative’ means that WFH has a positive/negative impact on developer productivity in terms of an aggregate value of a metric. “-” means there are no cases in which WFH values are significantly different from those of working onsite for a metric.

We find that for C++ projects, there are both positive and negative cases. For example, the mean of commit_count, release_count, and review_count are belong to positive cases while the sum of compile_count, lines_deleted, and lines_inserted are belong to negative cases. On the other hand, all the cases of Java projects belong to negative cases. This indicates that WFH has more negative impacts on Java projects than C++ projects.

WFH has both positive and negative impacts on developer productivity for C++ projects in terms of different metrics but has a negative impact on developer productivity for Java projects.

3.3 RQ3. The impact of project type

Methodology. We split the records in the dataset into three parts based on the project type, i.e., APP, SERVER, and SDK. For each part, we use the same approach as RQ1 and RQ2 to investigate the difference in developer productivity between WFH and working onsite for projects with different types.

Results. Table 6 presents the cases in which the WFH values are significantly different from those of working onsite, and the effect sizes are not negligible for APP, SERVER, and SDK projects. For APP projects, there are some positive and negative cases, which indicate that WFH might have both positive and negative impacts on developer productivity. For example, in terms of mean, median, sum and max of commit_count and review_count, their values for developers when WFH are significantly larger than these when working onsite; on the contrary, in terms of mean of build_count and compile_count, their values for developers when WFH are significantly less than these when working onsite.

For SDK and SERVER projects, most of the cases are negative, which indicates that WFH has a negative impact on developer productivity. This might be because SDK and SERVER projects usually

Table 6 The cases in which WFH values for APP, SDK, and SERVER projects are significantly different from ONSITE values

Type	Metric	Positive	Negative
APP	build_count	Max	Mean, min
	commit_count	Mean, median, sum, max	-
	compile_count	-	Mean
	lines_deleted	-	Sum
	lines_inserted	-	-
	release_count	Mean	-
	review_count	Mean, median, sum, max	-
SDK	build_count	Mean, max	Min
	commit_count	-	Min
	compile_count	-	-
	lines_deleted	-	Min
	lines_inserted	-	Min
	release_count	-	-
	review_count	-	Min
SERVER	build_count	-	Median, min
	commit_count	-	-
	compile_count	-	Mean, sum
	lines_deleted	-	Sum
	lines_inserted	-	Sum
	release_count	-	Mean
	review_count	-	-

Table 7 The cases in which WFH values for 2017 and 2018 projects are significantly different from ONSITE values

Year	Metric	Positive	Negative
2017	build_count	Max	Median, min
	commit_count	Mean, sum, max	-
	compile_count	-	Mean
	lines_deleted	Median, max	-
	lines_inserted	Median, max	Sum
	release_count	Mean	-
	review_count	Mean, median, sum, max	-
2018	build_count	-	Median, min
	commit_count	-	-
	compile_count	-	Mean, sum
	lines_deleted	-	Mean, sum
	lines_inserted	-	Sum
	release_count	-	-
	review_count	-	Sum

have more components than APP projects, which requires more collaboration and communication with the other developers. According to the feedbacks from Baidu (see Subsection 4.1), WFH has a negative impact on collaboration and communication, which decreases developer productivity.

WFH has both positive and negative impacts on developer productivity for APP projects and decreases developer productivity for SDK and SERVER projects.

3.4 RQ4. The impact of project age

Methodology. As the projects in the dataset are created in two different years, i.e., 2017 and 2018, we split the records into two groups based on the year in which a project is created. Then, we investigate the difference in developer productivity between WFH and working onsite for projects of different ages.

Results. Table 7 presents the cases in which WFH values are significantly different from those of working onsite, and the effect sizes are not negligible for 2017 and 2018 projects. As shown in Table 7, for projects created in 2017, there are more positive cases than negative cases. For example, the mean of commit_count, release_count, and review_count are belong to positive cases while only the mean of

Table 8 The cases in which WFH values for small and large projects are significantly different from ONSITE values

Size	Metric	Positive	Negative
Small	build_count	Max	Median, min
	commit_count	Mean, sum, max	-
	compile_count	-	Mean
	lines_deleted	Median	-
	lines_inserted	Median	-
	release_count	-	-
	review_count	Mean, median, sum, max	-
Large	build_count	Sum	Median, min
	commit_count	-	-
	compile_count	-	Mean, sum
	lines_deleted	-	Sum
	lines_inserted	-	Sum
	release_count	-	Mean, max
	review_count	-	-

compile_count is belong to negative cases. On the other hand, all the cases of projects created in 2018 belong to negative cases. According to the feedback from Baidu, compared to projects created in 2017, projects created in 2018 are less mature and might have more tasks and schedules. It is not easy for a project to complete some kinds of tasks or schedules when working at home, for example, recruiting a new developer usually takes more time when working at home, but a newer project usually needs more new developers than these older projects.

WFH has a positive effect on developer productivity for projects created in 2017 but a negative impact on developer productivity for projects created in 2018.

3.5 RQ5. The impact of project size

Methodology. As shown in Table 3, these eight projects have different numbers of developers. According to the feedback from Baidu, we regard the projects P4 and P5 with more than 20 developers as large projects, and the other 6 projects as small projects. Then, we split the records into two groups based on project size. Then, we investigate the difference in developer productivity between WFH and working onsite for projects with different sizes.

Results. Table 8 presents the cases in which WFH values are significantly different from those of working onsite, and the effect sizes are not negligible for small and large projects. We find that there are more positive cases than negative cases for small projects and most of the cases in large projects are belong to negative cases except for the sum of build_count. The reason might be that it is more difficult for a large project to adjust its structure and schedule after WFH and it is more difficult for large projects to collaborate and communicate with others when working from home.

WFH has a larger impacts on developer productivity for large projects than small projects.

3.6 RQ6. The productivity of individual developers when working from home

Methodology. For each developer in our dataset, we have two kinds of records, i.e., those of WFH and working onsite. Then, we investigate whether their productivity of WFH is significantly different from those of working onsite in terms of each metric. We also compute Cliff’s delta [23] to quantify the amount of difference. For the records of working onsite, we only use DATA_ONSITE according to the findings in RQ1.

Results. Table 9 presents the individual developers whose WFH productivity is significantly different from ONSITE productivity in terms of at least one aggregation metric. The second column in this table is the index of a developer in their project. Out of 139 developers in the whole dataset, the productivity of 21 developers when WFH is significantly different from when working onsite. On the other hand, the productivity of the majority of developers (84.8%) in terms of all the metrics is not significantly different from when working onsite.

Table 10 presents the number of developers for who WFH has a positive or negative impact on their productivity in terms of a metric. As shown in this table, only in terms of compile_count, there is no developer whose productivity when WFH is significantly larger than that when working onsite. We also

Table 9 Developer productivity of WFH is significantly different from that of working onsite in terms of different features

Project	Developer	build_count	commit_count	compile_count	lines_deleted	lines_inserted	release_count	review_count
P1	D1	0.567 (large)***	0.634 (large)***	X	0.324 (small)**	0.295 (small)*	X	0.452 (medium)***
P1	D2	X	-0.316 (small)*	-0.176 (small)*	X	X	X	X
P1	D3	X	X	X	X	X	X	0.445 (medium)***
P1	D4	-0.395 (medium)**	-0.358 (medium)**	X	X	0.272 (small)*	X	X
P1	D5	X	-0.083 (negligible)*	X	X	X	X	-0.335 (medium)**
P1	D8	X	X	X	-0.196 (small)*	-0.197 (small)*	0.148 (small)***	X
P2	D1	X	-0.667 (large)*	X	X	-0.667 (large)*	X	0.600 (large)*
P5	D5	-0.427 (medium)**	-0.296 (small)*	X	-0.290 (small)*	X	X	X
P5	D6	X	X	X	-0.293 (small)*	X	X	X
P5	D7	0.585 (large)**	0.530 (large)**	X	0.367 (medium)*	0.537 (large)**	X	X
P5	D11	X	-0.368 (medium)*	X	-0.405 (medium)**	-0.402 (medium)**	X	-0.299 (small)*
P5	D13	X	X	X	X	X	X	0.288 (small)**
P5	D14	X	X	X	X	X	X	0.214 (small)*
P5	D17	1.000 (large)**	0.636 (large)*	X	0.591 (large)*	-0.773 (large)**	X	X
P5	D20	X	X	-0.167 (small)*	X	-0.451 (medium)**	X	X
P5	D23	X	X	X	0.280 (small)*	X	X	0.372 (medium)**
P5	D24	0.351 (medium)*	0.385 (medium)**	X	X	X	X	-0.368 (medium)**
P5	D29	X	X	X	X	X	-0.688 (large)*	X
P6	D2	X	X	X	X	X	X	-0.342 (medium)*
P7	D2	X	-0.312 (small)*	X	-0.413 (medium)**	-0.264 (small)*	X	X
P8	D6	X	X	X	X	X	X	0.833 (large)*

* denotes p-value < 0.05; ** denotes p-value < 0.01; *** denotes p-value < 0.001.

Table 10 The number of developers whose WFH productivity is significantly larger or smaller than their ONSITE productivity in terms of each metric

Metric	#Positive	#Negative
build_count	4	2
commit_count	3	7
compile_count	0	2
line_deleted	3	6
line_inserted	3	6
release_count	1	1
review_count	7	4

notice that the productivity of several developers when WFH is significantly larger than when working onsite in terms of all metrics, e.g., D1 of the project P1 and D7 of the project P5. For these developers, the company should encourage them to work from home for more time since remote work can improve their productivity. On the contrary, some developers are less productive when working from home, e.g., D2 of the project P1 and D5 of the project P5. For these developers, remote work is not encouraged since their productivity decreases when working from home.

The productivity of most developers working from home is similar to that of working onsite. For a small portion of developers, WFH has different impacts on their productivity.

4 Discussion

In this section, we first present the feedback from Baidu, then provide implications of our findings. At the end of this section, we discuss some threats to validity.

4.1 Feedback from Baidu

Based on our findings, we perform a simple survey to get some feedback from developers in the studied projects. In the survey, we first collect some demographic information such as the developer role and the main programming language. Then, we ask responders whether they agree that WFH has an impact on productivity in a 5-point Likert scale (strongly disagree, disagree, neutral, agree, strongly agree). Finally, we have an open question to ask them about the factors that might affect developer productivity when working from home. Many of them agreed that WFH can have both positive and negative impacts on developer productivity. Some also agree that there is no difference in productivity when working from home. The followings are some of the feedback we collected.

WFH improves developer productivity.

- It is the first time for some developers to work from home, so they feel very excited and have a lot of energy to do their work.
- Developers can focus on their own work and not be disturbed by colleagues.

- After working from home, the company asked developers to write daily reports instead of weekly reports. Daily reports can help developers recall their daily work and push them to work harder in the second day if their tasks are not completed.

- WFH decreases the cost of transportation and saves a lot of time for developers.
- WFH might increase developers' working time because there is no switch between workplace and home and developers can work very early in the morning or very late in the evening.
- WFH gives developers better work-life balance so that developers can work in better condition.

WFH decreases developer productivity.

- There is much other stuff (e.g., looking after children or pets, cooking by themselves) to interrupt developers' work and take a lot of their time.
- Some developers without self-discipline cannot focus on work when working from home. Unlikely onsite, they might be too relaxed at home since there is no colleague around them.
- Although video conferencing tools or telephone are now very convenient for communication, the efficiency of collaboration still decreases due to WFH.

There is no difference in developer productivity when working from home.

- There are no barriers for many developers to complete their daily tasks (e.g., writing code, building projects, code review) when working from home.
- There is no difference of project schedule between WFH and working onsite since developers can know the schedule using an online project schedule tool.
- Current video conferencing tools are very powerful, for example, they usually support screen sharing. So, there is a very slight difference between meeting at a meeting room of the company and online.

4.2 Implications

WFH has different impacts on overall developer productivity. Many previous studies show that WFH has a positive effect on the productivity of workers [6–10,24,25]. Some other studies also show that WFH might have a negative impact on employee productivity. For example, Kazekami [25] found that long working time of WFH would decrease teleworker productivity. In our study, we use a quantitative analysis method to show that WFH has different impacts on developer productivity. From the feedback of Baidu, the difference in developer productivity might be caused by many reasons. We also find that the productivity of most developers is not significantly different from when working onsite (RQ6), so we think WFH can be considered as a choice of work arrangement for employees because WFH offers many other benefits except productivity, such as saving costs for the company and the flexibility of working time for developers.

A project needs to prepare for WFH according to its own characteristics. In this study, we find that developer productivity might be associated with the characteristics of a project including programming language and project type/age/size. For example, WFH might have a positive effect on developer productivity for small project but does not affect developer productivity for large projects (see RQ2 in Subsection 3.5). Thus, we believe that adopting the WFH policy for a project should be based on its own characteristics, e.g., programming languages and project size. When starting the WFH policy, a project needs to prepare some resources to reduce the risks of decreasing developer productivity. For example, a large project should consider the communication cost of WFH and prepare the relevant tools to facilitate the communication of teams.

Using different strategies of WFH for individual developers. We find that the productivity of most developers in this study does not change when working from home. Still there exist some developers whose productivity of WFH is different from these of working onsite. Thus, some approaches based on some development metrics can be used to identify whether the productivity of a developer increases or decreases when working from home. Once the productivity of a developer decreases, the project manager needs to identify the reasons behind it. If the developer is not suitable to work from home, he should be asked to work in the company. On the other hand, if the productivity of a developer increases, the project team should allow them to continue working from home.

For researchers, to improve individual developers' productivity when working from home, more empirical studies are required to investigate more factors that affect their productivity, e.g., personality, moods, and their working environment at home. Additionally, some machine learning models based on developers' daily activities can be built to predict whether their productivity will change when working from home.

4.3 Threats to validity

Threats to internal validity. First, there might exist errors in our code and experiment setting. We have written a python script to process and analyze the dataset provided by Baidu. We double-check our code, however, there may exist some errors that we do not notice. The second internal validity is that we use some quantitative metrics of software development (e.g., the number of builds and commits) in the dataset to measure developer productivity. These metrics such as lines of code have been used to measure developer productivity [12–14]. Hence, we think these metrics can potentially indicate the productivity of developers and we also use multiple aggregate values (such as mean and median) of these metrics by day to measure the productivity. Third, there might be many other factors (e.g., the workload in different times) that affect developer productivity. It is difficult to exclude all other factors in the study. To minimize this threat, we use the records of developers' activities of WFH and ONSITE in the same year (i.e., 2020) for comparison. Finally, the metrics in the dataset are standardized due to the security policy and privacy of Baidu. But we focus on the difference between the productivity of developers when working from home and working onsite. Thus, we believe that the comparison results using standardized values do not affect the findings in the study.

Threats to external validity relate to the generalizability of our findings. In this study, the dataset we used is from Baidu. The number of projects and developers is limited. Thus, it is unclear whether the same results still hold for other developers from other companies. However, since Baidu is one of the largest IT companies in China, we believe that our findings in this study have typical and common characteristics to some extent. Additionally, we analyze 138 working days of software development activities from 139 developers. These developers are from eight different projects with different characteristics such as programming languages and project types. Another threat to external validity relates to the generalizability of the metrics used to measure the productivity of developers. These metrics used in this study are very general and often used in software development [17, 26]. In the future, to reduce these threats, we plan to investigate more developers from different companies and consider more metrics.

5 Related work

In this section, we discuss related work in the fields of WFH and developer productivity.

5.1 Working from home

There are many studies in the literature that investigate the benefits and drawbacks of WFH [1, 27–31]. According to the literature, the main benefits of WFH for companies include saving costs of buildings and increasing productivity and job satisfaction of employees. For employees, WFH gives them more flexible working time and provides a better work-life balance. WFH can also offer benefits to some special kinds of persons, such as the disabled [2] and transgender developers [32]. The main disadvantages of WFH are the access to technology and the integration of telework with the company's strategy and organizational structure, as well as the teleworkers motivation and control [29]. Felstead and Henseke [30] also reported that telework makes employees it difficult to insulate the world of work from other aspects of life when both worlds collide and overlap.

Several studies have investigated the impact of WFH on productivity [6–10, 24, 25]. Among these studies, many of them reported that WFH has a positive impact on the productivity of teleworkers. For example, Coenen and Kok [24] found that telework has a positive effect on the performance of new product development through enabling knowledge sharing, cross-functional cooperation, and inter-organizational involvement. On the contrary, WFH might have a negative impact on productivity. For instance, Kazekami [25] found that appropriate telework hours increase labor productivity but when telework hours are too long, telework decreases labor productivity. WFH might decrease the efficiency of developer communication, which plays an important role in software development [5].

However, most of the previous studies for WFH are based on a qualitative analysis using surveys or interviews. The participants in these studies are general workers not only developers, and they do not have the experience of WFH for a long continuing time. In this study, we perform a quantitative analysis based on a lot of activity data of developers when working from home during the time of the COVID-19 pandemic. We focus on the impact of WFH on developer productivity and the potential factors affecting developer productivity.

5.2 Developer productivity

A lot of studies use developers' daily activities to investigate their productivity. For example, Perry et al. [33] found that many developers spend a lot of time on the communication with colleagues. Additionally, many studies reported that developers' work is fragmented and frequently interrupted, which has an important impact on their productivity [34–39]. For example, Sanchez et al. [40] found that work fragmentation is correlated to lower observed productivity and longer activity switches seem to strengthen the effect.

Developer productivity is often measured by software artifacts produced by developers in a certain time, e.g., submitted lines of code (LOC) [13,14], function points [41], completed tasks [42], and time to implement a requirement [43]. Meyer et al. [19] proposed a list of metrics to measure developer productivity and several ways to improve a developer's productivity through tool support. Some studies also investigate the factors affecting developer productivity, e.g., characteristics of workplace (e.g., privacy, noise) [44], programming languages and development tools [45], project switching [46], and developers' mood [47]. Additionally, personal factors might have an impact on productivity, for example, some developers feel more productive when communicating with others but some don't like to be interrupted when working [48].

Due to the outbreak of COVID-19, some researchers also start to investigate the effect of the pandemic on developers' productivity. For example, Ralph et al. [49] conducted a survey and found that the pandemic had a negative effect on developers' productivity.

In this study, we focus on the difference of developer productivity between WFH and working onsite. We measure developer productivity by several metrics based on developers' daily activities, which have been used in previous studies. We also investigate the factors affecting developer productivity when working from home, such as programming language, project type, and project size.

6 Conclusion

In this paper, we investigate the productivity of developers when working from home for a long time due to the COVID-19 pandemic. We use a quantitative analysis based on a dataset of developers' daily activities from Baidu. To compare developer productivity of WFH with that of working onsite, we use several metrics of software development in the dataset, such as the number of builds, commits, and inserted/deleted lines. We find that WFH has different impacts on developer productivity in terms of different metrics. Also, we investigate some factors affecting developer productivity when working from home including programming language and project type/age/size. Additionally, we find that a small number of individual developers have different productivity when working from home. In the future, we plan to extend our study by using more data from more developers and companies. We also want to build machine learning models to predict developer productivity based on developers' daily activities.

Acknowledgements This work was partially supported by National Key Research and Development Program of China (Grant No. 2018YFB1003904), National Natural Science Foundation of China (Grant Nos. U20A20173, 61902344), and Natural Science Foundation of Zhejiang Province (Grant No. LY21F020011).

References

- 1 Shin B, El Sawy O A, Sheng O R L, et al. Telework: existing research and future directions. *J Organ Comput Electron Commer*, 2000, 10: 85–101
- 2 Spark R. Accessibility to work from home for the disabled: the need for a shift in management style. In: *Proceedings of the 14th Web for All Conference on the Future of Accessible Work*, 2017. 1–4
- 3 Hill E J, Ferris M, Mårtinson V. Does it matter where you work? A comparison of how three work venues (traditional office, virtual office, and home office) influence aspects of work and personal/family life. *J Vocat Behav*, 2003, 63: 220–241
- 4 Meyer A N, Barton L E, Murphy G C, et al. The work life of developers: activities, switches and perceived productivity. *IEEE Trans Softw Eng*, 2017, 43: 1178–1193
- 5 Wolf T, Schroter A, Damian D, et al. Predicting build failures using social network analysis on developer communication. In: *Proceedings of the 31st International Conference on Software Engineering*, 2009. 1–11
- 6 Neufeld D J, Fang Y. Predicting telecommuter productivity. In: *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 2004
- 7 Neufeld D J, Fang Y. Individual, social and situational determinants of telecommuter productivity. *Inf Manage*, 2005, 42: 1037–1049
- 8 Baker E, Avery G C, Crawford J. Satisfaction and perceived productivity when professionals work from home. *Res Practice Human Resour Manag*, 2007, 15: 37–62

- 9 Laihonen H, Jääskeläinen A, Lönnqvist A, et al. Measuring the productivity impacts of new ways of working. *J Facil Manage*, 2012, 10: 102–113
- 10 Campbell K M. Flexible work schedules, virtual work programs, and employee productivity. Dissertation for Ph.D. Degree. Minneapolis: Walden University, 2015
- 11 Kamei Y, Shihab E, Adams B, et al. A large-scale empirical study of just-in-time quality assurance. *IEEE Trans Softw Eng*, 2013, 39: 757–773
- 12 Walston C E, Felix C P. A method of programming measurement and estimation. *IBM Syst J*, 1977, 16: 54–73
- 13 Devanbu P, Karstu S, Melo W, et al. Analytical and empirical evaluation of software reuse metrics. In: *Proceedings of the 18th International Conference on Software Engineering*, 1996. 189–199
- 14 Nguyen V, Huang L, Boehm B. An analysis of trends in productivity and cost drivers over years. In: *Proceedings of the 7th International Conference on Predictive Models in Software Engineering*, 2011. 1–10
- 15 Xia X, Bao L F, Lo D, et al. Measuring program comprehension: a large-scale field study with professionals. *IEEE Trans Softw Eng*, 2018, 44: 951–976
- 16 von Krogh G, Spaeth S, Lakhani K R. Community, joining, and specialization in open source software innovation: a case study. *Res Policy*, 2003, 32: 1217–1241
- 17 Bao L F, Xia X, Lo D, et al. A large scale study of long-time contributor prediction for GitHub projects. *IEEE Trans Softw Eng*, 2021, 47: 1277–1298
- 18 Melo C, Cruzes D S, Kon F, et al. Agile team perceptions of productivity factors. In: *Proceedings of AGILE Conference*, 2011. 57–66
- 19 Meyer A N, Fritz T, Murphy G C, et al. Software developers' perceptions of productivity. In: *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, 2014. 19–29
- 20 Paiva E, Barbosa D, Lima R, et al. Factors that influence the productivity of software developers in a developer view. In: *Proceedings of Innovations in Computing Sciences and Software Engineering*, 2010. 99–104
- 21 Zhou Y M, Leung H, Xu B W. Examining the potentially confounding effect of class size on the associations between object-oriented metrics and change-proneness. *IEEE Trans Softw Eng*, 2009, 35: 607–623
- 22 Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bull*, 1945, 1: 80–83
- 23 Cliff N. *Ordinal Methods for Behavioral Data Analysis*. London: Psychology Press, 2014
- 24 Coenen M, Kok R A W. Workplace flexibility and new product development performance: the role of telework and flexible work schedules. *Eur Manage J*, 2014, 32: 564–576
- 25 Kazekami S. Mechanisms to improve labor productivity by performing telework. *TeleCommun Policy*, 2020, 44: 101868
- 26 Yan M, Xia X, Lo D, et al. Characterizing and identifying reverted commits. *Empir Softw Eng*, 2019, 24: 2171–2208
- 27 Nilles J M. *Making Telecommuting Happen: A Guide for Telemanagers and Telecommuters*. New York: van Nostrand Reinhold, 1994
- 28 Kurkland N B, Bailey D E. The advantages and challenges of working here, there anywhere, and anytime. *Organ Dyn*, 1999, 28: 53–68
- 29 Pérez M P, Sánchez A M, de Luis Carnicer M P. Benefits and barriers of telework: perception differences of human resources managers according to company's operations strategy. *Technovation*, 2002, 22: 775–783
- 30 Felstead A, Henseke G. Assessing the growth of remote working and its consequences for effort, well-being and work-life balance. *New Tech Work Employ*, 2017, 32: 195–212
- 31 Aguilera A, Lethiais V, Rallet A, et al. Home-based telework in France: characteristics, barriers and perspectives. *Transpation Res Part A-Policy Practice*, 2016, 92: 1–11
- 32 Ford D, Milewicz R, Serebrenik A. How remote work can foster a more inclusive environment for transgender developers. In: *Proceedings of the 2nd International Workshop on Gender Equality in Software Engineering (GE)*, 2019. 9–12
- 33 Perry D E, Staudenmayer N A, Votta L G. People, organizations, and process improvement. *IEEE Softw*, 1994, 11: 36–45
- 34 Bailey B P, Konstan J A, Carlis J V. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In: *Proceedings of INTERACT*, 2001. 593–601
- 35 Chong J, Siino R. Interruptions on software teams: a comparison of paired and solo programmers. In: *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work*, 2006. 29–38
- 36 Czerwinski M, Horvitz E, Wilhite S. A diary study of task switching and interruptions. In: *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 2004. 175–182
- 37 Horvitz E C M C E. Notification, disruption, and memory: effects of messaging interruptions on memory and performance. In: *Proceedings of Human-Computer Interaction: INTERACT*, 2001
- 38 Parnin C, DeLine R. Evaluating cues for resuming interrupted programming tasks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010. 93–102
- 39 van Solingen R, Berghout E, van Latum F. Interrupts: just a minute never is. *IEEE Softw*, 1998, 15: 97–103
- 40 Sanchez H, Robbes R, Gonzalez V M. An empirical study of work fragmentation in software evolution tasks. In: *Proceedings of the 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 2015. 251–260
- 41 Albrecht A J. Measuring application development productivity. In: *Proceedings of IBM Applications Development Joint SHARE/GUIDE Symposium*, 1979
- 42 Minelli R, Mocchi A, Lanza M. I know what you did last summer-an investigation of how developers spend their time. In: *Proceedings of the 23rd International Conference on Program Comprehension*, 2015. 25–35
- 43 Cataldo M, Herbsleb J D, Carley K M. Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. In: *Proceedings of the 2nd ACM-IEEE International Symposium*

- on Empirical Software Engineering and Measurement, 2008. 2–11
- 44 DeMarco T, Lister T. Programmer performance and the effects of the workplace. In: Proceedings of the 8th International Conference on Software Engineering, 1985. 268–272
- 45 Boehm B W. Improving software productivity. *Computer*, 1987, 20: 43–57
- 46 Vasilescu B, Blincoe K, Xuan Q, et al. The sky is not the limit: multitasking across github projects. In: Proceedings of the 38th International Conference on Software Engineering, 2016. 994–1005
- 47 Khan I A, Brinkman W P, Hierons R M. Do moods affect programmers' debug performance? *Cogn Technol Work*, 2011, 13: 245–258
- 48 Meyer A N, Zimmermann T, Fritz T. Characterizing software developers by perceptions of productivity. In: Proceedings of ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2017. 105–110
- 49 Ralph P, Baltes S, Adisaputri G, et al. Pandemic programming: how COVID-19 affects software developers and how their organizations can help. 2020. ArXiv:2005.01127