Introduction

Follow-up comprehension questions engineers ask

Principal Investigator/Faculty Supervisor: Dr. Joanne M. Atlee, School of Computer Science, University of Waterloo, jmatlee@uwaterloo.ca
Student Investigator: Rafael F. Toledo, School of Computer Science, University of Waterloo, rftoledo@uwaterloo.ca

To help you make an informed decision regarding your participation, this letter will explain what the study is about, the possible risks and benefits, and your rights as a research participant. If you do not understand something in the letter, please ask one of the investigators prior to consenting to the study.

What is the study about?

You are invited to participate in a research study for a PhD thesis that investigates sequences of questions that software engineers ask to comprehend the program behaviour during software maintenance. Your participation will help us gain a better understanding of how engineers build a mental model of the program's behaviour and how they comprehend code through code queries. Program Comprehension questions help engineer in their manual inspection and understanding of an unknown codebase.

Your responsibilities as a participant

What does participation involve?

Participation in the study will consist of assigning frequency and difficulty levels to common questions asked when trying to understand an unknown codebase. The participants are also asked to identify questions that are commonly asked in sequence. The tasks are performed in an online survey that records only participants' answers. This study is meant to help us learn about the mental process of program comprehension; it is not intended to test your individual performance in any way.

The first set of survey questions asks about demographics and professional experience (e.g., age, gender, current job title, familiarity with your current codebase). After answering these, the second part of the survey lists a set of common program-comprehension questions and asks participants for five types of information associated with each question:

1. Indicate how difficult each question is using a 5 Likert scale ranging from "Extremely easy" to "Extremely hard".
2. Indicate how often each question is asked through a $100 prioritization exercise.
3. Indicate how useful each question is through a $100 prioritization exercise
4. Provide potential follow-up questions that you would pose after each of the listed questions.
5. Provide program-comprehension questions that you commonly use but were not included in the survey.

The study will be a single session and can last around 30 minutes. You will complete all steps via a Qualtrics online survey. Qualtrics has implemented technical, administrative, and physical safeguards to protect survey responses from loss, misuse, unauthorized access, disclosure, alteration, or destruction. However, no Internet transmission is ever fully secure or error-free. Please note: We do not collect or use internet protocol (IP) addresses or other information which could link your survey responses to your computer.

Who may participate in the study?

The study will involve software engineering professionals.

Your rights as a participant

Is participation in the study voluntary?

Your participation in this study is voluntary. If you withdraw from the survey (by not answering required questions), your responses will be excluded from our analyses, and your data will be destroyed; you may still choose to enter the draw for a gift card (see details below). If you complete the survey and choose to enter the draw, you will have two entries in the draw. You cannot request that your data be removed from the study after the survey is submitted because the anonymous nature of the survey precludes the researchers from identifying specific responses.

Will I receive anything for participating in the study?

In appreciation of your time, you will have the option to enter a draw to win a $50 CAD Amazon gift card (20% of participants will be awarded a gift card). At the end of the survey, you will be prompted to enter the draw. If you want to enter the draw the survey will direct you to a second survey intended to collect your email. Your email will be stored separately from your survey answers and will not be part of the analyzed data. If you begin answering the survey but do not reach the end of the survey to submit your email address, your participation in the study will end and you will not enter the gift card draw.

What are the possible benefits of the study?

Participation in this study may not provide any personal benefit to you. The study will benefit the advancement of science on software development tools.

What are the risks associated with the study?

No known or anticipated risks.

Will my identity be known?

The research team will not know which data is associated with your participation, since all data will be de-identified for analysis, reporting, and public data sharing.

Will my information be kept confidential?

Study data will be stored on secured servers (e.g., password protected University of Waterloo server for Qualtrics) and on the investigators' personal computers (in an encrypted folder). The data will be uploaded to a secure UW server immediately after the study session. The data will remain on secure servers at the University of Waterloo for a minimum of 7 years.
De-identified data, i.e., demographics and recorded answers will be accessible by the researchers as well as the broader scientific community. Only aggregate demographics data will be shared publicly. Survey responses (including the text answers) will be shared in the format of a spreadsheet that excludes individual respondents' demographics data. The data will be used only for the purposes associated with teaching, scientific presentations, publications, and/or sharing with other researchers, including the potential use for future studies. The dataset without identifiers may be shared publicly, perhaps on an online repository and/or made available to other researchers upon publication. The individual responses to the survey will be shared in the format of a spreadsheet, which will not include individuals' disaggregated demographics information collected by the survey. Your identity will be confidential. Anonymous quotations from the answers you provide during your session may be used in publications or presentations resulting from this study.

There will be no association between your identity and the research data. Your name will not appear in any public dataset or publication resulting from this study. Individual participants will be referred to using generic labels (P1, P2, …) along with relevant demographic information: age, gender, and professional experience. This information will be aggregated when reporting our results.
You will complete the study via an online survey hosted on Qualtrics. The data collected by this survey will be public on a GitHub repository. When information is transmitted or stored on the internet, privacy cannot be guaranteed.

Questions, comments, or concerns

Who is sponsoring/funding this study?

This study is funded/sponsored by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Has the study received ethics clearance?

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #46733). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

Who should I contact if I have questions regarding my participation in the study?

If you have any questions regarding this study or would like additional information to assist you in reaching a decision about participation, please contact Dr. Joanne M. Atlee by email at jmatlee@uwaterloo.ca. You can also contact any member of the research team listed on the first page of the consent form.

By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

Do you agree to participate in this study?

- ○ No
- ○ Yes

Do you agree to allow your responses to be available to other researchers with the understanding that your identity will not be disclosed?

- ○ No
- ○ Yes

Do you agree to allow your responses to be used in teaching, scientific presentations and/or publications?

- ○ No
- ○ Yes

Do you agree to the use of anonymous quotations in any presentation or publication that comes from this research?

- ○ No
- ○ Yes

Demographics

How would you describe yourself?

- ○ Man
- ○ Woman
- ○ Non-binary
- ○ _____ Prefer to self-describe
- ○ Prefer not to say

How old are you?

- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 55-64
- ○ 65 and older
- ○ Prefer not to say

What is your current job title?

How many years of experience do you have developing software?

- ○ 0-1
- ○ 1-2
- ○ 2-5
- ○ 5-10
- ○ 10-15
- ○ 15-20
- ○ 20-30
- ○ 30+
- ○ Prefer not to say

How many months of experience do you have working with your current codebase?

- ○ 0-1
- ○ 1-3
- ○ 3-6
- ○ 6-12
- ○ 12-24
- ○ 24+
- ○ Prefer not to say

How familiar are you with your current codebase?

- ○ Mostly unfamiliar
- ○ Somewhat unfamiliar
- ○ Somewhat familiar
- ○ Mostly familiar
- ○ Prefer not to say

List the percentage of work time (out of 100%) you spend on the following activities

Fixing bugs	0
Writing code	0
Planning	0
Editing code	0
Understanding code	0
Debugging	0
Reviewing code	0
Total	0

Frequency/Difficulty/Usefulness

This section asks about program-comprehension questions that you ask in the course of developing or maintaining an unfamiliar codebase.

How difficult are the following questions to answer?

	Not a program-comprehension question	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult	Prefer not to answe
What statement is responsible for updating this field, variable?	○	○	○	○	○	○	○
What data can we access from this object?	○	○	○	○	○	○	○
What are the arguments to this function?	○	○	○	○	○	○	○
Who broke the build?	○	○	○	○	○	○	○
Where is this variable or data structure being accessed/used?	○	○	○	○	○	○	○
Where is this method defined in the type hierarchy?	○	○	○	○	○	○	○

	Not a program-comprehension question	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult	Prefer not to answe
Where is this method being called?	○	○	○	○	○	○	○
What are the parts of this type?	○	○	○	○	○	○	○
Which types is this type a part of?	○	○	○	○	○	○	○
Does this type have any siblings in the type hierarchy?	○	○	○	○	○	○	○
How can we improve collaboration and sharing between teams?	○	○	○	○	○	○	○
What gets called when this method gets called?	○	○	○	○	○	○	○

	Not a program-	Extremely easy	Somewhat easy	Neither easy	Somewhat difficult	Extremely difficult	Prefer not to

	Not a program-comprehension question	Extremely easy	Somewhat easy	Neither easy nor difficult	Somewhat difficult	Extremely difficult	Prefer not to answe
	comprehension question			nor difficult			answe
How is control getting (from here to) here?	○	○	○	○	○	○	○
How is data put into this variable?	○	○	○	○	○	○	○
What data is being modified in this code?	○	○	○	○	○	○	○
Where does this information/data flow to?	○	○	○	○	○	○	○

In this exercise, suppose you were given $100 that you must distribute over the listed questions, assigning individual weights (i.e., number of $) to each question to represent how **often** you ask that question when trying to understand your codebase.

If you prefer not to answer, assign 100 to this option	0
What statement is responsible for updating this field, variable?	0
What data can we access from this object?	0
What are the arguments to this function?	0
Who broke the build?	0
Where is this variable or data structure being accessed/used?	0
Where is this method defined in the type hierarchy?	0
Where is this method being called?	0
What are the parts of this type?	0
Which types is this type a part of?	0
Does this type have any siblings in the type hierarchy?	0
How can we improve collaboration and sharing between teams?	0

What gets called when this method gets called?	0
How is control getting (from here to) here?	0
How is data put into this variable?	0
What data is being modified in this code?	0
Where does this information/data flow to?	0
Total	0

In this exercise, suppose you were given $100 that you must distribute over the listed questions assigning individual weights (i.e., number of $) to each question to represent the **usefulness** of that question in understanding your codebase.

If you prefer not to answer, assign 100 to this option	0
What statement is responsible for updating this field, variable?	0
What data can we access from this object?	0
What are the arguments to this function?	0
Who broke the build?	0
Where is this variable or data structure being accessed/used?	0
Where is this method defined in the type hierarchy?	0
Where is this method being called?	0
What are the parts of this type?	0
Which types is this type a part of?	0
Does this type have any siblings in the type hierarchy?	0
How can we improve collaboration and sharing between teams?	0
What gets called when this method gets called?	0
How is control getting (from here to) here?	0
How is data put into this variable?	0
What data is being modified in this code?	0
Where does this information/data flow to?	0

Total 0

Follow-up questions

After answering each of the following program-comprehension questions, what question would you be most likely to ask next?

What statement is responsible for updating this field, variable?

What data can we access from this object?

What are the arguments to this function?

Who broke the build?

Where is this variable or data structure being accessed/used?

Where is this method defined in the type hierarchy?

Where is this method being called?

What are the parts of this type?

Which types is this type a part of?

Does this type have any siblings in the type hierarchy?	
How can we improve collaboration and sharing between teams?	
What gets called when this method gets called?	
How is control getting (from here to) here?	
How is data put into this variable?	
What data is being modified in this code?	
Where does this information/data flow to?	

Open-ended question

Is there any frequent or useful program-comprehension question that you sometimes ask that was not listed in this survey?

Remuneration draw

Do you want to enter the $50 CAD Amazon gift card draw?

- ○ No
- ○ Yes