USAGE OF KNOWLEDGE SHARING COMMUNITIES FOR LEARNING

Stack Overflow Q&A community as a platform for professional development in programming

Saleh Moniri

Abstract

Purpose: The purpose of the study is to investigate the usage of the Stack Overflow as an informal learning method for expertise development.

Theory: Socially Situated Learning and Communities of Practice by Jane Lave and Etienne Wenger.

Method: The data used for analysis in this study is the Stack Overflow website and the same was studied from the Developer Survey provided by Stack Overflow.

Results: The results indicated overall satisfaction by participants as a result of their respective participation in the community. Besides, the analysis reported that skills and experience are the two most important concepts in terms of expertise development. Consequently, it was indicated that the more experienced users are, the more active they become in delivering knowledge and supporting newcomers in the community. However, the majority of the participants were found to be the receivers of knowledge. The result stressed out that the majority of participants would recommend the Stack Overflow as a Q&A community for software developers to others who are interested in developing their programming skills.

Foreword

This thesis was indeed one of the most challenging, yet rewarding and amazing learning experiences for me, which would have not been possible without my supervisor's continuous support and guidance. I would like to take this opportunity to express my deepest appreciation to Mr Markus Nivala, faculty member of Department of Education, Communication and Learning at the University of Gothenburg and thank him for being such a wonderful mentor. Also, I would like to thank Ms Mahshid Baghestani, faculty member of Business and Management at the University of Wollongong for her support and help in my thesis writing.

Table of content

1.0 Introduction

Nowadays, online question and answer communities with a significant number of participants are emerging as a learning platform. In such societies, users are allowed to discuss a particular topic of interest in order to share knowledge and support each other. Online communities provide an excellent opportunity for learning and skill development of participants across the globe. However, not all the communities can be categorized as a learning community and also not all the learning communities use similar methods and structures to support their participants. For instance, there are many online communities such as Quora, YahooAnswer, Reddit, which allow the users to participate in open discussion topics or ask general questions from different categories. On the other hand, there are several communities that are identified as more specialized community with specific types of participants and area of topics. GitHub, Google developers, developer Facebook, Freecodecamp, and Stack Overflow are examples of such communities that are popular in the field of computer programming and software development. Knowledge sharing is one of the key concepts to begin success in such communities where the main concern of participation is to discuss the software programming issues. According to Sarka and Ipsen (2017), software development, as a collaborative activity, requires communication and knowledge sharing to meet the needs of participants.

Established in 2008, Stack Overflow (SO) can be considered as an example of such online forums that has now become one of the most popular online communities for the individuals who both code and who would like to learn and improve their programming skills in an informal educational environment. According to the report of Stack Overflow, Stack Overflow database currently holds more than 14 million questions and over 19 million answers out of which, over 7.5 million developers found solutions to their respective issues from the available answers in the community (Stack Overflow, 2019).

Participating in the knowledge sharing communities for skill developments, leads the study to the theory of socially situated learning by Lave and Wenger (1991). In this theory, Lave and Wenger described the learning as involvement of an individual in the process of social activity rather than being a passive knowledge receiver. As the goal of this study is to investigate the online community (Stack Overflow Q&A community), the extension of the same theory, named as " Communities of Practice" by Wenger (1998) was also used to further elaborate and address the research question. In comparison to other similar learning communities, Stack Overflow is popular for being unique in types of questions and answers that are posted by participants. Stack Overflow strictly recommends the users to avoid posting the open ended topics or the questions that lead to discussion. Instead its emphasize is more on the types of questions which can possibly have a unique solution Accordingly, the older participants of the community, will not only not respond to the general and open discussion questions, but also they vote down the post to be listed and removed later on. This feature of the Stack Overflow is one of the reasons that makes the contents of the website more valuable from the learning perspectives when the participants of the community could find more unique and detailed information about their area of concern.

Over the last decade, many studies have done by researchers on online communities and social activities from learning perspectives. Similarly, in the last few years, the popularity of the Stack Overflow has brought attention to researchers to study and investigate this community from different aspects. Nores et al. (2019), Hwang et al., (2015), Borg et al. (2017), Wang et al. (2018), etc., have commonly named the Stack Overflow community in their researches.

Given such accountable volume of information on the website and the structure of the posts on Stack Overflow, it was important to study the community through different theoretical framework lenses to investigate the influence of participation in such learning community on skill and expertise enhancement. Therefore, in parallel to previous researches, in this paper, the profile of participants, the learning preferences of the users and participation in the community will be studied.

1.1 Aim of the study

The study aims to discover the influence of online communities on expertise development. To address this goal, Stack Overflow as one of the most popular communities in the computer programming and software development society, has been selected as a context of this study. Accordingly, the users' characteristics, the participators' learning preferences, and the usage of the Stack Overflow community by their participants were the main considerations of this research. In order to address these purposes, different categories of the Stack Overflow Developer surveys such as users' profiles including educational backgrounds and professional status, the preferred sources of informal learning, users' activities and participation, and their overall satisfaction rate of the community have been explored to study the influence of participating in such online communities on expertise development.

The study's structure is formed, considering below three questions:

1. What are the characteristics of the Stack Overflow users?
2. What are the participants' preferred informal learning methods?
3. How do the users participate in the Stack Overflow community?

2.0 Literature Review

This chapter provides an overview of previous studies about online communities and how such communities are used by their respective participants as an informal method of learning. In the following sections, brief information about different types of online communities based on previous research studies is firstly given. The section continues by looking at more precise studies on the Stack Overflow and its participants as well as the role of this community in expertise development.

2.1 Online communities

Early 1980, the ideas of discussion forums were conceptualized with the main objectives of creating a platform for informal discussions and communication perspectives (Rekha and Venkatapathy, 2014). By developing internet technology, online communities play an important role in satisfying individuals in both personal and professional needs via their social interactions and activities in various numbers of online communities. According to Ziegler et al. (2014), research on online communities began in mid-1990 with the focus on investigating and describing the matter of what happened in online societies.

Proter (2014), indicated two types of online communities, namely organization sponsored and peer initiated. Organizational communities with the identified users such as universities, continuing education programmes as well as formally organized groups are mostly goal oriented communities which set the predefined specific types of conversation. Accordingly porter (2014) defined peer initiated communities as online support groups with not officially recognized users or anonymous members where, unlike the organization sponsored communities, the participants themselves determined the contents of their discussion and activities in the community. In such online communities users are producers of the contents. Guan et al. (2017) divided the users into two main categories of questioners and answerers. Questioners seek knowledge while answerers are knowledge contributors of the community by providing the solutions to these questions. Gao and Chen (2010) described the comparison between two types of communities from learning perspective and based on curricular. They investigated and compared six differences of limited communities and informal open learning communities. They studied the role of instructors, content providers, formation of participants, common activities, the history of communities, participation evaluation and finally the target of participation for both types of communities' characteristics.

The usage of online communities for professional supports and inspirations are progressively increasing. Participants of such communities develop their skills and knowledge by interacting and sharing information with other professional participants of societies. Duncan-Howell (2009), studied how teachers made connections using online communities as a source of professional learning. She indicated that online communities provide a valuable form of professional learning source for teachers. She also reported that teachers pointed out various numbers of advantages such as access to wider experience and help from peers outside their work as well as the advantages of receiving more comments by several respondents on different subject matters and classroom situations.

In the Internet society, there are several types of online communities with different characteristics. Quora.com, Askville.com, Yahoo Answer, Reddit.com are examples of popular Q&A communities that allow their participants to contribute in general discussion topics. More specialized and unique

communities such as Stack Overflow, GitHub.com, CodeRanch.com, Google Developer, Freecodecamp.org are commonly interested societies for computer programming in which their participants discuss programming issues. Asking the question and receiving the answers is a common thing in all of those communities, but each society has its own structure and strategy to interact with their participants. For instance, Quora expand the domain of discussion through social media.

According to Ovadia (2011), the social element of Quora allows the non-Quora users to reach and participant in the discussion. The questioners in Quora can invite more people to answer the question via tools like Facebook and Twitter. Besides, the feature of following and followers of Quora allows users to track the posts by following the topics, questions, or the users. Patil and Lee (2016) had studied detecting the expert users in Quora based on their participation and the quality of their posts in the community. Part of their report indicated how non-experts participants use the social features of Quora to follow the experts in order to access the valuable posts and gain knowledge.

Since the context of this paper is Stack Overflow, in the next section, there will be a more detailed overview of previous studies about more specialized communities such as Stack Overflow.

2.2 Previous studies on Stack Overflow

Over the last few years, different studies by Schenk and Lungu (2013), Hwang et al. (2015), Borg et al. (2017), Wang et al. (2018), and etc., have commonly named the Stack Overflow community as one of the most popular forums in the domain of computer programming and software development. Established in 2008, Stack Overflow continually has expanded in terms of number of participants and amount of information that has been discussed in the community up till today. Stack Overflow allows the participants to cross the geographic boundaries in order to engage in knowledge sharing and skills development. Following sections provide a detailed discussion over associated literature about the users' characteristics of the Stack Overflow and the usage of this community from learning and educational perspectives.

2.2.1 Participants of Stack Overflow

Based on Geo-location Yahoo API, Schenk and Lungu (2013) discovered that the participants of Stack Overflow were from 189 countries with high level of involvement in discourses from North America and Europe. However, by relying on the Answer Score, the participants from Asia were more active in answering the questions, hence seen as great importers of information and knowledge to the community. Furthermore, the non-boundary capability of the Stack Overflow is providing unlimited access to its users to reach a more comprehensive source of knowledge and information that is shared by developers with different skills worldwide. On the same note, Hwang et al. (2015) have an interesting result in their research about knowledge sharing in online communities. They studied Stack Overflow's user participation in the discourse based on two factors of geographical zone and experience level of the participants. Additionally, they argued how the online communities like Stack Overflow influenced the knowledge sharing in the workplace and how an individual employee could have an access to the solutions that might not be otherwise found locally, hence to benefit from the information beyond the geographical boundaries. They also found that the participants who were new in the community were less active in knowledge sharing and instead played the observer roles who preferred to find their solutions in the existing posts. Furthermore, the newcomers' participation was

limited to a similar geographic location or level of expertise. However, the participants with higher reputation as well as the users who had been involved for a longer time in the community preferred to extend their contribution based on expertise and knowledge level and with less focus on the geo-location (Hwang et al., 2015).

Since the Stack Overflow is aimed at developing and improving the skills through knowledge sharing, the roles of experienced and elderly users become essential in the community. Accordingly, researchers had studied the link between the Stack Overflow reputation scores with the age and experience of the users. Kowalik and Nielek (2016) had realized that experienced users were more actively participating in answering the questions rather than posting new issues. This implies that the more experienced the users were, the higher the level of their willingness to share their knowledge and be supportive of others. Moreover, to examine the influence of experience on understanding the code content and the required time for detecting the errors, Nivala et al. (2016) used the eye-tracking debugging task to discover the visual process of developers. The result of the study reported that the experienced programmers spent less time understanding the codes, and they performed better in detecting the errors.

Furthermore, Kowalik and Nielek (2016) found out that senior users had earned more reputation scores, but the junior participants had a higher percentage in terms of accepting their answers by questioners, yet the difference was not huge. Moreover, the reason could be the availability level and the responding speed or even the strategy of answering the questions deployed by juniors. Additionally, it is interesting to know that gender played a vital role in answering the questions. Ahmed and Srivastava (2017) analysis on Stack Overflow showed that users preferred to respond more to the posts from the same gender group that was applicable to both females and males.

In addition, Marrison and Murphy (2013) demonstrated the effect of age on gaining programming skills and the expansion of knowledge in technologies. Similarly, they had discussed the relationship between age and ability to learn new technology of Stack Overflow members. They pointed out the link between the age and reputation scores of the users, and also with the help of 'Tag' feature in the Stack Overflow, they traced the participants based on the number of tags that they had been using in their posts. The result of the analysis showed a reduction of the tags number in the age of '30s, increasing in the ages of '40s to '50s and dispersion in the '60s and above. By investigating the tags that were representing the latest technologies and also the usage of those tags that have been followed by elderly users, they emphasized that older users had this ability to obtain knowledge and even this group of users were more willing to learn new technologies.

2.2.2 Usage of Stack Overflow for learning perspectives

Although several studies have investigated the Stack Overflow from different perspectives, not much research has been done on this community from a learning perspective and expertise development in programming through knowledge sharing. In fact, in some of those limited studies, the researchers emphasized that participation in the Stack Overflow had been an active important part in various aspects of expertise development. Smrithi and Venkatapathy (2014) indicated that contributors used the forum to obtain a more specific answer to the coding issues rather than opening the discussion topics in computer programming. Besides, Matei et al. (2018) indicated that the quality of contents in online knowledge showed that Stack Overflow had reached the stable stage in terms of quality, even though there is a restriction policy that limited the permission for editing the posts by other

participants. Stack Overflow is a competitive forum that provides just-in-time knowledge in the society that requires individual efforts in order to keep the quality of the learning. While, evaluating the behaviours of the technical users in Stack Overflow, Ahmad et al. (2017) pointed out that plagiarisms were found in the posts that might have had happened due to duplications in posts as well as individuals' efforts in answering the questions in the shortest possible time for the reasons of obtaining the reputation scores without any efforts.

Borg et al. (2017) studied the usage of active learning and self-training on Stack Overflow and more specifically on text mining program learning. They emphasized that self-education could successfully be combined with active learning. Based on their experience from the study and to retain the knowledge quality, they recommended continuous annotation involvements in parallel with the questioners. Likewise, they suggested designing the setting in a way that allowed overlapping to discover diversities in early stages. Stack Overflow provides two mechanisms for increasing the quality of the posts. Through encouraging their participants to be involved in the *Revision* process and *Reviewing* the posts, Stack Overflow assures the quality of the questions and the answers in the community. However, only the participants with more than 2000 reputation scores are allowed to review posts, while all the users are able to do the revision process. On the other hand, Wang et al. (2018) found out that in few cases, neither the revision's answers were understandable nor the contents were clear. Even in some cases there was a tendency for the revisions to be incorrect. Stack Overflow motivates their participants to revise the posts by awarding the 'Badge' points to their profiles, which means, whoever has more revisions will earn more badges. However, according to Wang et al. (2018), the current badging system that is designed to assure the quality of the posts is however failing. They found out that many participants did the revision just for the reason of increasing the badge points, and this quantitative consideration had an opposite effect on the revision system.

Moreover, Stack Overflow has been examined as a replacement of old standard learning requirements toward Practice Learning or Learning by Doing at the university level (Nores et al, 2019). The results showed that students who participated in the replacement course had used the technical terminology more properly in comparison with the students who were studied in the traditional learning system. The same study showed that the students' overall satisfaction level of experiencing, could place Stack Overflow at the core of one of the programming courses. Additionally, students had a better understanding of the programming structure and accessing to more precise and real code examples.

In parallel with the previous studies, this paper will explore the engagement of professional participants in the informal learning community in order to gain knowledge and support other participants. This continues with studying the preferences of using the Stack Overflow Q&A community as an informal learning source of self-study and discovering the users' satisfaction of participation in the community.

3.0 Theoretical framework

This section presents the theoretical framework that has been used for this research. Considering the aim of the study, which is investigating the notion of learning and expertise development through an online community, the theory of Socially Situated Learning by Jean Lave and Etienne Wenger is exerted to support the arguments in this paper. Lave and Wenger defined the concept of learning as the level of involvement in social activities and the role of the learner in the process of learning rather than being just a passive knowledge receiver. According to Lave and Wenger (1991), the incentive of Situated Learning is the intention of exploring how people could learn new knowledge or improve their skills in an informal training process and more precisely by involving and practicing in the communities where the knowledge sharing is based on participants' experience in different contexts and different geo-locations. Wenger (1998), had complemented the Situated Learning theory later as *Communities of Practice Learning, Meaning, and Identity,* which focused on the development of communities and also characterized the social participation as a learning process.

Given the aim of this paper which is about investigating an online community (Stack Overflow) where their users from different geographic location and different level of expertise are responsible for moderating the posts on the website, therefore it is believed that participants have significant roles. In such communities, not only the participants but also visitors can develop their skills by reading the existing posts and get the benefits from the shared knowledge. Having the learner placed in the center of the educational process, the theory of Socially Situated Learning by Lave and Wenger argued that an individual could gain knowledge by being part of the community, hence made the theory applicable for this paper.

3.1 Socially situated learning

In the Legitimate peripheral participation, Lave and Wenger provided a method that allowed participants to speak about activities, identities, artifacts, and communities of practice. The notion of this method argued the process of joining the learners in the communities by involving them in activities and practicing with other participators to develop their skills. Learners increasingly developed their skills proficiency by gaining knowledge from the more experienced participants in the community.

> *"Draw the attention to the point that learners inevitably participate in communities of practitioners and that the mastery of knowledge and skill requires newcomers to move toward full participation in the sociocultural practices of a community". (Lave and Wenger, 1991, p.29)*

The theory of socially situated learning fosters the domain of gaining knowledge through participating and being part of activities in the communities. According to Lave and Wenger (1991), peripheral participation is about being involved in the community. Subsequently, the concept of changing the position and perspectives of the learner participation in the community for developing their identities and shaping the membership types in the social world will be part of learner educational trajectories.

Wenger (1998) proposed a social participation theory of learning that concentrate on the conceptual structure, and it reached to a set of components that recommended a principle for better understanding and learning knowledge. *Meaning, Practice, Community,* and *Identity* were the four components that emphasized the conventions of what the essential requirements of learning are, concerning the nature of knowledge, knowing and knowers. Those components reflected from the Wenger's impressions toward the nature of learning process. He categorized his assumptions in four main concepts that were inspired by the nature of human in terms of gaining knowledge. Wenger described that 1) the human nature is *being social*, which is the central perspective of learning, 2) *knowledge* is a scale of competency, 3) *knowing* is a matter of involvement and engagement in the society, 4) *Meaning* represents the ability to experiment a meaningful notion that is produced through learning.

In the *Figure 1,* the components of a socially situated learning are shown where learning is the primary focus of this structure, and the other elements are intensely interconnected. According to Wenger's argument, due to the flexibility of structure and the correlation between components, each element could be switched and placed in the center as the main focus and the framework would still make sense.

Figure 1 *components of socially situated learning*

- **Meaning:** *a way of talking about our (changing) ability – individually and collectively to experience our life and the world as meaningful.*
- **Practice:** *a way of talking about the shared historical and social resources, frameworks, and perspectives that can sustain mutual engagement in action.*
- **Community:** *a way of talking about the social configurations in which our enterprises are defined as worth pursuing and our participation is recognizable as competence.*
- **Identity:** *a way of talking about how learning changes who we are and creates person histories of becoming in the context of our communities.*

(Wenger, 1998,p.5)

According to Wenger (1998), the concept of learning brought in the picture of the traditional education system into the mind of many people where classrooms, teachers, students, textbooks and homework were parts of the learning process. But conversely and based on Wenger experience, learning is an integral part of everyday lives and a segment of our daily involvement in communities.

By applying the theory of socially situated learning to this paper, the interaction level between the participants in the community and how their involvement level in the community influence their learning and knowledge development as a peripheral of participation will be studied. The worldwide participants of the Stack Overflow who are having different level of experiences are the main elements responsible of moderating the community by their activities to support other participants as well as gaining new knowledge and developing their skills. Besides, the impacts of socially situated learning components on the Stack Overflow participants' activities and level of the users' experience, given their role to be situated in the centre of the learning process and towards full participation in the community when shaping their identity, will be reviewed as part of this study.

3.2 Communities of practice

When it comes to the definition of communities, it involves a variety of social units with different sizes within which the participants of society have a common area of interest for being parts of that community. A community could be as small as members of a family or in the larger scale could be as prominent as of living neighborhood, schools, workplace, clubs and similar society where groups of people cooperate with shared interests. Communities expanded geographically in the virtual life, and also it gathers the participants with different characters, cultures, and knowledge who are participating in various numbers of online communities, beyond boundaries. Brown and Gray (1998) defined the community as a group that should not necessarily be authorized or identified, and also they mentioned that the members of such communities were collaborating for some period of time. Accordingly, they purposed that members of such groups were not necessarily team members or task force peer of real work, but they held common sense purposes wherein the notion of knowledge sharing retained them together.

In the notion of peripheral in the Socially Situated Learning, not all sorts of societies are capable of being presumed as the *Community of Practice*. According to Hildreth & Kimble (2000), the community of practice is recognized as a group where the members would share soft knowledge with organized support. In addition, participants and especially newcomers in the community of practice will learn how to associate in the society and contribute in the community to become an experienced and full member of their community (Aubrey and Riley, 2016).

> *A community of practice is a set of relations among persons, activity, and world over time and in relation with other tangential and overlapping communities of practice (Lave and Wenger,1991, p.98).*

For Wenger also, communities of practice do not include all the societies. Indeed, not all societies that everyone might call that a community has the capability of being used as a community of practice. Wenger (1998) argued that the communities of practice defined itself along three following dimensions:

1. **Mutual Engagement** (as participant work and support each other)
2. **Joint Enterprise** (a mediated and collective understanding of their activities and purpose)
3. **Shared Repertoire:** (Members employ a range of related manners, tools, words, ways of behaving and communicating)
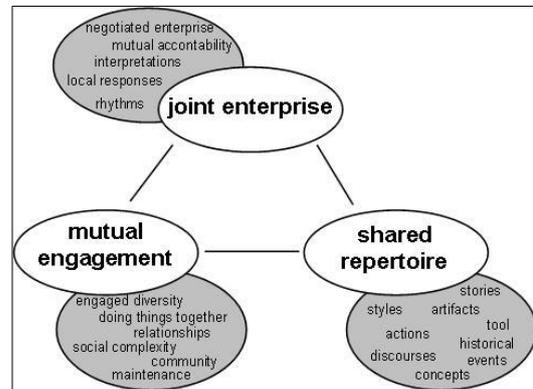
Figure 2 *Dimensions of practice as the property of a community (Wenger, 1998 p.73)*

The capability of the Stack Overflow community to be known as a community of practice will be studied by applying three dimensions of practice. Participants of the Stack Overflow play an important role in supporting each other by sharing knowledge in the community. The computer programming and software engineering matters are the main area of consideration in the Stack Overflow that gathers individuals who are interested in developing coding knowledge. Stack Overflow provides several motivational and communicational tools and features in the website for improving the users' participation and activities in the community. Accordingly, there is a possibility that the three elements of mutual engagement, joint enterprise, and shared repertoire can be applicable to the Stack Overflow community.

In the following sections, the theory of socially situated learning and communities of practice is used as a tool to analyse the data and study the findings, hence making it possible for the research questions to be addressed.

4.0 Methods

This section provides general information about the setting of this research, including the establishment background, features of the website, and the current status of the website. This is followed by a brief description of the target group (users). Finally, the data sources and collection procedure are described at the end of this section.

4.1 Context of the study (Stack Overflow)

Stack Overflow is a dynamic community which is moderated by its users who would like to develop the expertise of programming by sharing their knowledge and supporting other developers. One of the Stack Overflow' strategies is to build a professional and unique library of detailed answers to the programming and software development related topics. To achieve this goal, Stack Overflow encourages their users to have effective and active participation in the community by providing a various number of communicational, motivational and supportive tools.

Since 2008, Stack Overflow has been established, they have been constantly growing, and now they are listed as one of the most thriving Q&A communities in increasing the number of participants and helping developers to solve their coding issues and also to improve their programming skills. Stack Overflow Job is an additional facility of this community that partners with businesses and connect them with developers to help both sides to reach their goal in future development.

To improve the efficiency of the information in the community, Stack Overflow has provided a participation structure in the forum, and they are expecting to be considered by users while they are sharing their posts. For instance, one of the most important matters that it insists on is that the whole idea of this community is toward receiving answers instead of opening discussions. Thus, they highly recommended avoiding the opinion-based questions or the type of questions that could generate discussions rather than a unique answer and accordingly. Therefore, the users are warned that posts which require amendment might be closed until someone fixes that.

Stack Overflow is equipped with several supportive tools as well as functionalities to assist the users to enhance the value of their participation activities, and consequently, such facilities also will help Stack Overflow in collecting more advantageous information (Questions and Answers) in their database.

In the following sections, a brief description of several features of the Stack Overflow website will be firstly discussed followed by a broader observation of the community reputation in online society will be shared secondly.

4.1.1 Stack Overflow features
The purpose of this paper is not studying the technical matters or investigating the functionality of the Stack Overflow features, but to have a better picture of the features of the website for which some tools with a concise description are listed below.

- **Ask Question:** This is the main place that allows the users to share the programming issues with other participants to receive the advice or solution on their coding difficulties. As mentioned earlier, several structures have been designed to increase the productivity, and the accessibility of the posts on the website. For example, the users could refer to the user guidelines on how to post a question properly to receive more efficient response. In the process of publishing a new post, a user should accomplish these five steps: selecting the type of question, tagging, write a proper title, add some descriptions followed by the final step that is reviewing the issue that is going to be posted.

- **Tags:** This function of the website will be used in categorizing and classifying the questions, and it will help the participants to find the relevant topics about their issues from the existing posts. Since a question could be related to the several subjects, then each post could have up to 5 tags, and users are able to browse the topics of their interests by clicking on the tags list.

- **Vote:** Voting is a measuring tool to evaluate the quality of participant post. The more the relevant the answer is, the higher the chances of receiving votes and hence remaining on the top of the forum. On the other side, down vote could increase the risk of closing or deleting the question.

- **Reputation score:** Users obtain reputation by receiving votes on their questions, answers, and comments. These scores unlock new privileges and give the user more access and freedom on the site.

- **Editing or commenting:** To keep the quality of the answers and to improve the efficiency of the provided solutions by other developers, according to the restriction of reputation scores policy, users can revise and review the posts or even edit and repost the existing posts.

- **Accept answer:** The person who posts the question can accept one answer. Accepting the answer does not mean that the answer is correct or this is the best solution for that particular question, but it does mean that the user is satisfied with the marked response and it helps the person who brought up the issue. This method is one of the ways that is used by Stack overflow to trace and report that developers have received help 7.5 million times.

- **Users:** All profiles of the Stack Overflow members is listed in the Users page. Profile details, activities, reputation scores, career background, and programming skills are examples of the information that could be accessible by clicking on each member.
- **Chat rooms:** Is a live discussion room that is accessible for all the members in various area of discussion in the separate rooms. Although for gaining a permission to participate in the conversation, a user should have at least 20 reputation scores, however, the conversation in the room is available to be read by all the members. Participation in Chat rooms has less restriction than Q&A forums and users have more freedom to argue about some issue in this supportive communicational facility of the website.

- **Team:** Is a service that allows a group of people in a form of a team to have unlimited secure and private discussions. This feature can be used for team members of a project in some organizations and could have an educational usage in colleges or classrooms too.

- **Jobs:** Is a connection point for the businesses and the developers. The users have this opportunity to search for companies that are Stack Overflows partners and also they can search for job opportunities in this section. Members could increase the chance of being reached by employers via creating and adding the developer story in their profiles.

4.1.2 Stack Overflow in numbers

Given information in *Tables 1* and *2* are to demonstrate the importance of the Stack Overflow usage by professional programmers and their participation in one of the most popular community for software developer. The information provided in this section has been extracted from the Stack Overflow website and also from the annual reports of the Stack Overflow developer survey.

In the *Table 1,* the information is related to the traffic of the website and how big is the range of the usage and visits. According to Stack Overflow report, by January 2019, they 've had registered more than 10 million user accounts from all over the world out of which the professional developers contributed to the highest percentage followed by the second biggest category of the members who are college students. The statistic shows that each month, more than 50 million people either directly or through the search engine are reaching the Stack Overflow website and the website is loaded over 205 million times. Based on the activities of the visitors, as per the website, approximately over 21 million of these individuals are classified as professional developers and University-level students.

Table 1 *Stack Overflow Website traffic information*

Description	Quantity
Number of registered User as a Stack Overflow member	10 Million user account
Monthly visit of the website	205.2 Million times per month
The unique monthly visit (number of people who visit the website per month)	50.7 Million
Online developer at the time	Average of 51000 developers

According to the above information, it is fair to consider the Stack Overflow as a virtual society with a massive number of expert participants who are developing their programming skills via sharing knowledge and supporting each other. This consideration is even more significant as Stack Overflow announces that an average of 51000 online developers at the same time are participating in the website.

Another important strategy of the Stack Overflow is to build a unique library of answers to the real-life programming issues. In order to implement the strategy, all the Questions and Answers that are posted by participants will be stored in the Stack Overflow Databases. Currently, they have gathered more than 14 million questions that are answered over 19 million times. Those questions are related to

the real coding issues, and the provided solutions are being discussed in different ways and as per the respondents' skills. According to the analysis report of Stack Overflow website and as is shown in the *Table 2*, more than 7.5 million times the developers or more precisely, the person who has posted a question, has classified an answer as a satisfactory (ACCEPTED) solution from Stack Overflow.

Table 2 *Stack Overflow Data Information*

Type of Activity	Quantity
Number posted Questions	14+ Million Questions
Number of the responds (Answers)	19+ Million Answers
Number of the times developers got help	7.5+ Million times

Career development is another aim of Stack overflow that is designed to help businesses and developers find the proper choice for their goal. For this purpose, Stack Overflow has partnered up with more than 17,250 companies worldwide, which would demand a software developer in their organization.

– Companies which partner on Stack Overflow talent: ~ 17250 worldwide

4.2 Users and participants

For having a broader understanding of the Stack Overflow users, the platform has conducted several developer surveys between the years of 2011 to 2019. Based on the assessment of the developer reports, the user group contains members of the Stack Overflow and from a geographical perspective, the participants are from more than 170 different countries worldwide. The statics also shows that the United States of America along with India are the two countries that own the highest percentage of participants. The participants are from both gender groups with the majority of male. Most of the users are full-time employees who are satisfied with their career and current job as a software developer. The absolute majority of participants have the university level background, and mainly they had been studying in the majors of computer science, computer engineering, or software engineering. The studies show that the majority of the professional developers regardless of their experience, they would like to improve their skills by using both formal and informal educational system, alongside technological advancements. Relevantly, the Stack Overflow reports illustrate that a significant number of members are interested in developing their programming skills by registering in official online courses (e.g., a MOOC) or educating themselves by using informal learning tools.

4.3 Data collection

The required data for this paper is extracted from the developer survey data which are freely available in the Stack Overflow website for research purposes (https://insights.stackoverflow.com/survey). To study the latest users feedback and also explore a broader range of information, especially for the categories of users' participations in the community and learning preferences of the participants, all

the available dataset from 2011 till 2018, that were collected by the Stack Overflow have been reviewed. Based on the available categories of the survey questions and also the requirements of this study, it was decided to use the most recent available dataset of the Developer Survey 2017 and the Developer Survey 2018 accordingly. Hence the available online datasets provided by Stack Overflow are qualified and usable data for this research. According to Stack Overflow, the responses that had not met the qualification requirements were excluded from the final samples. For instance, the median response time to complete the entire questions of 2018 survey was 29.4 minutes, consequently, the responses under 5 minutes were excluded from the published version of dataset. Similarly, Stack overflow excluded the responses, which had not completed the questions that were asked to describe the respondents' developer kind.

The surveys are designed in the way that not all the questions appeal to individual respondents. There were many questions that were shown based on previous answers. For example, the questions about salary, size of the company and other job-related questions were only shown to those who had identified themselves as working in a job when responding job status questions. This is one of the reasons that in most of the analysis provided under the result chapter of this paper, the value of Not Applicable (NA) shows high number of percentages. Another reason for such considerable number of NA in the analysis is that most of the questions has options such as: I prefer not to answer or similar ones. Accordingly, in some cases the frequency of NA indicated a big number, and in order to have more meaningful results, it was decided not to drop the NA respondents from the analysis.

To handle the massive amount of data systematically, *IBM SPSS Statistics v.25* was used for data preparation, and analysis process. The selected items in the datasets contained a discrete data and the variables were in two formats of *nominal* and *ordinal* measure types. To report the analysis of such data, it was decided to use the frequency descriptive statistics (also known as descriptive analysis). The whole process of handling the data including the conversion process of the data from string to numeric format, coding the variables, analysis process, tables and graph preparations and more procedures were done with the help of SPSS. In several cases, recoding the existing coded variables was required. For this procedure to be carefully handled, the old codes of required variables was selected manually and then automatically recoded, using SPSS.

To address the research objective, the most relevant questions were selected from the two aforementioned datasets and used in systematic quantitative analysis. In the following sections, the structure of the data and the types of questions which were selected as well as the formulation and categorization of the variables from each dataset, will be described separately.

4.3.1 Developer Survey dataset 2017

In 2017 developer survey, the data was collected from developers from all around the world where they were asked to share their feedback on different categories such as favourite technologies, work preferences, learning and education in skills developments and suchlike information that leads to developing the expertise society. Stack Overflow used those data for different reasons like improvements of the website, educating the employers to have a better understanding of developers and also empowering developers by sharing the information about their respective industry and peers.

As reported by Stack Overflow, a total of 64,227 software developers from 213 countries had participated in the Developer Survey 2017 out of which only 51,392 responses were marked as useful data. Due to missing data, the remaining participants were eliminated from the analysis process. This survey had 99 questions, and those questions were provided to the participants based on the previous answers. However, due to ethical concerns, Stack Overflow did not release all the questions and answers in the public version of the dataset. So, the total numbers of 66 questions with 154 variables as responses have been reviewed for this study. The questions contained several items from different sections such as developer profiles, technology, work and participation in the community.

For this study, all the available 2017 data have been reviewed and three categories were the focus of this review. First, the developer profile was used to understand the users. The second category is the education and learning that allowed investigating the methods and tools developers preferred to use for skills development. Thirdly, the efficiency of the information and the usage of community by the developers were analysed for this paper. As it is shown in *Table 3,* Stack Overflow has categorized the questions in five different categories in the original datasets and allowed researchers to use either the entire data or part of it based on their studies. Since, not all the categories were relevant to the aim of this paper, several categories and items such as questions about advertisements, job seeking, types of hardware and technology and etc., were removed from the original dataset. This review has been done in two rounds. In the first round, 26 questions out of 66 items were selected for the evaluation. Out of these, 14 questions that were more relevant to the aim of this study then were analysed.

Table 3 *Developer Survey 2017 Dataset*

Description	*Original Dataset*	*First round of review*	*Final Review (accepted for analysis)*
Number of Participants	51,392	51,392	51,392
Number of Questions	66	26	14
Number of Variables	154	41	22

Question Categories	• Developer Profile
	  o Geography
	  o Developer Roles
	  o Experience
	  o Demographics
	  o Connection and competition
	  o Life outside work
	• Technology
	• Education
	• Work
	• Community
		• Developer Profile
		  o Geography
		  o Developer Roles
		  o Experience
		  o Demographics
		• Education
		• Work
		  o Company type
		  o Career values
		  o Job Status
		• Community
			• Developer Profile
			  o Geography
			  o Experience
			  o Demographics
			• Education
			  o Education status
			  o Educational attainment
			  o Types of education
			• Community
			  o Engaging on Stack Overflow

4.3.2 Developer Survey dataset 2018

As explained in the previous section (4.3.1), the same procedure used to review the dataset 2017 has been applied for the Dataset 2018. Since 2011, Stack Overflow has been collecting the developers' feedback in survey formats for analytical purposes. These surveys have similar categories, however depending on the studies and requirements, the design and questions were different for each year.

In the Developer survey 2018, approximately 121,600 people from 183 countries had participated, and from those 98,855 responses were qualified for analytical purposes. As shown in *Table 4,* out of 112 questions (*excluding ethical data*), 16 questions from three categories were selected for the analysis process. In parallel to the dataset 2017, the three categories of Developer profile, Education, and Community are the main focus of this study. Types of selected questions for the Developer profile are quite similar to the previous dataset, but the extracted data from Education and especially the Community categories are slightly different. For instance, the collected data from Community has more concentration on the features of the Stack Overflow in comparison to dataset 2017 that is more focused on the efficiency and quality of the information in the Stack Overflow.

Table 4 *Developer Survey 2018 Dataset*

Description	Original Dataset	First round of review	Final Review (accepted for analysis)
Number of Participants	98,855	98,855	98,855
Number of Questions	112	31	16
Number of Variables	129	33	18

Question Categories	• Developer Profile
	  ○ Geography
	  ○ Developer roles
	  ○ Experience
	  ○ Demographics
	  ○ Connection and competition
	  ○ Life outside work
	• Technology
	• Education
	• Work
	• Community
	  ○ Engaging together
		• Developer Profile
		  ○ Geography
		  ○ Developer roles
		  ○ Experience
		  ○ Demographics
		• Education
		• Work
		  ○ Company type
		  ○ Career values
		  ○ Job Status
		• Community
			• Developer Profile
			  ○ Geography
			  ○ Experience
			  ○ Demographics
			• Education
			  ○ Education status
			  ○ Types of education
			  ○ Preferred learning tools
			• Community
			  ○ Engaging on Stack Overflow

4.3.3 Type of questions for each category from the Developer Surveys

As described earlier, to address the aim of this study, the required data for analytical purposes were extracted from three categories. This section provides brief information about each group and also the content of questions that have been studied within each category. Each category was inspired by the purpose of this research and the required data based on the analytical needs of the paper have been selected.

- **Developer profile:** This category contains questions that provide information about demographics, geographic, professions status, knowledge and skills background of the Stack Overflow participants.

    The substance of questions includes:
    - *Gender*
    - *Current location of the participant*
    - *Professional Status*
    - *Coding experience*

- **Education:** The information extracted from the data of this category could be divided into three parts, such as the educational background, current learning status, and the learning method preferences.

    The questions of this category consist:
    - *Highest formal education*
    - *The influence of formal education on the success of developer*
    - *Current study status*
    - *Informal educational preferences*
    - *Types of self study tools*

- **Community:** The collected data from the community section of the developer survey provides the developer feedback on the general usage of the community and the feature of the Stack Overflow website. This information allows the analytical investigation of the usage of the online community in order to understand how developers improve their skills by participating in the community.

    The content of the questions are:

    - *Membership status*
    - *Activities and participation status*
    - *Overall satisfaction and recommending to others*
    - *The quality and efficiency of the information*
    - *Feedback on website features (relevant to this study)*

5.0 Results

IBM SPSS Statistics v.25 and descriptive statistics were used to analyse the data. As explained in section *"4.3 Data Collection"* of this paper, due to types of conducted data (Nominal and Ordinal) it was realistic to use descriptive statistics. Consequently, for the reason of not having a scale or measurable variables, findings of this study are reported in percentage as types of descriptive analysis. Accordingly, the information is presented in forms of text, graphs, and tables. To avoid the duplication in the analysis process, it was decided to use the most recent data *(dataset 2018)* with the size of *n= 98855* respondents for the similar variables in both datasets. Additionally, for the cases of using another dataset *(dataset 2017)* in the analysis process, the sample size is *n=51392* respondents. Following sections contain the findings for the users' characteristics, types of informal learning preferences, followed by the influences of user participation in the stack overflow on skills development.

5.1 User characteristics

Online communities by nature, connect people from different demographical, geographical, knowledge and skills in one place with common interests. With respect to communities with participation restriction rules, there are accountable amount of knowledge sharing and discussion communities in online societies. In such communities every person, regardless of their gender, age or even knowledge level are welcome to register for participation. Stack Overflow is a learning-based online community with a large number of participants whose users' profiles will be described in this section. Based on the datasets 2018 (*n=98855*), the frequencies of

Figure 3 *Gender Frequencies*

the demographic data shows that participants are from both gender groups with the majority of male category for 60% *(Figure 3).* The statistics indicate that participants of the latest dataset are from 184 countries worldwide. Top three countries with the highest number of participants in 2018 are Unites States with the majority of 21% placed on top, followed by India as the second highest contributor of 14%, followed by Germany as third country with the 6.5% of the users. Since one of the elements of expertise measurement is the level of experiences, and as of focus of the study relay on professional users activities, *Figure 4* shows the distribution of participants' coding experiences. The range of years indicated in the chart includes the total number of years that participant has coding experience for both professional and educational period. As shown
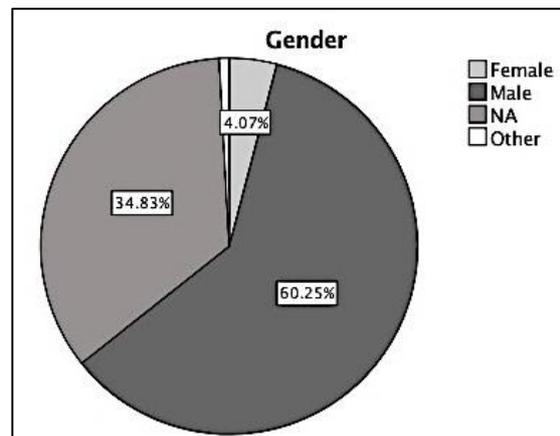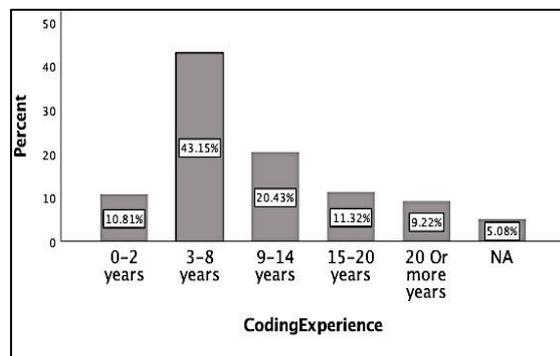
Figure 4 *Years of participants coding experince*

in the graph with the sample size of *n=98855*, the frequency of *42561* participants belong to the group with the coding experience between 3 to 8 years that represents the highest percentage (43%) of the participants. The second group consists of 20% of contributors with higher coding experience between 9 to 14 years. Third, considered as the biggest group of users based on their level of experiences, are 15 to 20 years, followed by junior programmers with 0 to 2 years, and finally the group of participants that identified their coding experiences as 20 years or more.

5.1.1 Users educational background

From the educational background perspective, the study of data with the size of *n=98855* shows that the majority of 44% of participants are holding a bachelor's degree. In addition, by adding the percentage of the higher education categories (Professional degree, Doctoral degree, Master degree, Bachelor degree and Associate degree), provided by the users, it could be seen that a total number of 72% of participants have the university level background *(Table 5)*. According to Stack Overflow report, participants had been studying mostly in the majors of computer science, computer engineering, or software engineering.

Table 5. *The relation between Employment Status and higher level of formal education*

		Employment Status														
		Full-time		Part-time		Freelancer/Self-employed		Not employed/but looking for work		Not employed/Not looking for work		Retired		NA		Total	
		Row N%	Column N%	Row N%	Column N%	Row N%	Column N%	Row N%	Column N%	Row N%	Column N%	Row N%	Column N%	Row N%	Column N%	Row N%	Column N%
Formal Education Level	Doctoral degree (Ph.D, Ed.D., etc.)	84%	2.6%	2.5%	1.0%	8.4%	2.0%	2.5%	1.0%	0.7%	0.4%	1.0%	10.1%	1.4%	0.8%	100%	2.2%
	Master's degree (MA, MS, M.Eng., MBA, etc.)	83%	25.3%	2.9%	11.4%	9.3%	21.4%	2.8%	10.2%	0.8%	4.3%	0.1%	14.1%	0.6%	3.7%	100%	21.6%
	Bachelor's degree (BA, BS, B.Eng., etc.)	81%	50.0%	4.1%	33.4%	7.9%	37.2%	4.8%	36.1%	1.4%	15.1%	0.1%	27.8%	0.9%	10.5%	100%	44.2%
	Professional degree (JD, MD, etc.)	76%	1.6%	4.8%	1.3%	12%	1.9%	3.4%	0.8%	1.0%	0.3%	0.6%	4.0%	1.5%	0.6%	100%	1.5%
	Associate degree	71%	3.0%	7.8%	4.3%	9.7%	3.1%	7.0%	3.6%	2.1%	1.5%	0.3%	4.4%	1.6%	1.4%	100%	3.0%
	Secondary school	36%	4.5%	14%	22.7%	12%	11.9%	15%	22.5%	20%	42.5%	0.2%	7.9%	4.0%	10.2%	100%	9.1%
	Primary/elementary school	22%	0.5%	7.4%	2.3%	12%	2.2%	14%	3.9%	38%	15.2%	0.4%	3.1%	6.6%	3.1%	100%	1.7%
	Some college/university study without earning a degree	62%	10.3%	9.4%	20.4%	13%	16.9%	8.4%	16.9%	4.9%	13.9%	0.3%	16.7%	1.5%	4.8%	100%	11.8%
	I never completed any formal education	51%	0.5%	6.0%	0.8%	18%	1.3%	7.3%	0.9%	11%	1.8%	1.7%	5.3%	5.9%	1.2%	100%	0.7%
	NA	27%	1.6%	3.1%	2.4%	4.6%	2.0%	5.8%	4.2%	4.9%	4.9%	0.4%	6.6%	54%	63.7%	100%	4.2%
	Total	71%	100.0%	5.4%	100.0%	9.4%	100.0%	5.9%	100.0%	4.2%	100.0%	0.2%	100.0%	3.6%	100.0%	100%	100.0%

The statistics show that 71% of the participants are currently not enrolled in a formal or degree-granting program. These results indicate that the majority of participants who are not students are the main contributors in the community. Second group of participants who identified their current formal study status are a group representing 19% of participants who are full-time students and another 6% of users indicated that they are currently registered in formal education as part-time students.

To explore the Influence of formal education as one of the key factors in becoming a successful developer, another important element that was investigated during the analysis process was the information about the participants' opinion on the subsequent impact of formal education on their success. As demonstrated in *Table 6*, the percentage of users who agreed that formal schooling had important effect on their success as a developer is lower in comparison to the other groups who did not agree on the same.

Table 6 *Users' assessment on influence formal education on being a success as a developer*

		Column N %
Importance of formal education	Very important	7.2%
	Important	11.4%
	Somewhat important	12.2%
	Not very important	9.3%
	Not at all important	5.2%
	NA	54.6%

As mentioned in the *Table 5,* 3% out of the group had never completed any formal education as well as 47% of users who had not earned any official university certificates, totally representing the employed people (86%). The given information emphasizes the importance of skills and experience roles in parallel to formal education as key successes in the field of programming and software development. Next section will explain more about the employment status of the participants and the influence of education and experience on the occupation.

5.1.2 Employment status

From the career perspective and position at the workplace, Stack Overflow reported that the majority of the users were identified as software developers (full-stack, back-end, and front-end), and the other two large groups had classified themselves as desktop application developers and mobile application developers who were satisfied with their career and current job as a developer.

To study the relation between the level of formal education and the employment status, *Table 5* presents the influences of each variable on one another and the results are shown in percentages. Additionally, total percentage of the contributors for each group is provided separately. The result in *Table 5* shows that the majority of users are full-time employees and more precisely, total of 86% of the participants are working as full-time, part-time, and freelancer. On the other hand, 6% of the users are not currently employed, though are looking for jobs. Lastly, 4% are neither employed, nor searching for the job positions.

5.2 Preferences of Informal learning

In the survey, developers were asked to select the type of non-degree education they had used or participated in. The variety of options, which were given to the developers, covered a wide area of informal learning in both online and physical environments. Participators were asked to select as many as the following options that they had attended:

- *Participated in a Hackathon*
- *Participated in a full time developer training program or bootcamp*
- *Received on-the-job training*

- *Taking part time in person course*
- *Participating in online coding competitions*
- *Contributed to open source software*
- *Taught yourself new language, framework, or tool without taking a formal courses*
- *Taking an online course in programming (e.g. a MOOC)*

In this paper, the last two options that were more relevant to the research area are studied. First, the group of participants who had selected self-taught education and attempted online courses, were separated from others. The result shows that 60% of participants indicated at least one of those two options as a choice for selection. In the second step, only the group of participants who had selected self-learning education through participating in the online communities were studied. The reason for this process was to identify the percentage of the users who had chosen the Q&A community of Stack Overflow as a self-learning source. To find out the result, first, the participants who had selected the online communities (both selection of Stack Overflow and other communities) were separated from the following given options:

- *The official documentation*
- *Pre-scheduled tutoring or mentoring sessions with a friend or colleague*
- *The technology's online help system*
- *A book or e-book from O'Reilly, Apress, or a similar publisher*
- *Tapping your network of friends, family, and peers versed in the technology*
- *Internal Wikis, chat rooms, or documentation set up by my company for employees*
- *A college/university computer science or software engineering book*
- *Online developer communities other than Stack Overflow (forums, listservs, IRC channels, etc.)*
- *Questions & answers on Stack Overflow*

In the final step, the percentage of the users who had selected the Stack Overflow as one of their selections was calculated. The result indicated that the Stack Overflow community was chosen by 25% of participants as one of their choices of collections.

Following sections provide information about the results of the choices of informal learning *(5.2.1),* followed by the analysis of the preferred sources of self-learning *(5.2.2).*

5.2.1 Types of Informal learning

According to the data model explained in the previous section (5.2), the valid data that represents the two types of informal learning and was collected from the multi-response questions, was divided into five parts. Three groups of participants had a unique selection of choices. First, those who selected only self-taught study as an informal learning method. Second group belonged to the users who selected only online courses, and the third group included the participants who had chosen just the two options of Self-taught and participating in the online courses. In addition to aforementioned three groups of participants, the fourth and fifth groups included the users who had selected at least one of those two options (Self-taught Study or Online Courses) along with other methods described in section 5.2.the result of dataset 2018 analysis with *n=98855* in *Table 7,* demonstrates that overall 60% of the participants had selected one of the two options of self-taught study and participating in online courses as an informal learning method. Breaking down the result, it indicates that the selection of self-taught

learning has the highest preference of choices. As displayed in *Table 7*, the sum of percentage of the users who had selected Self-taught study as one of their choices of selection for non-degree education is 39%.

Table 7 *Preferred types of non-degree education*

	Frequency	Percent	Valid Percent	Cumulative Percent
Only Self-taught Study	6900	7.0	7.0	7.0
Only Online Courses (MOOC)	1245	1.3	1.3	8.2
Only Self-taught & Online Courses (MOOC)	11807	11.9	11.9	20.2
At least Once Self-taught	19972	20.2	20.2	40.4
At laest Once Online Course (MOOC)	19980	20.2	20.2	60.6
NA	30895	31.3	31.3	91.9
Non of those two options	8056	8.1	8.1	100.0
Total	98855	100.0	100.0	

5.2.1.1 Impact of experience on selecting the Self-taught study

For studying the impact of experience on of type of informal education preferences, the data is re-formulated to compare the selection of Self-taught study with other options based on level of coding experiences. As demonstrated in *Table 8,* first column represents the data that includes the percentage of the group who had at least once selected the self-taught study as an item in their choices. In the second column of the table, other selections of participants are stored. Accordingly, each row represents the percentage of selection within the category of coding experience.

Table 8 *The assortment of Self-taught learning on each group of Code Experience*

		Total Self-Taught As Informal							
		Total Once Self-Taught		Not Selected Self-Taught		NA		Total	
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Coding Experience	0-2 years	27.0%	7.5%	28.8%	10.5%	44.2%	15.3%	100.0%	10.8%
	3-8 years	39.4%	43.4%	29.5%	43.0%	31.1%	42.9%	100.0%	43.1%
	9-14 years	46.8%	24.4%	30.8%	21.3%	22.4%	14.7%	100.0%	20.4%
	15-20 years	46.9%	13.6%	34.8%	13.3%	18.3%	6.6%	100.0%	11.3%
	20 Or more years	47.1%	11.1%	37.8%	11.8%	15.2%	4.5%	100.0%	9.2%
	NA	0.5%	0.1%	0.8%	0.1%	98.7%	16.0%	100.0%	5.1%
	Total	39.1%	100.0%	29.6%	100.0%	31.3%	100.0%	100.0%	100.0%

By comparing the results given in the first and second columns for each category of coding experience, it can be seen that, except for the first category (0-2), the preference of self-taught study is higher in the experienced participants in comparison with other types of non-degree educational methods. This difference is more remarkable in the category of 9-14 in which 47% of participants of this group preferred self-education methods, while 30% were willing to use other methods. These results illustrated that the level of experience influenced the preference for selecting the method of learning. The more the experienced the participants are, the higher chances that they would prefer to use the Self-taught learning method in comparison to other informal learning methodology.

5.2.2 *Self-taught learning sources*

As mentioned earlier in section 5.2, the results of studying the two sources of Self-taught learning will be discussed. This section provides information about the participants who had selected self-taught methods of informal learning and accordingly selected the given options (Stack Overflow and other online Communities) of which represent the participating in the online communities as a source of learning. Similar to the previous section *(5.2.1)*, the valid data for the online communities selection is separated from the whole choice of selection and spliced into different items as shown in *Table 9*. As provided below, a total of 27% of participants have selected at least one of the options of online communities as self-learning sources. Furthermore, the statistics show that 31% of participants have never selected participating in online communities as a learning tool.

Table 9 *Preference on the usage of online communities for both choices of Stack Overflow and other online communities*

	Frequency	Percent
Only Stack Overflow	604	.6
Only Online Communities other than Stack Overflow	308	.3
Only Stack Overflow & Only Online Communities other than Stack Overflow	1515	1.5
At least Once Stack Overflow	11666	11.8
At least Once Only Online Communities other than Stack Overflow	2015	2.0
At least Once Stack Overflow & Only Online Communities other than Stack Overflow	10502	10.6
NA	41501	42.0
OtherSelections	30744	31.1
Total	98855	100.0

To calculate the Stack Overflow's participants, the above categories are divided into two categories of Stack Overflow and Other selection. The Stack Overflow category includes all the variables, including Stack Overflow as an option. The second category contains other selections of the choices including the online communities other than Stack Overflow. The results of this analysis are displayed in form of a graph. As seen in *Figure 5,* almost 25% of the participants have selected the Stack Overflow community as one of their selections of informal education sources.
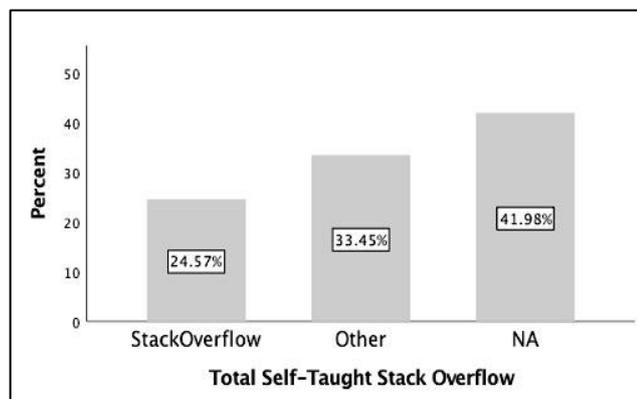

Figure 5 *Indicating the percentage of self-learning sources. Comparison of Stack Overflow with other sources*

5.2.2.1 Influence of experience level on selecting the Stack Overflow

In order to identify the level of experience for the participants who have selected Stack Overflow, *Table 10* is drawn to indicate the detailed information about this matter. According to *Table 10*, the majority of the users who have chosen Stack Overflow, have 9 to 14 years of coding experience. Besides, the other two groups with higher level of experience (15-20 and 20 and more) also have a reasonable percentage of choosing the stack overflow community. On the other hand, the participants with a minimum of 0 to 2 years of coding experiences are the smallest group of this category.

Table 10 *The assortment of Stack Overflow percentages on years of experience*

		StackOverflow		Other		NA		Total	
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
CodingExperience	0–2 years	17.0%	7.5%	20.6%	6.7%	62.4%	16.1%	100.0%	10.8%
	3–8 years	25.8%	45.3%	30.5%	39.4%	43.7%	44.9%	100.0%	43.1%
	9–14 years	29.8%	24.8%	38.7%	23.6%	31.6%	15.4%	100.0%	20.4%
	15–20 years	27.7%	12.8%	46.0%	15.6%	26.2%	7.1%	100.0%	11.3%
	20 Or more years	25.6%	9.6%	53.5%	14.7%	21.0%	4.6%	100.0%	9.2%
	NA	0.2%	0.0%	0.3%	0.0%	99.5%	12.0%	100.0%	5.1%
	Total	24.6%	100.0%	33.5%	100.0%	42.0%	100.0%	100.0%	100.0%

Generally, 25% of the participants who have selected Stack Overflow are spliced into different coding experience categories. As shown in *Figure 6*, each bar has three parts that each of them represents the percentage of users who chose the Stack Overflow comprises the entire group of participants *(n=98855)*. For instance, in the group that represents the second category of experience with 43% of participants, 11% of them use Stack Overflow online community as a self-learning source. Also, 13% of the same group preferred to use other learning sources, and almost 19% did not respond to the question.
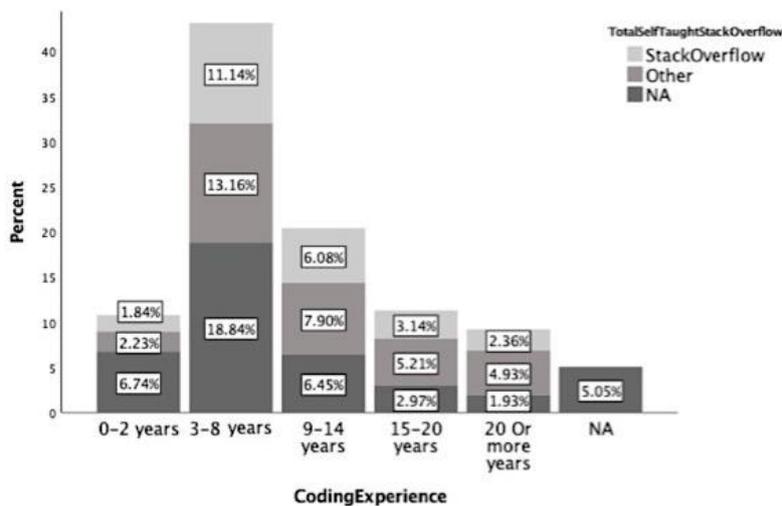

Figure 6 *Total percentage of choosing Stack Overflow on years of coding experience*

The result of this analysis indicated that the preference of using the Stack Overflow community as one of the sources for the self-taught learning are more common for the group of participants with a higher level of experience.

5.3 Stack Overflow participation data overview

This section provides the findings of the study on how users are participating in the Stack Overflow community. It starts by describing users membership status and how frequently they are visiting and participating in the community. The report continues by analysing the data to discover the usage of different types of activities in the community and also investigating the interest level of participating in the hypothetical tools on Stack Overflow. Finally, it provides the result of the analysis on the overall satisfaction of using the Stack Overflow by participants.

5.3.1 Members participation status on the Stack Overflow

To describe the participation status of the users, the data that represents the account registration of the participants has been investigated. Dataset 2018 with the sample size of n=98855 has been studied for this section. The results show that 68% of the participants registered a user account on the website, while 7% of participants are visiting the community as guest users. As Seen in *Table 11*, even the users who have not identified themselves as an account holder of the Stack Overflow had a remarkable percentage of visiting times in the community. Based on the analysed data, 43% of participants consider themselves as the Stack Overflow members. This means, due to types of the activities and participations in the community, not all users who are identified as account holders recognized themselves as an active participant user.

Table 11 *Status of having Stack Overflow account and the percentage of frequent visit of the website.*

| | | Having an account on Stack Overflow | | | | | | | | | |
| | | Yes | | No | | I'm not sure / I can't remember | | NA | | Total | |
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Status of the frequent visit of Stack Overflow	Multiple times per day	91.7%	32.6%	5.9%	21.1%	2.3%	18.8%	0.1%	0.1%	100.0%	24.1%
	Daily or almost daily	89.2%	33.2%	7.5%	27.9%	3.2%	27.2%	0.1%	0.1%	100.0%	25.3%
	A few times per week	84.0%	21.5%	11.2%	28.7%	4.7%	27.9%	0.0%	0.0%	100.0%	17.4%
	A few times per month or weekly	79.7%	10.5%	13.7%	18.0%	6.5%	19.6%	0.1%	0.0%	100.0%	8.9%
	Less than once per month or monthly	73.1%	1.7%	16.0%	3.7%	10.6%	5.7%	0.3%	0.0%	100.0%	1.6%
	I have never visited(before today)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	1.9%	100.0%	0.4%
	NA	1.8%	0.6%	0.2%	0.6%	0.1%	0.7%	97.9%	97.9%	100.0%	22.3%
	Total	67.9%	100.0%	6.8%	100.0%	3.0%	100.0%	22.3%	100.0%	100.0%	100.0%

For having a better picture of user participation, the engagement level provided by users has been studied. In the questionnaire it was clearly described that active participation means being involved in asking, answering, voting for, or commenting on posts. The results indicated that 29% of the users have at least a weekly active involvement in the community. As demonstrated in *Table 12*, if the 26% of monthly participants is added to the previous group, which is recognized as weekly participants, the Stack Overflow contains almost 55% of active participators per month.

To discover the relation between participation and years of coding experience, the two relevant categories of the dataset is studied. The result indicated that the more experienced users are, the more actively they participate in the community. As Seen in *Table 12*, in each category of daily, weekly and monthly participation, as the level of experience increases the percentage of participating in the same group is also increased.

Table 12 *Percentage of Participating in the community on coding experience*

		Coding Experience													
		0-2 years		3-8 years		9-14 years		15-20 years		20 Or more years		NA		Total	
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Participation In Stack Overflow	Multiple times per day	7.3%	1.5%	40%	2.0%	19%	2.0%	14%	2.7%	19%	4.4%	0.0%	0.0%	100%	2.2%
	Daily or almost daily	7.3%	2.7%	43%	3.9%	23%	4.4%	14%	4.8%	13%	5.7%	0.1%	0.1%	100%	4.0%
	A few times per week	7.1%	5.1%	44%	7.8%	23%	8.8%	14%	9.7%	12%	9.9%	0.1%	0.1%	100%	7.8%
	A few times per month or weekly	7.1%	9.9%	43%	15.0%	24%	17.7%	14%	19.1%	11%	18.7%	0.1%	0.2%	100%	15.0%
	Less than once per month or monthly	7.1%	17.1%	44%	26.3%	25%	32.2%	14%	31.5%	10%	29.4%	0.1%	0.4%	100%	26.1%
	I have never participated in Q&A on Stack Overflow	18%	19.0%	48%	12.9%	17%	9.7%	9.2%	9.3%	7.4%	9.3%	0.1%	0.2%	100%	11.5%
	NA	14%	44.8%	41%	32.1%	15%	25.0%	7.7%	22.9%	6.2%	22.5%	15%	99.1%	100%	33.5%

5.3.2 Types of activity in the community

The type of activities by participants in the community is studied for this section. In the survey, the participants were asked to indicate how often they had done each of these activities such as: find the answers, ask a question, write answers to others question, and copy the code examples into their codebases. The conducted data have been studied from datasets 2017 with the sample size of *n=51392* for this section and the results are presented as below.

Analysis of the dataset 2017, presented in *Table 13*, indicated the percentage overview of Write Answer which shows low-toned participation especially on daily and weekly bases activities. 38% of participants indicated that they had not written any answers at all. Similarly, the data conducted from ask question activity shows the low percentage of participation. Higher percentage (41%) of users specified that they had never asked questions in the Stack Overflow community.

Moreover, the percentage on the result of activity that includes the users who copied the codes and pasted it into their codebase is slightly growing in this category compared to the other two pervious categories. Regardless of the ethical matters of copyright, which is beyond the study of this paper, investigating the data for this category emphasized the availability of useful information in the community. In fact, the result for the category that represents a finding of an answer in the community highlighted the percentage increment for users feedbacks. Besides, in the same category the result shows that only 1.2 % of participants did not find their answers in the community.

Table 13 *Indicated percentage of participation on each type of activities by users*

		Percent %				Percent %
	At least once each day	.8			At least once each day	.3
	At least once each week	2.6			At least once each week	.8
	Several times	9.1		Ask	Several times	7.4
Write Answer	Once or twice	15.0		Question	Once or twice	16.2
	Haven't done at all	37.7			Haven't done at all	40.7
	NA	34.7			NA	34.6
	Total	100.0			Total	100.0

		Percent %				Percent %
	At least once each day	10.3			At least once each day	2.0
	At least once each week	25.5			At least once each week	7.4
	Several times	24.2		Copied	Several times	19.9
Found Answer	Once or twice	5.6		Code	Once or twice	20.2
	Haven't done at all	1.2			Haven't done at all	16.4
	NA	33.3			NA	34.0
	Total	100.0			Total	100.0

These results stressed-out that the information quality of the existing posts in the community helps the participants to find and use the available solutions before asking a new question. Consequently, this could be one of the reasons that participants reported lower activity on asking questions. Accordingly, the number of posted questions influenced the users' participation in writing the answers.

5.3.3 Usage of the hypothetical tools

Stack Overflow provides different facilities to support their users for gaining knowledge and career developments. In the developer survey, Stack Overflow has created hypothetical tools based on the website facilities for evaluating the overall result of the website facilities. In the questionnaire participants were asked to rate their interest on participating on those hypothetical tools. In order to explore the preferences of interest on using such tools, based on the participants experience level, the three of those tools, which were relevant to the learning have been examined. Overall, participating in Programming Blog Platforms and also Peer mentoring system has more preference rating than another tool, which provides a private area for the participants who are new to programming.

As can be seen in *Table 14* the dataset 2018 represents the values indicated that the participants with the fewer years of experiences were more interested in participating in the community as hypothetical tool that represented the private area for who are new to programming. For instance, the first row of the table which represents "Extremely interested", the number indicated under each column for that particular row is constantly decreasing in each category of coding experience from 0 to 2 (16.8%) till (2.5%) which was identified by the participants with 20 and more years of experiences.

Similarly, as the experience level is going higher the value that signifies the percentage for the variable represents 'not at all interested' on this tool is increased too. This means that the more experienced the participants are, they less agreed with the concept of using the community as a private area for new programmers.

Table 14 *Rating the of interest participating in private area for new to programming as a hypothetical tool*

		CodingExperience													
		0–2 years		3–8 years		9–14 years		15–20 years		20 Or more years		NA		Total	
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Interest In Private Area For New To Programming	Extremely interested	24%	16.8%	55%	9.7%	14%	5.1%	4.8%	3.3%	3.0%	2.5%	0.1%	0.1%	100%	7.6%
	Very interested	15%	18.0%	54%	16.5%	19%	12.2%	7.7%	9.0%	5.0%	7.2%	0.1%	0.3%	100%	13.2%
	Somewhat interested	8.8%	13.0%	47%	17.5%	22%	17.5%	12%	16.6%	9.4%	16.3%	0.1%	0.2%	100%	16.0%
	A little bit interested	6.1%	7.6%	41%	12.8%	25%	16.6%	15%	18.1%	13%	18.6%	0.1%	0.2%	100%	13.5%
	Not at all interested	3.0%	5.8%	32%	15.1%	27%	27.4%	19%	35.3%	18%	41.1%	0.0%	0.1%	100%	20.5%
	NA	14%	38.9%	42%	28.4%	15%	21.3%	6.9%	17.7%	4.5%	14.3%	17%	99.1%	100%	29.1%
	Total	11%	100.0%	43%	100.0%	20%	100.0%	11%	100.0%	9.2%	100.0%	5.1%	100.0%	100%	100.0%

The overall result indicated in *Table 15* also shows that the level of experience has an inverse effect on the preference level of using the community as peer-mentoring system. But there is a slight change in comparison with *Table 14* for each group of the experience category. By comparing the rate of preference in each group within the same category, it was indicated that the preference for using a mentoring system has a fair balance in interest of using for each category of experience group. According to the shown results, the overall percentage that represents the interest of using the community as a peer mentoring system is increased by more experienced participants.

Table 15 *Rating the of interest participating in peer mentoring system as a hypothetical tool*

		Coding Experience													
		0–2 years		3–8 years		9–14 years		15–20 years		20 Or more years		NA		Total	
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Interest In Peer Mentoring System	Extremely interested	17%	11.8%	52.1%	9.2%	18%	6.8%	8.1%	5.5%	4.9%	4.1%	0.1%	0.1%	100%	7.6%
	Very interested	12%	14.7%	49.6%	15.5%	21%	14.0%	10%	11.9%	7.3%	10.7%	0.1%	0.1%	100%	13.5%
	Somewhat interested	9.2%	18.4%	44.1%	22.0%	23%	24.1%	13%	25.1%	11%	24.9%	0.1%	0.3%	100%	21.6%
	A little bit interested	7.3%	9.3%	40.2%	12.8%	24%	16.0%	15%	18.1%	14%	20.7%	0.1%	0.1%	100%	13.8%
	Not at all interested	5.2%	6.9%	35.7%	11.9%	25%	17.9%	17%	21.9%	16%	25.6%	0.1%	0.2%	100%	14.4%
	NA	14%	38.9%	42.1%	28.4%	15%	21.2%	6.8%	17.5%	4.4%	14.0%	17%	99.1%	100%	29.1%
	Total	11%	100.0%	43.1%	100.0%	20%	100.0%	11%	100.0%	9.2%	100.0%	5.1%	100.0%	100%	100.0%

The overall rating for Participating in the community as a Programming-oriented blog platform has a slightly higher percentage in comparison with the other two hypothetical tools. As demonstrated in *Table 16*, overall percentages in the first three rows are higher with slight change in each group of experience category. But by exploring the row which represent the values of 'not being interested' by

participants, illustrated that as the years of experience goes higher the percentage of disagreement on accepting the community as programming blog platform is also increased.

Table 16 *Rating the of interest participating in programming blog platform as a hypothetical tool*

| | | CodingExperience | | | | | | | | | | | | |
| | | 0–2 years | | 3–8 years | | 9–14 years | | 15–20 years | | 20 Or more years | | NA | | Total | |
		Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %	Row N %	Column N %
Interest In Programming Blog Platform	Extremely interested	17%	14.5%	55%	12.0%	17%	8.0%	6.5%	5.4%	4.2%	4.2%	0.1%	0.2%	100%	9.4%
	Very interested	12%	19.0%	50%	20.2%	21%	18.1%	9.6%	14.7%	6.9%	12.9%	0.1%	0.2%	100%	17.3%
	Somewhat interested	8.4%	14.3%	43%	18.3%	23%	20.7%	14%	22.4%	12%	23.3%	0.0%	0.2%	100%	18.4%
	A little bit interested	6.4%	7.6%	38%	11.3%	25%	15.4%	16%	17.7%	14%	19.9%	0.1%	0.2%	100%	12.7%
	Not at all interested	4.6%	5.6%	32%	9.8%	26%	16.5%	19%	22.2%	18%	25.4%	0.0%	0.1%	100%	13.0%
	NA	14%	39.0%	42%	28.5%	15%	21.3%	6.8%	17.6%	4.5%	14.3%	17%	99.1%	100%	29.2%
	Total	11%	100.0%	43%	100.0%	20%	100.0%	11%	100.0%	9.2%	100.0%	5.1%	100.0%	100%	100.0%

5.3.4 Users satisfaction on Stack Overflow

In order to discover the users satisfaction on participating in Stack Overflow, three items has been examined.

- How the information are helpful *(dataset 2017)*
- The overall Participants satisfaction with Stack Overflow *(dataset 2017)*
- How likely participants recommend Stack Overflow to others *(dataset 2018)*

Table 17 *Answer and code assessment*

In the Developer survey, participants were asked to rank the level of their satisfaction on the answers and code examples shared in the Stack Overflow Community. As demonstrated in the *Table 17*, the majority of the users either agreed or strongly agreed on the helpfulness of code examples and answers they have got through the Stack Overflow community.

		Value
The answers and code examples on Stack Overflow are helpful	Strongly disagree	0.2%
	Disagree	0.4%
	Somewhat agree	6.4%
	Agree	32.1%
	Strongly agree	27.1%
	NA	33.9%
	Total	100.0%

Although the results are based on individual contentment's on the answers and solution for their issues, it is important to be aware that this analysis could not be used for evaluating the quality of the answers. The result showed that participants were satisfied with the solution based on their requirement.

Overall Users satisfaction with participating in the stack overflow as a community for sharing knowledge and skills development as a second item has been studied. The statistics emphasized the high general satisfaction of using the Stack Overflow community by participants. As displayed in *Table 18*, 18% of the participants indicated the highest score of ranking. This rating is continued by 19% and 20% as the other second and third most top scores. On the other hand, the indicated dissatisfaction percentage on the same table shows the big gap between the highest and lowest measurement scale. Accordingly, the users' absolute satisfaction is reported by the result of this analysis.

Table 18 *Stack Overflow Satisfaction result*

Case Processing Summary

		Marginal Percentage
Stack Overflow Satisfaction	0(Lowest)	0.1%
	1	0.1%
	2	0.1%
	3	0.3%
	4	0.4%
	5	1.4%
	6	2.8%
	7	9.2%
	8	19.9%
	9	19.1%
	10(Highest)	17.8%
	NA	28.8%
Valid		100.0%

The last item that has been exanimated for this section was the data represented the result of how likely participants would recommend the Stack Overflow to friend or colleague. The participant were asked to rate their feedback on a scale from 0 to 10, with "10" represents the "very likely" to recommend. The statistics indicated that the majority of the participants were very likely to suggest the Stack Overflow to others. As demonstrated in *Table 19*, the participants who have selected ten as the highest rate of ranking are 53% of the population. In this set of data, there is a significant gap between the highest rating score and the rest of the measurement score. As can be seen, the result of this analysis indicated that the majority of the participants are very likely would recommend the overall participation in the community to others.

Table 19 *Stack Overflow Recommending to others result*

Case Processing Summary

		N	Marginal Percentage
Recommending Stack Overflow to a friend or colleague	0 (Not Likely)	270	0.3%
	1	124	0.1%
	2	191	0.2%
	3	294	0.3%
	4	313	0.3%
	5	1183	1.2%
	6	1320	1.3%
	7	3887	3.9%
	8	7667	7.8%
	9	8450	8.5%
	10 (Very Likely)	52790	53.4%
	NA	22366	22.6%
Valid		98855	100.0%
Missing		0	
Total		98855	

5.4 Summery of the results

This section represents a summary of the results that were studied in details. In order to provide a better outlook of the findings, *Table 20* demonstrates a concise version of the results based on the research questions as below.

Table 20 *Summery of the results*

Research Question	Summary of the results
Q1. What are the characteristics of the Stack Overflow users?	• Participants were from both genders with majority of male category (60%) • Participating from 184 countries worldwide • 43% of participants were having 3 to 8 years of coding experience • 72% of participants had the university level background • 25% of participants were currently official student • The smaller group of the participants agreed on the effect of formal education on their success as a developer • 86% of the participants employed with the majority in the field of software development
Q2. What are the participants' preferred informal learning methods?	• 60% of participants indicated the selection of self-taught education and attempted online courses • 39% of participants Self-taught study were one of their choices of selection for non-degree education • Experienced participants preferred to use the Self-taught learning method in comparison to other informal learning methodology • 27% have selected online communities as self-learning sources • 25% of the participants have selected the Stack Overflow community as an informal learning source • The preference for using the Stack Overflow is higher for experienced participants
Q3. How do the users participate in the Stack Overflow community?	• 68% of the participants had registered a user account • 43% of participants consider themselves as the Stack Overflow members • Stack Overflow contains 55% monthly active participators • More experienced users are more active on participating in the community • 38% never wrote the answer to others posts • 41% never asked the question in the community • Majority of the participants found a solution to their issues • The participants with few years of coding experience agreed on using community as a private place for new programmers • More experienced participants agreed on using community as peer mentoring system • As the participants experience level goes up, the preferences of using the community as programming blog platform goes down • The majority of the participants find the information in the community useful • The users' absolute satisfaction has reported • 53% of the participants are very likely would recommend the community to others

6.0 Discussion

The main consideration of this study is to explore the impact of participation in the knowledge sharing communities on expertise development. Consequently, the three factors of users, learning preferences, and participation in the community, with the use of socially situated learning theory and communities of practices, have been studied.

The overall result of this study indicated that users are satisfied with the shared contents and their participation in the community. Given the notion of socially situated learning, the results indicated that the participators of the Stack Overflow community are satisfied by finding solutions, supporting other participants, and gaining knowledge through interacting with the experienced users in the community, hence locating themselves in the center of learning process. Besides, the results reported how the level of experience affected the types of activities in the community. Participants with fewer years of coding experience preferred to use the community as a private place for practicing and skills improvement. In addition, as the level of expertise increases, it changes the preferences of the participants to use the community as peer mentoring system. This matter indicated the process of developing the users' identity and shaping their participation status by moving toward full participation in the community.

Referring to four components of the socially situated learning (Meaning, Practice, Community, and Identity), results showed how the Stack Overflow participants contributed in the community by interacting with others in sharing knowledge, finding a solution, and experiencing the content. Since Stack Overflow is a specialized community that allows the participants to raise a specific question in the field of computer programming, the participants are skilled enough to be engaged in the community and use the contents properly. Moreover, significant number of monthly visitors of the community supports the Wenger argument about daily involvements in the communities as a concept of learning.

Further, the results stressed-out how three dimensions of communities of practice defined by Wenger proved the capability of Stack Overflow to be recognized as a community of practice. The results indicated the mutual engagement by the Stack Overflow participants through supporting each other toward skills development in computer programming purposes. Stack Overflow provided a variety of communicational and motivational features in the community that allowed the participants to employ those tools based on their requirements for interacting in the community.

In the following sections, the results of this paper will be discussed based on the three following research questions:

1. *What are the characteristics of the Stack Overflow users?*
2. *What are the participants' preferred informal learning methods?*
3. *How do the users participate in the Stack Overflow community?*

6.1 Users characteristics

In order to explore the users' characteristics of the Stack Overflow community, different variables of the users' profiles have been studied. The indicated results were similar to the results of previous studies from perspectives of demographic, geographic, and educational background. The study of this paper demonstrated that the majority of the Stack Overflow participants are from the male gender group that was also reported previously by Ahmed and Srivastava (2017) and Ford et al. (2016). The majority of participants have a university level education and participating from different countries. Similarly, Schenk and Lungu (2013) also reported the same with a slight change due to using a more accurate technique (Geo-location API) that the users from 189 countries participated in the community.

Another important factor in user identification that has been studied in this paper was the level of coding experience. The concept of user experience itself has been considered previously in many researches including Hwang et al. (2015) and Kowalik and Nielek (2016), but in each study the particular variable formulates the level of experience based on the requirements of the study. Similarly, in this paper, the range of coding experience was sorted in five different categories that indicated the significant number of participants has 3 to 8 years of programming skills.

Since the aims of this study are based on informal education, it was interesting to study the influence of formal learning as a key success towards becoming a developer. Accordingly, the impact of formal education on participants' progress as an additional finding of this paper has been studied. The unexpected result indicated that a small group of participants agreed on the importance of formal education on their success of being developers. This implies that, with the focus on the skills level, experience is an essential factor in parallel with formal education in the field of computer programming.

The implication of the study provides a general overview of the characteristics of the users that it is required in order to address the other two research questions. Besides, to explore the incentives of situated learning, it was important to understand the participants' characteristics especially from the knowledge background and geographic perspectives. According to Lave and Wenger (1991), the incentive of situated learning is the intention of exploring the process of skills developments by participating in an informal training process and being involved in the communities where the participants are experiencing different contexts and different geo-locations.

Based on the results of this paper and existing literature, it can be seen that the findings of this paper have a close link to the described participants' characteristics in the theory of situated learning.

6.2 Preferred informal learning methods

The Preferences of using informal learning methods show a reasonable percentage in the result of this paper. The majority of participants indicated that they would like to attempt an online course as well as gaining knowledge through other self-study methods. Since one of the objectives of this research is the usage of online communities as a source of self-taught learning, results indicated that 25% of participants use Stack Overflow as a learning source. Furthermore, the effect of experience on the selection of types of informal learning source had been studied. The Result section of this paper reported that more experienced participants would likely prefer using Stack Overflow community as

an informal learning source in order to develop their professional skills by interacting with other participants in the community.

This result is consistent with the results of previous literature where Kowalik and Nielek (2016) indicated that the experienced participants are active members of the community. Additionally, as reported by Marrison and Murphy (2013), skilled programmers learn by engagement in the community. Since 25% of the Stack Overflow participants are recognized as students, a group of junior programmers also indicated their preferences of using the Stack Overflow as an informal learning source. This result referred to the theory of situated learning that defined learners should practice and experience meaningful content in real life. The contents of Stack Overflow provide an opportunity for the participants to learn and develop their programming skills through real-life examples. Obviously, finding a proper solution and the way of behaving in the community is part of the learner responsibilities. However, as Stack Overflow is known as a peer-initiated community, there is always a possibility to receive support from other members of the community in order to shape the learner identity and find the right path of participation toward skills developments.

Proper usage of contents and exploring the solutions are one of the reasons that the percentage of Stack overflow usage by experienced users indicated continual increase. Knowing and knowledge are the two factors described by Wenger which made the skilled participants more comfortable compared to juniors in order to participate and track the solutions in the community. Similarly, Nivala et al. (2016), Kowalik and Neilek(2016), and Borg et al. (2017) have reported the positive influence of experience on learning and finding solutions through different studies.

As described above, the findings of this section were similar to existing studies in terms of general concepts. As per given reasons and indicated in the results section of this paper, 39% of the participants preferred to use self-taught learning as an informal educational method. In addition to previous studies' findings, this paper reported the participation preferences in online communities and especially the Q&A community of the Stack Overflow as a source of self-learning methods. Further, it has been discovered that the experienced participants have a higher percentage in using the Stack Overflow for skills developments and learning perspectives.

6.3 Participation in the community

By exploring the interaction of participants in the community toward expertise development, the main findings of this paper indicated that the majority of participants were registered as a user in the community. From the activities perspective, more than half of the participants identified themselves as active monthly participators who were involved in producing the content in the community. Accordingly, there were a small group of participants who did not have a user account, but they recognized themselves as Stack Overflow participants. This fact indicated that both registered users and visitors were having access to the shared information in the community. Importantly, the majority of the participants agreed that the contents in the community were helpful for solving programming issues. Consequently and satisfactorily they highly recommended the usage of the Stack Overflow to people who were interested in coding and computer programming.

Results of activities in this study interpreted the roles of participants when using the contents in the community. The analysis indicated that only a small group of participants wrote answers or asked a

new question in the community. Instead, participants were more active in searching for solutions and using the codes example on their visits to the community. Since the number of participants of the community increases constantly along with a lower population for experienced users, it fitted the Wenger's argument about limited involvement of newcomers and their observer roles in the communities. Based on the findings, the majority of the participants play only a peripheral role while a smaller group of users seem to be legitimate participants. Similarly, Wang et al. (2018) have reported active participation in revision and reviewing posts by experienced users and Smrithi and Venkatapathy (2014), reported that contributors used the forum to obtain a more specific answer to the coding issues.

To discuss the participation results in a more specific way and based on the expertise and proficiency, the usage of different hypothetical tools on experience has been studied. As expected, the result showed that experienced participants were more involved in participating in the community by interacting on posts. Since the experienced participants showed more interest in using the community as a peer mentoring system, they played supportive and instructional roles in the community. This opportunity represents the component of meaning that improves the process of learning by users' interaction to experiment a meaningful notion in supporting each other in the community. In parallel to the findings of this paper, Duncan-Howell (2009) reported that online communities provided a valuable form of professional learning and similarly Nores et al. (2019) reported how participating in the Stack Overflow improved the students' knowledge in using technical terminology by learning through real-life code examples.

Based on the study, majority of the participants agreed that the information in the community was helpful. Participants indicated their satisfaction with the solutions provided to their issues in the community and it accordingly increased their level of satisfaction in overall participation in the community. Consequently, a significant number of recommendations for using the Stack Overflow to others have been reported. The available data limited the study to evaluate the quality of the contents, and unfortunately the assessment of satisfaction will not prove the quality of the information. However, Matei et al. (2018) reported that Stack Overflow had reached the stable level in terms of content quality. Although, based on the dataset of this study, it was not possible to prove that either. The results of this research could only prove the level of user satisfaction with the given solution, but did not represent the quality of the answers.

6.4 Limitation of the study

The objective of the study was to understand how participation in the Stack Overflow as learning community influences the expertise development. To approach the aim of the study, the results of the developer survey have studied. One of the most challenging tasks of this research was the number of limited and types of available data relevant to learning and activities on the community in the datasets. Further, to collect more data and gathering different types of variables to avoid the limitation of presenting the result by percentages as descriptive analysis, the additional survey has designed. But due to the website restriction policy there was a limitation in accessing the participants in order to collect more data. As a result, all the analysis had been done based on limited data provided by Stack Overflow.

Another limitation lies in the study was testifying the reliability of the data. As the data were based on the results of online self-doing survey, verifying the indicated information by the participants has a low possibility. Especially this applies in the cases were participants required to evaluate themselves such as indicating the coding experience. Each individual might have different scale of measurement while evaluating his or her experience (one hour per day and full-time a day both count as daily coding experience).

Finally, the time allocated for work on this research limited the study for broader exploration on the older available datasets. Besides, on the time of doing the research for this paper, the 2019 dataset were not published in the website.

Based to this reasons, the results of this paper cannot be generalized and further research are recommended by studying more datasets.

6.5 Suggestions for future research

The sample of the study was base on the participants of the developer surveys, which is a small sample size in comparison to the registered number the Stack Overflow users. For more precision of the results and have a better picture of the users, and their participation in the community, a more significant sample of participants, recommended for the future studies. Besides, there is a possibility to increase the reliability of the results by using more technical methods such as geo-location API, Tag detectors, analysis the users profile systematically, etc.

If there is a possibility of extending the allocated time for the project and the awareness of technical issues to access the Stack Overflow participants, designing a survey and collecting more data with a focus on learning is recommended for the future researches.

In addition, due to the capability of the community of Stack Overflow and the considerable number of participants and the traffic of the website, the variety of possible study can be recognized. One of the interesting future studies that have detected during the analysis of this paper is the relationship between experiences and prefers types of employment. During studying the data, it has been identified that as the participants' level of experience goes up, the percentage of the employment status for the category that represents the self-employed and freelancer increased. This could be an interesting study to investigate how the level of experience influences the preferences of selecting types of employment.

7. Conclusion

This study attempted to explore the influence of online communities on expertise developments. In order to address the aim of the study, the Stack Overflow Q&A community as one of the most popular online communities in the field of software development, was studied. Two sets of data that were derived form Stack Overflow developer survey were used to reach the study objectives. The Socially situated learning by Lave and Wenger (1991) and its extensions the Communities of Practice by Wenger (1998) were among the most applicable theories, the implications of which were helpful in addressing the aims of the study

As per the finding of this research, it was indicated that majority of participants were satisfied with their participations in the community. It was also seen how influential the level of experiences was when participating in the community. Additionally, the results reported that the experienced participants were more active in producing content in the community and also they played supportive roles when helping other members.

Application of the theory of socially situated learning to the participants of the stack overflow discovered that the majority of the participants who were having fewer years of coding experience, played the peripheral role by observing and finding solutions from the existing posts. The users' participation role was changed as the level of experience was increased. The more experienced users, representing a smaller group of the participants, seemed to be legitimate participants in the community.

The community of Stack Overflow is an example of peer-initiated society where unknown participants are responsible for moderating the contents shared by other participants. Stack Overflow provides a variety of tools and features to improve communicational and motivational interactions of the participants in order to develop knowledge. The result of this paper reported the three dimensions of practice, namely as mutual engagement, joint enterprise, shared repertoire as the properties of a community that were fitted to the Stack Overflow community and made it to be known as a community of practice.

The outcomes of this research emphasized the influence of online communities on skills development for learning perspective and how important that is. Informal learning requires self-consistency, if one wishes to either gain new knowledge or improve the existing ones. Based on the study, experienced participants are using contents more properly and they participate in the community confidently, while newcomers required more time to find their positions in the society. Another considerably important fact is a concept of knowledge evaluation. Unfortunately, Online communities do not provide exact scales of knowledge evaluation. This has made, the motivation and self-consistency of the participants among the most important factors when participating in such communities.

Reference list

Ahmed, T., Srivastava, A. (2017). Understanding and evaluating the   behavior of technical users. A study of developer interaction at Stack Overflow. *Human-centric Computing and Information Sciences,* 7:8. DOI 10.1186/s13673-017-0091-8

Aubrey, K., Riley, A. (2016). *Understanding & Using Educational theories*. (1st ed.). Los Angeles: SAGE.

Borg, M., Ros, R.,   Bjarnason, E., &   Lennerstad I. (2017).   On Using Active Learning and Self-training when Mining Performance Discussions on Stack Overflow. *EASE'17 Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering* Pages 308-313

Brown, J.S., Gray, E.S. (1998). The people are the company. *Fast Company online: http://www.fastcomany.com/online/01/people.html*

Dewey, J. (1986). Experience and Education. *The Educational Forum, 50:3,* 241-252, DOI: 10.1080/00131728609335764

Duncan-Howell J. (2009), Teachers making connections: Online communities as a source of professional learning. *British Journal of Educational Technology, Vol 41 No 2 2010* 324–340 doi:10.1111/j.1467-8535.2009.00953.x

Ford, D., Smith, J., Guo, P.J., & Parnin, C. (2016). Paradise Unplugged: Identifying Barriers for Female Participation on Stack Overflow. *New York, NY, USA ISBN: 978-1-4503-4218-6* DOI: 9N7897-18-14-5405303-4-422118-6/16//111

Gao, D., Chen, X. (2010), On-Line Learning Community Based on Curricular. *Springer-Verlag Berlin Heidelberg* LNCS 6248

Guan, T., Wang, L., Jin, J., & Song X. (2017). Knowledge contribution behavior in online Q&A communities: An empirical investigation. *Computers in Human Behavior* 81 137-147

Hildreth, P., Kimble, C. (2000). Communities of Practice in the Distributed International Environment. *Journal of Knowledge Management; Kempston* Vol. 4, Iss

Hwang, E., Singh, P.V., &Argote, L. (2015). Knowledge Sharing in Online Communities: Learning to Cross Geographic and Hierarchical Boundaries. *Organization Science* 26(6):1593-1611

Kowalik, G., Nielek, R. (2016). Senior Programmers: Characteristics of Elderly Users from Stack Overflow. *Lecture Notes in Computer Science,* vol 10047. Springer, Cham

Lave, J., Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: University Press.

Matei, S.A., Abu Jabal, A., &Bertino, E. (2018). Social collaborative determinants of content quality in online knowledge production systems: comparing Wikipedia and Stack Overflow. *Social Network Analysis and Mining, Springer Vienna* 8: 36. https://doi.org/10.1007/s13278-018-0512-3

Morrison, P., Murphy, E. (2013). Is programming knowledge related to age? An exploration of stack overflow. *In 2013 10th IEEE Working Conference on Mining Software Repositories (MSR)*, pages 69–72, May 2013

Nivala, M., Hauser, F., Mottok, J., &Gruber, H. (2016). Developing visual expertise in software engineering: An eye tracking study. *IEEE Global Engineering Education Conference (EDUCON),* 978-1-4673-8633-3/16/$31.00 ©2016 IEEE

Nores, M.L., Fernández, Y. B., Torres, J.F., Arias, J.J.P., Solla, A.G, &Cabrer, M.R. (2018). Experiences from placing Stack Overflow at the core of an intermediate programming course. *Wiley Periodicals,* Inc DOI: 10.1002/cae.22109

Ovadia, S.(2011). Quora.com: Another Place for Users to Ask Questions. *Behavioral & Social Sciences Librarian, 30:3,* 176-180. DOI: 10.1080/01639269.2011.591279

Patil, S., Lee, K., (2016). Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors. *Springer,* DOI 10.1007/s13278-015-0313-x

Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation for future research. *Journal of Computer Mediated Communication, 10*

Rekha, V.S., Venkatapathy, S. (2014). Understanding the Usage of Online Forums as Learning Platforms. *Organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014).* doi:10.1016/j.procs.2015.02.074

Sarka, P., Ipsen, C. (2017). Knowledge sharing via social media in software development: a systematic literature review. *Knowledge Management Research & Practice, 15:4*, 594-609. DOI: 10.1057/s41275-017-0075-5

Schenk, D., Lungu, M. (2013). Geo-locating the knowledge transfer in StackOverflow. *In: Proceedings of the 2013 international workshop on social software engineering.* ACM, pp 21–24

Sowe, S.K., Stamelos, I., &Angelis L. (2008). Understanding knowledge sharing activities in free/open source software   projects: An empirical study. *Journal of Systems and Software.* 81.3 (2008): 431-446.

StackOverflow. (2019). 2019 stackoverflow developer survey. Available from: http: //stackoverflow.com/research/developer-survey-2019

StackOverflow. (2018). 2018 stackoverflow developer survey. Available from: http: //stackoverflow.com/research/developer-survey-2018

StackOverflow. (2017). 2017 stackoverflow developer survey. Available from: http: //stackoverflow.com/research/developer-survey-2017

Wang, S., Chen, T., &Hassan, A. (2018). How Do Users Revise Answers on Technical Q&A Websites? A Case Study on Stack Overflow, *IEEE Advancing Technology for Humanity 0098-5589* (c) 2018 IEEE. DOI 10.1109/TSE.2018.2874470

Wenger, E. (1998). *Communities of practice : Learning, meaning, and identity*. New York, N.Y.: Cambridge University Press

Ziegler, M.F., Paulus, T., &Woodside, M. (2014). Understanding Informal Group Learning in Online Communities Through Discourse Analysis. *Adult Education Quarterly*, Vol. 64(1) 60–78