Stack Overflow – Informal learning and the global expansion of professional development and opportunities in programming?

Markus Nivala
Department of Education,
Communication and
Learning
University of Gothenburg
Gothenburg, Sweden
markus.nivala@gu.se

Alena Seredko
Department of Education,
Communication and
Learning
University of Gothenburg
Gothenburg, Sweden
alena.seredko@gu.se

Tanya Osborne
Department of Education,
Communication and
Learning
University of Gothenburg
Gothenburg, Sweden
tanya.osborne@gu.se

Thomas Hillman
Department of Applied
Information Technology
University of Gothenburg
Gothenburg, Sweden
thomas.hillman@gu.se

Abstract—The purpose of this study is to examine if, and to what extent, online Community Question Answering platforms expand the opportunities for professional development in programming. Longitudinal and cross-sectional analyses of Stack Overflow Developer Surveys were used to examine users' geographical distribution, gender, experience, professional status, platform usage and education. In order to study differences between the countries with the largest number of respondents, the developer survey data was combined with indicators of human development, gender equality and educational attainment. The results show that the Stack Overflow community has expanded to some extent, both in terms of wider geographical distribution and the programming expertise of users. However, the community reflects and fails to mitigate the apparent gender disparity in the field of programming. Furthermore, participation seems to be conditioned by formal education, especially in developing countries. In general, participation patterns in Stack Overflow seem to be heavily influenced by local conditions in different countries.

Keywords—*informal learning, programming, stack overflow, equality*

I. INTRODUCTION

Informal lifelong learning through networked online communities and the timely issue of fostering programming skills across the population represent key parts of the information communication technology (ICT) literacy aspect of 21st Century Skills [1] and policy initiatives at national and transnational levels [2]. The purpose of this study is to examine if and to what extent a particular type of online community, Community Question Answering (CQA) platforms, expand the opportunities for expertise development in programming using Stack Overflow as a case. In particular, the aim is to study how users with different backgrounds are represented in the community and to discuss the possible obstacles for participation in informal learning of programming.

Expertise in programming, and its development, has traditionally been studied at the individual level by, for example, comparing novices to experts in terms of their knowledge and reasoning strategies [3-5]. In the last decade, however, “additional emphasis has been placed on the idea of considering learning as an interactive process of ongoing participation in various cultural practices and in shared learning activities” [6, p. 176]. Indeed, studies have shown that programming expertise is connected to better communication and cooperation skills as well as to the

presence of highly functional personal networks [4]. Studying expertise development in its social context is especially pertinent to programming. Large-scale software development projects are often (multinational) collaborations between largely autonomous teams that are engaged in joint problem-solving processes [7]. Furthermore, the online platforms that support such collaboration are becoming more and more important as resources for expertise development [8], meaning that programming expertise increasingly resides and develops in networked communities.

Stack Overflow is a popular online community where more than 50 million “professional and aspiring programmers visit Stack Overflow each month to help solve coding problems, develop new skills, and find job opportunities” [9]. Unlike most discussion forums, Stack Overflow is structured as a CQA platform for professional and enthusiast programmers that is focused on finding solutions to authentic programming problems. This follows the model of earlier non-programming focused CQA platforms like ProfNet, WikiAnswers and Yahoo! Answers with a cross pollination of features between platforms, leading to a distinct structure that differentiates CQA from other types of knowledge sharing platforms.

As is the case with most current CQA, the Stack Overflow community and its content is moderated by users and organised based on user assigned subject area tags. The quality and relevance of content is moderated by voting questions and answers up or down in a process that affects their visibility to the community. Votes that members receive for posed questions or given answers in turn determine the members' ‘reputation score’. The higher the reputation, the more privileges a member has access to (e.g. acquiring status as a community moderator). Furthermore, it has been argued that participation in the community, as indicated by reputation score, enables users to obtain recognition and social prestige among their peers [10, 11], and to develop their identities as experts in the community [12]. Moreover, the reputation score may serve as a way to demonstrate expertise when applying for a job [13]. Therefore, Stack Overflow can be seen as a potential platform for professional development and opportunities.

However, past research has shown that there are considerable barriers to participation and engagement. For example, Oliveira, Andrade and Reinecke [14] studied Stack Overflow participation patterns in terms of posing questions, providing answers, commenting on posts and editing content. The results show that, compared to users from Eastern countries, users from Western countries are more likely to

textually contribute to the community. Similarly, Calefato, Lanubile and Novielli [15] found that the best chance of getting a successful answer to a question is to post it during office hours in North America, further highlighting Western dominance on the platform. However, it has been argued [14] that instead of West vs. East or a particular country's economic viability, analyses should consider national culture (individualist vs. collectivist) and the general English proficiency of residents.

In addition to geographical differences, gender has been identified as an important factor in terms of Stack Overflow participation. It is difficult to make generalised claims about the representation of women on global platforms such as Stack Overflow because there are significant differences in the uptake of computer science professions across countries, with substantially more women participating in computer science study and occupations in Asian settings [16, 17]. Analysis of a sample of user profiles [18] suggests that women account for only 7% of Stack Overflow users. Even in US contexts, with typically low female participation in programming, 18% of the programmers in Silicon Valley are female [19], and 25% of the computer science workforce in general are female [20]. Therefore, the number of estimated female users of Stack Overflow suggests that the platform underrepresents the number of women occupied in programming or programming adjacent occupations. Furthermore, those women who do have a user account typically post less questions and answers as well as having shorter active careers on the platform and achieving lower maximum reputation scores. One possible explanation for the continued underrepresentation of women on Stack Overflow that has been argued for is based on a general sociological assumption that people often interact more with those of their own gender [21, 22]. While this argument could account for lower per capita levels of interaction in the form of questions, answers and comments for women, it does not account for the very low proportion of accounts [18] that have been identified as belonging to women. Unpacking this issue, Ford, Smith, Guo, and Parnin [23] found that women are more likely than men to doubt whether they have a sufficient level of expertise to contribute, and that they tend to feel inhibited by the massive user base and competitive consequences of the Stack Overflow reputation score system. The authors also found that these factors contributed to a kind of vicious circle where women participate less and report feeling that they are restricted in their ability to participate by having limited awareness of platform features [23].

While earlier research on Stack Overflow has addressed issues of participation in relation to some aspects such as gender differences, there is still much to unpack about the ways in which access to developing programming expertise through the knowledge base on the platform is differently accessible for different members of this global community. Seeking to contribute to unpacking these issues, this study is guided by the question, how are opportunities for expertise development in programming accessible in the case of Stack Overflow?

II. METHODS

A. Datasets

Stack Overflow Developer Surveys (SODS) from 2011 to 2018 [24] were used as the main source of data. The sample size across the surveys has been increasing alongside the growth of the platform. The sample size ranges from 2812 (in 2011) to 988855 (in 2018). In addition to demographic data, the surveys include information about respondents' professional status, experience, education and use of Stack Overflow. The variables that were extracted from the annual Stack Overflow Developer Surveys are listed below. In order to improve readability, some of the variables have been renamed for the purposes of this paper. In these cases, the original variable names are given in parentheses:

2011, $N = 2812$:

Continent (What Country or Region do you live in?)

Programming experience (How many years of IT/Programming experience do you have?)

2015, $N = 26086$:

Employment status

Gender

2017, $N = 33618$:

Platform usage (StackOverflowFoundAnswer, StackOverflowCopiedCode, StackOverflowAnswer)

Importance of education (EducationImportant)

2018, $N = 988855$:

Country

Gender

Employment status (Employment)

Programming experience (YearsCoding)

Professional programming experience (YearsCodingProf)

Mean salary (ConvertedSalary)

Education (FormalEducation)

The Education variable was recoded into two categories: 1) without professional degree and 2) with professional degree (associate/bachelor's degree or higher). In the original survey data, the Programming experience and Professional programming experience variables were measured with increments of two years. In the current analysis, the highest value of each two-year increment was used. While the imprecision that originates from the original survey data should be taken into account when examining the means, the main use of these variables here is to relate them to the other variables rather than to examine them independently.

In addition to SODS, data from the United Nations (UN) Human Development Index 2017 (HDI) [25] and Gender Development Index 2017 (GDI) [26] were used to examine the societal differences between the respondents' countries. The Human Development Index (HDI) summarizes three developmental measures: life expectancy, education (mean and expected years of schooling) and gross national income per capita. Gender Development Index (GDI) measures the gender gap in HDI, that is, "the ratio of the HDIs calculated separately for females and males" [26].

Data from UNESCO Institute for Statistics [27] were used as a measure of a country's overall educational attainment. For this purpose, the "Educational attainment: at least completed short-cycle tertiary (ISCED 5 or higher), population 25+ years, both sexes (%)" data from the years 2009-2019 were chosen, as these correspond closely to the

above-mentioned recoding of the Education variable. For each country, the latest available value was used.

Finally, the Global Gender Gap Report 2018 by the World Economic Forum was used to obtain information about the ratio of female and male graduates in ICT (information and communication technologies) subjects [28].

The data sources are indicated in the tables as SODS (Stack Overflow Developer Survey), UN (United Nations) and UNESCO (UNESCO Institute for Statistics).

B. Analysis

The longitudinal (SODS 2011-2018 datasets) and cross-sectional (SODS 2018, HDI, GDI, UN, and UNESCO datasets) analyses of the survey data focus on respondents' geographical distribution, gender, experience, professional status, platform usage, and education. In order to obtain a reasonable sub-group sample sizes, especially for questions about gender and educational background, more detailed analyses were carried out for countries with more than one thousand respondents ($N = 20$). Analyses that pertain only to these 20 countries are marked with, for example, " $(N = 20)$ ". Otherwise, full SODS datasets were used for the analyses with Spearman's correlation coefficient (r_s) used to examine the connections between different variables.

In addition, it should be noted that there have been considerable changes made to survey items between years. Therefore, the analysis summarizes the comparable results between different years when possible (e.g. geographical and gender distribution, programming experience, employment status). Yet, the main analytical focus here is on cross-sectional analyses of the available data.

III. RESULTS

A. Geographical and gender distribution

In 2011, most of the respondents were from North America (44%) or Europe (38%). However, the percentage of respondents from Asia has increased from 9% in 2011 to 28% in 2018, while in 2018 North America and Europe account for 25% and 37% of participants respectively. Africa, Oceania and South America are still relatively invisible, but the total amount of respondents from these areas has increased from 5% in 2011 to 10% in 2018. Table 1 shows the number of respondents, HDI ranking, HDI, percentage of females and GDI for countries with more than 1000 respondents. It seems that while some developing countries such as India and Pakistan are becoming more prominent in this particular programming community, there is still a large gender disparity across all nations.

From 2015, which was the first year when information about gender was collected, to 2018, only 5–7% of the respondents have identified themselves as females, which is in line with the percentage found by analysing the user profiles [18]. Interestingly, there were no significant correlations between percentage of females and HDI, HDI ranking or GDI ($N = 20$). For example, India and Pakistan, countries among with the lowest HDI and GDI, have more than double the number of female respondents compared to Switzerland with the highest HDI. Similar discrepancies can be observed in terms of percentage of female participants and GDI. This seems to suggest that female participation in this

community is not necessarily directly related to the factors behind these measures.

TABLE I. GEOGRAPHICAL DISTRIBUTION (2018), HDI, % FEMALE AND GDI

SODS			UN		
Country	N	% female	HDI ranking	HDI	GDI
USA	20309	9	13	0.924	0.992
India	13721	7	130	0.640	0.841
Germany	6459	5	5	0.938	0.967
UK	6221	6	14	0.922	0.960
Canada	3393	9	12	0.926	0.986
Russia	2869	6	49	0.816	1.019
France	2572	5	24	0.901	0.987
Brazil	2505	4	79	0.759	0.992
Poland	2122	6	33	0.865	1.006
Australia	2018	6	3	0.939	0.975
Netherlands	1841	5	10	0.931	0.966
Spain	1769	5	26	0.891	0.979
Italy	1535	2	28	0.880	0.967
Ukraine	1279	7	88	0.751	0.993
Sweden	1164	4	7	0.933	0.992
Pakistan	1050	7	150	0.562	0.750
China	1037	8	86	0.752	0.955
Switzerland	1010	3	2	0.944	0.987
Turkey	1004	5	64	0.791	0.922
Israel	1003	7	22	0.903	0.975

B. Programming experience, professional status and education

In terms of programming experience, the percentage of novice programmes (with less than two years of programming experience) has doubled from 5% (2011) to 11% (2018). On the other hand, while only 54% of the respondents were employed full-time in 2015, in the 2018 SODS this grew to 71%. However, employment status is not necessarily related to employment in computer sciences. The majority of the respondents (79%) reported that they have professional programming experience. On average, male respondents had 8 years ($SD = 5.6$) and female respondents had 6 years ($SD = 6.7$) of professional programming experience. While the mean salary was 96 477 USD/year for males, for females it was 88 251 USD/year. Yet, the variance in salaries was considerable ($SD = 201\,243$ for males, $SD = 188\,578$ for females). Indeed, median salaries were more equal (males 56 100 and females 54 336 USD/year) across genders. Although it is conceivable that the difference in experience explains the salary gap between the genders, the correlation between experience and salary was relatively low ($r_s = 0.148$, $p > 0.001$).

Examining the Global Gender Gap Report from the World Economic Forum for the represented countries with

more than 1000 respondents, the average ratio between female and male graduates in ICT subjects is 0.24 ($SD = 0.18$, $Min = 0.12$, $Max = 0.79$, $N = 17$, data from China, Pakistan and Russia missing) [28]. However, India (ratio = 0.79) and Turkey (0.54) are notable exceptions. There was a positive correlation between percentage of females on Stack Overflow (Table 1) and female/male ratio of graduates in ICT subjects ($r_s = 0.557$, $p = 0.015$, $N = 17$). The correlation between the ratio of female to male ICT graduates and GDI was reasonably high ($r_s = -0.410$), but it was not statistically significant ($p = 0.102$), possibly due to the small number of countries ($N = 17$). Considering the fact that GDI did not correlate with the number of female participants on Stack Overflow, it seems that instead of general measures of human development or gender equality, women's participation in ICT education emerges as a stronger predictor of Stack Overflow participation.

Furthermore, it seems that respondents with different levels of professional experience use the platform in a similar manner. In 2017, the highest rated (scale 1-5, 1 = "haven't done at all", 5 = "at least once a day") types of platform usage were:

- "Found an answer that solved my problem" ($M = 3.6$, $SD = 0.9$)
- "Copied a code example and pasted it into my codebase" ($M = 2.4$, $SD = 1.1$)
- "Written an answer to someone else's question" ($M = 1.7$, $SD = 0.9$)

Correlations between professional programming experience and the above-mentioned usage ratings were statistically significant ($p < 0.001$), but relatively small ($r_s = -0.146$, -0.110 , and 0.052 , respectively). Furthermore, in 2015, 45% of the respondents answered that they use Stack Overflow "to get help for my job" and 13% indicated that they cannot do their job without it. These results show how Stack Overflow has become an essential part of many participants' professional lives.

In 2017, only 41% of the respondents considered their formal education to be important or very important to their career success. Yet, in terms of access or participation in informal learning, a formal degree still seems to be an important factor, especially in developing countries. Table 2 shows the percentage of respondents with and without a professional degree (2018) as well as the country's HDI ranking and general educational attainment (at least Bachelor's or equivalent) [27]. Although the HDI includes the education index, educational attainment was added to the analysis as it excludes the other HDI dimension, namely, life expectancy and GNI per capita.

In general, the countries that have the lowest HDI ranking, that is, Pakistan, India and China, also had the lowest percentage of respondents without a professional degree. Furthermore, these countries also have low levels of educational attainment, which seems to indicate that only the small, educated minority of the population is participating in the community. However, France and Italy are interesting exceptions. Both countries have reasonably high HDI ranking, yet France had one of the lowest (10%) and Italy the highest (43%) percentages of respondents without professional degree. This is especially peculiar as France, compared to its European counterparts such as Germany, the

Netherlands and Spain, has a similar level of general educational attainment, whereas for Italy it is much lower. When France and Italy are removed from the data as outliers, the correlation between HDI ranking and respondents without professional degree by country was negative and reasonably high ($r_s = -0.47$, $p < 0.05$, $N = 18$). Although the correlation between educational attainment and respondents without a professional degree shows an emerging trend ($r_s = 0.400$, $N = 19$), it was not statistically significant ($p = 0.111$). These results highlight the complex interplay between different local conditions.

TABLE II. PERCENTAGE OF RESPONDENTS WITH AND WITHOUT A PROFESSIONAL DEGREE (2018), HDI RANKING AND EDUCATIONAL ATTAINMENT

Country	SODS		UN	UNESCO
	% with degree	% without degree	HDI ranking	Educational attainment
Australia	73	23	3	43
Brazil	64	33	79	17
Canada	70	26	12	48
China	74	17	86	9
France	87	10	24	30
Germany	62	33	5	26
India	85	9	130	10
Israel	63	33	22	47
Italy	55	43	28	14
Netherlands	70	28	10	33
Pakistan	81	13	150	9
Poland	69	28	33	25
Russia	67	29	49	62
Spain	72	24	26	31
Sweden	63	34	7	33
Switzerland	68	26	2	39
Turkey	76	20	64	19
Ukraine	73	24	88	na
UK	72	26	14	44
US	77	21	13	45

IV. DISCUSSION

The results show that Stack Overflow has grown from a relatively regional community (North America/Europe) to a global one and that developing countries are becoming more and more visible in this particular online programming community. Similarly, the user base seems to have also widened with regard to programming experience, that is, novices are becoming more involved. However, while more novices are participating, the fact that such a large percentage of respondents reported full-time employment and professional programming experience shows that, in addition to novices, the community continues to strongly attract professional developers.

The overall growth of the platform and the subsequent diversity in user expertise levels and platform usage can be seen to be positive in terms of access to knowledge sharing, but it may also have led to phenomena that decrease the educational value of Stack Overflow. Srba and Bielikova [29] have argued that simultaneous to platform growth, the quality of the posed questions and provided answers has decreased. Furthermore, their analysis shows that already between the years 2011 and 2014, the number of novices and what they term ‘reputation collectors’ (users who answer as many low quality or uninteresting questions as possible to gain reputation) has increased. At the same time, they found that the number of expert users who provide good answers has decreased. Additionally, most users in their analysis only asked questions and did not provide any answers. Therefore, it can be questioned if, or for how long, Stack Overflow can provide possibilities for professional development especially at higher levels of expertise.

It can also be argued that instead of professional development or learning, Stack Overflow is, as indicated by the most popular usage types (SODS), mainly used as a reference for quick copy and paste solutions to programming problems. Indeed, Fischer et al. [30] analysed 1.3 million Android applications and found that 15.4% of the applications contained security-related code snippets that very likely originated from Stack Overflow, and 97.9% of these were at least partially insecure. Moreover, based on their analysis of the up- and down-voting patterns of platform users, they conclude “that the currently deployed feedback system is insufficient for providing reliable and precise security estimation to the user” [30, p. 113]. This finding parallels similar concerns over quality problems that have been observed in relation to code snippets on Stack Overflow that are outdated, harmful or violate their original license [31].

In terms of the affordance of access to opportunities for expertise development in programming, the results of the present study highlight other concerns. At least based on SODS, the Stack Overflow community reflects and fails to mitigate the apparent gender disparity in the field of programming. However, this gender disparity is not evenly distributed with the noteworthy feature that gender disparity by country is not related to more general measures of gender equality in a country (GDI). In fact, in many individual cases (e.g. Pakistan and India), the data seems to suggest that the exact opposite is true. This resonates with the results of Stoet and Geary who found that “countries with lower levels of gender equality had relatively more women among STEM graduates than did more gender-equal countries” [32, p. 590], a phenomenon they refer to as a ‘gender-equality paradox’. The authors argue for a range of individual conditions that might account for the issue that girls generally outperform boys in measures of school age STEM attainment, but participate in post-secondary STEM education at a rate of less than 25% in the highest GDI countries. The results from the present study show that women are represented on Stack Overflow at a much lower rate of 5-7% despite the potential for anonymity on the platform that might otherwise be seen as a way of mitigating certain structural social factors inhibiting participation. The underrepresentation of women is perhaps even more striking if one considers that women who have a user account have also been found to be less likely than men to pose and answer questions [18].

The SODS data shows that Stack Overflow follows the pattern of a higher proportion of woman participants in lower GDI countries, but at a much lower rate than that of formal STEM postsecondary education. This result is in stark contrast to the gender distribution in some of the largest technology companies with women holding about twenty percent of the technical roles at Google, Facebook and Microsoft [33]. One might expect that, in comparison to a potentially anonymous online platform, the structural social factors that cause gender disparity would be more pronounced in professional life. On the other hand, that gender disparity is lower in major technology companies shows that active diversity policies have an effect that could also be argued for in relation to other issues of diversity. This also suggests that rather than fostering greater access for women to develop programming expertise, the conditions created on Stack Overflow may actually contribute to exacerbating an already unequal situation. However, it should be noted that the issue of diversity has been acknowledged by Stack Overflow owners, Stack Exchange, who report that they are actively working to make the platform more inclusive in terms of programming experience, race, gender, sexual orientation and religion [34].

Turning to another aspect of access to the expertise development opportunities on Stack Overflow, the SODS data indicates that, especially in developing countries with low GDI rankings, one of the aspects that affects participation on the platform is formal education. This raises interesting questions about possible barriers that regulate a person’s ability to make use of informal learning platforms. Data from SODS indicates that almost half of community members do not consider their formal education to be important for their career development, suggesting that programming might be an accessible career in which people without formal higher education can succeed.

At the same time, much of the policy discourse about online informal learning platforms, such as that from the World Bank [1], has championed the opportunities they provide for greater access to knowledge and learning for those without formal higher education. However, despite the potential of informal avenues for learning programming and the optimism of policy makers, the results of this study indicate that in the case of Stack Overflow, this promise is not greatly realised, especially in developing countries.

One possible explanation for the disparities in participation between countries that is argued by Oliveira, Andrade and Reinecke [14] is the requirement for a high level of English language competency. Indeed, based on the Education First English Proficiency Index (EF EPI) 2019 [35], fifteen out of the twenty countries that were included in the current analysis have, on average, a medium to native level of English competency. However, while the argument may hold for overall participation levels, the educational systems and social conditions of each country are different, as evidenced by the large difference in participation profiles for France and Italy (EFI EPI rankings 31 and 36, respectively). What is clear, though, is that in all examined countries the large majority of the participants have a higher education degree. In this sense, rather than being a place for those who have not had access to formal higher education to develop expertise, such an education can be seen to be a condition of participation for most.

While there are trends that can be identified from the SODS data, it only reaches a small part of the Stack Overflow user base. Therefore SODS, along with the current study, has some considerable limitations. However, at least in terms of gender balance, the survey seems to be in agreement with other methodological approaches [18] and the survey data have yielded consistent results over the years, indicating sufficient reliability. Furthermore, the survey provides information (education, preferences for platform usage, etc.) that is not available or easily obtained from, for example, user profiles. The survey also captures visitors who do not have an active Stack Overflow user account. For example, in 2018, the percentage of survey respondents without an active user account was 13%. While their engagement with and contribution to Stack Overflow is obviously limited, it is yet conceivable that the community may still be an important resource to them. Whether this kind of limited engagement has potential to translate to expertise development or professional opportunities remains an open question.

Furthermore, the sample size for the more detailed analyses presented here was rather small and, therefore, the conclusions of these analyses should be considered to be tentative. Yet, the results highlight the importance of studying local conditions and factors that might foster or hinder participation in informal online learning. Therefore, in addition to large scale quantitative analyses of the SODS, user accounts and behaviours, future research should also utilize case-based analyses of participant's profiles and platform activity in different countries. In addition, the platform should be studied in more detail with regard to different learning trajectories and how the platform's "interactive process of ongoing participation in various cultural practices and in shared learning activities" [6, p. 176] contributes to them.

In conclusion, while Stack Overflow, with its massive user base and volume of programming related knowledge, offers opportunities for expertise development, there are still considerable barriers to participation. Many of these barriers are not necessarily related to the platform itself. Instead, they are the result of different local and individual conditions. However, past research has shown that some of the barriers are indeed related to platform mechanics and how different user groups perceive them. Our results confirm these findings and shed light on various factors that may contribute to these barriers. Moreover, some platform mechanics may not necessarily be conducive to supporting expertise development. As this study highlights, any discussions about the potential of informal learning in programming, or in any domain, should not mistake mere access in the technical sense for opportunity. Instead, the afore-mentioned local and individual barriers, as well as factors that foster participation and learning, must be taken into account.

REFERENCES

- [1] World Economic Forum, *New Vision for Education: Unlocking the Potential for Technology*, Geneva: World Economic Forum, 2015.
- [2] European Commission, *ICT for Work: Digital skills in the workplace*, Brussels: European Commission, Directorate-General of Communications Networks, Content & Technology, 2014.
- [3] A. Robins, J. Rountree, and N. Rountree, "Learning and teaching programming: A review and discussion", *Computer science education*, 2003, vol. 13, pp. 137–172.
- [4] S. Sonnentag, C. Niessen and J. Volmer, "Expertise in software design", in *The Cambridge handbook of expertise and expert performance*, K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman, Eds. Cambridge: Cambridge University Press, 2006, pp. 373–387.
- [5] V. Popovic and B.J. Kraal, "Expertise in software design: Novice and expert models," *Proceedings of Studying Professional Software Design*, 2010.
- [6] H. Gruber, T. Palonen, M. Rehl and E. Lehtinen, "Understanding the nature of expertise: Individual knowledge, social resources and cultural context," *Turku: Finnish Educational Research Association (FERA)*, 2007, pp. 227–250.
- [7] N.G. Leveson, "Role of software in spacecraft accidents," *Journal of spacecraft and Rockets*, 2004, vol. 41, pp. 564–575.
- [8] J. Moreno-León, G. Robles and M. Román-González, "Examining the relationship between socialization and improved software development skills in the Scratch Code Learning Environment," *J. UCS*, 2016, vol. 22, pp. 1533–1557.
- [9] Stack Overflow. [Online] Available: <https://www.stackoverflowbusiness.com/talent/about>
- [10] A. Halavais, K.H. Kwon, S. Havener and J. Striker, "Badges of friendship: Social influence and badge acquisition on stack overflow," *47th Hawaii international conference on System Sciences*, IEEE, 2014, pp. 1607–1615.
- [11] A. Zagalsky, D.M. German, M.A. Storey, C.G. Teshima and G. Poo-Caamaño, "How the R community creates and curates knowledge: An extended study of stack overflow and mailing lists," *Empirical Software Engineering*, 2018, vol. 23, pp. 953–986.
- [12] N. Oliveira, M. Muller, N. Andrade and K. Reinecke, "The exchange in StackExchange: Divergences between Stack Overflow and its culturally diverse participants," *Proceedings of the ACM on Human-Computer Interaction*, 2018.
- [13] A. May, J. Wachs and A. Hannák, "Gender differences in participation and reward on Stack Overflow," *Empirical Software Engineering*, 2019, pp. 1–23.
- [14] N. Oliveira, N. Andrade, and K. Reinecke, "Participation differences in Q&A sites across countries: Opportunities for cultural adaptation," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*, Gothenburg, Sweden, 2016, pp. 1–10.
- [15] F. Calefato, F. Lanubile, N. Novielli, "How to ask for technical help? Evidence-based guidelines for writing questions on Stack Overflow," *Information and Software Technology*, 2018, vol. 94, pp. 186–207.
- [16] M. Othman and R. Latih, "Women in computer science: no shortage here!" *Communications of the ACM*, 2006, vol. 49, pp. 111–114.
- [17] Y. L. A. Wong, "Diverse Global Entryways: How Young Men and Women from China, India, and the United States Make the Decision to Enter Computer Science." M.A., University of California, Santa Barbara, 2018.
- [18] B. Vasilescu, A. Capiluppi and A. Serebrenik, "Gender, representation and online participation: A quantitative study of StackOverflow," *2012 International Conference on Social Informatics*, Lausanne, 2012, pp. 332–338.
- [19] J. P. John and M. Carnoy, "The Case of Computer Science Education, Employment, Gender, and Race/Ethnicity in Silicon Valley, 1980–2015." *Journal of Education and Work*, 2019, vol. 32, no. 5, pp. 421–35.
- [20] J. Ehrlinger, E. A. Plant, M. Hartwig, J. Vossen, C. Columb, and L. Brewer, "Do Gender Differences in Perceived Prototypical Computer Scientists and Engineers Contribute to Gender Gaps in Computer Science and Engineering?," *Sex Roles*, 2018, vol. 78, no. 1, pp. 40–51.
- [21] T. Ahmed and A. Srivastava, "Understanding and evaluating the behavior of technical users. A study of developer interaction at StackOverflow," *Hum. Cent. Comput. Inf. Sci.*, vol. 7, no. 1, p. 8, Dec. 2017.
- [22] D. Ford, A. Harkins, and C. Pamin, "Someone like me: How does peer parity influence participation of women on stack overflow?," in *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Raleigh, NC, 2017, pp. 239–243.
- [23] D. Ford, J. Smith, P. J. Guo, and C. Pamin, "Paradise unplugged: identifying barriers for female participation on stack overflow," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering - FSE 2016*, Seattle, WA, USA, 2016, pp. 846–857.

- [24] Stack Overflow Annual Developer Survey. [Online]. Available: <https://insights.stackoverflow.com/survey/>
- [25] United Nations Human Development Index. [Online]. Available: <http://hdr.undp.org/en/content/human-development-index-hdi>
- [26] United Nations Gender Development Index. [Online]. Available: <http://hdr.undp.org/en/content/gender-development-index-gdi>
- [27] UNESCO Institute for Statistics, Learning Outcomes. [Online]. Available: <http://uis.unesco.org/en/topic/learning-outcomes>
- [28] World Economic Forum, The global gender gap report 2018. [Online]. Available: <https://www.weforum.org/reports/the-global-gender-gap-report-2018>
- [29] I. Srba and M. Bielikova, "Why is Stack Overflow failing? Preserving sustainability in community question answering," *IEEE Software*, vol. 33, no. 4, pp. 80–89, Jul. 2016.
- [30] F. Fischer et al., "Stack Overflow considered harmful? The impact of copy paste on Android application security," in 2017 IEEE Symposium on Security and Privacy (SP), 2017, pp. 121–136.
- [31] C. Ragkhitwetsagul, J. Krinke, M. Paixao, G. Bianco, and R. Oliveto, "Toxic code snippets on Stack Overflow," *IEEE Transactions on Software Engineering*, pp. 1–1, 2019.
- [32] G. Stoet and D.C. Geary, "The gender-equality paradox in science, technology, engineering, and mathematics education," *Psychological science*, 2018, vol. 29, no. 4, pp. 581–593.
- [33] K. Silveira, S. Musse, I. Manssour, R. Vieira, and R. Prikladnicki, "Reinforcing diversity company policies: Insights from StackOverflow Developers Survey," in *Proceedings of the 21st International Conference on Enterprise Information Systems*, Heraklion, Crete, Greece, 2019, pp. 119–129.
- [34] Stack Overflow. [Online]. Available: <https://stackoverflow.blog/2019/10/10/iterating-on-inclusion/>
- [35] Education First, English Proficiency Index. [Online]. Available: <https://www.ef.com/wwen/epi/>