How Developers Acquire FLOSS Skills

Ann Barcomb, Michael Grottke, Jan-Philipp Stauffert, Dirk Riehle, and Sabrina Jahn

Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
{ann.barcomb,michael.grottke,jan-philipp.stauffert}@fau.de,
dirk@riehle.org, sabrina.jahn@fau.de
WWW home page: http://www.fau.de

Abstract. With the increasing prominence of open collaboration as found in free/libre/open source software projects and other joint production communities, potential participants need to acquire skills. How these skills are learned has received little research attention. This article presents a large-scale survey (5,309 valid responses) in which users and developers of the beta release of a popular file download application were asked which learning styles were used to acquire technical and social skills. We find that the extent to which a person acquired the relevant skills through informal methods tends to be higher if the person is a free/libre/open source code contributor, while being a professional software developer does not have this effect. Additionally, younger participants proved more likely to make use of formal methods of learning. These insights will help individuals, commercial companies, educational institutions, governments and open collaborative projects decide how they promote learning.

Keywords: competencies, informal learning, non-formal learning, open source, skills, software developer

1 Introduction

Free/libre and open source software (FLOSS) is important to the economy, with many companies now relying on FLOSS components not only internally but also as the basis for their commercial offerings. Fostering FLOSS talent is critical for companies, which now support an estimated 50% of FLOSS development [1]. Governments have also become increasingly concerned not only with FLOSS adoption but with building capacity for FLOSS development as a means of promoting innovation. It is therefore in the interests of companies, governments and FLOSS communities to know which skills or competencies[1] are necessary for FLOSS development and how they can be trained.

Contributing code to a FLOSS project requires both technical skills and social skills such as the ability to coordinate with others, to clearly articulate an

[1] Some authors distinguish between competencies and skills while others do not. In this paper the terms are used interchangeably.

argument [2], to give constructive feedback and to comply with social rules [3]. Basic information and communications technology (ICT) skills are described as including "the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet" [4].

People can acquire skills through formal, non-formal and informal learning, often making use of multiple methods in acquiring a single skill [5]. Formal learning follows a course structure and results in certification, non-formal learning is structured but is not formally certified, while informal learning is neither structured nor formally recognized.

Although the skills necessary for FLOSS development have been identified, it is unclear which learning methods are being used to acquire them, especially by people who are not involved in FLOSS development. Our research shows that being a FLOSS code contributor and being a professional software developer have different effects on the learning methods used, and that age is an important factor in how FLOSS skills are acquired. We make use of an internet survey of users and developers of the beta release of a popular file download application to statistically validate the learning methods employed in the acquisition of FLOSS skills by group.

This paper makes three contributions. We confirm that the fact that someone is a FLOSS code contributor tends to make it more likely for the person to have used informal learning to acquire FLOSS skills, as discovered by Ghosh et al. [2]. We determine that being a FLOSS code contributor and being a professional software developer have different effects on the learning methods employed. Finally, we demonstrate that age affects the learning methods used.

2 Related Work

We identified three topics related to our research: the learning style preferences of individuals and groups, FLOSS and ICT skills acquisition, and comparisons between FLOSS code contributors and professional software developers.

Studies have examined the relationships between learning preferences and many other factors, such as hemisphericity ('right'- or 'left'-brain), social preferences, chronobiology and culture [6]. Age as an explanatory factor has recently attracted attention due to popular press claims that immersion in technology has created a generation of 'digital natives' with a uniquely self-directed and interactive learning style. Little support has been found for the premise in the literature [7]. A preference for informal learning was found in young students [8], but a preference for the informal techniques was observed irrespective of age in students aged from under 20 to over 30 [7]. In a study of Canadian adults, older people were found to make greater use of independent learning [9].

FLOSS code contributors strongly prefer informal methods such as reading source code and weakly prefer non-formal methods such as participating in workshops over formal study [2]. Contributors already possess many FLOSS skills before joining the community [10], but younger cohorts do improve skills significantly within the community [11]. In contrast with these studies, our research looks at FLOSS code contributors in relation to others.

The learning styles being used to acquire technical skills, regardless of preference, are largely informal and non-formal. One of the largest surveys of skills acquisition indicates that the majority of Europeans acquired their ICT skills informally, through learning by doing, informal assistance from others, self-study materials, or through non-formal courses [4]. In a survey of Canadian adults, computer skills related to employment were among the skills most frequently acquired informally [9]. These studies focus on how skills are acquired, but they do not examine differences between groups. Our research shows that the same skills may be acquired differently by different groups of people.

FLOSS code contributors are often professional software developers or otherwise employed in the ICT sector [12]. They often already possess technical skills before joining a FLOSS project [13]. However, there are some demographic differences between FLOSS code contributors and professional software developers: FLOSS code contributors are more likely to be male [12] and come from North America or north-western Europe [14]. In terms of motivations, most FLOSS volunteers do not differ significantly from paid FLOSS code contributors [15], with most being motivated by need rather than altruism [16]. In FLOSS projects, a small percentage of people contribute most of the code [17], just as in other joint production communities. Top contributors in such projects participate in fundamentally different ways than others [18], and have different motivations [16]. Our research compares the effects of being a FLOSS code contributor and being a professional software developer and establishes that there are differences in learning methods used in the acquisition of FLOSS skills.

3 Theory Development and Hypotheses

We consider three methods of learning—formal, non-formal and informal—which differ on two key attributes: whether it is structured and if it includes certification. Formal learning "refers to the education received from a recognized education center that leads to a certification" [5]. "Non-formal learning is provided by any organised, structured and sustained educational activity... but typically does not lead to certification" [4]. "Informal learning is undertaken on one's own, either individually or collectively, without either externally imposed criteria or the presence of an institutionally authorized instructor" [9].

FLOSS code contributors expressed a strong preference for informal learning and a weaker preference for non-formal learning compared to the formal method of learning [2]. In this, they differ from respondents of another large survey (of government employees) [19]. Because the effectiveness of a learning method depends on how well it matches a person's learning style [6], we expect that FLOSS code contributors will make less use of formal learning than other respondents.

Hypothesis 1_a: Being a FLOSS code contributor makes it more likely that FLOSS skills have been acquired via informal learning methods.

Hypothesis 1_b: Being a FLOSS code contributor makes it more likely that FLOSS skills have been acquired via non-formal learning methods.

It has been demonstrated that there are differences between prolific contributors and ordinary contributors in open collaborative projects [16, 18] but there is no indication that the average FLOSS code contributor differs from professional software developers. Indeed, many FLOSS code contributors work in the ICT sector [12]. Therefore we anticipate that being a professional software developer exhibits the same effects on the learning methods used for skills acquisition as being a FLOSS code contributor.

Hypothesis 2_a: Being a professional software developer makes it more likely that FLOSS skills have been acquired via informal learning methods.

Hypothesis 2_b: Being a professional software developer makes it more likely that FLOSS skills have been acquired via non-formal learning methods.

The majority of studies which examined the relationship between age and a preference for learning methods found no indication that age affects preferences [7, 20]. While a preference for a particular method does not require that a person acquire a skill using that method, informal methods of acquiring technical skills are readily accessible and we expect that people will make use of their preferred methods of learning when the opportunity exists. We do not expect that age will affect the learning methods used in the acquisition of FLOSS skills.

Hypothesis 3_a: Age does not influence the extent to which FLOSS skills have been acquired via informal learning methods.

Hypothesis 3_b: Age does not influence the extent to which FLOSS skills have been acquired via non-formal learning methods.

4 Data Sources and Research Method

4.1 Data Sources

The primary data source for this paper is an online questionnaire conducted between December 2013 and January 2014 [21]. The survey was distributed to users of JDownloader 2 Beta via a link in the client interface. JDownloader[2] is an open source download management tool used by about 20 million people. A subset of users run the beta version.

The questionnaire was developed as a split survey with seven parts. In all parts, participants were asked common demographic questions and about their FLOSS participation. The distinct portion in each part related to the acquisition of FLOSS skills identified in prior work [2, 3, 22]. The questions connected with each of the skills examined (see Table 1 in Section 5) were ordered randomly and each appeared in two survey parts. Participants were randomly directed to one of the seven parts, resulting in a different number of responses for each question.

A second data source is the FLOSS 2013 survey [12] on demographics of FLOSS participants, which was conducted in late 2013.

[2] `http://jdownloader.org/`

4.2 Survey Reliability

An estimated 200,000 people use JDownloader 2 Beta, of which a total of 26,853 people started to answer one of the survey parts and 5,878 continued to the end.

We compared the completed responses to incomplete responses, and found that people who completed the survey were more likely to have engaged in software development and to have participated in FLOSS. Although the questionnaire stressed that the survey was intended for a broad audience, the focus on ICT may have discouraged some respondents.

We further eliminated responses where the participant failed to answer follow-up questions or where we suspected age misreporting because the response was outside the expected age range (born 1930–2000) of our population, leaving a total of 5,309 responses for the combined survey.

Internal reliability of the survey was demonstrated by computing Cronbach's alpha for the original versions of the skills questions against the control versions. All results were in the range of 0.6 to 0.9, which is considered acceptable.

4.3 Survey Representation

In our survey, we provided several ways for people to describe their FLOSS participation. We compared our respondents who selected from the five options which had close representations in FLOSS 2013 by gender, age and income. This comparison involved all FLOSS participants, not just FLOSS code contributors.

Using Pearson's Chi-squared test for gender we determined that there were differences, with our FLOSS participants being less likely to be female (1.4% compared to 11.1%).

To determine age in FLOSS 2013, needed for a t-test, we used the year of initial FLOSS participation and age at the time. Interval values with a range were adapted with random numbers from a uniform distribution within the range. Unbounded intervals were adapted as follows: "before 1960" was set to 1960, "10 or younger" was given a distribution from the set $\{8, 9, 10\}$ and "55 or older" followed a distribution from $\{55, \ldots, 65\}$. A Welch two sample t-test showed with 95% confidence that our sample differed. Our sample was younger, as expected from the JDownloader population.

Income was expressed in intervals in both surveys. For a t-test, we converted the observations to values drawn from a uniform distribution within the interval limits and adjusted values to cover the same length of time. The results were consistent with a younger sample: our group had a lower average income.

4.4 Survey Design and Modeling Approach

Learning Style. Participants were asked to gauge their mastery of each skill shown in Table 1, by moving a slider between the extremes of "I am not skilled at all" and "I am very skilled." The maximum value corresponded with 10,000 but the numeric value was hidden from the participant. Participants who indicated some measure of skill were subsequently asked to evaluate to what extent various methods were used to acquire the skill. Five options were presented: 'learning in school, university or apprenticeship' (formal); 'reading a book or online tutorial' (informal); 'observing other people perform the activity or the result of their work' (informal); 'participating in workshops or advanced training courses' (non-formal); and 'learning by doing' (informal). Learning styles were also displayed as unnumbered sliders with an effective range of 0–100 and a visible range of "nothing at all" to "all." Informal learning was favored by all participants for all skills, accounting for 62–80% of learning, while non-formal learning had a range of 9–16% and formal learning from 11–26%.

FLOSS Code Contributors. People who answered positively to the question "Have you ever participated in a FLOSS project?" and subsequently selected one or both of the participation options 'code contributions' and 'project founder' were categorized as FLOSS code contributors. The binary variable `FCC` was used to indicate if a respondent is a FLOSS code contributor. We observed that FLOSS code contributors differed from the rest of our sample by being younger (by 1.5 years) and less likely to be female (1.5% versus 3.4%).

Professional Developers. Professional software developers were classified by their selection of the answer "I work or worked in software development as part of my job" to the question "Have you worked in software development?" Based on this classification we created an indicator variable, `Prof`. It should be noted that all possible combination of values for the variables `FCC` and `Prof` occurred, giving us four different groups. The smallest group size was 196.

Age. Age was operationalized based on the year of birth, variable `YoB`, reported in the questionnaire by the respondents.

Technical Knowledge. We created a control variable, `TechK`, to describe an individual's technical knowledge. It contained the sum of the self-estimates of the mastery of technical skills (2, 5, 7, 11, 12 and 13) in Table 1.

Modeling Approach. For each skill, the vector $\boldsymbol{y} = (y_a, y_b, y_c)$ observed for a participant was assumed to follow a Dirichlet distribution with expectation $\boldsymbol{\pi} = (\pi_a, \pi_b, \pi_c)$ and precision ϕ, using the alternative parameterization proposed by Maier [23]. Here, y_a, y_b and y_c represent the observed relative learning acquired through informal, non-formal and formal learning styles, respectively, while π_a, π_b and π_c denote the corresponding expected values. Choosing formal learning as the reference category, the parameters were modeled to depend on the explanatory variables and the control variable as follows:

$$\ln\left(\frac{\pi_j}{\pi_c}\right) = \beta_{0j} + \beta_{1j} \cdot \texttt{FCC} + \beta_{2j} \cdot \texttt{Prof} + \beta_{3j} \cdot \texttt{YoB} + \beta_{4j} \cdot \texttt{TechK}, \quad j \in \{a, b\},$$
$$\ln \phi = \gamma_0 + \gamma_1 \cdot \texttt{FCC} + \gamma_2 \cdot \texttt{Prof} + \gamma_3 \cdot \texttt{YoB} + \gamma_4 \cdot \texttt{TechK}.$$

After estimating this model based on all observations available for a certain skill, we tested the null hypotheses $H_0(i_j) : \beta_{ij} \leq 0$ for $i \in \{1, 2\}$ and $j \in \{a, b\}$ using t-test statistics. If the related p value of such a test statistic was smaller than the chosen significance level α, then the null hypothesis could be rejected, indicating support for the respective alternative hypothesis, which is one of the Hypotheses 1_a–2_b formulated in Section 3. In contrast to this, the simple Hypotheses 3_a and 3_b formed the null hypotheses $H_0(3_j) : \beta_{3j} = 0$ with $j \in \{a, b\}$.

5 Results

When fitting the above Dirichlet regression model for each one of the 17 skills listed in Table 1, only those respondents could be taken into account who met the following conditions: they all had some mastery of the respective skill, they allocated a non-zero value to at least one learning method for acquiring this skill, they gave their year of birth, and they answered the questions necessary for determining their technical knowledge as well their membership in the FLOSS code contributor and professional software developer groups. Table 2 lists the sample sizes N available for the 17 models, as well as the p values p_{ij} obtained when testing the six null hypotheses $H_0(i_j)$, $i \in \{1, 2, 3\}$, $j \in \{a, b\}$. Results significant at a level α of 5% are shown in bold type.

Hypothesis 1_a was supported for skills 1–3, 5, 10–13, and 15. Descriptions of the skills can be found in Table 1. As expected, the fact that a person is a FLOSS code contributor tends to increase his/her use of informal learning methods in the acquisition of some FLOSS skills.

Skill #	Description
1	to evaluate the work of others
2	to work on own software module alone
3	to communicate with many different target groups
4	to understand English, especially technical discussion
5	to document code
6	to clearly articulate an argument
7	to understand different software architectures
8	to show respect for the work of others
9	to follow discussions on mailing lists
10	to communicate without offending others
11	to write code in a way that can be reused
12	basic/introductory programming skills
13	to acquaint yourself with code from others
14	to maintain contact with a community
15	to coordinate own work with the work of others
16	to change criticized behavior
17	to understand and work with people from different cultures

Table 1: Skills (skills shown in gray had control questions)

Skill #	N	p_{1a}	p_{1b}	p_{2a}	p_{2b}	p_{3a}	p_{3b}
1	964	**0.0115**	0.5337	0.6664	0.0945	**0.0200**	**0.0000**
2	758	**0.0019**	0.1479	0.9986	0.9698	**0.0032**	**0.0000**
3	475	**0.0498**	0.8459	0.3033	0.4610	**0.0072**	**0.0122**
4	815	0.3313	0.8751	0.0847	0.0532	**0.0003**	**0.0000**
5	697	**0.0095**	0.2109	0.8809	0.9791	**0.0057**	**0.0000**
6	916	0.1645	0.0588	0.8712	0.4882	**0.0361**	**0.0000**
7	790	0.1832	0.6709	0.9999	1.0000	**0.0000**	**0.0000**
8	343	0.2783	0.2505	0.0640	0.4892	**0.0060**	**0.0043**
9	395	0.1483	0.4034	0.2757	0.1426	**0.0125**	**0.0020**
10	636	**0.0071**	0.2223	0.4565	0.1272	**0.0189**	**0.0000**
11	709	**0.0030**	0.0654	0.7705	0.6665	**0.0011**	**0.0000**
12	1046	**0.0000**	**0.0000**	1.0000	0.9956	**0.0028**	**0.0000**
13	591	**0.0133**	0.1295	0.2402	0.5800	0.1011	**0.0020**
14	350	0.0967	0.2230	0.1388	0.2587	0.1805	0.0641
15	751	**0.0144**	0.1769	0.6464	0.3472	**0.0005**	**0.0000**
16	527	0.3298	0.2965	0.4763	0.5724	**0.0227**	**0.0005**
17	390	0.3763	0.2838	0.3794	0.3056	**0.0158**	0.2393

Table 2: Sample sizes and test results

Hypothesis 1_b was not supported, except for skill 12. Being a FLOSS code contributor does not make it more likely that non-formal learning methods have been used to acquire FLOSS skills.

Hypothesis 2_a and **Hypothesis 2_b** were not supported. The fact that a person is a professional software developer does not tend to increase his/her use of informal or non-formal learning methods in the acquisition of FLOSS skills.

Hypothesis 3_a was rejected except for skills 13 and 14. **Hypothesis 3_b** was rejected except for skills 14 and 17. For most FLOSS skills, age influences the use of informal and non-formal methods of learning. More specifically, our results indicated that for all skills where age has an effect, being older is associated with a higher likelihood of having acquired FLOSS skills informally and non-formally.

6 Discussion and Limitations

The learning methods used by FLOSS code contributors were expected, but the fact that not all skills showed this increased tendency toward informal learning suggests that future work is needed to determine why this variation exists.

There are important implications of the finding that in the acquisition of skills necessary for FLOSS development, being a professional software developer has a different effect from being a FLOSS code contributor. It has generally been assumed that the pool of potential FLOSS code contributors consists of all software developers, but our results suggest that there may be fundamental differences between professional software developers who contribute to FLOSS projects and those who do not. Future research should examine the extent of these differences, in order to determine if it is possible to encourage FLOSS participation among software developers or if only a certain type of person—one who makes greater use of independent learning and exploration—is likely to become a FLOSS code contributor.

The effects of age on skills acquisition may reflect the availability of learning, rather than preferences. Previously, there were fewer formal options for acquiring FLOSS skills. This suggests it may not be futile to try to teach FLOSS skills, since they can be acquired through more formal methods. Future research could examine not only the methods by which skills were acquired, but the extent to which the skills were mastered. It should be noted that in our sample the age effect tended to offset the higher preference for informal learning among the FLOSS code contributors, because the FLOSS code contributors were younger.

Our sample was one of convenience optimized for response rate and is not representative of the general population, FLOSS code contributors, or software developers (see Section 4.3). Although this may limit the general applicability of our findings, the results are relevant for young adults, a group which is one of the most important targets for increasing FLOSS skills. Furthermore, as all our respondents came from the same population, our observations about the relative use of informal learning methods by different groups likely remain true.

7 Conclusions

In this paper, we presented a statistical analysis of a survey on the learning methods employed in the acquisition of FLOSS skills. We found that—unlike being a professional software developer—the fact that someone is a FLOSS code contributor tends to make it more likely for the person to have used informal learning methods to master a number of skills. Moreover, age strongly predicted differences in learning methods, with younger people proving more likely to make use of formal learning.

Our results provide some indication of how companies, FLOSS projects and governments can promote the acquisition of FLOSS skills, but also demonstrate the need for further research on how and to what extent FLOSS skills are learned.

Acknowledgments

The authors would like to thank Thomas Rechenmacher and JDownloader for promoting the survey and Nicole Kimmelmann for her contribution to the survey design.

References

1. Riehle, D., Riemer, P., Kolassa, C., Schmidt, M.: Paid vs. volunteer work in open source. In: Proc. 47th Hawaii Int. Conf. System Sciences. pp. 3286–3295 (2014)
2. Ghosh, R.A., Glott, R., Krieger, B., Robles, G.: Free/libre and open source software: Survey and study. Tech. rep., International Institute of Infonomics, University of Maastricht (2002), http://flosspols.org/deliverables.php

3. Kimmelmann, N.: Career in open source? Relevant competencies for successful open source developers/Karriere in Open Source? Relevante Kompetenzen für erfolgreiche Open Source Entwickler. it-Information Technology 55(5), 204–212 (2013)
4. Ala-Mutka, K.: Review of learning in ICT-enabled networks and communities. Tech. Rep. 24061, Institute for Prospective Technological Studies (2009)
5. Galanis, N., Mayol, E., Alier, M., Garcia-Peñalvo, F.J.: A social framework for supporting, evaluating and validating informal learning. In: Proc. 2nd Int. Conf. on Technological Ecosystems for Enhancing Multiculturality. pp. 589–594 (2014)
6. Dunn, R., Beaudry, J.S., Klavas, A.: Survey of research on learning styles. California Journal of Science Education 2(2), 75–98 (2002)
7. Lai, K.W., Hong, K.S.: Technology use and learning characteristics of students in higher education: Do generational differences exist? British Journal of Educational Technology (2014), to appear
8. Hong, K.S., Aziz, N.A.: Technology use and digital learning characteristics among Malaysian undergraduates. Sains Humanika 2(1), 117–124 (2014)
9. Livingstone, D.W.: Exploring the icebergs of adult learning: findings of the first Canadian survey of informal learning practices. (1999)
10. Ghosh, R., Glott, R.: Flosspols: skills survey interim report: 32. MERIT, University of Maastricht, Maastricht (2005)
11. Glott, R., Meiszner, A., Sowe, S.K.: FLOSSCom phase 1 report: Analysis of the informal learning environment of FLOSS communities (2007), http://kn.open.ac.uk/public/getfile.cfm?documentfileid=12042
12. Arjona-Reina, L., Robles, G., Dueas, S.: The FLOSS2013 free/libre/open source survey (2014), http://floss2013.libresoft.es
13. Fang, Y., Neufeld, D.: Understanding sustained participation in open source software projects. Journal of Management Information Systems 25(4), 9–50 (2009)
14. Takhteyev, Y., Hilts, A.: Investigating the geography of open source software through GitHub (2010), http://www.takhteyev.org/papers/Takhteyev-Hilts-2010.pdf
15. Lakhani, K.R., Wolf, R.G.: Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In: J. Feller, B. Fitzgerald, S.H., Lakhani, K.R. (eds.) Perspectives on Free and Open Source Software, pp. 3–22 (2005)
16. Shah, S.K.: Motivation, governance, and the viability of hybrid forms in open source software development. Management Science 52(7), 1000–1014 (2006)
17. Kagdi, H., Hammad, M., Maletic, J.I.: Who can help me with this source code change? In: Proc. IEEE Int. Conf. Software Maintenance. pp. 157–166 (2008)
18. Panciera, K., Halfaker, A., Terveen, L.: Wikipedians are born, not made: a study of power editors on Wikipedia. In: Proc. ACM 2009 Int. Conf. Supporting Group Work. pp. 51–60 (2009)
19. Grundmann, S.T.: Making the Right Connections: Targeting the Best Competencies for Training. DIANE Publishing (2011)
20. Byrne, P., Lyons, G.: The effect of student attributes on success in programming. In: ACM SIGCSE Bulletin. vol. 33, pp. 49–52. ACM (2001)
21. Jahn, S.: Teaching open source competency (2014), http://osr.cs.fau.de/2014/04/02/final-thesis-teaching-open-source-competency/, bachelor thesis
22. Kimmelmann, N.: private communication (2013)
23. Maier, M.: DirichletReg: Dirichlet regression for compositional data in R. Tech. Rep. 125, Vienna University of Economics and Business (2014)