Development of a questionnaire to study motivations of Drupal contributors

Dani Nordin, Adi Meir
Bentley University, Research Methods Class
November 21, 2013

Introduction

The purpose of this pilot is to investigate the motivations of contributors to the Drupal (drupal.org) open-source project. Our goal is to assess differences in motivating factors between those who contribute code to non-code contributors. To inform future research on non-contributors, we also collect motivational data related to choosing not to contribute.

Some co-variables might also have effect on the different motivations of code and non-code contributors, and these are measured in our study. These include gender, place of residence, years of experience, and time devoted to Drupal projects. The effect of these co-variables on the relation between motivation and type of contribution will be further examined in future analysis and research.

The focus on code vs. non-code contributions is unique to our study, as the majority of research on open source contributors focuses on developers only (Krogh, Haefliger, Spaeth, & Wallin, 2012). For example, although David's (2003) study mentions many types of contributions aside from writing and debugging code, the majority of questions are still biased towards developers, e.g. "How important were the following factors when you first started developing open source/free software" (David, Waterman, & Arora, 2003).

The predictive validity of this survey lies with its ability to find out what will motivate contribution to the project. Since the Drupal project survives based on regular contribution of all types of skills, our research can help focus on issues that need to be publicized as benefits of contributing to the community. Since we hypothesize that people who contribute code may have different motivations from those who contribute other skills, we focus on non-code contributors as the comparison group for code contributors.

Note that some information was gathered from non-contributors also. The questionnaire that non-contributors saw was different in many respects from the questionnaire that contributors saw. Data was collected from non-contributors solely for the purpose of another study. It is not analyzed or compared to data collected from contributors in any way.

Sample

The questionnaire was promoted via various social media channels related to the Drupal community: 1) Nordin's Twitter feed (1,493 followers, approximately 20% in Drupal community, est. $n = 300$); 2) Nordin's Facebook stream (332 friends, approximately 20% in Drupal community, est. $n = 67$); 3) the Women in Drupal Facebook group ($n = 256$); 4) the Design 4 Drupal Facebook page ($n = 221$); 5) Nordin's Google Plus profile ($n = 30$). This approach yielded an estimated reach of 844 potential individuals, but included significant overlap among the networks. Respondents also shared the link to the survey several times via social media. Hence, the exact reach is impossible to quantify.

Responses were voluntary and offered no reward. Forty-eight hours of data collection yielded 127 responses. Twenty-six (20.5%) partially complete questionnaires and one response (0.8%) from a person who was not working with Drupal were not included in the analysis, resulting in a sample size of 100. Contributors to Drupal, who are the subject of our study, were 78% of this sample ($n=78$).

Our decision to conduct the pilot via social media was the result of our aim to reach as many respondents as possible under time pressure. Nordin's experience as a known voice in the Drupal community gave us

unique access to a significant percentage of actual Drupal users in a short period of time. The ability of respondents to share the survey via retweets expanded this reach.

Despite an impressive reach for a pilot, this was a convenience sample, and it possesses the drawbacks of a convenience sample. Its reach is limited to a circle within the Drupal community, and it is not random. On top of that, participants within the circle of reach were self-selected, since the survey was published and they chose whether or not to participate. Hence, it is not representative of the population of Drupal contributors. This introduces a possible confounding factor which we cannot measure. Any result we got might have been the result of specific characteristics of people who chose to participate. These respondents are people who have the time and energy, who want to participate, and who are using social networks. This results in low external validity.

As we consider this a pilot survey, we were willing to accept these tradeoffs, and use the data collection to inform a revision of the design for a future survey, which will be distributed via more representative distribution lists.

Question and Variables Design

Assessing participant's relationship to Drupal
Questions 1–3 (see Appendix A) focused on establishing the respondent's relationship to Drupal, both in length of engagement, and the respondent's primary assumed role, e.g. developer, designer, etc. Terminology for role options is specific to the Drupal community, increasing face validity of the questionnaire, and providing information on potential co-variables.

The complexity inherent in Drupal "roles" caused much discussion among the researchers. Someone who considers themselves primarily a "designer" may also easily be a front-end developer (called "themer" in Drupal), site builder, or UX designer. This prevented us from sorting respondents cleanly into "developer" and "non-developer" roles. This is illustrative of the complexity of the Drupal community at large; many who work with Drupal find themselves taking on multiple roles. Our ultimate decision to use a single-choice list, which would make analysis easier, also caused problems for respondents; feedback suggested that respondents had a difficult time choosing a single option, since they would often step into multiple roles during the course of a given project. This is a threat to the criterion validity of this item that has implications on internal validity, since we are unaware of the "real" category (code vs. non-code) our participants belong to. Future research will require adjustment of this question.

Another important issue came about in the question of the respondent's gender (question 9). As gender may be a potential confounding variable in the motivation of individuals, it was decided to request this information. However, the issue of heteronormative gender is extremely sensitive in the Drupal community, which includes a significant number of female and GLBT individuals, both in the general community and in the community leadership, e.g Angela Byron (Druckman, 2011).

After much discussion and reflection, it was decided to use the optional question "what is your preferred gender pronoun?" with a series of options that included the option to select a unique answer or to "prefer not to answer" (see Appendix A). We also included questions for the respondent's age and country of residence, as those were deemed potentially relevant to motivation and/or attitude.

Independent Variable: Type of Contribution
Type of contribution to Drupal was measured using question 4 (see appendix), in which the respondents were asked to select which contributions they have made. These options were adapted from the FLOSS-US study of open-source developers (David et al., 2003), and customized to be more specific to the terminology and activities of the Drupal community. This increased the construct validity of our factors, and corrected for a bias towards developers inherent in the FLOSS study.

Prior to analysis, the question was recoded into a dichotomous variable, receiving a value of 1 if any code contribution was made (writing and maintaining modules, writing patches or core development), a value of 2 if non-code contributions (documentation, UX/UI Design or other) were made, and a missing value if no contribution was made.

Dependent Variables: Motivation Factors

Potential motivations were assessed using questions 5a and 5b (see appendix). Question 5b is in fact a set of 11 factors; users were asked to rate each factor on its importance to his or her Drupal contribution. Importance was rated on a scale of 1 to 100, using a slider (see Figure 1). The same format, with different statements, was used in question 4b (shown in Figure 1), which assessed motivations not to contribute.

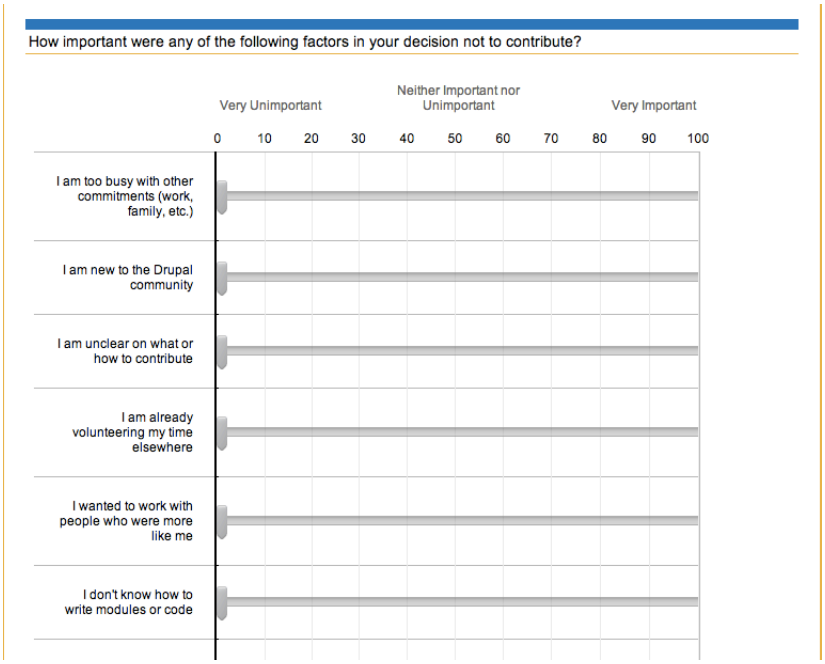


Figure 1. Questions about motivation factors were assessed using a slider.

Question 5d was devoted to assessing economic motivation, and asked participants if they earn money as a result of their contributions. Prior to analysis, answers were coded as follows: 0 - does not earn money, 1 - directly earns money (paid for the contribution), 2 - indirectly earns money ("got a job because of my experience contributing to Drupal" or "got a job because of contacts within the Drupal community").

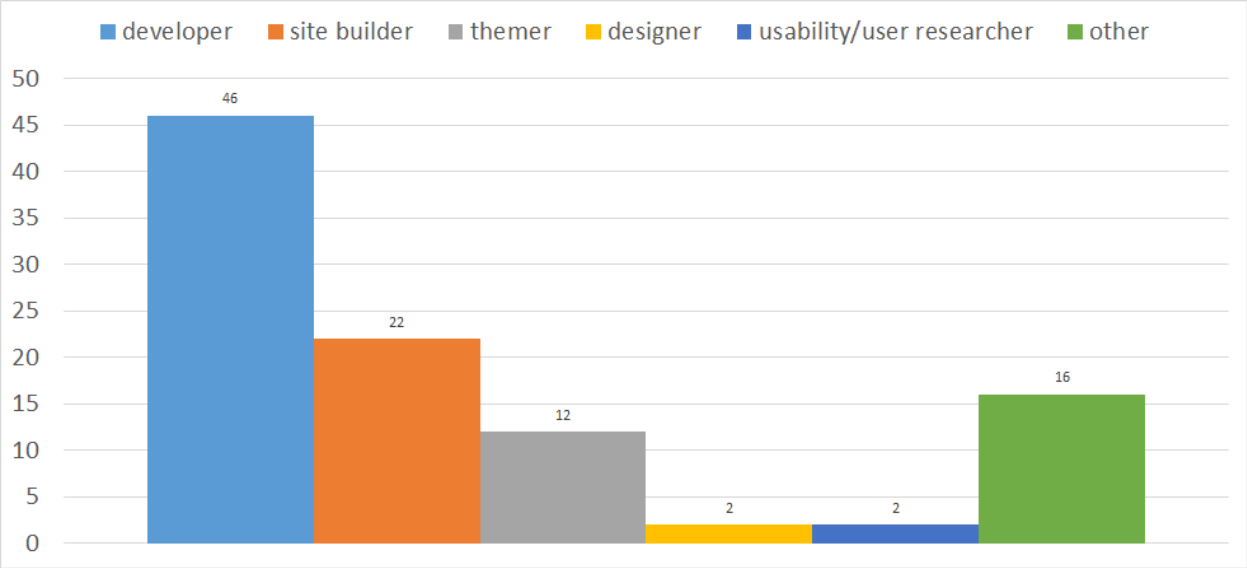
Questions 5a and 5b, shown only to users who had answered "Yes" to being contributors, were both adapted from the FLOSS-US study (David et al., 2003), and rewritten to contain terminology specific to Drupal. Questions were adapted overall to give equal weight to code vs. non-code contributions.

Results

All 100 valid questionnaires were analyzed. The distribution of occupation can be seen in graph 1. Almost half of the respondents (46%) were developers, 22% site builders, 12% themers, 16% "others" and a few were usability/user researcher (2%) and designers (2%). Respondents were 35% female (*n=35*), 51% male (*n=51*), with 3 respondents choosing one of the other gender pronouns. 9 respondents chose "prefer not to answer" for the gender pronoun. The majority of respondents were 25-40 years old (*n=56*), with the bulk of other responses (*n=39*) over 41. 4 respondents were 18-24 years old. The majority of respondents

(64%, *n=63*) were from the United States, with the second most common country being the United Kingdom (12%, *n=12*).

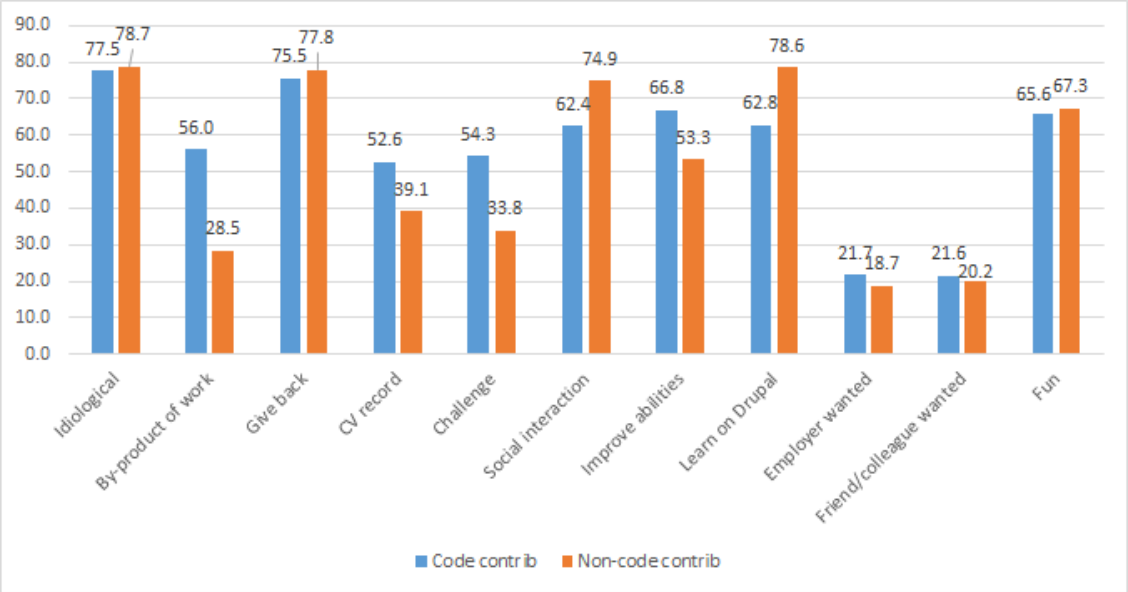
Graph 1. Primary Drupal Role



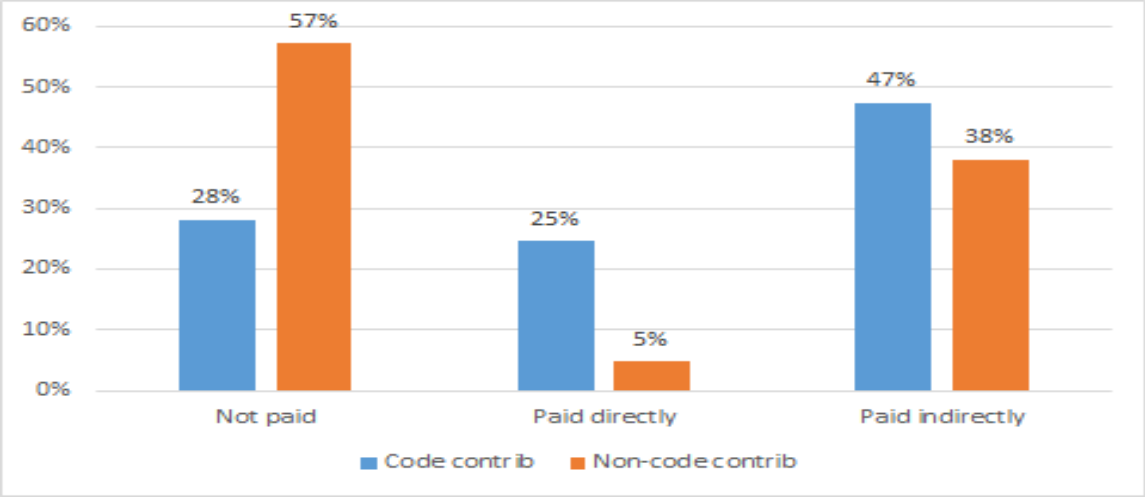
Out of the total (*n=100*), 57 reported contributing code to the Drupal project, 21 reported non-code contributions, and 22 reported not contributing. Focusing on contributors (*n=78*), motivations for contributions were compared between the two types of contributions. The three most important motivators for coders were open-source ideology, the desire to give back to the community and the desire to improve their abilities as programmers. Interestingly, for non-code contributors, the three most important were open-source ideology, the desire to learn more about Drupal and the desire to give back to the community. Hence, open-source ideology and the desire to give back to the community seem to be about equally motivating for both groups; learning more about Drupal and improving one's abilities also seems to be important to both, although there is some difference between the two groups (See Graph 2).

Indeed, motivations were generally similarly rated among code and non-code contributors, on all measures except "by-product of work" and "challenge." On these two motivations, coders rated both points higher than non-coders. This suggests that coders are more likely to contribute because their contribution is related to work they are already doing, or because they believe that contributing to Drupal gives them more control over the development challenges they are able to take on.

Graph 2: Mean score of importance of each motivation, code vs. non-code contributors



Graph 3: Economic Motivations, code vs. non-code contribution.



The payment motivation, examined in a separate question and analyzed in graph 3, shows a greater difference. While most non-code contributors reported contributing without payment (57%), most who contribute code reported being paid for their contribution, either directly or indirectly. The indirect payment is prevalent in both groups, but the direct payment is prevalent only for code contributors (25%) and barely exists for non-code contributors (5%). These results are interesting, but should be further analyzed using more rigorous statistical analyses.

Discussion

Contrary to expectations, our analysis found that most types of motivations being assessed affected both code and non-code contributors to a similar degree. The motivating factors of the contribution being a part of some other project, handling interesting challenges, and being paid to contribute were the only motivating factors that were different and higher for code vs. non-code contributors.

Validity & Sensitivity

Since our questions were modeled after the FLOSS study, and many of the core concepts are similar to those used in the FLOSS study, it is possible to compare results against what David (2003) found. If the characteristics are found to be similar, than our results might also be generalized to all open source projects. Thus, these results might be valid for other communities. At the same time, it should be acknowledged that the Drupal community has its own characteristics and terminology.

The survey format naturally has low ecological and external validity because, rather than observing experience directly, we ask respondents to reflect back on their experiences and report that data. However, we have tried to increase the ecological validity by using terminology common to the Drupal community. That said, the question regarding their motivations presents a construct validity challenge, since it asks them to reflect back on motivations for their initial contribution. This question will be corrected in future research, to reflect more recent experiences.

We were also concerned that the different nature of the two types of contributions might mean that respondents' answers would relate to fundamentally different concepts (e.g. writing code vs. writing documentation or providing user support). This was a potential threat to the internal validity, especially since the construct might not be valid for both populations. We dealt with this by focusing on motivational factors that are known to be common in open-source projects, such as open-source ideology and the ability to work on exciting challenges (for a review, see Krogh et al., 2012). Interestingly, our results showed that many of the factors that are most common in the motivations of open-source developers (e.g. open-source ideology, the desire to enhance current skills) were equally relevant to non-code contributors.

Using a well-established tool, as well as elaborating all types of motivations that are known to be relevant to Drupal contributors, increased both the construct validity of the questionnaire and the concept validity of the questions. Allowing a large scale for rating the importance of each motivation created a potentially highly sensitive measure. In practice, however, it presented problems for interpretation during analysis, and it is unclear whether respondents understood it all in the same way, and that is a threat to the construct validity.

Another area for improvement was in the questions related to economic motivation. It was reported by a few respondents that these questions seemed to be focused more for those who work as W2 wage holders at a private company. This made it difficult for some to answer the questions accurately. As the Drupal community includes a large volume of independent workers, future research using this instrument should be careful to take this into account. Hence, our results regarding income as a motivating factor have lower criterion validity than we intended.

On a positive note, the questions related to types of contributions (5a), non-economic motivations for contributing to Drupal (5b) and the question related to motivations for not contributing (4c) appear to have good criterion validity, and no problems choosing items were reported.

Opportunities for Future Research

This study provided an interesting first look at what motivates users to contribute to Drupal. Additionally, the open-ended question collected from non-contributors (Question 3b, see Appendix A) gave us some interesting insights on non-contributors that will contribute to a second survey. Future research should consider breaking the survey up into multiple questionnaires, which can be delivered to different populations, to increase the response rate and more tightly focus the data collection process.

Feedback on the questionnaire design from the pilot will be incorporated into a revised set of questionnaires, which will be distributed via the Design 4 Drupal mailing list (n = approx. 1500), of

which Nordin is a co-maintainer. We will also be reaching out to colleagues in the Drupal Association (https://association.drupal.org/), which has a wide reach in the global Drupal community.

Appendix A: Survey Questions

The survey was delivered online via Qualtrics software. Some questions were shwn to users only if a certain answer was selected; these are noted below. Note that the last option of question #1, "I do not work with Drupal," was used as a screener to ensure that all respondents were current Drupal users. If selected, the respondent was taken to a message that the survey was only open to current Drupal users, and ended the survey.

1. How long have you been using Drupal?

a) Less than 1 year
b) 1-3 years
c) 4-6 years
d) more than 6 years
e) I do not work with Drupal

2. What would you consider your primary role in your work with Drupal?

a) developer
b) site builder
c) themer
d) designer
e) usability testing and user research
f) other (project management, documentation, etc)

3. Have you ever contributed code or other volunteer work to the Drupal project?

a) Yes
b) No

If No to Question 3

4a. Have you ever considered contributing to Drupal?

a) Yes
b) No

4b. If yes, please say, in your own words, what has held you back from collaborating on Drupal.

a) Free text entry field

4c. How important were the following factors in your decision not to contribute?

a) I am too busy with other commitments (work, family, etc.)
b) I am new to the Drupal community
c) I am unclear on what or how to contribute
d) I am already volunteering my time elsewhere
e) I wanted to work with people who were more like me
f) I don't know how to write modules or code
g) It is too difficult to collaborate on my type of work in the Drupal community (please elaborate)

If Yes to Question 3

5a. What types of contributions have you made to Drupal since you started contributing?
(check all that apply)

a) writing and maintaining modules
b) writing patches
c) core development
d) reporting bugs and feature requests to the issue queue
e) providing support via IRC or the issue queue
f) UX/UI design
g) running Drupal events and meetups
h) documentation

5b. How important were the following factors when you first started contributing to Drupal?

a) I thought open source was the best way for software to be developed
b) I needed to write or update a module for a particular project
c) I wanted to give something back to the community
d) I thought it would attract employment opportunities
e) I liked the challenge of contributing to a large software project
f) I wanted to interact with like-minded web developers and designers
g) I saw it as a way to become a better developer
h) I wanted to learn more about how Drupal worked
i) My employer wanted me to contribute to Drupal
j) A friend/colleague wanted me to contribute to Drupal
k) I have fun collaborating on Drupal

5c. Do you earn money from contributing to Drupal, either directly or indirectly?

a) No
b) Yes, directly: I am paid to work on core/contrib development on the job
c) Yes, directly: I am paid to work on other contributions (user testing, documentation, etc.) on the job
d) Yes, indirectly: I got my job because of my experience contributing to Drupal
e) Yes, indirectly: I got my job because of contacts within the Drupal community

5d. What is the **typical number of hours** per week you spend on work related to Drupal contribution?

a) 0-10 hours
b) 11-20 hours
c) 21-30 hours
d) 31-40 hours
e) more than 40 hours

5e. When do you work on Drupal contrib? (check all that apply)

a) Before I go to work
b) At work, during work hours
c) At work, but off work hours
d) After work
e) On weekends
f) I'm unemployed right now, so whenever I want

Shown to all users: Optional questions

6. What is your preferred gender pronoun? This data is used for demographic purposes only. (pronoun source: http://www.gsafewi.org/wp-content/uploads/What-the-heck-is-a-PGP1.pdf)

a) She/her
b) He/him
c) Ze/hir/zir
d) E/Ey/Em
e) Per
f) Hu/Hum
g) Something Else
h) I prefer not to answer

8. What age were you on your last birthday? This data is used for demographic purposes only.

a) Under 18
b) 18-24
c) 25-40
d) 41-55
e) Over 56
f) I prefer not to answer

7. In which country do you reside?

a) Choose a country

References

danigrrl, T. U. (2013, November 8). Twitter / danigrrl: If I want to be inclusive on ... *twitter.com*. Retrieved November 10, 2013, from https://twitter.com/danigrrl/status/398991322624696320

David, P. A., Waterman, A., & Arora, S. (2003). FLOSS-US the free/libre/open source software survey for 2003. *Unpublished (online) available at URL: http://www. stanford. edu/group/floss-us/report/FLOSS-US-Report. pdf*.

Druckman, K. (2011, February 24). Angela Byron on Drupal 7 | Linux Journal. *Linux Journal*. Retrieved November 10, 2013, from http://www.linuxjournal.com/content/interview-angie-byron-drupal

Krogh, von, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. *MIS Quarterly*, *36*(2), 649–676.