UNDERSTANDING CONTRIBUTOR BEHAVIOUR WITHIN
FREE/LIBRE/OPEN SOURCE SOFTWARE COMMUNITIES: A
SOCIALIZATION PERSPECTIVE

KÉVIN DANIEL ANDRÉ CARILLO

A THESIS

submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy in

Information Systems

Victoria University of Wellington
2014

This PhD thesis is licensed under a **Creative Commons Attribution-Noncommercial-Share Alike 3.0 License**.

ABSTRACT

Attracting a large number of new contributors has been seen as a way to ensure the survival, long-term success, and sustainability of Free/Libre/Open Source Software (FLOSS) communities. However, this appears to be a necessary, but not a sufficient condition, as the well-being of FLOSS communities also relies on members performing behaviours that nurture and protect the community.

Despite a large body of research on FLOSS communities, few studies have been undertaken to explore the influence of a participant's socialization experience on their contribution behaviour. In addition, there has been relatively little research that has adopted a community-level view of FLOSS community participants' contribution that goes beyond the mere notion of writing lines of code.

The purpose of this study is to develop and rigorously test a socialization model that predicts contributor behaviour in the FLOSS community context.

Drawing upon theories of socialization and citizenship behaviours from organizational behaviour research, this research develops and tests two separate but related research models. The first model proposes the direct impact of socialization factors on two performance-related dependent variables: task performance and community citizenship behaviours. The second model hypothesizes a mediating effect of two proximal socialization variables, social identification and social integration, between the socialization factors and the dependent variables.

An exploratory study involving eleven FLOSS community leaders, managers, and experienced members was first conducted, to investigate the key variables that characterize FLOSS community newcomer socialization experience as well as the various instances of citizenship behaviours that are specific to the FLOSS community context. The analysis of the interview data revealed the existence of six socialization variables: task segregation, task purposefulness, interaction intensity, mentoring, joining structuredness, and supportiveness. Two sets of FLOSS community citizenship behaviours (CCB) were identified drawing on the citizenship behaviour literature. The first set, labelled CCB-I, comprised citizenship behaviours directed towards the benefit of individuals. The second set, CCB-P, included citizenship behaviours directed towards the benefit of the project. The findings were integrated in the two conceptual models.

Subsequently, a research instrument was developed, following an extensive purification process that consisted of card sorting and expert review rounds, and a survey pretest. A pilot study assessed responses from 46 FLOSS contributors from two large FLOSS communities. Overall, the scales demonstrated high reliability and showed adequate construct validity. The analysis of the pilot study suggested the existence of a third CCB dimension, named CCB-C, that characterizes citizenship behaviours that are oriented towards the benefit of a project's community.

The main study was based on an online survey involving 327 respondents from twelve large FLOSS communities. Using Partial Least Squares (PLS), the collected data was used to test the two models. The results showed the overall superior predictive capability of the model hypothesizing the mediating effect of both social identification and social integration.

Task performance was found to be directly predicted by task purposefulness as well as by interaction intensity and supportiveness through the mediation of social identification. Meanwhile, CCB was found to be impacted by the direct effect of task segregation and task purposefulness, and by interaction intensity and supportiveness through the mediation of both social identification and social integration. The existence of the third CCB dimension, CCB-C, was confirmed.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the many people who helped me during this venture.

To Sid Huff, my primary supervisor, thank you for having been an extraordinary mentor. Your wisdom, experience, and diligence have infused me with an energy and faith in doing 'high-quality' research I can only hope will continue to grow.

To Brenda Chawner, my secondary supervisor, thank you for your constant support, availability, and energy for pushing me to always surpass myself.

My partner Karine and son Arthur - for your love and for having been the sunshine that helped me go through difficult times and complete this journey.

To my friends and colleagues, Eusebio Scornavacca and Jean-Gregoire Bernard - for your support and insights.

To my fellow PhD students and colleagues from the School of Information Management - for their involvement and valuable support.

To the many FLOSS passionates who believed in my research and opened the door of the FLOSS world to me.

To Victoria University of Wellington for having given me the opportunity to embark on this wonderful, life-changing journey.

To my parents, family, friends.

CONTENTS

1	INTRODUCTION	1
1.1	Research problem, questions and research objectives	2
1.1.1	Problem statement	2
1.1.2	Research questions	3
1.2	Research Design	3
1.3	Significance of research	4
1.3.1	Academic value	4
1.3.2	Practitioner value	4
1.4	Delimitations	5
1.5	Thesis outline	5
2	LITERATURE REVIEW AND INITIAL RESEARCH MODELS	7
2.1	Chapter Outline	7
2.2	Theorization of the FLOSS environment under study	7
2.2.1	Commons-based peer production	8
2.2.2	Beyond the bazaar stereotype	9
2.2.3	Theoretical boundaries: Community-organization hybrids	9
2.3	Free/Libre/Open Source Software communities	11
2.3.1	A condensed history of the Free/Libre/Open Source Software movement	11
2.3.2	Free Software vs. Open Source Software	12
2.3.3	From software glitch to the notion of FLOSS community	13
2.3.4	FLOSS Project artifacts	13
2.3.5	Community roles	14
2.3.6	Relationship between core and non-core developers	17
2.3.7	Multiple roles and non-programming/non-technical tasks in FLOSS communities	17
2.3.8	Shift from a FLOSS project-centered to a community-centered focus	18
2.4	Individual contribution behaviour and performance in FLOSS communities	20
2.4.1	Contribution and performance in FLOSS research	20
2.4.2	A reconceptualization of performance and contributor behaviour	23
2.5	Towards the importance of newcomer socialization	28
2.6	Socialization literature and FLOSS socialization	30
2.6.1	Differences between FLOSS socialization and organizational socialization	30
2.6.2	The socialization literature tradition	32
2.6.3	Past FLOSS research about socialization	33
2.6.4	Van Maanen and Schein's socialization tactics	36
2.6.5	Refinements of Van Maanen and Schein's model by Jones (1986)	38
2.6.6	Aspects of Van Maanen and Schein's socialization tactics in large and complex FLOSS projects	39
2.6.7	Overall socialization framework in organizational behaviour	41
2.6.8	Socialization and performance	43
2.7	Summary of literature review and research gaps	48

2.8	Preliminary research models	49
2.8.1	Conceptualization of the main notions	50
2.8.2	Preliminary research models	51
3	RESEARCH DESIGN	53
3.1	Theoretical perspective of this research	53
3.1.1	Ontology	53
3.1.2	Epistemology	54
3.2	Perspectives about methodology selection	55
3.3	Chapter summary	56
4	QUALITATIVE ASSESSMENT OF FLOSS NEWCOMER EXPERIENCE AND CONTRIBUTOR PERFORMANCE	57
4.1	Introduction	57
4.2	Methodological choices	57
4.2.1	Interviews vs. focus groups	57
4.2.2	Recorded semi-structured interviews	58
4.2.3	Sampling	59
4.3	Data analysis procedure	60
4.3.1	Data reduction	60
4.3.2	Data display	63
4.3.3	Conclusion drawing and verification	65
4.4	Interview participants' characteristics	66
4.4.1	Interview procedure	66
4.5	Results	68
4.5.1	Relevance of performance operationalization	68
4.5.2	Findings about FLOSS community citizenship behaviours	69
4.5.3	Interview findings about socialization in FLOSS communities	79
4.6	A FLOSS socialization model	88
4.7	Chapter summary	90
5	CONCEPTUAL MODELS AND RESEARCH HYPOTHESES	91
5.1	Conceptual definitions	91
5.2	Dual investigation and conceptual models	92
5.3	First research model	92
5.4	Second Conceptual Research Model	96
5.4.1	Overall description	96
5.4.2	Social identification	97
5.4.3	Relationship between socialization experience and social identification	99
5.4.4	Social integration	101
5.4.5	Relationship between socialization experience and social integration	103
5.5	Chapter Summary	106
6	METHODOLOGY	107
6.1	Methodological choices	107
6.1.1	Survey method	107
6.1.2	Online survey versus regular survey	108
6.2	1st Phase	109
6.3	2nd Phase	109
6.3.1	Item generation	111

6.3.2	Expert panel	123
6.3.3	Card sorting	131
6.3.4	Survey design	143
6.3.5	Pretest	145
6.3.6	Pilot study	146
6.3.7	Refinements before survey administration	165
6.4	3rd Phase	167
6.4.1	Survey administration and data collection	167
6.4.2	Dataset release under a FLOSS license	172
6.4.3	SEM technique : Partial least squares	173
6.5	Chapter Summary	174
7	RESULTS	175
7.1	Respondent demographics	175
7.2	Dataset characteristics	177
7.3	Validation and measurement of CCB as a second-order construct	179
7.4	Measurement model evaluation	180
7.4.1	Criteria for measurement model evaluation	180
7.4.2	Exploratory Factor Analysis	182
7.4.3	Measurement model 1: Results	186
7.4.4	Measurement model 2: Results	187
7.4.5	About the dimensionality of CCB	189
7.5	Structural model evaluation	189
7.5.1	Criteria for structural model evaluation	189
7.5.2	Structural Model 1: Results	200
7.5.3	Structural Model 2: Results	201
7.6	Subsequent analyses	202
7.6.1	CCB estimation: Repeated items vs. Two-Step	206
7.6.2	Full mediation vs. partial mediation	206
7.7	Chapter Summary	207
8	DISCUSSION AND INTERPRETATION OF THE FINDINGS	209
8.1	Introduction	209
8.2	Answering the research question	209
8.3	FLOSS socialization under the social scientist's theoretical microscope	211
8.3.1	Task segregation	212
8.3.2	Task purposefulness	212
8.3.3	Mentoring	213
8.3.4	Interaction intensity	213
8.3.5	Joining structuredness	213
8.3.6	Supportiveness	213
8.4	It's not only about writing code.	213
8.5	Modelling the socialization phenomenon	215
8.6	Yes, FLOSS socialization matters.	216
8.6.1	Predictors of task performance	216
8.6.2	Predictors of Community Citizenship Behaviours	218
8.6.3	Predictors of social identification	219
8.6.4	Predictors of social integration	220
8.7	Contributions of the Study	220
8.7.1	Theoretical contributions	220
8.7.2	Theoretical conjectures	223

8.7.3	Practical contributions	227
8.8	Limitations	228
8.9	Directions for Future Research	229
8.10	Summary	230

APPENDICES 233

A	HUMAN ETHICS COMMITTEE APPROVALS	235
B	QUALITATIVE DATA ANALYSIS	241
C	JONES' (1986) MEASUREMENT SCALES	249
D	CARD SORTING PROCEDURE	251
E	TERMS OF THE ODC OPEN DATABASE LICENSE (ODBL)	255
F	SURVEY INSTRUMENT USED FOR THE MAIN SURVEY	257

BIBLIOGRAPHY 267

LIST OF FIGURES

Figure 1	Initiation and growth cycle of a FLOSS project (adapted from Al Marzouq, Zheng, Rong & Grover, 2005)	14
Figure 2	Onion-shape structure of FLOSS communities (adapted from Crowston et al., 2005, and Crowston and Howison, 2005)	16
Figure 3	Onion-model of FLOSS community organizational hierarchy (Jensen and Scacci, 2005, 2007)	16
Figure 4	Contributing within the GNOME community	18
Figure 5	Contributing within the KDE community	19
Figure 6	Koha bug reporting guidelines	26
Figure 7	KDE community code of conduct	27
Figure 8	Development tools for the GNOME platform	32
Figure 9	Jones' (1986) classification of socialization tactics	39
Figure 10	Linux kernel newbie community	41
Figure 11	Example of collective socialization in the KernelNewbie community	42
Figure 12	Saks et al. (2007)'s model of socialization tactics and newcomer adjustment	43
Figure 13	Bauer et al. (2007)'s model of newcomer adjustment during organizational socialization	43
Figure 14	First conceptual model	51
Figure 15	Second conceptual model	52
Figure 16	Miles and Huberman's (1994) data analysis components	60
Figure 17	FLOSS socialization experience - Model	90
Figure 18	First conceptual model	94
Figure 19	Second conceptual model	97
Figure 20	Methodology overview	107
Figure 21	Instrument development process	110
Figure 22	Survey invitation in Perl blog post	149
Figure 23	Message on Perl Foundation's website	149
Figure 24	LibreOffice mailing list invitation	151
Figure 25	Survey invitation page in researcher's blog	170
Figure 26	Survey invitation in a project's 'planet'	171
Figure 27	Survey invitation in Linux.com	171
Figure 28	PLS Model building for second-order constructs- Repeated Indicators and Two-Step approaches	180
Figure 29	Structural Model 1	200
Figure 30	Structural Model 2	203
Figure 31	FLOSS socialization framework	212
Figure 32	Performance-related measures	215
Figure 33	Summary of the results	217
Figure 34	<i>Low number of contributors / Low complexity</i> - Conjectured socialization models	224
Figure 35	<i>Low number of contributors / High complexity</i> - Conjectured socialization models	225

Figure 36	<i>High number of contributors / Low complexity</i> - Conjectured socialization models 226
Figure 37	Non-software commons-based peer production projects - Conjectured socialization models 227

LIST OF TABLES

Table 1	Van Maanen and Schein's socialization dimensions (1979) 37
Table 2	Open coding - Coding examples 62
Table 3	Axial coding - Coding examples 64
Table 4	Interviewees' characteristics 67
Table 5	Interview findings - Instances of CCB-I behaviours 70
Table 6	Interview findings - Instances of CCB-P behaviours 74
Table 7	From <i>formal tactics</i> to <i>task segregation</i> : definitions 81
Table 8	From <i>collective tactics</i> to <i>interaction intensity</i> : definitions 82
Table 9	<i>Task purposefulness</i> and <i>joining structuredness</i> : definitions 86
Table 10	From <i>serial tactics</i> to <i>mentoring</i> : definitions 87
Table 11	From <i>investiture tactics</i> to <i>supportiveness</i> : definitions 89
Table 12	Summary of conceptual definitions 93
Table 13	Pool of items before expert panel and card sorting procedure -1 113
Table 14	Pool of items before expert panel and card sorting procedure -2 114
Table 15	Pool of items before expert panel and card sorting procedure -3 115
Table 16	Task performance measurement items 117
Table 17	Pool of CCB-I items before expert panel and card sorting procedure -1 120
Table 18	Pool of CCB-P items before expert panel and card sorting procedure 121
Table 19	Organizational identification scale 122
Table 20	Social integration scale 123
Table 21	Refined <i>task segregation</i> items - Expert panel 125
Table 22	Refined <i>interaction intensity</i> items - Expert panel 126
Table 23	Refined <i>task purposefulness</i> items - Expert panel 126
Table 24	Refined <i>joining structuredness</i> items - Expert panel 127
Table 25	Refined <i>mentoring</i> items - Expert panel 128
Table 26	Refined supportiveness items - Expert panel 128
Table 27	Refined Task Performance items - Expert panel 129
Table 28	Refined CCB-I items - Expert panel 130
Table 29	Refined CCB-P items - Expert panel 132
Table 30	First sorting round: Individual judge's construct labels 134
Table 31	Inter-judge agreement 135
Table 32	Landis and Koch's Kappa statistics nomenclature 136
Table 33	Items Placement Score – Round 1 136

Table 34	Inter -judge agreement – Socialization	138
Table 35	Items Placement Score – Round 2	139
Table 36	Task Purposefulness Items - Rephrasing	139
Table 37	Inter -judge agreement – Performance	141
Table 38	Items Placement Score – Socialization	141
Table 39	Rephrasing after first round of pretest	147
Table 40	Demographics of the respondents - pilot test	150
Table 41	Reliability assessment - Socialization experience constructs	157
Table 42	Results of EFA - Socialization constructs	159
Table 43	Reliability assessment - Social identification and social integration	160
Table 44	Results of EFA - Social identification/integration	160
Table 45	Results of EFA - CCB Dimensions	162
Table 46	Reliability assessment - CCB dimensions	164
Table 47	Reliability assessment - Task performance	164
Table 48	Results of EFA - CCB / Task Performance	166
Table 49	Modified layout of the consent information page	167
Table 50	FLOSS project description	169
Table 51	Respondent demographics	176
Table 52	Roles/Tasks performed by respondents	177
Table 53	Outer model evaluation strategy	181
Table 54	Results of EFA - Socialization constructs	184
Table 55	Social identification/integration	185
Table 56	Results of EFA - CCB Dimensions	186
Table 57	MODEL 1- consistency reliability, indicator reliability, convergent validity of socialization constructs	188
Table 58	MODEL 1- consistency reliability, indicator reliability, convergent validity of performance constructs	189
Table 59	MODEL 1 - Cross-loadings 1	190
Table 60	MODEL 1 - Cross-loadings 2	191
Table 61	MODEL 1 - Fornell–Larcker criterion	191
Table 62	MODEL 1 - Item loading issues	192
Table 63	MODEL 2- Consistency reliability, indicator reliability, and convergent validity of socialization constructs	192
Table 64	MODEL 2- Consistency reliability, indicator reliability, and convergent validity of remaining constructs	193
Table 65	MODEL 2 - Cross-loadings 1	194
Table 66	MODEL 2 - Cross-loadings 2	195
Table 67	MODEL 2 - Fornell–Larcker criterion	195
Table 68	MODEL 2 - Item loading issues	196
Table 69	Loadings of CCB dimensions	196
Table 70	Summary of the criteria for structural model evaluation	197
Table 71	MODEL 1 - Results summary	201
Table 72	MODEL 1 - Cohen's (1988) pseudo F-test (f^2 effect size)	201
Table 73	MODEL 1 - Predictive relevance (Q^2 and q^2)	201
Table 74	MODEL 2 - Results summary	203
Table 75	MODEL 2 - Cohen's (1988) pseudo F-test (f^2 effect size)	204
Table 76	MODEL 2 - Predictive relevance (Q^2 and q^2)	204
Table 77	Overall model quality (GoF)	204

Table 78	MODEL 1 - Hypothesis tests	205
Table 79	MODEL 2 - Hypothesis tests	205
Table 80	Model test with direct and indirect paths	207

INTRODUCTION

The prominence garnered by Free/Libre/Open Source Software (FLOSS) projects in a broad range of applications is now unquestionable. Operating systems such as the Linux kernel, GNU, or FreeBSD projects, server and client software like Apache Web Server, Eclipse, Joomla, Mozilla Firefox, or else PHP and Python FLOSS programming languages are all testaments to the overall attractiveness and success of FLOSS development as a viable alternative to the conventional proprietary model of producing software (Hahn, Moon & Zhang, 2008). A key feature that differentiates FLOSS projects from traditional proprietary software projects is the reliance on an active and dedicated community consisting of talented individuals whose skills tap into a wide spectrum of software-related domains.

Despite the success of the overall FLOSS movement, research has demonstrated that FLOSS projects are characterized by high project abandonment rates (Stewart, Ammeter & Maruping, 2006). Reasons that explain project failure vary but it certainly makes the FLOSS movement less attractive for contributors and poses certain risks for adopters and users of FLOSS applications (Chengalur-Smith, Sidorova & Daniel, 2010). Past FLOSS research has highlighted the fact that communities suffer from a lack of new developers' enrolment (Hahn et al., 2008) and a lack of sustained participation (Fang & Neufeld, 2009). As a result, large FLOSS projects have now reached a point where ensuring survival, long-term success, and sustainability has become crucial.

However, ensuring the continuing enrolment of newcomers is not, in and of itself, a guarantee that new members will perform well and help to sustain their FLOSS communities. A FLOSS community that attracts a large number of new contributors from which a majority does not successfully or satisfactorily carry out project tasks may not prosper in the long run and may even disappear. One of the key factors for maintaining the strength of FLOSS communities concerns the manifestation of citizenship behaviours among members, behaviours such as altruistic sharing and helping (Hars & Ou, 2002; Kuk, 2006; Sowe, Stamelos & Angelis, 2008; Wu, Gerlach & Young, 2007). If such citizenship behaviours begin to fade, the survival of the community may be seriously jeopardized.

Because of the specificity of development processes and the technical complexity of the multiple projects within each FLOSS community, newcomers may be confronted with community entry obstacles before engaging in project participation. Communities have traditionally believed that access to the source code repository and to the interaction logs (such as mailing lists or forums) was sufficient for prospective contributors to attain a necessary level of performance, and engage in citizenship behaviours. FLOSS communities are becoming aware that such efforts are not sufficient, and unless more is done, their sustainability is endangered. Adopting a wide range of strategies, they have launched various initiatives to attract new participants but also to ensure that appropriate contributor behaviours are generated from new contributors. Some

FLOSS communities have created sub-communities of newcomers that create more collective socialization experiences such as KernelNewbies for the Linux Kernel project or GNOME Love. Formal mentoring programs such as the Google Summer of Code have also been developed for the same reason. From 2005 and 2012, the Summer of Code program has brought together more than 3000 student participants and mentors from more than 100 countries worldwide to take part in various FLOSS projects. Mentoring was found to be an effective way of bridging the gap between core developers and “newbies” (or novices) by emphasizing the acceptance and integration of new developers within FLOSS communities, and by pairing newcomers with role models who demonstrate good community practices. Other communities specifically recommend that new members perform a pre-defined series of peripheral tasks before engaging into more complex project work.

Furthermore, to date there has been no careful scientific investigation of the effectiveness of such initiatives. There is then a need for FLOSS communities and academic researchers to collaborate in order to help communities design appropriate socialization practices that generate satisfactory contributor behaviours. The first step to undertake on this research avenue, is to clearly understand which socialization factors have an impact on contributor behaviour and which do not.

The socialization literature within the organizational behaviour domain has provided a wealth of theoretical insights over the past 30 to 40 years. These findings provide a useful starting point for deriving a FLOSS-specific socialization model. Furthermore, the concept of participation within FLOSS communities has mostly been viewed from a code-centric perspective, using project archival data. Participation within a FLOSS community goes well beyond writing lines of code, fixing bugs, and submitting patches. A richer conceptualization is necessary to understand better the deeper functioning of FLOSS communities. This research aims to address these gaps.

1.1 RESEARCH PROBLEM, QUESTIONS AND RESEARCH OBJECTIVES

1.1.1 *Problem statement*

FLOSS communities are directly dependent on the capacity of the members to successfully carry out tasks in line with community performance expectations. The success of FLOSS communities is also tied to the initiatives undertaken by members to nurture and protect the community through various non-task related efforts such as altruistic behaviours of sharing and helping (Hars & Ou, 2002; Kuk, 2006; Wu et al., 2007).

One of the unique features of FLOSS communities is that members shape their own integration: they are free to enrol or not into any socialization programmes provided by the communities. They are also free to get involved with other newcomers or not. The experience through which an individual goes during the socialization phase is unique to each individual based on the behaviour and interaction undertaken by newcomers with the software project and the rest of the community. As a result, it is not enough for FLOSS communities to make sure socialization resources and initiatives are available for potential newcomers. Communities need to identify the practices that generate behaviours which are aligned with a community’s values and goals.

Accurately assessing the impact of an individual's socialization process on his/her behaviour within a community requires thoroughly understanding and conceptualizing participation. Past FLOSS research has suffered from a too restrictive performance-driven perspective towards participation, the measurement of which was based on archival data available from project repositories. The question here is not to debate whether this approach is appropriate or not, but rather to assess whether such a limited view of participation is sufficient for the actual needs of FLOSS communities. The performance-driven view may omit essential aspects of member participation that are crucial for FLOSS communities, but which cannot be captured by conventional measures such as the number of lines of code written by a contributor.

To accurately assess the impact of socialization, there is a need to adopt a more encompassing view of member participation by revising and reconceptualizing the concept and by using an approach that is more consistent with the actual functioning of FLOSS communities. A thorough review of the organizational behaviour literature identified numerous studies which have examined the concept of citizenship behaviours and which appeared particularly insightful when attempting to reconceptualize the notion of participation in the FLOSS community context. In particular, the socialization literature suggested the importance of potential mediators (also called socialization proximal outcomes) between socialization factors and contributor behaviour. These factors need to be investigated, and their measurement validated in the FLOSS community context.

1.1.2 *Research questions*

The overarching research question guiding this study is:

What socialization factors influence contributor behaviour in FLOSS communities?

The review of the FLOSS and broader organizational behaviour literatures around socialization and citizenship behaviours resulted in the following subsidiary research questions:

1. What are the key factors that characterize the socialization experience of newcomers in FLOSS communities?
2. What are the different instances of citizenship behaviour in the FLOSS community context?
3. What are the important proximal socialization factors in the FLOSS community context?
4. What is the influence of such proximal socialization factors on contributor behaviour?
5. What socialization factors influence the identified proximal socialization factors?

1.2 RESEARCH DESIGN

This research relies on a three-phase mixed-method approach (Creswell & Clark, 2007) that consists of an overall quantitative study in which a qualitative component is em-

bedded. The first phase involved the collection of interview data to explore the newcomer experience and citizenship phenomena, which led to the design of two alternative conceptual models in line with the findings from the review of the literature. The second phase includes the design and validation of a questionnaire instrument to test the models. The final phase includes the collection of data from FLOSS community members, using an online survey, to quantitatively test (using Partial Least Squares analysis) the research models.

1.3 SIGNIFICANCE OF RESEARCH

1.3.1 *Academic value*

This research project addresses the lack of strong theoretical foundations in the study of FLOSS communities and practices. The review of the FLOSS literature revealed the extensive use of approaches and conceptual models that are not theoretically well-grounded. This research drew insights from the socialization literature and integrated results from FLOSS socialization studies to conceptualize FLOSS participant socialization using strong theoretical insights. To do so, this research adds to the FLOSS stream of research by generating two generic FLOSS socialization conceptual models that, we hope, will trigger further research efforts in the FLOSS community context.

Furthermore, the notion of participation within FLOSS communities has been restrictive due to a lack of theoretical considerations and an extensive use of archival data. By introducing the notion of FLOSS community citizenship behaviours, this project re-conceptualized participation using insights from organizational behaviour literature and past FLOSS studies to provide a more encompassing and FLOSS-specific perspective that reflects more the FLOSS reality.

This study also contributes to construct specification and measurement by developing a socialization framework based on six newly-designed theoretical constructs. So far, the definition and measurement of performance within FLOSS communities had been oversimplified. The previously used constructs have been mainly software-centric and measured using archival data. This study developed and validated a survey questionnaire instrument that measured individual performance (using two dependent variables: task performance and community citizenship behaviours) in a much broader sense that is more in line with the actual functioning of FLOSS communities. This study also illustrates the effective use of primary (survey) data in understanding the FLOSS phenomenon.

1.3.2 *Practitioner value*

This research project addresses two important issues faced by FLOSS communities. On the one hand, this research helps FLOSS communities in understanding the socialization factors that contribute to the successful socialization of new members. It thus helps communities to tailor proper socialization initiatives that help to develop contributors who behave more in line with the values and sustainability of communities.

On the other hand, this study also helps FLOSS communities to better understand the inner individual mechanisms that influence contributor behaviour. In spite of the importance of the spread of citizenship behaviours to ensure the survival of a community, the investigation of the individual factors that lead to citizenship behaviours is a novel approach in FLOSS research. In addition, the finer understanding of performance provided in this research also benefits communities in helping them to better understand the full range of ways in which individuals contribute. On the other side of the coin, this research benefits FLOSS participants themselves, by providing insights into the practices with which they should engage when joining a new FLOSS community.

Finally, this research indirectly contributes to virtual communities more generally. Both socialization and contributor behaviour are concerns within a wide range of virtual communities (not just FLOSS communities). The study's findings may be to some extent extrapolated to the domain of all virtual communities.

1.4 DELIMITATIONS

This research focuses solely on the socialization experience and contributor behaviours of FLOSS community members. The notion of socialization has strong organizational connotations. As a result, this research particularly applies to FLOSS groups that have some organizational aspects that are usually present within communities that consist of more than a handful of programmers collaborating on some FLOSS-licensed software project. Adopting a socialization approach assumes the presence of a minimum of structural and hierarchical features that characterizes FLOSS projects of a certain size. This project will consistently use the term FLOSS community (instead of project, group, or any other term) to emphasize this important assumption throughout this research.

1.5 THESIS OUTLINE

This chapter has outlined the research questions and objectives, the proposed research design, its anticipated value to academics and industry, as well its delimitations. Chapter 2 reviews the relevant literature with the purpose of establishing the theoretical boundaries and foundations of this research; it ends with the development of two preliminary conceptual models. Chapter 3 details the proposed research design. Chapter 4 presents the results of the qualitative phase of this research, the goals of which were to delineate key socialization factors in FLOSS communities and to identify important instances of FLOSS community citizenship behaviours. Chapter 5 combines the literature streams analyzed in Chapter 2 and the results of Chapter 4, to extend and modify the two conceptual FLOSS socialization research models to their final, testable state. Chapter 6 outlines the research methodology used for this study. Chapter 7 presents the evaluation of the measurement and structural models as well as the findings of the main survey. Finally, Chapter 8 discusses the findings, and summarizes the main contributions of this research, and its limitations.

LITERATURE REVIEW AND INITIAL RESEARCH MODELS

2.1 CHAPTER OUTLINE

This chapter reviews and discusses the literature relevant to this study. In order to develop a firm understanding of the accumulated literature in this field, several research streams have been considered for this review. The chapter concludes with an introduction of the two initial conceptual models that guided this research.

The chapter begins with the theorization of the FLOSS environment under investigation followed by a delimitation of the theoretical boundaries of the research. It continues with a brief summary of the FLOSS history that provides some background information about FLOSS communities. Then, the key characteristics of FLOSS projects and communities are discussed. In the following subsection, the concepts of contributor behaviour and performance are reviewed in FLOSS research, leading to a re-conceptualization of the concept of performance. In the next subsection, the justification of the use of a socialization approach as a relevant strategy to tackle the research problem, is provided. The subsequent subsection reviews the socialization tradition in the domains of organizational behaviour and FLOSS research, accompanied by a review of the socialization concepts emphasized and used in this research. In addition, a subsection summarizes the findings and research gaps drawn from the review of the literature and consequently derives the subsidiary research questions that led this research project. Finally, the last subsection combines the results into two initial conceptual models that were refined into testable models at a later stage of this research (discussed in Chapter 5). Some additional literature is discussed in Chapter 5 in the context of the development of the specific hypotheses that were tested in this research.

2.2 THEORIZATON OF THE FLOSS ENVIRONMENT UNDER STUDY

FLOSS development is an alternative software production mode that significantly differs from the traditional firm-based proprietary model. At the theoretical level, FLOSS communities are software-producing social entities that have unique characteristics and mechanisms which significantly differentiate them from organizations; thus rendering the IS body of knowledge about conventional software development inapplicable to some extent. Besides, '*FLOSS project*' shall be taken as an umbrella term since the extensive variation between the different FLOSS development initiatives prevents from making any generalisations about the way projects work. FLOSS projects significantly differ along a variety of attributes. For instance, FLOSS project size ranges from one to several thousands contributors; the way a handful of programmers work on a software project cannot be similar to the way a community of thousands of individuals collaborate on a project involving a much broader spectrum of tasks such as marketing, design, or licensing. Similarly, project complexity can vary from simple utilities such as a password manager or a plain text editor to much more modular projects like operating systems or programming platforms.

As a result, it is important to clearly define the theoretical boundaries of the IT artifact under investigation (Orlikowski & Iacono, 2001) which will allow to determine the generalizability scope of the results as well as the degree of generalizability to other related environments. The following two subsections introduce the theoretical context of the social entities under study. The section ends by delimiting the theoretical boundaries of this research: *large* and *complex* projects are conceptualized as software-producing community-organization hybrids relying on commons-based peer production.

2.2.1 Commons-based peer production

FLOSS projects can be seen as a particular type of commons where individuals collectively collaborate to produce software artifacts that are non-rival public goods (English & Schweik, 2007). The term “commons” refers to “a particular institutional form of structuring the rights to access, use, and control resources” (Benkler, 2006, p. 60) that can be understood as the opposite of the term ‘property’. Indeed, the notion of ‘property’ relies on the principle that law allows people or entities to have the authority to determine how a given resource can be used (for instances it can be sold, rented, or given away). The asymmetric allocation of power that determines the notion of ownership towards resources is the basis of the conventional transaction-based view of markets and firms (Coase, 1937; Williamson, 1979).

Since Garrett Hardin’s (1968) famous article “The Tragedy of the Commons”, there has been a widespread interest in understanding how natural resources seen as ‘commons’ can be successfully managed (Feeny, Berkes, McCay & Acheson, 1990). Indeed, Hardin predicted the over-harvesting and eventual destruction of common-based resources such as water, fish stock, forests, or the atmosphere. By treating software production inputs and outputs as public goods, the extrapolation of Hardin’s claim to the FLOSS movement raises concern about the viability of the FLOSS production model.

The term “commons-based peer production” was introduced by Yochai Benkler (2002; 2006) to characterize an alternative model of production that consists of generating information-based products through digitally networked environments. ‘Peer production’ is a form of social production that is an alternative to other forms of production such as contract- or market-based, managerial-firm based, or else state-based production (Benkler & Nissenbaum, 2006). It relies on two main assumptions. First, peer production is *decentralized* as contributors collaborate without relying on any form of market pricing or managerial hierarchy. Authority to participate and contribute mainly resides within individual agents. Second, *social cues and motivations* govern the coordination of agents instead of prices or commands (Benkler & Nissenbaum, 2006). ‘Commons-based’ emphasizes that both inputs and outputs of the production process can be freely accessed, shared, and used.

FLOSS development has been seen as the “quintessential instance of commons-based peer production” (Benkler, 2006, p. 63). The FLOSS production model has now spread to various instances of information, knowledge, and cultural goods production (Benkler & Nissenbaum, 2006; Berry & Moss, 2008) which vary along several structural characteristics. According to Yochai Benkler, each instance of commons-based

peer production has three main attributes that characterize the produced objects. *Modularity* refers to the extent to which a product is organized into modules and sub-modules. *Granularity* characterizes the sizes of the project's modules. The more fine-grained or small in size, modules are, the more a project can draw from a large pool of contributors. Finally, *low-cost integration* must be enabled through efficient quality control mechanisms on modules and also efficient integration mechanisms of the modules.

This research project conceptualizes the FLOSS production model using the notion of commons-based peer-production as defined by Yochai Benkler (2002; 2006).

2.2.2 *Beyond the bazaar stereotype*

Early works on FLOSS development made use of a 'bazaar' metaphor to describe the functioning of FLOSS projects. The term was coined by Eric Raymond in an essay he wrote based on his observations of the Linux kernel development process (Raymond, 1999). He described it "as a great babbling bazaar of differing agendas and approaches [...] out of which a coherent and stable system could seemingly emerge only by a succession of miracles" (p. 2). More recent studies have criticized the overall lack of structure and organization that the bazaar metaphor implies when characterizing FLOSS projects. Fitzgerald (2006), for instance, studied the gradual shift of the FLOSS projects from 'bazaar-like' to more structured entities. Fitzgerald qualifies the overall development process as Open Source Software (OSS) 2.0 and highlights the use of planning purposive strategies, complex analysis and design, organized implementation subphases, and the increase of the number of paid developers to justify his claim. This research follows Fitzgerald's line of thought by arguing that FLOSS development now stands somewhere between two ends of a spectrum: unstructured bazaar-like entities and structured organisational settings.

2.2.3 *Theoretical boundaries: Community-organization hybrids*

This research project focuses on a particular category of FLOSS projects: *large* and *complex* projects. Such entities thus inherit features that characterize the nature of commons-based peer production that is the digitally-enabled co-creation of information-based products through a production process relying on the use of inputs and outputs that are treated as non-rival public goods. In addition, FLOSS production is thus inherently *decentralized* and *social cues and motivations* are key drivers that govern how such projects function (more details are provided in further subsections).

This research more precisely focuses on *complex* software projects such as operating systems, desktop environments, or programming platforms, that are characterized by a high level of modularity. Linux desktop environment projects (e.g. KDE or GNOME) for instance, consist of a vast collection of tools, more or less independent from each other, that address all functions of operating systems such as application launching, file handling, or window and task management. The high granularity of such commons-based projects pertains to the nature of software itself as contributions can start from modifying a single line of code up to submitting patches or entire software modules. Furthermore, as it will be developed further in this thesis, contri-

butions in large and complex FLOSS projects are not limited to code-related tasks as such projects involve a wide array of tasks. Documentation and translation are for instance important areas from which granularity is also high as contributors can start from correcting a single typo error, to writing a documentation paragraph, or even translating an entire set of documentation pages.

The term *large* characterizes the population of contributors instead of the software artifact itself. This research focuses on FLOSS projects relying on a large number of contributors which justifies the community-perspective that is adopted. For instance, a majority of projects hosted on the FLOSS forge SourceForge¹ have between one and five contributors. Such projects may or may not be instances of commons-based peer production but they do not belong to the category of entities that are under investigation in this research. The overall organizational-community approach adopted in this research is inappropriate to study small projects hosted on Sourceforge that have a handful of contributors. Furthermore, some FLOSS projects may be complex but not large and would lay outside the scope of this research. The Linux distribution, Linux Mint², is a complex software project that involves a rather small number of contributors (in total, 22 individuals contributed within the last year as of 14 February 2014) according to Ohloh³, a website tracking down the activity of the main FLOSS projects. The Ohloh website reports a number of several hundreds to several thousands contributors for complex projects such as GNOME, KDE, or Ubuntu for instance, thus justifying to consider them as *large* projects. Even though there exists no number threshold to distinguish a group from a community, the appropriateness of using an organizational-community approach is questionable when considering 'smaller' projects such as LinuxMint since some of the organizational assumptions would be invalid in such contexts.

Moreover, large and complex FLOSS projects rely on a population of contributors that function in a rather structured and organized manner that contradict the implications of the chaotic bazaar-view. As mentioned by Fitzgerald (2006), the software production process, from the planning to the implementation phase, relies on procedures that are used in conventional IT organizations that produce and maintain software, suggesting the presence of some degree of organization and authority. Besides, a majority of large and complex FLOSS projects now rely on either non-for-profit organizations (also called 'foundations') and/or corporate sponsors that take an active role in determining the goals and direction of the projects. Finally, the commonly accepted view about the peripheral participation of newcomers that assumes that socialization entirely depends on newcomer motivation and initiatives is another erroneous belief in the case of large FLOSS projects. Indeed, a number of projects now rely on organized joining processes and more and more events and programs are being run to attract and help integrate newcomers. As a result, large and complex FLOSS projects are neither 'bazaars' nor conventional organizations but rather lay in a spectrum between 'bazaar-like' entities and organizations. They shall thus be seen as community-organization hybrids that have a mix of features characterizing either communities or organizations.

¹ <http://sourceforge.net/>

² <http://www.linuxmint.com>

³ see <http://www.ohloh.net/p/linuxmint>, as of 11 February 2014

In conclusion, this research investigates the influence of socialization on contributor performance in large and complex FLOSS projects. The theoretical particularities of this research resides in the fact that such entities can be seen as community-organization hybrids that produce complex software artifacts through commons-based peer production processes, which clearly delimitates the theoretical boundaries of this research. Moreover, the dual nature (organization and community) of large and complex FLOSS projects strongly compromises the applicability of traditional organizational research as well research on digitally-enabled communities when investigating issues related to socialization, contribution, and individual performance.

The following sections and subsections describe key features and mechanisms that characterize large and complex FLOSS projects in order to provide a deeper understanding of the social entities that were studied during this research. The term *FLOSS communities* is consistently used throughout this thesis to refer to the population of contributors that is associated with large and complex FLOSS projects.

2.3 FREE/LIBRE/OPEN SOURCE SOFTWARE COMMUNITIES

FLOSS communities are complex social entities that are influenced by the origin of the FLOSS movement and its core beliefs. This subsection starts with some background information about the FLOSS movement. Then, it describes key characteristics of FLOSS communities and their members. The subsection concludes with considerations about the gap between FLOSS researchers and practitioners.

2.3.1 *A condensed history of the Free/Libre/Open Source Software movement*

The Free/Libre/Open Source Software (FLOSS) movement is the descendant of the hacker tradition which started at Massachusetts Institute of Technology where some students started calling themselves “hackers” in the beginning of the 1960’s (Himanen, 2002). The term “hacker” itself existed before that period, it referred to individuals who “program enthusiastically.”

The hacker culture is said to have been born within the Tech Model Railroad Club (TMRC), a student club which progressively shifted to computers and programming (Raymond, 1999). Their influence was spread far wider after 1969, the first year of the ARPAnet, the first transcontinental, high-speed computer network, which later became the Internet. The use of the ARPAnet network allowed isolated hacker groups, which had developed their own ephemeral local culture, to unite as a “networked tribe” (Raymond, 1999). The hacker culture then became a “new way of life, with a philosophy, an ethic and a dream” (Levy, 1984, p. 27).

In the 1970s and early 1980s, realizing the economic potential of the nascent software industry, software companies started distributing only binary copies of computer programs in order to prevent computer programmers from studying and modifying their software (Moerke, 2000; Simon, 1996). In 1980, the US copyright law was extended to computer programs, strengthening the control of software companies on source code and computer programs. In 1983, Richard Stallman, a hacker from the MIT Artificial Intelligence Laboratory, announced the launch of the GNU project as a means to ex-

press his frustration towards the overall change in computer and software culture. He then founded the Free Software Foundation (FSF) in October 1985.

Free/Libre/Open Source Software community practices are strongly influenced by the values and norms inherited from the early hacker movement. Such values have shaped the key characteristics but also the social dynamics of FLOSS communities.

In 1998, The label 'open source' was adopted during a session held at Palo Alto, California, in reaction to Netscape's January 1998 announcement of a source code release for its Internet browser called Navigator. The term open source was chosen with the deliberate intention to keep some distance from the ideological and confrontational connotations of the term "free software" and in order to attract the attention of corporations from the software industry. The Open Source Initiative was founded the same year by Bruce Perens and Eric S. Raymond, to promote open source software. Ever since both the Free Software Foundation and Open Source Initiative have gained increasing momentum and attracted numerous individuals from throughout the world. Academics and practitioners who wanted to address both movements commonly referred to the Free/Open Source Software movement (or FOSS).

The term *libre software* started being used some years later as the result of the disambiguation of the term *free* as highlighted in the famous statement from Richard Stallman: "Free software is a matter of liberty, not price. To understand the concept, you should think of free as in free speech, not as in free beer" (Stallman, 2002, p. 43). The first use of the acronym FLOSS is often attributed to Rishab Aiyer Ghosh, an Open Source Initiative board member and researcher. In recent years, the label Free/Libre/Open Source Software movement (or FLOSS) has been progressively acknowledged by both academics and members of the Free Software and Open Source Initiative to address this overall shift in the software landscape.

2.3.2 *Free Software vs. Open Source Software*

The Free Software Foundation (FSF) has been maintaining a definition that helps determining whether a software is a free software or not⁴. The definition is centered around four core beliefs about software freedom that are:

- The freedom to run the program, for any purpose.
- The freedom to study how the program works, and modify it.
- The freedom to access to source code and the freedom to redistribute copies.
- The freedom to distribute copies of modified versions of the program to others, allowing to help and contribute to the community.

The definition of free software denotes an overall emphasis on software users and freedom towards the use of software. The FSF regularly evaluates the licenses that potentially qualify as free software and keeps records of those licenses on its website⁵. The Open Source Initiative has also maintained a definition of open source software⁶. The definition addresses various aspects centered around the overall notion

⁴ <http://www.fsf.org/philosophy/free-sw.html>

⁵ <http://www.gnu.org/licenses/license-list.html>

⁶ <http://www.opensource.org/docs/osd>

of software distribution, summarized in 10 clauses. It specifies issues related to free redistribution, access to source code, rights to generate derived works, and absence of discrimination towards individuals or fields. Licensing aspects include distribution, non-restrictiveness, and technology neutrality. OSI regularly evaluates the licenses that comply with the OSI definition and maintains a list of licenses on its website⁷.

2.3.3 *From software glitch to the notion of FLOSS community*

The embryo of a typical FLOSS project is a software idea in the mind of one developer, answering a specific software need. FLOSS projects are often initiated to fulfill a need that cannot be met by proprietary software applications (Nelson, Sen & Subramaniam, 2006). As a result, projects are usually developed alone or by a very small group before being accessible to other potential contributors through the Internet (Crowston & Howison, 2006). The founders' ideas are then expressed into working code, generally submitted to one of the FLOSS repositories, and potentially start attracting more developers and users. If the project is successful, it then further develops quickly and gathers features and capabilities that in turn attract additional developers and users (Crowston & Howison, 2006).

English and Schweik (2007) found that a majority of FLOSS projects disappear after a short amount of time due to a lack of user and developer interest, thus leading to project abandonment. An analysis of the Sourceforge repository projects confirmed that most FLOSS projects fail: 58% do not move beyond the alpha developmental stage, 22% remain in the planning phase, 17% remain in the pre-alpha phase, and some become inactive (S.-Y. T. Lee, Kim & Gupta, 2009). Similarly, Stewart, Darcy and Daniel (2006) found that most FLOSS projects do not even last a year.

By gaining more attention and participation, a FLOSS software usually becomes more usable and in turn the number of individuals joining the community keeps increasing (provided the software continues meeting the developers' interest). A growth cycle starts that feeds both the community and development of the software. As the community grows, it becomes more diverse by encompassing additional roles that are less programming-oriented (Al Marzouq et al., 2005). FLOSS project roles are further described in section 2.3.5. The initiation phase and growth cycle of FLOSS projects is represented in Figure 1, which is based on works from Al Marzouq et al. (2005).

2.3.4 *FLOSS Project artifacts*

Due to the diversity of FLOSS projects, there is not a single development environment. Nonetheless, it is still possible to highlight the most recurrent development tools that are used in a majority of projects. In order to handle the assembling and versioning of source code submitted by FLOSS developers, a revision and version control system such as Concurrent Versions System (CVS) (Fogel, 1999), is usually used. The source code repository and the revision control system of a FLOSS project give access to the complete history of files and source code. Git has been among the most common distributed revision control systems. Git was initially designed and developed by Linus Torvalds for managing the development of the Linux kernel (Cornford, Shaikh & Ci-

⁷ <http://www.opensource.org/licenses>

Figure 1: Initiation and growth cycle of a FLOSS project (adapted from Al Marzouq et al., 2005)

borra, 2011). The use of software utilities to automate testing, compiling and bug reporting is also a common practice. Users' wishes of most wanted features may also be integrated in the bug tracking system. Bugzilla is among the most commonly used bug tracking and testing software (Raghunathan, Prasad, Mishra & Chang, 2005). It was written for the nascent Mozilla.org project to replace Netscape's proprietary bug tracking system.

Interactions among the contributors of a FLOSS project can be enabled through various Internet-related communication means such as mailing lists, FAQs, forums, message boards, Internet Relay Chat (IRC), and instant messaging. A number of FLOSS projects are hosted on "software forges" (Riehle et al., 2009) such as GNU Savannah or SourceForge which integrate all the above components in one single platform. However, large FLOSS projects tend to have their own independent platform and website such as it is the case for Linux distribution projects such as Debian⁸, Gentoo⁹, or Fedora¹⁰.

2.3.5 Community roles

Early research studies of the FLOSS phenomenon that examined communities often focused on identifying the various roles played by each member and the tasks performed within projects. Past research has mainly focused on the software development process, omitting other important roles that are less programming-oriented or

⁸ <http://www.debian.org>

⁹ <http://www.gentoo.org>

¹⁰ <http://fedoraproject.org>

even non-technical. In addition, the emphasis on software development also led to neglecting the consideration that participants may take part in several projects and sub-projects, assuming different roles. These limitations are further explained in section 2.3.7.

Early research articles made the assumption that in FLOSS projects, the people who contribute code to a project are always users of the code produced (Gacek & Arief, 2004). In these studies, FLOSS contributors were mapped within a developer/non-developer categorization (Gacek & Arief, 2004; G. L. Lee & Cole, 2003; Sen, 2007). Among developers, researchers introduced another dichotomous classification which sees developers as being either core or non-core developers (Fielding, 1999; G. L. Lee & Cole, 2003; Mockus, Fielding & Herbsleb, 2002). Combining both notions of core/non-core developers and non-developers, Nakakoji, Yamamoto, Nishinaka, Kishida and Ye (2002) and Ye and Kishida (2003) were among the first to introduce a more exhaustive list of roles performed by FLOSS members by identifying eight distinct role categories within an onion-like model. Gacek and Arief (2004) preferred a hierarchical view, mostly referring to the same roles mentioned in Nakakoji et al. (2002) and Ye and Kishida (2003).

Further attempts to understand the social structure of FLOSS projects extended the 'onion-like' model (Crowston, Heckman, Annabi & Masango, 2005; Crowston & Howison, 2005, 2006) (see Figure 2). The authors proposed the following categories: passive users, active users, co-developers, and core developers (within which were included initiator and release coordinator). The authors highlighted that the outer layer has a "virtually unknowable boundary" (Crowston & Howison, 2005, p. 7) consisting of passive users of the software who do not speak on either the project's discussion lists or forums.

Based on the integration of the results of the above works, Jensen and Scacchi (2005, 2007) proposed a more comprehensive model that focuses instead on a project community rather than the software project itself (see Figure 3). This approach is more in line with the one adopted in this research project and will be further detailed.

One of the limitations of the proposed models concerns their static nature. Indeed, it was shown that the core group of a FLOSS project does not persist for long periods of time; new generations take the lead over after a certain amount of time (Herraiz, Robles, Amor, Romera & González Barahona, 2006). Furthermore, the onion model tends to assume that a new contributor, starting as a user, has to work his way through all layers of the onion model in order to reach the core developer level. By studying the concurrent versions system (CVS) used within the GNOME community, Herraiz et al. (2006) found that there was no typical joining pattern for new members, limiting the predictive power of the onion model. Jensen and Scacchi (2007) confirmed the shortcomings of the onion model in terms of role migration and advancement processes. A relevant strategy to encompass the dynamic nature of FLOSS communities within a model is to rely on a socialization perspective (Ducheneaut, 2005; Fang & Neufeld, 2009; Qureshi & Fang, 2010). This is the position adopted in this research to investigate the main research question; it will be elaborated in greater detail later in the literature review.

Figure 2: Onion-shape structure of FLOSS communities (adapted from Crowston et al., 2005, and Crowston and Howison, 2005)

Figure 3: Onion-model of FLOSS community organizational hierarchy (Jensen and Scacci, 2005, 2007)

2.3.6 *Relationship between core and non-core developers*

The onion-shape structure of FLOSS communities implies that the number of core-developers of a FLOSS project is generally small (Aberdour, 2007). This aspect has been confirmed by academic research for individual FLOSS projects such as Apache (Mockus, Fielding & Herbsleb, 2000), Linux (G. L. Lee & Cole, 2003), or FreeBSD (Dinh-Trong & Bieman, 2005). Crowston and Howison (2006) recommend a number from 3 to 10 core members in order for a FLOSS community to be 'healthy'.

Among the core members of a FLOSS project, an individual may play the role of project leader (such as is the case with Linus Torvalds for the Linux project), who is often the person who has initiated the project. He or she is responsible for the vision and overall direction of the project (Nakakoji et al., 2002). Core developers are usually in charge of maintaining the high modularity of projects (Aberdour, 2007) and control the architecture and direction of development (Mockus et al., 2002). In general, they have been with the project for a long time (occasionally since the project's inception) and have made significant contribution to the system (Amrit & van Hillegersberg, 2010). Core members/contributors, also called maintainers (Robles, Gonzalez-Barahona & Michlmayr, 2005; Shaikh & Cornford, 2003) are usually in charge of submitting most of the code (Crowston & Howison, 2005). For instance, a case study from Mockus et al. (2002) revealed that development within the Apache community was quite centralized with only about 15 developers contributing more than 80 percent of the code for new functionality.

Contributing developers (also known as peripheral developers or co-developers), occasionally contribute new features and functionality to the system (Amrit & van Hillegersberg, 2010). Frequently, the core developers review their code before inclusion in the code base. Their main task is usually to submit patches of computer code and bug reports that may be tested and reviewed by either core members or else by peripheral members depending on the overall size and structure of the FLOSS community. Subsequently, core members usually select and retain some of the submitted patches to produce an official release, which then goes through more cycles of tests, and code improvements (G. L. Lee & Cole, 2003). By displaying interest and ability, co-developers may move from the periphery to the core. The periphery of a FLOSS project is usually an obscure zone in which passive and active users (who test new releases, write documentations), bug reporters, bug fixers, and peripheral developers interact.

2.3.7 *Multiple roles and non-programming/non-technical tasks in FLOSS communities*

As noted in Sections 2.3.5 and 2.3.6, past FLOSS research has largely focused on tasks and roles that are directly related to the software development process. This approach is justified for projects of small size or else for those that are close to the initiation phase during which, in general, only developers are involved. Nonetheless, the omission of non-programming and non-technical roles appears questionable when a FLOSS project grows to the point that the project gains the support of an active community that takes part in every aspect of the project. In such cases, individuals may be assigned one or several roles (Gacek & Arief, 2004; Scacchi, 2007) and may be in charge of various tasks that may go well beyond bug reporting, patch submission, or

simply writing code.

Some research efforts have identified some of the roles that are performed by individuals within a FLOSS community (Jensen & Scacchi, 2007). In studying role migration within the Mozilla, Apache, and NetBeans communities, Jensen and Scacchi (2007) identified the presence of a wide set of tasks such as quality assurance roles, source code versioning roles (e.g. CVS manager, CVS committer, etc), project planning, usability, licensing, and marketing roles.

To date, large FLOSS communities try to involve more and more contributors into documentation and translation tasks, graphical and other artistic tasks, and even project promotion through the organization of various events. For instance, large communities such as KDE and GNOME encourage newcomers to contribute in translation, documentation, accessibility, or project promotion tasks (see Figure 4 and 5).

Figure 4: Contributing within the GNOME community

Get Involved

The GNOME Project is a diverse international community which involves thousands of contributors, many of whom are volunteers. Anyone can contribute to GNOME: the project involves a wide range of activities, not just programming.

So, what do you want to do?

<p>Write</p> <p>Have a good command of the English language? GNOME is always looking for writers, whether it is to create user and developer guides for our software, write content for the GNOME websites, produce articles for the GNOME Journal, or just to blog (and microblog!) about the great stuff that is always happening in GNOME.</p> <p>Get started with writing documentation Get involved in the marketing team</p>	<p>Design</p> <p>Graphics and user interface design are an important part of the GNOME Project. You can help to design user interfaces and in so doing make GNOME software easy, satisfying and efficient to use. You can also assist our designers in the creation of icons, websites, clothing and other merchandise.</p> <p>Get involved in GNOME Design</p>
<p>Translate</p> <p>The GNOME Project is proud to be available in many different languages, and this is a key part of our mission to make our software available to as many people as possible. If you have a good command of English and another language, you can participate in one of the many GNOME translation teams, so that GNOME's software, documentation and websites can be used by people from all over the world.</p> <p>Get started with translating GNOME</p>	<p>Code</p> <p>There are many coding tasks up for grabs in GNOME, both small and large. You don't have to be an expert to start contributing code. GNOME software is written in many different languages, so it is easy to feel right at home, and we have some incredible programmers from whom you can learn.</p> <p>Install the latest code and find tasks Learn about GNOME technology</p>
<p>Test</p> <p>Making high quality software requires lots of testing. Reporting errors or possible improvements is a valuable way to help with the project and is a great way to start out. If you like living dangerously, you can also run our latest development code and help to provide feedback on changes as they happen.</p> <p>Install the latest code</p>	

Source: <http://www.gnome.org/get-involved/> (accessed on 19 June, 2013)

2.3.8 *Shift from a FLOSS project-centered to a community-centered focus*

The previous sections have demonstrated that past FLOSS research has provided insights in understanding the various tasks and roles performed by individuals in

Figure 5: Contributing within the KDE community

Source: <http://www.kde.org/community/getinvolved/> (accessed on 19 June, 2013)

FLOSS software development practices. By focusing on the technical side of the development of given FLOSS projects, past research has generated a vast array of findings that help in explaining various project/group related issues such as FLOSS project success (Grewal, Lilien & Mallapragada, 2006; S.-Y. T. Lee et al., 2009; Subramaniam, Sen & Nelson, 2009) or project/team performance (Colazo & Fang, 2010; Stewart & Gosain, 2006). An extensive literature review of FLOSS research published in 2010, indicated that around 60% of the identified papers had adopted a group/project level of analysis (Crowston, Wei, Howison & Wiggins, 2011).

The nature of large FLOSS projects presupposes a community of individuals contributing to the project (Fang & Neufeld, 2009). In other words, a FLOSS project cannot exist without the support of an active community. As pointed by Brian Behlendorf, cofounder of Apache:

To us, the primary benefit of using Apache was using it as part of the community at large rather than just using it for the 1s and 0s it represented. (Moody, 2002, p. 161)

It thus seems that FLOSS project success strongly depends on the well-being of the community of participants involved in the project. This research investigates the individual factors that lead to contributor behaviours which positively affect the sustainability and success of FLOSS communities.

Before being able to effectively address important aspects of community success and sustainability, it is important to first re-conceptualize the notion of contributor behaviour. The purpose of this research is to study the individual behaviours that engender the effectiveness and well-being of FLOSS communities. Therefore individual contribution must be conceptualized at the community level instead of the traditionally used software/project level. This logically affects both the definition and understanding of a contributor's behaviour and actions within a FLOSS community. Writing lines of code or reporting and fixing bugs comprise but a small fraction of tasks that can be performed to contribute to a FLOSS community. In turn, re-conceptualizing contributor behaviour must be accompanied by a reconsideration of the concept of individual performance which has been overall considered from a project development perspective in past FLOSS research. Addressing individual performance with respect to a FLOSS community must go hand-in-hand with a conceptualization of the notion of contributor behaviour.

Therefore, the following subsection reviews past research that has investigated either contributor behaviour and/or individual performance in the context of FLOSS communities and other forms of virtual communities. Then, a following subsection introduces the conceptualization adopted in this research based on insights from seminal works in the organizational behaviour field.

2.4 INDIVIDUAL CONTRIBUTION BEHAVIOUR AND PERFORMANCE IN FLOSS COMMUNITIES

FLOSS studies have relied on a wide range of perspectives when assessing the behaviour of FLOSS community contributors. Some authors have adopted a technocentric approach by assessing member participation within a project through programming related measures such as number of lines of code written in a project source code (Roberts, Hann & Slaughter, 2006). Others have chosen a broader community-based perspective such as knowledge sharing activities (e.g. Kuk, 2006). Because of such diverging understandings and conceptualizations of member behaviour, there is no commonly accepted way to assess individual performance in FLOSS research. This subsection first reviews FLOSS studies that investigated various individual behavioural aspects of FLOSS contributors, starting from joining projects, to contributing, and performance. Then, we propose a reconceptualization of the notion of contributor behaviour based on seminal works in the organizational behaviour field.

2.4.1 *Contribution and performance in FLOSS research*

FLOSS practices started attracting the attention of IS researchers when they realized that a totally different software development methodology could engender high quality software such as the Linux operating system or the Apache web server. As a result, a number of research projects investigated individual contribution and performance

by focusing on software-centered metrics such as the number of lines of code written or the number of source code contributions. This approach is well exemplified by the following citation from [Roberts et al. \(2006\)](#) who studied the links between individual motivations, participation, and performance:

Although [FLOSS] volunteers provide many different kinds of valuable services to their respective projects [...], a principal participation behaviour consists of authoring and maintaining the software, i.e., writing lines of software code. ([Roberts et al., 2006](#), p. 992)

The authors measured individual participation using two items: the cumulative number of patches submitted and accepted into the software revision control system, and the cumulative number of lines of code submitted and accepted.

In spite of the overall tendency to solely focus on software development when measuring contributor behaviour and performance, some early works in FLOSS research highlighted the importance of non-programming tasks. For example, [Gacek and Arief \(2004\)](#) suggested that individuals could contribute in the following ways: reporting bugs, suggesting new features, reviewing code, modifying code (fixing bugs and implementing new feature), and making decisions.

A qualitative study of the phpMyAdmin community that empirically assessed participation through the lens of legitimate peripheral participation theory, used an approach consistent with Gacek and Arief's view by identifying two categories of contributions: conceptual (advising others) and practical contributions (improving the code) ([Fang & Neufeld, 2009](#)).

Other studies have examined participation by studying interactions with FLOSS artifacts such as mailing lists or a bug tracker system, suggesting that individuals may contribute by writing code and submitting patches, but also by exchanging information with other community members and tracking bugs. For example, in studying the factors leading to the sustained participation from a community-of-practice perspective, [Fang and Neufeld \(2009\)](#) captured participation activity in terms of frequency of mailing list submissions, frequency of tracker submissions, and frequency of CVS commits. Similarly, [Dahlander and O'Mahony \(2011\)](#) studied the phenomenon of progression through task-based authority within FLOSS communities and operationalized project interactions as technical contributions, technical communication, and coordination work. [Xu, Jones and Shao \(2009\)](#) proposed a broader view of contribution behaviour by defining developer performance as the amount and quality of knowledge contribution. The definition suggested a shift from the traditional source code perspective to a more knowledge-oriented view. However, performance was eventually measured in terms of the number of function points made and accepted into the project during an observed time period; a measure that has obvious software coding connotations.

Some researchers have addressed the notion of individual performance from a more community-centered perspective, by capturing performance in terms of rank or role promotion ([Okoli & Oh, 2007](#); [Roberts et al., 2006](#)). This approach goes one step further in relating the outcome of individual contributions to the community instead of just the software development process.

From software to task focus

Another stream of research adopted a task-focused approach, as opposed to the traditional programming-oriented view of FLOSS participation. Researchers have addressed this aspect by considering task effort (Hann, Roberts & Slaughter, 2004; Ke & Zhang, 2010), task performance (Ke & Zhang, 2009), or else task completion (Chou & He, 2010; Stewart & Gosain, 2006). This perspective is particularly insightful since the emphasis on tasks lessens the bias of considering only programming-related jobs. It addresses all contributions within a FLOSS community.

An early work operationalized task effort in terms of the time a participant devotes to their FLOSS contributions (Hann et al., 2004). Even though measuring effort solely according to time considerations may appear to be simplistic, this approach provides an important suggestion: FLOSS communities may be especially interested in having members who dedicate a lot of their time in contributing to the community. Ke and Zhang (2009) introduced a conceptualization of task effort as goal commitment and effort intensity. Ke and Zhang extended their previous work by introducing a more encompassing view of the concept (Ke & Zhang, 2010). Task effort was measured in terms of time commitment, task persistence, and effort intensity. In addition to confirming the importance for communities to have participants dedicating time to contribute, this work goes a step further in suggesting that effort is also a matter of overcoming difficult problems (task persistence) and working to one's full capacity (effort intensity).

Finally, studies that considered task completion have also provided some valuable insights in unveiling the behaviours that benefit FLOSS communities. Stewart and Gosain (2006) explored the impact of the adherence to components of the FLOSS ideology on effectiveness in FLOSS teams captured in terms of team effort and task completion. In extrapolating the idea of task completion from team to individual level of analysis, it is important to recognise that communities benefit from people who actually 'do the job'. This statement is even more relevant by bearing in mind that tasks are mostly self-assigned in FLOSS projects (Crowston, Wei, Li, Eseryel & Howison, 2005); communities thus have little control in allocating resources and monitoring progress. Chou and He (2010) used a similar approach, and found that expertise integration affects both team size and team effort, which in turn jointly influence task completion.

Knowledge sharing

FLOSS research has identified sharing and learning as important motivators for individuals to join FLOSS projects (Subramanyam & Xia, 2008; Wu et al., 2007). The findings may have encouraged researchers to adopt a more knowledge-oriented approach in contrast to the more common software development perspective used when addressing contribution issues in the FLOSS context. A large volume of research has framed the participation of individuals within virtual communities in terms of knowledge contribution and sharing (Chiu, Hsu & Wang, 2006; Ma & Agarwal, 2007; Wasko & Faraj, 2005). It is likely that the interest by FLOSS researchers to look at knowledge exchange within FLOSS communities was triggered by a desire to extend the findings of virtual community research.

Kuk (2006) explored the impact of strategic interactions (from a participation inequality perspective) among community members on knowledge sharing and confirmed how participation inequality triggers it. An analysis of the logs of the KDE mailing lists allowed him to identify three categories of knowledge sharing activities:

- reusing the existing public domain knowledge, including reuse of the existing design features and codes and code libraries;
- revealing personal uniquely acquired experience and knowledge, including submissions of new codes and libraries for new design features and applications;
- recombining new knowledge by integrating with the existing knowledge, including combining and rewriting existing codes into new codes.

A similar approach was taken by Sowe et al. (2008), who analyzed the posting and replying behaviours within Debian community mailing lists. G. L. Lee and Cole (2003) studied the Linux kernel community and designed a knowledge creation model based on the community of practice literature. Compiling the knowledge-sharing and self-efficacy literatures, Endres, Endres, Chowdhury and Alam (2007) proposed a self-efficacy model which looks at knowledge sharing behaviours as its dependent variable.

Helping and sharing behaviours are common and valued in FLOSS communities (Kuk, 2006; Wu et al., 2007). More and more community members dedicate some of their time helping newcomers, by way of various forms of interactions such as formal or informal mentoring (which will be examined further in section 2.6.6). Communities strongly rely on those kinds of behaviour to assure their survival and longevity.

The foregoing discussion has shown that a focus on software development when studying contribution behaviour within FLOSS communities is restrictive and generally excludes important behaviours that are necessary for the well-being of FLOSS communities. Focusing on the concept of task helps in tackling this issue by identifying key aspects such as task effort, persistence, intensity, task completion, or even helping and sharing that are not captured when examining the number of lines of code or software patches submitted by a community member. This reflection opens a new avenue in understanding individual contribution in a more encompassing and meaningful way. The next subsection introduces key insights from the organizational behaviour body of literature that will help in re-conceptualizing the notion of individual behaviour.

2.4.2 *A reconceptualization of performance and contributor behaviour*

Chester Barnard (1938)'s idea of "willingness to cooperate," and Daniel Katz's (1964) distinction between dependable role performance and "innovative and spontaneous behaviours" have both influenced social scientists in conceptualizing individual performance within organizations. Katz was among the pioneers in suggesting that individual performance be viewed as a dual component comprised of in-role and extra-role performance. Katz stated that in addition to being induced to enter and remain within the system, individuals must successfully perform two types of behaviour. First, they must carry out their role assignments in a dependable fashion. Second, they must

manifest innovative and spontaneous activities in achieving organizational objectives which go beyond role specifications. In the early 1980's, the term 'organizational citizenship behaviour' (OCB) was adopted for Katz' category of extra-role behaviour, Bateman and Organ (1983) being among the first studies to introduce the term.

In-role and extra-role individual performance

According to Katz, dependable role performance (or in-role performance) is the most conspicuous behavioural requirement of the organization. Assigned roles must be carried out and must meet some minimal level of quantity and/or quality of performance. Katz argues that compliance with the legitimate rules of an organization primarily influences only reliable role performance. He thus emphasizes that performance is assessed according to the norms, standards and rules of a given organization. In the FLOSS context, the notion of role is vague since individual roles are often self-assigned (Crowston, Wei et al., 2005) and may involve tasks from various domains such as quality assurance tasks, source code creation tasks, and source code versioning tasks, but also usability, marketing, and licensing tasks as well (Jensen & Scacchi, 2005). It is then more pertinent to focus on the notion of task itself rather than role (echoing past FLOSS research that looks at the notions of task effort and task completion) when assessing individual performance. Simply said, FLOSS communities associate a member's level of performance with the extent to which this person gets the work done (by appropriately performing project tasks).

Reflecting on the behavioural requirements that characterize employee performance, Katz (1964), noted that another type of performance exists, extra-role performance, that pertains to innovative and spontaneous activities. The notion of extra-role performance introduced by Katz led to an entire stream of research in the field of organizational behaviour literature, which led to the introduction of the concept organizational citizenship behaviour (OCB). Dennis W. Organ, one of the pioneers and leading researchers in OCB research, defined organizational citizenship behaviours as:

"individual behaviour that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization. By discretionary, we mean that the behaviour is not an enforceable requirement of the role or the job description, that is, the clearly specifiable terms of the person's employment contract with the organization; the behaviour is rather a matter of personal choice, such that its omission is not generally understood as punishable." (1988, p. 4)

2.4.2.1 Organ's view of citizenship behaviour

The most acknowledged operationalization of organizational citizenship behaviour was introduced by Organ (Organ, 1988; Organ & Konovsky, 1989) who proposed five distinct dimensions that characterize the construct:

- Altruism: voluntary actions that help another person with a work-related problem;
- Conscientiousness: behaviour indicating that employees accept and adhere to the rules, regulations, and procedures of the organization;

- Civic virtue: behaviour indicating that employees take an active interest and participate in the life of their organization;
- Sportsmanship: a willingness on the part of employees to tolerate less than ideal circumstances without complaining and making problems seem bigger than they actually are;
- Courtesy: gestures that help others avoid a problem, such as checking prior to committing to an action that will affect them (providing advance notice to people who need such information).

Later, Organ subsequently expanded this model to include two other dimensions: peacekeeping and cheerleading (Organ, 1990a). However, later empirical research (MacKenzie, Podsakoff & Fetter, 1991; N. P. Podsakoff, Whiting, Podsakoff & Blume, 2009; P. M. Podsakoff & MacKenzie, 1994) found difficulty in clearly distinguishing altruism, courtesy, peacekeeping, and cheerleading, and suggested they all belong to an overall helping dimension.

Some aspects of Organ's OCB dimensions seem to apply to the context of FLOSS communities. For instance, helping behaviours captured by Organ's notion of altruism, appear to play an important role in the functioning of FLOSS projects. Past FLOSS research has already recognized the importance of altruistic behaviours (Hars & Ou, 2002; Wu et al., 2007). Sportsmanship taps into some of the aspects that have been addressed in the studies that investigated task effort, intensity, and persistence (Ke & Zhang, 2009, 2010).

The concept of conscientiousness (also called compliance) has not been investigated in FLOSS research. However, FLOSS communities have sets of rules, regulations, and procedures that members must follow when contributing. For instance, the Koha community utilizes a large array of compliance mechanisms, including coding, enhancement request, submission, request-for-comments (RFC), and bug reporting guidelines (see Figure 6). Also, many communities adopt codes of conduct that is clearly visible on the community websites – e.g., the KDE, Ubuntu, and Debian communities (see Figure 7). This suggests that compliance with the community formal and informal rules, a core component of conscientiousness, may be an important behavioural requirement within FLOSS communities.

Courtesy behaviours may also be important within FLOSS communities. FLOSS projects are usually developed by loosely organized teams of individuals located around the world (Wu et al., 2007) and in which tasks are mostly self-assigned (Crowston & Scozzi, 2004). Anti-courtesy behaviours may thus quickly lead to a high number of conflicts among community members and dysfunction in the integration of the various project components/modules.

Finally, civic virtue behaviours pertain to the participation of community members to the events and administration of the project. Communities such as the Debian community rely on democratic authority mechanisms and thus encourage members to take part in the decisions to be made within the community:

Together, the Developers may:

- 1) Appoint or recall the Project Leader.

- 2) Amend this constitution, provided they agree with a 3:1 majority.
- 3) Make or override any decision authorised by the powers of the Project Leader. (Debian constitution, article 4.1)¹¹

Furthermore, a number of communities organize a multitude of events and conferences to fulfill various goals such as discussing future features and direction. For instance, GNOME and KDE communities jointly organize an annual Desktop Summit, a conference that gathers a variety of contributors of the GNOME and KDE communities. The Wordpress community organizes frequent Wordcamp¹² events throughout the world. These events draw from both casual users and core developers, and encourage them to participate, share ideas, and get to know each other. In a similar manner, a number of events with various purposes are organized by the Linux Foundation throughout the world¹³. For instance, the Linux Foundation End User Summit focuses on Linux users, the Linux Foundation Collaboration Summit is organized to discuss the most current Linux development issues, while the Linux Foundation Member Legal Summit assembles members to debate ways of advancing the legal defense infrastructure for Linux.

Figure 6: Koha bug reporting guidelines

Source: http://wiki.koha-community.org/wiki/Bug_Reporting_Guidelines (accessed on 19 June, 2013)

2.4.2.2 William and Anderson's (1991) view towards OCB

The second major conceptualization of organizational citizenship behaviours was developed by Williams and Anderson (1991). The authors organize OCBs into categories on the basis of the target or direction of the behaviour. In particular, they distinguish

¹¹ Source: <http://www.debian.org/devel/constitution> (accessed on 17 May, 2011)

¹² see <http://central.wordcamp.org/>

¹³ <http://events.linuxfoundation.org/events>

Figure 7: KDE community code of conduct

Source: <http://www.kde.org/code-of-conduct/> (accessed on 19 June, 2013)

behaviours that are directed toward the benefit of other individuals (or OCB-I), and behaviours directed toward the benefit of the organization (called OCB-O). All of Organ's OCB dimensions can be captured by Williams and Anderson's conceptual scheme. Altruism, courtesy, peacekeeping, and cheerleading tap into the OCBI category whereas conscientiousness, civic virtue and sportsmanship are captured in the OCBO dimension (N. P. Podsakoff et al., 2009).

One of the key benefits of Williams and Anderson (1991)'s model is that it is broad enough to encompass additional OCB dimensions that have been introduced in other works. For instance, OCB-I incorporate Graham's (1991) interpersonal helping, Van Scotter and Motowidlo's (1996) interpersonal facilitation, and Farh, Earley, and Lin's (1997) helping coworkers and interpersonal harmony. In a similar manner, OCB-O captures Graham's (1991) organizational loyalty; Borman and Motowidlo's (1997) endorsing, supporting, and defending organizational objectives; Van Scotter and Motowidlo's (1996) job dedication; or LePine and Van Dyne's (1998) voice behaviour.

Another advantage of Williams and Anderson (1991)'s view towards OCB, is that it is able to include other types of citizenship behaviours that are specific to the context under study. Behaviours that pertain to the unique way FLOSS communities function may qualify for being instances of citizenship behaviours directed towards individuals or else towards the project and its community. There is a gap in FLOSS research about the specific behaviours that can be seen as instances of citizenship behaviours in the FLOSS context. Organ's operationalization of OCB risks omitting other types of behaviours that are important for the functioning of a FLOSS community.

Intuitively, the notion of sportsmanship introduced by Organ, does not seem to rigorously fit within the FLOSS context since there is in general, no deadline, tasks are usually self-assigned, and there is no hierarchical structure in FLOSS projects. However, the idea of sportsmanship does tap into some of the aspects that have been addressed in the studies that investigated task effort, intensity, and persistence (such as Ke and Zhang 2009; 2010). Furthermore, FLOSS communities value the promotion of the project by members to potential users and new contributors (as can be seen, for instance, in KDE's GetInvolved page in Figure 5). This may mean that other OCB aspects such as voice behaviour (Lepine & Van Dyne, 1998) or loyal boosting (Moorman & Blakely, 1995) are important components of FLOSS community citizenship behaviours and shall be included in this study.

2.4.2.3 FLOSS community citizenship behaviours: CCB-I and CCB-P

Williams and Anderson's framework has the merit of simplicity while incorporating the various types of citizenship behaviour dimensions that the literature has provided. It includes some of Organ's aspects that are relevant to the FLOSS context while being able to add other OCB-I and OCB-O behaviours identified by other OCB researchers that are meaningful and valuable to the functioning of FLOSS communities. As a result, the William and Anderson framework was adopted in this research to conceptualize and operationalize FLOSS Community Citizenship Behaviour (or CCB) in terms of CCB-I (or community citizenship behaviours oriented towards the benefits of other individuals), and CCB-P (or community citizenship behaviours oriented towards the benefits of the project and its community).

The review of the FLOSS literature pointed out the relevance of some of the OCB concepts such as helping, courtesy, voice behaviour, or loyal boosting, but did not definitively identify all potential behaviours that pertain to FLOSS community citizenship behaviour. Therefore there was a need in this research to collect qualitative data from FLOSS practitioners to help understand the nature of CCB and the various behaviours that characterize it.

2.5 TOWARDS THE IMPORTANCE OF NEWCOMER SOCIALIZATION

This subsection introduces the socialization perspective adopted in this research.

There is no question about the overall success of FLOSS initiatives such as Apache, Linux or Ubuntu which all have a very active community and have proven to produce high quality software. Nonetheless, such success stories are but the tip of the FLOSS iceberg. A vast majority of FLOSS projects fail to take off and are abandoned (Hahn et al., 2008). For instance, Amrit and van Hillegersberg (2010) reported that the Sourceforge portal (the largest FLOSS repository) contained around 160,000 registered projects (by March, 2009). Only 17 % of the projects (around 17,000 projects) had a stable version of their software and only 1.52 % had reached a mature status. Other researchers have also acknowledged that most Sourceforge projects become inactive soon after registration on the software platform (Chengalur-Smith & Nevo, 2010). These findings triggered the interest of researchers in gaining a deeper understanding of the factors responsible for FLOSS project success and community sustainability.

One of the most acknowledged reasons for the failure of FLOSS projects is the inability of the project to bring together a critical mass of developers (Hahn et al., 2008; Lerner & Tirole, 2001; T. O'Reilly, 1999; von Krogh, Spaeth & Lakhani, 2003). In FLOSS research, Kevin Crowston's determination of FLOSS success measures included number of contributors as a measure of FLOSS project success (Crowston, Annabi & Howison, 2003; Crowston, Howison & Annabi, 2006; Crowston & Scozzi, 2004). Similar insights were proposed by Stewart, Ammeter and Maruping (2006) who indicated that attracting and motivating contributors is a key factor for FLOSS project success. Chengalur-Smith et al. (2010) examined the factors that influence the long-term sustainability of FLOSS projects. The authors developed a project sustainability model based on organizational ecology and tested it using archival data from 2,772 projects registered in SourceForge. The authors hypothesized that a project's ability to attract and retain developer resources would have a positive effect on its future sustainability. They found that project demographic characteristics such as age and size, as well as the niche size occupied by the project, influence the project's ability to attract and retain user and developer resources in the future, which, in turn, was found to be a predictor of future project sustainability.

The importance of attracting and retaining new individuals within FLOSS communities is in line with Katz's early seminal work which argues that it is essential for organizations that "people must be induced to enter and remain within the system" (Katz, 1964, p. 132). However, simply being present, and performing well, within a social entity are two obviously different notions. As Katz said: "people may, of course, be within the system physically but may be psychological absentees" (p. 132). A FLOSS community would not benefit by having a majority of members being inactive or not successfully carrying out project tasks. Continuous induction of new people is thus a necessary condition but by no means a sufficient one to ensure an organization's or a community's effectiveness. In line with Katz' statement, induction in FLOSS projects must go hand in hand with generating performing behaviours and contribution quality from newcomers.

FLOSS research has provided rich insights into the factors influencing people to join and stay in FLOSS communities, thus addressing Katz' first requirement. Nonetheless, there is a major gap in the literature to tackle the remaining two behavioural requirements: performing well in assigned tasks and manifesting innovative and spontaneous initiatives. There is no guarantee that a newcomer who has decided to join a community and has engaged in lurking activities, will eventually contribute by carrying out tasks with a sufficient level of performance. It is possible that the phase through which an individual progresses from a community outsider to an insider, also called socialization, will have an influence on project contribution type, amount and quality. FLOSS communities have the potential to facilitate this experience by implementing effective strategies, social practices, and technological artifacts. However in order to gain insights into which strategies, practices and tools might be most effective, it is important to examine the socialization factors that generate adequate contributor behaviour. The following section reviews the socialization literature as well as past FLOSS works that adopted a socialization perspective, and then presents the theoretical foundations adopted in this research.

2.6 SOCIALIZATION LITERATURE AND FLOSS SOCIALIZATION

This subsection first compares FLOSS socialization and organizational socialization. It then reviews the socialization literature in FLOSS research. It then introduces the two main socialization models that have been used in previous research. After having assessed the pertinence of each socialization model in the FLOSS context, the subsection ends with a review of important factors characterizing and assessing the impact of socialization on newcomers. These factors are then reviewed in both FLOSS and online community research.

2.6.1 *Differences between FLOSS socialization and organizational socialization*

Because large and complex FLOSS projects are organization-community hybrids, the socialization phenomenon in this particular context is characterized by a combination of aspects pertaining to both organizations and communities. However, FLOSS socialization also has unique characteristics that pertain to the specificities of its production mode (commons-based peer production), the nature of the good that is produced: highly complex and modular software artifacts, and also the size of the project in terms of number of contributors.

A first aspect that differentiates FLOSS socialization from organizational socialization is that the two phenomena have different spatial-temporal constraints. Commons-based peer production involves generating information-based products through digital networked environments. Contributors are thus geographically dispersed and interactions between newcomers and the produced software artifact are computer- / Internet-mediated, but also the interactions between newcomers and other project contributors, and interactions between the project as a whole and potential or engaged newcomers (through the project website, IRC channels, discussion forums, voice/video conversations and so forth).

Since authority is decentralized (a feature inherited from commons-based peer production), both experienced and new contributors are free to decide how they will participate and contribute. As a result, a significant difference from traditional organizations is that FLOSS projects rely on *decentralized* socialization mechanisms allowing new contributors to shape their own socialization experience as opposed to traditional employee socialization. Indeed, within a FLOSS community, newcomers are free to choose the way to adjust to the community and learn the ropes. For instance, a community such as Apache provides formal mentorship programmes (either in-house¹⁴ or else through the Google Summer of Code), but newcomers are free to get directly involved in projects. Similarly, the Linux kernel community has created a specific community for newbies¹⁵ which encourages newcomers to learn collectively and interact. Nonetheless, participation to such communities is not a requirement in order to become involved in some of the community projects. As a consequence, socialization in FLOSS communities is shaped by newcomers themselves. However, FLOSS communities rely on more and more structured and organized newcomer programs and initiatives. For instance, newcomers are free to join GSoC (provided their application

¹⁴ see <http://community.apache.org/mentoringprogramme.html>

¹⁵ see <http://kernelnewbies.org>

is accepted) but taking part to the program is a formal process during which for instance the mentee is evaluated twice by the mentor to ensure appropriate progress (enabling the mentee to get paid). Other examples can be found in projects such as Debian or Gentoo, which rely on a complex and structured joining process requiring formal mentoring and online tests to be taken. As a result, decentralization in the FLOSS socialization context is not synonymous of absence of organizational features as socialization experience may vary along two ends of a spectrum, from a totally solitary experience to a much structured and project- / community-driven experience.

Furthermore, following principles inherent to commons-based peer production, social cues and motivations are mechanisms that govern FLOSS socialization whereas socialization in organizations is usually characterized by managerial commands and authority. For instance, research has shown that FLOSS projects are characterized by a gift culture (Bergquist & Ljungberg, 2001) and that providing feature gifts were common mechanisms that were used by newcomers when joining a project (Von Krogh, 2003). Besides, early works on FLOSS socialization showed that newcomer socialization could be seen as an iterative process of situated learning (with “thinking” and “doing” activities), and identity construction through community recognition (identity-regulation) and self-perception (identity-work) (Fang & Neufeld, 2009).

The more complex a FLOSS project is the more skills are required for individuals to be able to contribute. In other words, barriers of entry are higher in complex projects. Contributing to large projects like GNOME involves prior knowledge about the programming platform and the various tools and libraries that exist (see Figure 8). User interface contributions, for instance, are performed through a GNOME-dedicated UI editor called Glade that generates scripts that are then incorporated within the software artifact. Smaller projects may just require programming skills in languages such as C, C++, or Java. As a consequence, FLOSS socialization in large and complex projects is particularly challenging for newcomers and requires considerable learning efforts.

Another important difference between organizational and community contexts pertains to the nature of the boundaries of each type of entity. Traditional forms of organizations are characterized by fixed boundaries and new employees are purposefully selected and recruited by the organizations themselves. Boundaries in FLOSS communities are much more porous and permeable because the decision to join is in the hands of newcomers themselves. Moreover, the boundary that determines whether an individual is or is not a project member is blurry as there is no official ‘membership’ status. Instead, projects are surrounded by a cloud of project lurkers engaged into some degree of peripheral participation that can lead to the socialization of some of them. As a result, FLOSS communities are unable to clearly know the number of project newcomers as well as their identity.

In conclusion, socialization in large and complex FLOSS projects is a complex phenomenon that inherits some of the particularities of both organizational and online community socialization but that also has characteristics of its own. The following subsections review the socialization literature tradition as well as past FLOSS socialization research.

Figure 8: Development tools for the GNOME platform

Source: <https://developer.gnome.org/platform-overview/stable> (accessed on 20 November, 2013)

2.6.2 *The socialization literature tradition*

Organizational socialization has a long history in academic research particularly in the fields of applied psychology (Louis, Posner & Powell, 1983), sociology (Kanter, 1968), and management (Buchanan II, 1974). In an organizational context, socialization has been characterized as the transformation from outsider to participating and effective insider (Feldman, 1976). Socialization has been understood as the process through which individuals acquire knowledge about and adjust to their work context (C. D. Fisher, 1986; Van Maanen & Schein, 1979); but also as “the process of learning the behaviours and attitudes necessary for assuming a role in an organization” (Morrison, 1993b, p. 173). A number of socialization studies were conducted in the sixties and seventies in an attempt to better understand the integration of new employees within organizations. These research efforts essentially focused on helping companies to reach higher direct or indirect returns from heavy investments in recruitment, selection, and training (D. G. Allen, 2006). Ever since, the socialization literature has provided a rich array of results and theoretical perspectives that have the potential to help FLOSS communities in better grasping the key elements of members’ socialization experience.

2.6.3 *Past FLOSS research about socialization*

FLOSS research about how individuals external to the community are progressively socialized (with more or less success) into a FLOSS community is scarce (Ducheneaut, 2005). The review of the FLOSS literature identified a handful of studies, but none integrated theoretical insights from the socialization literature.

von Krogh et al. (2003) is among the earliest works about FLOSS socialization. A case study methodology was used to investigate community joining and member specialization during the creation of Freenet, a project aimed at developing a decentralized and anonymous peer-to-peer electronic file sharing network. The authors were particularly interested in understanding the initial contribution of Freenet joiners and their specialization within the project. First, the authors found that joiners who followed 'joining scripts' (in terms of level and type of activity) were more likely to become project members (i.e. gaining access to the CVS repository). Second, the authors found that newcomers derived learning benefits from specializing into a domain of contribution (based on prior experience). Third, the specialization of newcomers was found to be related to contribution barriers in the project. Finally, the authors found that feature gifts (providing entire project modules instead of contributing to an existing one) by newcomers emerge from the newcomers' prior domain knowledge and user experience.

The approach adopted by Ducheneaut (2005) was particularly insightful for this research. Reacting to important drawbacks of the common approach used by researchers which considers FLOSS communities as static entities rather than dynamic, Ducheneaut (2005) conceptualized socialization as a combination of two components. First, it consists of an individual learning process based on the construction of identities. Second, participation, socialization, and community reproduction are seen as inherently political. The author used a single case (the Python project) and designed what he called a "computer-aided ethnography". Based on Actor-Network theory, a visualization software was designed so as to represent the Python project as a socio-technical network (both including actors and artifacts within the same network). The results of the analysis, based on data from the CVS repository and the project mailing lists, identified four socialization trajectories; the author particularly focused on one of the trajectories, providing detailed insights about the socialization process of a single representative member. The steps followed by this member to become a core member are meticulously described and the political nature of bug resolution and later module addition is clearly explained.

Project joining processes were also studied in the context of the GNOME community (Herraiz et al., 2006). The results showed that two groups with different joining patterns could be identified. Volunteers tended to follow a step-by-step joining process, while hired developers usually experienced a "sudden" integration. This study found results that contradict the progression mechanisms forecasted by the traditional onion model (Crowston & Howison, 2005; Jensen & Scacchi, 2005). The authors found that most of the developers had committed a change in the CVS before they ever sent a bug report. Such findings confirm Ducheneaut (2005), who criticized the static nature of the onion-model and argued that it over-simplifies the integration mechanisms of

new contributors, thus highlighting a research gap in the FLOSS literature.

A situated learning perspective provided some interesting results in analyzing FLOSS communities as instances of communities of practice. Fang and Neufeld (2009) investigated the factors leading to the sustained participation of FLOSS community members based on predictions from Jean Lave and Etienne Wenger's legitimate peripheral participation theory (1991). They designed a longitudinal case study using the myPhpAdmin project and collected data from project documentations, mailing list messages, tracker messages, and CVS log files. The authors concluded that legitimate peripheral participation in FLOSS communities is an iterative process of situated learning (with "thinking" and "doing" activities), and identity construction through community recognition (identity-regulation) and self-perception (identity-work). Sustained participation is then generated when this process is positively reinforced.

Dahlander and O'Mahony (2011) characterized FLOSS communities as entities driven by lateral authority mechanisms (as opposed to vertical authority in conventional organizations). The authors examined the behaviours that predict an individual's progression to successive positions of lateral authority and how progression, once achieved, affects subsequent behaviour in the GNOME project. The authors investigated the antecedents predicting two successive stages of role progression: community member and board director. The results showed that although technical contributions are initially important to become a legitimate community member, coordination work was found to be more critical at a subsequent stage. The authors concluded that after having gained authority, individuals spent more time coordinating project work. One of the strengths of the article was its application of quantitative analysis in understanding role migration. Nonetheless, participation was again conceptualized with a programming-oriented bias that omits other types of participation within the GNOME community.

Qureshi and Fang (2010) published one of the most thorough and recent attempts to explore the phenomenon of socialization in FLOSS communities. Drawing insights from social resource theory, they studied the temporal socialization trajectories of 870 new contributors who became core members (whom they call "joiners") in 62 FLOSS projects hosted on the Sourceforge platform. The authors conceptualized the joiners' socialization within FLOSS communities as a dual process of developing and leveraging various social resources. They found that, first, joiners' socialization with core developers generally follows a non-linear socialization trajectory. Second, individual joiners begin with different initial levels of socialization and follow different patterns. Third, four trajectory classes were identified from which initial socialization and progression rate vary. Finally, statistical evidence was found to associate each class with a different time to attain the core developer status.

2.6.3.1 *Methodological concerns*

A large proportion of the previous research on FLOSS socialization has been exploratory in nature. Most of the articles reviewed here used a qualitative research strategy relying on a small number of individuals and usually a single FLOSS project. This approach had the merit of providing some in-depth understanding about an unexplored area: human behaviour in the FLOSS socialization context. However, focusing

on a single community and a limited number of subjects constrains the development of a broader and more generalizable comprehension of the phenomenon. There is a gap between understanding the inner socialization mechanisms in a few instances of socialization experience, and providing recommendations to FLOSS communities to facilitate the effective integration of newcomers. Furthermore, archival data from FLOSS project artifacts such as code repository and mailing lists were used in most cases. This raises the issue that researchers drew conclusions only based on the data that were available on the project platforms, thus limiting the scope of their conclusions. Indeed, interactions through emails, community forums, the various IRC channels, or even phone calls (or physical meetings) were simply omitted when testing the various hypotheses derived by the authors despite the potential importance of such interactions. A more global approach addressing an individuals' entire socialization experience may provide a more encompassing view of the phenomenon. An appropriate strategy to implement such approach would be to rely on the use of primary data, gathered from the individuals themselves. The use of a confirmatory approach would also help to improve the validation of the provided results and would potentially allow predictions. Finally, as noted previously (see section 2.3.8), the role played by communities in influencing the various socialization experiences has been totally omitted in most cases.

2.6.3.2 *Theoretical concerns*

The various theoretical foundations that were used to study FLOSS socialization appeared to be only partially satisfactory considering the approach adopted in this research as well as the nature of the social entities under investigation. Exploratory studies such as von Krogh et al. (2003) or Herraiz et al. (2006) relied on a grounded approach leading to a deeper comprehension of the socialization phenomenon but no generalizable model that could allow predictions. Apart from grounded theory research, past FLOSS socialization relied on the use of lateral authority theory (Follett, 2003), social resource theory (Lin, Ensel & Vaughn, 1981), legitimate peripheral participation (Lave & Wenger, 1991), and actor-network theory (Latour, 1987). Overall, all theories were used to focus on software artifact development thus omitting the community layer surrounding FLOSS projects as well as all types of contributions that benefit a project's community apart from code-related tasks. By adopting such software-centric view, authors have focused on the linear progression of members from being peripheral lurkers towards becoming core developers (Ducheneaut, 2005; Fang & Neufeld, 2009; Herraiz et al., 2006; von Krogh et al., 2003), or foundation members or board directors (Dahlander & O'Mahony, 2011). As a consequence, such approaches would be restrictive in the context of large and complex FLOSS projects as they would leave aside valuable code-related contributors who did not become core developers even though they are considered as valuable project contributors but also all the other contributors who strongly benefit a project's community and thus ensures its survival, success, and long-term sustainability.

Theories such as lateral authority theory focuses solely on code-related contributions and thus omits the social side of the socialization phenomenon. Lateral authority theory can be used to explain how developers provide signals through their code contributions that trigger peer recognition and help them move towards further developer-related socialization stages. Lateral authority is indeed a key mechanism that charac-

terizes newcomer progression but the theory omits other important social and technical aspects that must be taken into consideration when attempting to derive an encompassing socialization model predicting newcomer behaviour based on his/her socialization experience.

Some of the reviewed theories only studied the social side of FLOSS socialization. Social resource theory explains that certain newcomers move further or faster than other individuals because of various interaction patterns with recognized project members. The theory can be used in the context of large and complex projects as it could be applied to all project contributors without necessarily focusing on code-related tasks. However, interactions with the software artifacts as well as the nature of the tasks that are performed would be omitted.

Actor-network theory (ANT) is among the candidate theories that consider both types of interactions and thus appears more encompassing. Ducheneaut (2005) relied on ANT's principles by considering that project members interact both with social and material components. However, the author did not directly apply the theory itself as ANT is mainly used to provide a deep and detailed understanding about a given phenomenon but it does not allow to identify measurable constructs and causal relationships connecting them. As a consequence, ANT could be used to study in depth the socialization experience of certain individuals in a given large and complex FLOSS project but ANT cannot help derive a predictive and generalizable FLOSS socialization model.

Finally, legitimate peripheral theory (LPP) also considers interactions with both social and technical components. Relying on the idea that FLOSS socialization is totally decentralized and that FLOSS projects are instances of communities of practice, the theory explains how participation, situated learning, and identity construction are inter-related and co-evolve. However, LPP provides a thorough understanding of only one type of socialization experience wherein FLOSS projects are seen as passive entities and where individuals entirely shape and decide how they want to participate. Such assumption is erroneous in the case of large and complex FLOSS projects as it was earlier demonstrated in this chapter that these entities are community-organization hybrids in which socialization can have aspects that pertain to organizational contexts. In short, large and complex FLOSS projects cannot be considered as instances of communities of practice which thus seriously compromises the validity and applicability of LPP in the context of large and complex FLOSS projects.

Overall, the review of past FLOSS socialization research has highlighted the need for stronger and more encompassing theoretical foundations when studying socialization in the context of large and complex FLOSS projects.

2.6.4 *Van Maanen and Schein's socialization tactics*

If one had to choose a single study that has been the most insightful and most widely used in socialization research, Van Maanen and Schein's work (1979) would be among the most likely candidates to be selected. The primary concern of the authors was to understand why certain patterns of thought and behaviour are passed from one generation of organizational members to the next whereas others are not. To answer

the question van Maanen and Schein looked into the role played by organizations in the socialization phase of their employees. Socialization processes vary considerably ranging from:

A relatively quick, self-guided, trial-and-error process to a far more elaborate one requiring a lengthy preparation period of education and training followed by an equally drawn out period of official apprenticeship. (p. 3)

The initial work from van Maanen and Schein (1979) offered a theoretical explanation about how the various socialization techniques used by organizations have an influence on the role orientation of new employees (in terms of degree of innovativeness). According to their theory, newcomers respond to their roles differently because the socialization tactics that are used have an influence on how information is received (Jones, 1986). The authors suggested that the tactics organizations use once newcomers enter the organization could be classified into six dimensions: collective-individual, formal-informal, sequential-random, fixed-variable, serial-disjunctive, and investiture-divestiture. The initial definition of each dimension is provided in Table 1.

Dimension name	Definition from Van Maanen and Schein (1979)
COLLECTIVE (VS. INDIVIDUAL)	<i>the tactic of taking a group of recruits who are facing a given boundary passage and putting them through a common set of experiences together.</i>
FORMAL (VS. INFORMAL)	<i>those processes in which a newcomer is more or less segregated from regular organizational members while being put through a set of experiences tailored explicitly for the newcomer.</i>
SEQUENTIAL (VS. RANDOM)	<i>the degree to which the organization or occupation specifies a given sequence of discrete and identifiable steps leading to the-target role.</i>
FIXED (VS. VARIABLE)	<i>the degree to which the steps involved in a socialization process have a timetable associated with them that is both adhered to by the organization and communicated to the recruit.</i>
SERIAL (VS. DISJUNCTIVE)	<i>a process in which experienced members of the organization groom newcomers who are about to assume similar kinds of positions in the organization.</i>
INVESTITURE (VS. DIVESTITURE)	<i>the degree to which a socialization process is constructed to either confirm or disconfirm the entering identity of the recruit.</i>

Table 1: Van Maanen and Schein's socialization dimensions (1979)

2.6.5 *Refinements of Van Maanen and Schein's model by Jones (1986)*

Van Maanen and Schein's path-breaking conceptual typology has received strong support in academic research for the last 30 years. The model is still used in current research works (Bauer, Bodner, Erdogan, Truxillo & Tucker, 2007; Lopez & McMillan-Capehart, 2009; Saks, Uggerslev & Fassina, 2007; Simosi, 2010). However, researchers have mostly relied on the later advancements developed by Jones (1986) who refined Van Maanen and Schein's dimensions and regrouped them into broader categories.

Jones (1986) examined the impact of socialization tactics on role orientation and personal adjustment seen in terms of role conflict, role ambiguity, and intention to quit. His study also investigated the moderating effect of self-efficacy between socialization tactics and role orientation. The results confirmed that different patterns of socialization had indeed a significant influence on both role orientation and personal adjustment. The moderating effect of self-efficacy was also empirically verified.

Jones (1986) developed six 5-item self-report scales to measure each of the six socialization dimensions. He then grouped these six tactics into three factors on a single institutionalization continuum (see Figure 9). After having computed a factor analysis of the six dimensions and realized high positive intercorrelations among some of the tactics, Jones argued for the existence of three broader factors (the following definitions were provided by Bauer et al. (2007) in a meta-analysis of socialization research):

1. Context tactics (collective vs. individual and formal vs. informal): Learning task requirements as part of a group and having formal training before starting the actual job.
2. Content tactics (sequential vs. random and fixed vs. variable): Clear stages exist for training, and there is a clear timetable for role adjustment.
3. Social tactics (serial vs. disjuncture and investiture vs. divestiture): Receiving positive feedback and identity affirmation from organizational insiders and having a trusted insider to guide them within the organization.

Overall, research has highlighted the predominant role played by social tactics (as compared to content and context) in influencing newcomer adjustment the most significantly (Cable & Parsons, 2001; Jones, 1986; Saks et al., 2007; Simosi, 2010).

Figure 9: Jones' (1986) classification of socialization tactics

Tactics concerned mainly with:	CONTEXT ASPECTS	INSTITUTIONALIZED	INDIVIDUALIZED
		Collective	Individual
		Formal	Informal
	CONTENT ASPECTS	Sequential	Random
		Fixed	Variable
	SOCIAL ASPECTS	Serial	Disjunctive
		Investiture	Divestiture

Research in the domain of socialization has relied on Jones' works to conceptualize and measure the dimensions of socialization. The multi-dimensionality of Jones' model has generated three overall perspectives in dealing with socialization. Some researchers have considered socialization tactics as a unidimensional spectrum that consists of institutionalized (collective, formal, sequential, fixed, serial, investiture) and individualized (individual, informal, random, variable, disjunctive, and divestiture) tactics (Ashforth & Saks, 1996; Gruman, Saks & Zweig, 2006). Others have considered the six original tactics separately (D. G. Allen, 2006; Saks & Ashforth, 1997b). Finally, an intermediate approach has been to consider socialization characterized by Jones' three dimensions: content (sequential vs. random, fixed vs. variable), context (collective vs. individual, formal vs. informal), and social (serial vs. disjunctive, investiture vs. divestiture) (Cable & Parsons, 2001; Saks et al., 2007). The justifications for selecting one of the three perspectives vary with a set of theoretical, practical, or empirical reasons provided by authors.

2.6.6 Aspects of Van Maanen and Schein's socialization tactics in large and complex FLOSS projects

Even though van Maanen and Schein's socialization theory was developed in the context of traditional organizations, it was considered to be a promising candidate to study FLOSS socialization in large and complex FLOSS projects. Indeed, the six dimensions describing socialization cover a wide array of socialization types that go well beyond standard organization-based socialization. The sub-categorization of context, content, and social socialization aspects introduced by Jones, allows to break down socialization into three general domains that can be applied to FLOSS projects. Besides, the other sub-categorization provided by Jones that differentiates institutionalized versus individualized socialization tactics strongly echoes the dual nature of large and complex FLOSS projects: organization-community hybrids. Van Maanen and Schein's socialization theory thus encompasses the two ends of the FLOSS socialization spectrum: from a socialization experience that is entirely driven by the newcomer (individualized) to a socialization that is totally managed by a FLOSS project and its

community (institutionalized).

FLOSS communities have launched a wide array of initiatives that facilitate the induction of newcomers. Some of them exhibit characteristics of van Maanen and Schein's socialization tactics.

KernelNewbie¹⁶ is a Linux Kernel sub-community that involves both newcomers and experienced members of the Linux community. It has a specific mailing list, an IRC channel, a forum, and a wiki for participants to interact. The sub-community relies on three projects: KernelMentors, KernelJanitors, and KernelProjects that help assign appropriate tasks to newcomers.

The creation of such a sub-community is an example of a *formal* socialization mechanism. Indeed, the Linux Kernel community has explicitly tailored initiatives to help new members in becoming involved into the community. A sub-community dedicated to newcomers, such as KernelNewbie, encourages new members to interact together and provide some help that they may not get from more experienced individuals. Figure 11 illustrates a post on the KernelNewbie forum in which a newcomer requests some help about one of the kernel programming tutorials. A response was posted by another newcomer who encountered the same problem. This exemplifies some aspects of the *collective* dimension of socialization. In particular, it illustrates *context* socialization aspects of socialization according to Jones' categorization.

Furthermore, KernelNewbie clearly encourages the mentoring of newcomers by experienced participants thus highlighting aspects of *social* aspects of socialization tactics. The Google Summer of Code is another example. From 2005, the program has brought together nearly 2500 student participants and 2500 mentors from 98 countries worldwide in various FLOSS projects. Mentoring was found as a solution to bridge the gap between core developers and "newbies" (or novices) by enhancing the acceptance and integration of new developers within FLOSS communities and by linking newcomers to role models that demonstrate good community practices. Similarly, communities such as Apache¹⁷ or Ubuntu¹⁸ have implemented formal mentoring programmes.

Finally, research suggests that newcomers have two different joining patterns (Her-
raiz et al., 2006). The authors found that volunteers tend to follow a step-by-step joining process, while firm-sponsored developers usually experience a "sudden" integration. The results were also confirmed by Shibuya and Tamai (2009). This supports the importance of *content* aspects of socialization, particularly the sequential dimension in which individuals go through a fixed series of tasks instead of getting involved in random ones. Resources provided by FLOSS communities to guide newcomers usually specify a sequence of tasks to be performed before engaging into full participation. For instance, the GNOME website provides a step-by-step guide on how to start contributing to GNOME¹⁹ and indicates what should be done before starting to submit a patch. They suggest tasks such as reading the website, joining relevant mailing lists and IRC channels, studying the library API documentation, consulting documentation

¹⁶ <http://kernelnewbies.org/>

¹⁷ <http://community.apache.org/mentoringprogramme.html>

¹⁸ <https://wiki.ubuntu.com/MOTU/Mentoring>

¹⁹ <http://live.gnome.org/GnomeLove/HowToStart%20En>

Figure 10: Linux kernel newbie community

Source: <http://kernelnewbies.org> (accessed on 20 May, 2011)

on modules/libraries, going through module tutorials, and so forth.

In conclusion, it seems that individuals who join FLOSS communities experience different types of socialization which are consistent with some of Van Maanen and Schein's socialization dimensions. While their overall model may not fully apply to the FLOSS context because of its strong organizational connotations, it appears that Van Maanen and Schein's work provides good insights in understanding the experience of newcomers within FLOSS communities. This research therefore adopts Van Maanen and Schein's theory as a guiding theoretical lens that is used to investigate the socialization phenomenon in FLOSS communities and to identify the key aspects that characterize FLOSS socialization.

2.6.7 Overall socialization framework in organizational behaviour

Social scientists have investigated the organizational adjustment phenomenon by either examining variables that are more "proximal" to the employee adjustment process or aspects that are more global (or "distal") indicators of newcomer adjustment (Kammeyer-Mueller & Wanberg, 2003; Saks et al., 2007). Socialization research relies on the assumption that the socialization methods employed by organizations (e.g. orientation programs, socialization tactics, proactive tactics, mentoring) have an impact on immediate or "proximal" socialization outcomes (e.g. learning, task mastery, social integration, role conflict, role ambiguity, and person-job and person-organization fit). Proximal outcomes have in turn, an influence on distal (or longer-term) outcomes such as job satisfaction, organizational commitment, organizational identification, organizational citizenship behaviour, job performance, stress, and turnover (Johns & Saks, 2001).

Figure 11: Example of collective socialization in the KernelNewbie community

Source: <http://forum.kernelnewbies.org/read.php?6,214> (accessed on 20 May, 2011)

Proximal socialization outcomes are the results of an employees' interactions during the socialization process and may thus be regarded as the contents of socialization (Chang, Chang & Jacobs, 2009; Chao, O'Leary-Kelly, Wolf, Klein & Gardner, 1994). In brief, some researchers have investigated the direct effect of socialization on distal outcomes, whereas others have argued that the effect of socialization on distal adjustment outcomes is mediated by proximal adjustment outcomes (Saks et al., 2007). The model adopted in the meta-analysis of the socialization compiled by Saks et al. (2007) comprises an overall theoretical framework of socialization that has been acknowledged and empirically verified by socialization researchers (see Figure 12). The model assumes both a direct effect of socialization on distal outcomes and a mediation by proximal outcomes.

1995; Ashford & Black, 1996; Wanberg & Kammeyer-Mueller, 2000), organizational commitment (Adkins, 1995; N. J. Allen & Meyer, 1990b; Bauer & Green, 1998), and turnover (Adkins, 1995; D. G. Allen, 2006; Saks & Ashforth, 1997b; Wanberg & Kammeyer-Mueller, 2000).

A review of the socialization literature found a lack of research in behavioural outcomes other than performance and turnover (Saks & Ashforth, 1997b). The authors concluded that more emphasis should be placed on other types of behaviours such as organizational citizenship behaviours. Section 2.4.2 showed that citizenship behaviours were extra-role behaviours that positively contribute to the effectiveness of organizations. The literature review performed in this research confirmed Saks and Ashforth's statement. This research project proposes to adopt a wider understanding of performance in a FLOSS socialization context and will thus contribute to the socialization body of literature.

2.6.8.2 *Influence of newcomer socialization on individual performance*

Early socialization research works raised the importance of measuring socialization outcomes in light of performance-related behavioural factors such as carrying out role assignments dependably, remaining with the organization, and innovation/spontaneous cooperation (Feldman, 1976, 1981; Porter, Lawler & Hackman, 1975). Early socialization works from Daniel Charles Feldman adapted Katz's categorization in the context of socialization (1981). The author proposed a socialization model where both in-role and extra-role performance are indicators of an individual's socialization progress.

The socialization literature has relied on various theoretical approaches such as organizational socialization tactics (Ashforth & Saks, 1996; Jones, 1986; Van Maanen & Schein, 1979), newcomer sense-making behaviours Ashford (1986); Morrison (1993b); Wanberg and Kammeyer-Mueller (2000), or newcomer experiences (Boswell, Shipp, Payne & Culbertson, 2009; Chan & Schmitt, 2000). Others have studied socialization from a training perspective (Klein & Weaver, 2000; Saks, 1995) or in the particular case of mentoring practices (T. D. Allen, Eby & Lentz, 2006; Noe, 1988; Wanberg, Kammeyermueller & Marchese, 2006). Whichever perspective is taken, the socialization body of research has a long tradition in investigating factors that lead to higher newcomer performance.

Research that relied on the conceptual work from Van Maanen and Schein (1979), operationalized socialization tactics using three different strategies inferred from the early two-dimensional categorization of Van Maanen's socialization tactics by Gareth Jones (1986). Some researchers have considered socialization tactics as a unidimensional spectrum that consists of institutionalized (collective, formal, sequential, fixed, serial, investiture) and individualized (individual, informal, random, variable, disjunctive, and divestiture) tactics (Ashforth & Saks, 1996; Gruman et al., 2006). Others have considered the six tactics separately (D. G. Allen, 2006; Saks & Ashforth, 1997b). Finally, an intermediate approach has been to consider socialization characterized by three dimensions: content (sequential vs. random, fixed vs. variable), context (collective vs. individual, formal vs. informal), and social (serial vs. disjunctive, investiture vs. divestiture) aspects (Cable & Parsons, 2001; Saks et al., 2007). All three approaches

have been used in studying the impact of socialization tactics on individual performance.

For instance, a longitudinal field study investigated the impact of socialization tactics on various proximal and distal outcomes (Ashforth & Saks, 1996). The authors found that self-appraised performance was associated with more individualized socialization. However, conflicting results were provided by Saks et al. (2007) in their meta-analysis of socialization research: they concluded that institutionalized socialization tactics were positively related to job performance. In their meta-analysis, Saks et al. (2007) also compiled the regression coefficients concerning relative predictive strength of context, content, and social tactics on performance and respectively found coefficient of 0.147, 0.115, and 0.175 (all significant with $p < 0.01$).

Ashforth, Saks and Lee (1997) investigated the dimensionality of the operationalization of Van Maanen and Schein's tactics and collected data from business school graduates after 4 months ($n = 295$) and 10 months ($n = 222$). Among the set of results, the authors found investiture tactics (correlation coefficient of 0.20) and social tactics (coefficient of 0.10) to be correlated with performance. The authors concluded the six-factor model had a better fit than the 3-factor model when studying factors such as stress symptoms, role innovation, or performance.

Furthermore, an extensive review of the socialization literature within the domain of virtual communities concluded that by using formal, sequential and collective socialization tactics, new members were likely to learn better how to behave in a community and contribute more (Kraut, Burke & Riedl, 2011). The authors acknowledged the presence of various socialization initiatives within online communities and encouraged further research efforts in this area.

2.6.8.3 *Proximal socialization outcomes*

The socialization literature has provided a wide array of proximal socialization outcomes as predictors of distal outcomes. A thorough review of the literature revealed three broad categories of proximal factors: cognitive, identity-related, and relationship-related factors.

2.6.8.4 *Cognitive factors*

Some studies have assessed newcomer adjustment in terms of learning associated to assigned tasks and roles (Adkins, 1995). One of the primary challenges facing newcomers is learning how to perform their job (Feldman, 1976; Louis, 1980; Morrison, 1993a). To do so, newcomers need to acquire necessary job skills and knowledge. Second a related requirement for newcomers concerns clarifying their role in the organization (Feldman, 1976; Louis, 1980; Morrison, 1993a). Achieving role clarity is equivalent to understanding the behaviours that others expect in relation to a given role (Morrison, 1993a). Since the 1990s, there has been a consensus in the literature that learning is the core of organizational socialization, revealing the extent to which newcomers are socialized (Chao et al., 1994; Saks & Ashforth, 1997a). Later advancements in the socialization stream of research re-centered the socialization debate around the notion of learning in which newcomer learning is argued to be "at the heart of any organizational socialization model" (Cooper-Thomas & Anderson, 2005,

p. 117).

Some studies from the socialization literature included both task mastery (or competence) and role clarity as proximal socialization outcomes (Adkins, 1995; Morrison, 1993a; Wanberg & Kammeyer-Mueller, 2000). Other research projects have investigated different learning-related factors such as perceived self-efficacy (Bauer et al., 2007; Bauer & Green, 1994, 1998; Jones, 1986; Major & Kozlowski, 1997) or performance efficacy (Bauer & Green, 1998).

2.6.8.5 *Identity-related factors*

According to self-categorization and social identity theory (Tajfel & Turner, 1986; Turner, Hogg, Oakes, Reicher & Wetherell, 1987), an individual's identity has two main components: (1) a personal identity that entails the characteristics which differentiate the individual from the group; and (2) one (or several) social identity which refers to categorization of the self into one (or several) social group(s).

In a socialization context, newcomers are concerned with building both situational (Saks & Ashforth, 1997a) and self definitions (Ashforth & Mael, 1989). In doing so, people's conceptions of the self which are learned by interpreting the responses of others in situated social interactions, are affected during their socialization phase (Van Maanen & Schein, 1979). As a result, since socialization tactics vary in terms of social interactions (e.g. individual vs. collective or serial vs. disjunctive), conceptions of the self are impacted differently depending on the predominant socialization dimensions experienced by a newcomer. Ashforth and Mael (1989) argue that socialization influences identification, which in turn has an impact on internalization. Following the authors' seminal work in analyzing socialization through the lens of social identity theory, socialization researchers have found some empirical support for the important role played by organizational identification (as a proximal outcome) on various distal outcomes.

For example, social identification was found to be one of the most influential proximal outcomes of organizational socialization (Saks & Ashforth, 1997b), having an impact on distal outcomes such as job performance or OCB. In other studies serial tactics were found to have an influence on organizational socialization (Yi & Uen, 2006) as well as institutionalized socialization tactics (Ashforth, 1998; Ashforth & Saks, 1996) and investiture tactics (Ashforth et al., 2007). Organizational socialization was also treated as a distal outcome and was found to be affected by newcomer learning (Ashforth et al., 2007).

2.6.8.6 *Relationship-related factors*

The development of interpersonal relationships in the workplace has been shown to be important in providing social support when adjusting to a new organizational environment (C. D. Fisher, 1985). Social integration has been shown to be an important outcome of employee socialization (Gruman et al., 2006; Kammeyer-Mueller & Wanberg, 2003; Wanberg & Kammeyer-Mueller, 2000). Social integration is about the development of relationships with peers and co-workers (Morrison, 1993a). Institutionalized tactics and self-efficacy were found to have an impact on social integration (Gruman et al., 2006). Social integration was also found to be influenced by various

types of proactive socialization behaviours such as information seeking (Morrison, 1993a) feedback seeking (Gruman et al., 2006), and relationship building (Wanberg & Kammeyer-Mueller, 2000). Using a social capital perspective, Morrison (2002b) concluded that social integration was negatively related to friendship network size and positively related to friendship network tie strength.

2.6.8.7 Literature gap - Relationship between socialization and citizenship behaviours

The review of the socialization literature revealed that few research projects attempted to bridge the socialization literature and the in-role/extra-role performance literature. Task-related performance has been used as the main measure of newcomer performance even though it is well-acknowledged that extra-role employee performance facilitates organizational productivity (P. Podsakoff, MacKenzie & Ahearne, 1997), thus highlighting the need to identify the socialization factors that engender citizenship behaviours.

This research gap was pointed out by earlier research works (G. Chen, 2005; G. Chen & Klimoski, 2003; Cooper-Thomas & Anderson, 2006). For instance, G. Chen and Klimoski (2003) noted that prior socialization studies examined performance by only looking at task performance (Morrison, 1993a; Saks, 1995) thus ignoring broader conceptualizations of performance which have been demonstrated to be more representative of the construct (P. M. Podsakoff, Mackenzie, Paine & Bachrach, 2000; Welbourne, Johnson & Erez, 1998).

A more recent theoretical work from Cooper-Thomas and Anderson (2006) introduced a meta-level socialization model based on an extensive review of organizational socialization research. The authors reflected on the multi-faceted nature of individual socialization success and derived a multi-level learning-focused model based on an overview of past socialization research. Acknowledging that individual performance is a core success indicator, Cooper-Thomas and Anderson identified a large research gap the socialization literature concerning measuring the impact of socialization on employee extra-role performance.

Nearly a decade before Cooper-Thomas and Anderson's theoretical work, Saks and Ashforth (1997a) made the same claim by realizing the overall lack of research that studies behavioural outcomes of socialization other than turnover. The authors then suggested the relevance of studying organizational citizenship behaviours as an outcome of socialization and included OCB in their socialization model as a key individual level outcome.

As a result, there is a research gap in understanding the impact of newcomer socialization from a more encompassing view of performance that goes beyond the mere notion of task performance. This lack of research appears particularly important in the FLOSS community context as this chapter has demonstrated that the functioning and sustainability of FLOSS communities strongly rely on extra-project behaviours (such as helping other members or promoting the project to outsiders). This research proposes to adopt the well-acknowledged theoretical view that sees performance as being the combination of in-role and extra-role performance. In this study, the two behavioural dependent variables are: *task performance* and *community citizenship be-*

haviours (CCB). The term *task performance* was preferred to in-role performance since contrary to organizations, FLOSS project contributors are not assigned specific roles but rather perform multiple roles and freely choose how they want to contribute to the software project. The term *Community citizenship behaviours* (CCB) was derived from *organizational citizenship behaviours* (OCB), a term that has been commonly used by social scientists since the beginnings of extra-role behaviours research (Organ, 1988; Organ & Konovsky, 1989).

2.7 SUMMARY OF LITERATURE REVIEW AND RESEARCH GAPS

This chapter has presented an overview of the various streams of research that were explored to gain a thorough understanding about the existing bodies of knowledge that surround the main inquiry of this research. This section summarizes the main findings and research gaps that were identified during the review of the literature. The subsidiary research questions that were generated based on the outcomes of the literature review are also introduced.

The purpose of the literature review chapter is two-fold. First, it introduced the notion of FLOSS community. Since this research proposes to study socialization and contributor behaviour in the FLOSS context, it is important to first have a clear understanding of the functioning of such social entities. Second, it established the theoretical foundations that will be used in exploring the overall research question of this study:

What socialization factors influence contributor behaviour in FLOSS communities?

Section 2.3 began with a brief presentation of the FLOSS movement and presented its main principles. Then, it described various aspects of FLOSS communities such as details about the creation of a FLOSS project, the artifacts that are commonly used, and the various tasks and roles covered by the community members. The section concludes with claiming an overall need to bridge the gap between academic research and FLOSS practitioners by exploring the issues such as FLOSS community success and sustainability.

Section 2.4 reviewed individual contribution behaviour and performance in the FLOSS context and addressed some drawbacks in past and current research investigating such issues. The section highlighted a need for FLOSS researchers to reconceptualize contribution behaviour by adopting a broader view about tasks that are performed within FLOSS communities. It then suggested drawing insights from Katz (1964)'s works about individual behavioural requirements in organizations, and introduced the notions of task performance and community citizenship behaviours (respectively derived from the notions of in-role and extra-role performance). Acknowledging the overall lack of empirical evidence about the nature and different types of citizenship behaviours, the following subsidiary research question was then derived:

1- What are the different instances of citizenship behaviour in the FLOSS community context?

Section 2.5 justified the advantages of using a socialization perspective in exploring this issue. Section 2.6 provided an overview of socialization research in organizational

behaviour and FLOSS streams of research. Then, Van Maanen and Schein (1979)'s model of socialization and its revision by Jones (1986) were introduced, reviewed, and then analyzed within the FLOSS context in an attempt to justify their theoretical relevance for this study. Van Maanen and Schein's theoretical work was shown to provide interesting insights towards grasping the various facets of an individual's FLOSS community newcomer experience. However, this chapter raised some concern about the direct applicability Van Maanen and Schein's socialization theory due to important differences between traditional organizational contexts and FLOSS communities. The review of the literature did not clearly identify the key aspects that characterize a FLOSS community member's newcomer experience. The section concluded that a preliminary phase of this research should explore FLOSS community newcomer experience through the collection of qualitative data, using Van Maanen and Schein's work as a theoretical lens. This led to the following subsidiary question:

2- What are the key factors that characterize the socialization experience of newcomers in FLOSS communities?

The overall socialization framework adopted in organizational behaviour research was then presented in terms of socialization factors, proximal outcomes, and distal outcomes. The various types of proximal socialization outcomes were reviewed. The section concluded with the identification of a research gap concerning the exploration of the impact of socialization on a wider perspective of performance encompassing both in-role and extra-role behaviours. This led to the following subsidiary research questions:

3- What are the important proximal socialization factors in the FLOSS community context?

4- What is the influence of such proximal socialization factors on contributor behaviour?

5- What socialization factors influence the identified proximal socialization factors?

The following subsection integrates the above findings and derives the two initial research models that guided the first phase of this research project.

2.8 PRELIMINARY RESEARCH MODELS

This subsection introduces the two initial research models that were derived from the findings of the literature review. The literature review revealed two overall approaches when studying the impact of socialization on performance: a direct effect between socialization factors and individual performance, and a mediated effect through the influence of proximal socialization outcomes. The evidence collected through the review of the relevant streams of research did not support a confidential conclusion about the most appropriate approach. As a result, it was decided to test and compare two models respectively representing the direct and indirect effect of socialization factors. Both models were later refined into testable models after the first data collection phase of this research (first stage of the three-phase research methodology).

2.8.1 *Conceptualization of the main notions*

The review of the extant literature supported the following definition of FLOSS community socialization experience as well as the conceptualizations of the two dependent variables: task performance and FLOSS community citizenship behaviours (or CCB).

2.8.1.1 *Socialization experience*

This research conceptualizes socialization according to [Bauer et al. \(2007\)](#)'s definition. FLOSS newcomer socialization is defined as *the process by which newcomers make the transition from being community outsiders to being insiders*. This thesis builds on the socialization tradition in the organizational behaviour literature, in particular the seminal work of John Van Maanen and Edgar Schein ([1979](#)). Their socialization model has received a large amount of empirical support since they first proposed it.

Since this research project adopts a view of socialization from the newcomer perspective, each community member has the potential to experience a unique socialization process the outcome of which varies from individual to individual. As a result, it was decided to use the term *socialization experience* in this research in order to emphasize that the socialization of FLOSS newcomers is to some extent facilitated by communities but also shaped by the newcomers themselves.

2.8.1.2 *Dependent variables: Task performance and CCB*

This research conceptualizes FLOSS community contributor performance as task performance and community citizenship behaviour. The approach adopted in this research is strongly inspired by the earlier works from Katz ([1964](#); [1966](#)).

2.8.1.3 *Task performance*

According to Katz, dependable role performance is the most conspicuous behavioural requirement of the organization ([Katz, 1964](#); [Katz & Khan, 1966](#)). Assigned roles must be carried out and must meet some minimal level of quantity and quality of performance. Katz argues that compliance with the legitimate rules of an organization primarily influences only reliable role performance. He thus emphasized that performance is assessed according to the norms, standards and rules of a given organization. In the FLOSS context, the notion of role is more complex than in a traditional organization since individuals are often self-assigned ([Crowston, Wei et al., 2005](#)) and may have to perform roles from various domains such as quality assurance roles, source code creation roles, and source code versioning roles, but also usability, marketing, and licensing roles ([Jensen & Scacchi, 2007](#)). In this study, *task performance* is defined as *the extent to which an individual carries out assigned tasks by meeting some minimal level of performance defined by the FLOSS community to which he/she belongs*. Tasks may be either self-assigned, formally or informally assigned by other community members to the individual. Performance criteria encompass both technical and organizational aspects of tasks performed in community projects.

2.8.1.4 Community citizenship behaviours

Reflecting on the behavioural requirements necessary to perform innovative and spontaneous activities (Katz, 1964), Katz and Kahn (1966) noted that organizations require more than dependable assignment performance in order to be effective. They said:

“No organizational planning can foresee all contingencies within its own operations, can anticipate with perfect accuracy all environmental changes, or can control perfectly all human variability.... An organization which depends solely upon its blueprint of prescribed behaviour is a very fragile social system” (1966, p. 338).

In the context of FLOSS communities, this research project introduces the notion of FLOSS community citizenship behaviour (CCB) following Organ’s earlier definition. CCB is defined as *behaviours that are discretionary, not directly or explicitly recognized by the formal reward system of the community and that in the aggregate promote the effective functioning of the community.*

2.8.2 Preliminary research models

The findings from the literature review suggest the conceptual model presented in Figure 14. It assumes the existence of a direct influence of FLOSS community socialization experience on contributor performance (socialization distal outcome). The review of the performance-related literature applied to the FLOSS community context has revealed the pertinence to adopt a wider view towards individual performance consisting of task performance and community citizenship behaviour.

Figure 14: First conceptual model

Furthermore, the socialization literature has also demonstrated the importance of proximal socialization factors in explaining distal outcomes. This suggests that in the FLOSS community context, socialization factors may have an influence on contributor behaviour through the mediation of FLOSS-specific proximal outcomes. It thus suggests the conceptual model from Figure 15.

Because of the exploratory nature of this research due to the lack of empirical evidence in past FLOSS socialization research, this research investigates both models (direct and mediated influence of socialization on performance) in an attempt to iden-

tify which model better describes the actual FLOSS community context.

Chapter 3 introduces the research design and the methodology that were implemented to carry out the investigation. Chapter 4 presents the results of the qualitative phase of this research which allowed to identify key FLOSS socialization factors. Chapter 5 integrates the findings from the literature review and the results from the qualitative phase to generate the two conceptual models.

Figure 15: Second conceptual model

RESEARCH DESIGN

This chapter begins with an explanation of the philosophical paradigm adopted in this research, followed by some prior insights about methodology selection.

3.1 THEORETICAL PERSPECTIVE OF THIS RESEARCH

When opting for a particular theoretical perspective and appropriate methodology in a research project, it is important to first clearly identify ontology and epistemology. Both ontology and epistemology underlie the methodology choice. Ontology comes from the domain of philosophy, “that branch of philosophy which deals with the order and structure of reality in the broadest sense possible” (Angeles, 1981). Wand and Weber (1993) defines ontology as “a branch of philosophy concerned with articulating the nature and structure of the world” (p.220). Epistemology, in contrast, is not about the world itself or the nature of phenomena, but rather on the nature of knowledge itself. It concerns “the nature of human knowledge and understanding that can possibly be acquired through different types of inquiry and alternative methods of investigation” (Hirschheim, Klein & Lyytinen, 1995, p. 20).

3.1.1 *Ontology*

Objectivism and constructionism are the two main ontologies used in social science research. From an objectivistic point of view, the world is comprised of distinct objects with properties independent of the inquiring observer (Hirschheim et al., 1995, p. 58). According to objectivism, if two observers understand a phenomenon differently, then at least one of them is wrong due to human imperfection (lack of training, errors of judgement, etc.). In contrast, the constructivist view posits that realities are apprehended solely in multiple mental constructions that are socially and experientially generated (Guba & Lincoln, 1994). The notion of reality is thus relative and specific to the individual persons who form the mental constructions.

Objectivist research relies on the use of measurements to assess realities. Quantitative methods such as surveys or experiments as well as statistical analysis techniques are often used to compute generalisable results. Constructivist research is more inclined to use qualitative methods whose overall goal is to develop a rich description and understanding based on careful observations of individuals.

This research project aims at generating generalisable results about socialization experience and contributor behaviour within FLOSS communities. The findings are expected to provide insights about behavioural phenomena that occur in ‘all’ FLOSS communities to a certain extent. It thus assumes a view of reality in which FLOSS communities are distinct objects with rules and properties that are independent of observers. This research assumes an objectivist ontology.

3.1.2 *Epistemology*

Epistemological assumptions are not concerned with the nature of reality but rather with the nature of knowledge itself and with proper methods of inquiry (the means by which knowledge can be obtained) (Iivari, Hirschheim & Klein, 1998). Epistemologies can be mapped into a positivism-antipositivism spectrum (Iivari et al., 1998). Antipositivism is also called interpretivism.

Positivist researchers seek to explain and predict phenomena in the social world by searching for common patterns and causal relationships among constituents of phenomena (Iivari et al., 1998), typically investigated with structured instrumentation (Orlikowski & Baroudi, 1991). The primary purpose of positivist studies is to develop and test theories in an attempt to increase predictive understanding of phenomena (Orlikowski & Baroudi, 1991).

Antipositivist (interpretivist) researchers believe that the social world can solely be understood from the viewpoint of the individuals who are directly involved in the phenomena under investigation (Iivari et al., 1998). Interpretivist studies strive to understand phenomena by observing and capturing the meanings that participants assign to them (Orlikowski & Baroudi, 1991). In contrast to positivism, interpretivism does not aim at generalizing the findings from a particular setting to a population. Instead, the intent is to gain a deep understanding of the structure of a phenomenon, which in turn is believed to inform about other settings that have some similarities with the studied phenomenon (Orlikowski & Baroudi, 1991). In interpretivist studies, theory is considered as an end product and studies are not aimed at making predictions (Gregor, 2006).

More recently, the dichotomous view positivism/interpretivism has raised skepticism among researchers. Some scholars have argued for supplanting positivism in favour of a new conception, post-positivism, in which knowledge acquisition in the social sciences differs from the pure physical science model (Hirschheim, 1985). Post-positivists and positivists believe in the existence of an external reality independent of the observer. However, post-positivism differs from positivism by claiming that some aspects of human experience such as feeling or thinking are not directly observable. As a consequence, an important tenet of post-positivism is the belief in what is called 'methodological pluralism', the assertion that there is not just one correct method of science but many (Hirschheim, 1985). Post-positivism encourages researchers to view their comprehension of truth from a probability angle rather than through certainty claims (Mertens, 2009).

This study mainly investigates the socialization factors that influence contributor behaviour within FLOSS communities. The wording itself suggests that this project strives to identify common rules and relationships that exist among individual-level characteristics within FLOSS communities independently of the given context or observer. Therefore, this project adopts an overall positivist approach in investigating a given phenomenon. Nonetheless, the project respects the idea of 'methodological pluralism'. This study has drawn insights from a wealth of research projects from the FLOSS literature that relied on various perspectives including qualitative research and the interpretivist epistemology. Furthermore, the chosen methodology presented

in this chapter relies on an initial qualitative component (interviews). Consequently, this research adopts a post-positivist approach in investigating the social phenomena that are FLOSS community socialization and FLOSS contributor behaviour. Theoretical models are derived from past research; empirical evidence was collected to examine relationships between variables, thus reflecting an objectivist ontology and a post-positivist epistemology.

3.2 PERSPECTIVES ABOUT METHODOLOGY SELECTION

One of the core characteristics of IS research is its overall diversity (Sidorova & Evangelopoulos, 2008). Evidence of such diversity in the IS field includes the wide array of research areas being investigated, the wealth of theoretical perspectives being used as well as a broad range of methods being employed (Robey, 1996). In the mid-nineties, opinions about such diversity varied. While some researchers believed that diversity was a serious threat that would eventually harm the discipline by jeopardizing its legitimacy (Benbasat & Weber, 1996), other researchers believed that on the contrary, diversity was a strength rather than a weakness (Robey, 1996). The latter view argued that a field characterized by a high degree of diversity overall promulgates creativity and attracts talented researchers across a wide range of disciplines (Robey, 1996).

More recently, the diversity of IS research has come to be recognized (Sidorova & Evangelopoulos, 2008) as an essential strength of the discipline (Venkatesh, Brown & Bala, 2013). Referring to the term of 'disciplined methodological pluralism' previously introduced by Landry and Banville (1992), Robey (1996) refused to acknowledge a dominant paradigm but rather recommended that IS researchers adopt theories and methodologies that are derived from the nature and content of research questions. Other leading IS scholars have also recently reiterated the urgent need for relying on methodological pluralism when conducting IS research (Venkatesh et al., 2013).

The choice of methodology in this research was inspired by the above recommendations. The main research question of this research suggests the use of a quantitative approach, as the purpose of this research is to identify the socialization factors that have an influence on contributor behaviour. It thus suggests the need to collect a large amount of cross-community data in order to statistically demonstrate the positive or negative influence of such factors on a performance-related measure adapted to the FLOSS context. However, a thorough review of the literature revealed an insufficient amount of evidence to confidently identify, in advance, potential socialization factors. In addition, the lack of clear understanding of the broad notion of individual performance in a community-based environment suggested the need to conduct a preliminary qualitative investigation. Consequently, a mixed-method approach appeared to be the most appropriate way to conduct this research.

In spite of encouraging signs about the spread of methodological diversity in IS studies, research that relies on a mixed-method approach is still scarce (Venkatesh et al., 2013). Partisans of mixed methods research believe in the equal value of both quantitative and qualitative approaches in providing a deeper understanding about phenomena of interest. Creswell and Clark (2007) suggested four overall categories of mixed-methods designs:

1. *Triangulation*: merging qualitative and quantitative data to understand a research problem.
2. *Embedded*: using either qualitative or quantitative data to answer a research question within a largely quantitative or qualitative study.
3. *Explanatory*: using qualitative data to help explain or elaborate quantitative results.
4. *Exploratory*: collecting quantitative data to test and explain a relationship found in qualitative data.

The research design adopted in this project belongs to the second category: embedded design. It proposes collecting qualitative data to answer two subsidiary research questions within a largely quantitative study. While the qualitative component of this research is intended to help understand the nature of socialization experience and citizenship behaviours in the FLOSS community context, this study is overall quantitative in nature.

3.3 CHAPTER SUMMARY

The chapter began with an explanation of the choice for an appropriate ontology and epistemology. The chapter ends by explaining the choice for an 'embedded design' mixed-methods approach to effectively investigate the research question of this research.

QUALITATIVE ASSESSMENT OF FLOSS NEWCOMER EXPERIENCE AND CONTRIBUTOR PERFORMANCE

4.1 INTRODUCTION

Because of the scarcity of research about both FLOSS community newcomer experience and community citizenship behaviours (as emphasized in Chapter 2), it appeared necessary to collect qualitative evidence about both phenomena. The overall goal of this phase of the research was two-fold. First, it was intended to delineate the key factors that categorize newcomer experience in the context of large and complex FLOSS projects. Second, the notion of community citizenship behaviour, inherited from the organizational citizenship behaviour literature, had to be examined in the FLOSS context in order to identify which citizenship behaviours are critical to the functioning and sustainability of large and complex FLOSS projects. Semi-structured interviews were thus conducted with FLOSS experts who had experience in FLOSS contribution and community management. The overall goals of the interviews were to re-investigate the definitions of each of the socialization and performance-related constructs, re-assess their relevance, as well as to provide a basis for the generation of appropriate measurement items.

4.2 METHODOLOGICAL CHOICES

In line with beliefs of post-positivist social scientists, O'Leary (2004, p. 150) indicates that "collecting credible data is a tough task, and it is worth remembering that one method of data collection is not inherently better than another." Therefore, the choice of a particular data collection method must be motivated by its fit with the research goals and its advantages relative to other candidate collection methods. The decision to include a preliminary qualitative phase in this primarily quantitative research project was motivated by Strauss and Corbin's claim (1990) that there is "no primacy of either mode of doing research (qualitative versus quantitative)" (p. 27). A research approach should be designed not as an end in itself but rather as the best tool to achieve a given research goal.

4.2.1 *Interviews vs. focus groups*

The first phase of the selected methodology consisted of collecting qualitative data in order to further understand the socialization and citizenship phenomena in the FLOSS context. Conducting interviews or focus groups were the two main available means to collect such data.

Interviewing is a means of collecting data through "attempts to understand the world from the subjects' point of view, to unfold the meaning of peoples' experiences, to uncover their lived world prior to scientific explanations" (Kvale, 1996, p. 1). Kvale added that the interview method "sees the centrality of human interaction for knowl-

edge production, and emphasizes the social situatedness of research data.” (Kvale, 1996, p. 14). Focus groups are a type of group interviews that is particularly suitable for obtaining a consensus on a given topic (Rosemann & Vessey, 2008).

The main objective of this phase of the research was not to try to attain any common understanding among FLOSS practitioners about the phenomena under investigation. It was rather to collect data as context-specific and rich as possible in order to gain a thorough understanding of the newcomer experience and citizenship phenomena. As a result, the collection of qualitative data through the use of semi-structured interviews was selected as it was considered to be the most efficient strategy to satisfactorily capture the complexity and subtleties of FLOSS newcomer experience and FLOSS community citizenship behaviour.

4.2.2 *Recorded semi-structured interviews*

According to Mason (2002), there are three main types of qualitative interviews: in-depth or intensive, semi-structured, and loosely structured or unstructured interviews. All three types involve an “interactional exchange of dialogue”, a relatively informal style, and being “thematic, topic-centered, biographical, or narrative” (p. 62).

Reflecting on the best way to conduct interviews to generate rich data, Bogdan and Biklen (1992) state that:

Good interviews are those in which the subjects are at ease and talk freely about their points of view.... Good interviews produce rich data filled with words that reveal the respondents’ perspectives (p. 97).

Besides, Bernard (1988) also recommends letting the interviewer lead the conversation in a rather unstructured manner, and to use probing questions whose purpose is to refocus the conversation when needed at natural points in the conversation. Considering these recommendations, conducting semi-structured interviews appeared the most appropriate strategy considering that the objectives of this part of the research was to get a deeper and richer understanding of newcomer socialization and community citizenship behaviour in the context of large and complex FLOSS projects. It was then decided to conduct interviews during which informants would freely reflect on both notions in order to generate rich and context-specific data. However, it was also decided to use probing questions whenever the researcher would feel the need to get additional insights about certain aspects that were not discussed such as a given Maanen and Schein’s dimension or a given type of citizenship behaviour.

The recording of interviews is a debated issue among qualitative researchers. Indeed, Patton strongly recommends the “indispensable” use of a tape recorder (1990, p. 348) when conducting interviews, whereas Lincoln and Guba do not recommend recording because of the intrusiveness of recording devices that can also encounter technical problems. It was decided to follow Patton’s recommendations since the use of recordings had the obvious advantage of capturing data in a lot more faithful manner than by writing notes, allowing the researcher to fully focus on effectively conducting the interviews.

4.2.3 Sampling

In qualitative research, purposeful sampling is the dominant strategy to select interview respondents as it allows to generate data from information-rich cases (Patton, 1990). According to Patton (1990), the term 'purposeful sampling' refers to the deliberate selection of cases "from which one can learn a great deal about issues of central importance to the purpose of the research" (p. 169). In this research, two sampling strategies were combined when identifying and selecting potential interviewees. In a first phase, an invitation to participate to the research project was sent on a mailing list which is followed by FLOSS practitioners in New Zealand. Four individuals responded favorably to the invitation.

During the numerous informal conversations the researcher had with FLOSS practitioners, it was often mentioned that researchers involved in FLOSS research rarely 'give back' to FLOSS communities by sharing and disseminating their results outside the academic sphere. This perspective is particularly important, as interactions among FLOSS practitioners are frequently governed by "gift-giving" mechanisms (Bergquist & Ljungberg, 2001). By not contributing back to FLOSS communities, some previous researchers have thus broken the moral contract that had been established with the research subjects when they agreed to take part in a research project. As a result, researchers are now struggling to gain the support and cooperation of FLOSS practitioners when conducting research. A way of counteracting this tendency and gaining the trust of potential participants is by being referred by FLOSS project members with high status and reputation. This is the strategy that was adopted for identifying and recruiting respondents for this phase of the research.

Initially, five interviews were planned. However, preliminary coding of the interview data revealed that in spite of the overall validity of Van Maanen and Schein's (1979) approach towards socialization, at a deeper level of the theory, some conceptual and operational changes had to be undertaken on the original constructs. The interviews also showed that FLOSS community citizenship behaviours are a complex phenomenon, confirming the validity of most of the five types of citizenship behaviours: altruism, conscientiousness, civic virtue, and courtesy (Organ, 1988, 1990a; Organ & Konovsky, 1989; Organ & Ryan, 1995) but they also revealed new forms of citizenship behaviours that pertain specifically to the FLOSS context. Six additional interviews were therefore conducted (a total of 11 interviews were conducted altogether). 10 individuals were interviewed using video conferencing, while one person was interviewed face-to-face.

The second phase employed a theoretically motivated purposeful sampling method in selecting interview participants. The four individuals who took part in the first interview phase were asked to identify individuals who had some solid experience as FLOSS contributors and expertise in FLOSS community management. It was also crucial that respondents had a good understanding of FLOSS community newcomer experience. These individuals were then invited by email to take part to this research project. Seven individuals agreed to participate and were thus interviewed. Since the analysis of the interview data revealed that theoretical saturation had been reached after the eleven interviews, it was decided not to conduct any additional interviews.

Figure 16: Miles and Huberman's (1994) data analysis components

4.3 DATA ANALYSIS PROCEDURE

Qualitative research is an ever-evolving field of inquiry which is characterized by an ongoing debate over what constitutes quality and rigor. As a result, researchers have recommended a wide array of methods and techniques to guide researchers when analyzing qualitative data. Denzin and Lincoln (2000) criticized the multiplication of approaches:

An embarrassment of choices now characterizes the field of qualitative research. There have never been so many paradigms, strategies of inquiry, or methods of analysis for researchers to draw upon and utilize. (p. 18)

In this research, Miles and Huberman (1994)'s data analysis strategy was followed considering that the authors' recommendations are well acknowledged in qualitative research. The authors called their approach 'transcendental realism' and suggested that data analysis shall be conducted along three concurrent streams of activities that constantly interact throughout the analysis: data reduction, data display, and conclusion drawing and verification.

4.3.1 Data reduction

Data reduction is the process of "selecting, focusing, simplifying, abstracting, and transforming the data that appear in written up field notes or transcriptions" (Miles & Huberman, 1994). As Miles and Huberman explain:

Data reduction is not something separate from analysis. It is part of analysis. The researcher's decisions—which data chunks to code and which to pull out, which evolving story to tell—are all analytic choices. Data reduction is a form of analysis that sharpens, sorts, focuses, discards, and organizes data in such a way that "final" conclusions can be drawn and verified. (p. 11)

The main task in data reduction is the coding of the interview data. Strauss and Corbin (1990)'s two-phase coding strategy involving open coding and axial coding, was used. This decision was motivated by the 'grounded theory' flavour of the overall approach implemented during that phase of the research since the goal was to unveil key dimensions that characterize FLOSS socialization and CCB.

4.3.1.1 Open coding

Open coding involves “breaking down, examining, comparing, conceptualizing, and categorizing data” (Strauss & Corbin, 1990). The purpose of open coding is to identify and tentatively name conceptual categories allowing to group the observed phenomena or concepts. The outcome shall provide descriptive multi-dimensional categories which provide a preliminary framework for analysis (Strauss & Corbin, 1990). These categories are then gradually modified or replaced during the subsequent stages of analysis: axial coding.

Van Maanen and Schein’s socialization framework as well as Organ’s OCB dimensions were used as theoretical lenses to investigate FLOSS socialization and community citizenship behaviour. The initial set of codes used during open coding was comprised of van Maanen and Schein’s socialization dimensions as well as Jones’ sub-categorization. Three codes represented *context*, *content*, and *social* aspects, each aspect being sub-divided into the two associated dimensions from van Maanen and Schein. Similarly, CCB codes consisted of the five OCB dimensions from Organ. In addition, Some ‘descriptive codes’ were used in the beginning of the analysis following Miles and Huberman’s recommendations to help capture information about the cases. Finally, new codes emerged while the data was being analyzed (see Appendix B for the list of codes, their properties, and coding examples). Each new code was added to the initial set of codes (see Table 2 for coding examples).

For instance, the notion of *mentoring* clearly appeared as an important issue when discussing FLOSS socialization as interviewees highlighted its practice at both informal and formal levels. The code SOCIALIZATION/SOCIAL/MENTORING was then added. Similarly, the careful analysis of the data identified *being friendly* and *advocating projects* as potential instances of citizenship behaviour in the FLOSS context. The codes CONTRIBUTION/CCB/FRIENDLINESS and CONTRIBUTION/CCB/ADVOCACY were then created.

4.3.1.2 Axial coding

Axial coding, aims at reassembling data that was broken up during the open coding process (Strauss & Corbin, 1990). It involves re-examining the categories identified during open coding, comparing and combining them in new ways as the researcher begins to assemble the ‘big picture’.

At the end of the open coding phase, a careful examination of the codes allowed to generate axial codes along five distinct scenarios. A first type of axial code resulted in strictly reusing the open code along with its properties since the code appeared to capture an important concept. For instance the code ‘INFORM PROJECT’ defined as ‘keeping the community informed about one’s work and progress’ was retained since the code generated during open coding captured well the concept having been identified by the researcher.

A second category of axial codes consisted of retaining the code label determined during open coding. However, the scope of the definition was extended as it appeared that the initial definition was too narrow in the sense that it excluded important related aspects that could be all grouped under a same umbrella code. During the open cod-

Table 2: Open coding - Coding examples

Coding sample - Initial codes
<p>SOCIALIZATION/CONTENT/SEQUENTIAL</p> <p><i>...that precise path is probably the only way into a very technical group where source code is the important thing. So, your ability to submit a good bug report is your understanding of the computer. Once you've shown that, you can submit a good patch which is leaved on the service of it because you've shown that you understand the computer so the patch is probably valid. And then the next step in is a patch for a non-bug, a feature patch. It probably wouldn't be looked as a feature unless you've already shown that you don't break code when you submitting fixes.</i></p>
<p>CONTRIBUTION/CCB/ALTRUISM</p> <p><i>Some people believe that helping a new contributor is not an altruistic thing but instead it is a desperate act of self-promotion in order to gain a new contributor. So, it is radically not altruistic. If I can pay this person in kindness, then I buy a new contributor, and my project will have doubled in contributors because I am the only person right now.</i></p>
Coding sample - New codes
<p>SOCIALIZATION/SOCIAL/MENTORING</p> <p><i>The other thing that there is more formal programmes, there is of course the Summer of Code which people have to connect mentors to new contributors in a project. That's quite active. I had my own informal mentoring network and learnt what was going on and who is who. I think it does happen a lot.</i></p>
<p>CONTRIBUTION/CCB/FRIENDLINESS</p> <p><i>Other kinds of good contributors for a community have been people who just hang out on the IRC channels and are friendly</i></p> <p><i>I have only participated in projects where the expectations are to be friendly with people.</i></p>
<p>CONTRIBUTION/CCB/ADVOCACY</p> <p><i>Somebody advocating the project outside, that could be perhaps an aspect of citizenship. Somebody who is promoting the project to the business world, even to his friends, that could be one aspect.</i></p>

ing phase, the researcher had identified CONTRIBUTION/CCB/MAINTENANCE as an instance of CCB behaviour defined as 'performing maintenance tasks'. However, the initial definition was code-centric and excluded other types of tedious tasks that were also found to be valuable for the interviewees. As a result, the code CONTRIBUTION/CCB/MAINTENANCE was retained but its associated definition was reworded: 'keeping project resources to a good standard'.

Some open codes were discarded during axial coding since the interviewees clearly indicated their irrelevance in the FLOSS context. This was the case for the code SOCIALIZATION/CONTENT/FIXED for instance, as all interviewees acknowledged that the notion of timetable was foreign in FLOSS projects. The same issue arose with the code CONTRIBUTION/CCB/SPORTSMANSHIP. Most interviewees noted that the idea of tolerating difficult work conditions without complaining had nothing to do with FLOSS development since contributors usually manage their own commitment level.

The re-examination of some open codes resulted in modifying the scope and label of some initial codes. This happened for some of the socialization characteristics that were identified while thoroughly examining the associated open codes. The open code CONTRIBUTION/CCB/ALTRUISM resulted into creating the axial code CONTRIBUTION/CCB/INTERPERSONAL HELP because helping other contributors had indeed been shown to be an important instance of CCB behaviour; but the altruistic motivation governing helping actions was found to have been seriously questioned by a number of interviewees.

Finally, the last category of axial codes consisted of merging certain open codes. CONTRIBUTION/CCB/BEING NICE, CONTRIBUTION/CCB/CODE OF CONDUCT COMPLIANCE, and some aspects of CONTRIBUTION/CCB/CONSCIENTIOUSNESS were merged into the axial code CONTRIBUTION/CCB/CULTURE COMPLIANCE to capture behaviours that relate to 'behaving in a way that complies with the behavioural norms of a community'. Axial coding samples are provided in Table 3 and the complete list of axial codes along with the matching with the original open codes can be found in Appendix B.

4.3.2 *Data display*

Data display used along the data reduction phase provides "an organized, compressed assembly of information that permits conclusion drawing..." (p. 11). According to Miles and Huberman, a display could be some extended piece of text, a diagram, a chart, or a matrix. The graphical means used to generate data displays is not important as long as it helps the researcher arrange and think about ways to organize the data. Data displays are especially used to identify patterns and interrelationships stemming out of the data. The purpose of this phase of the research was simply to identify key aspects characterizing FLOSS socialization and community citizenship behaviours, but relationships among concepts were not sought. However, data displays were used during the axial coding phase in order to identify common themes, overlapping concepts, as well as necessary changes in the codes.

Table 3: Axial coding - Coding examples

CASE 1: Open code retained as final axial code
Open/ Axial code = CONTRIBUTION/CCB/INFORM PROJECT <i>I think it should be actively encouraged to put your hand up and say "I am doing too much work, help!"</i>
CASE 2: Label of open code is retained but scope is modified/extended
Open code = CONTRIBUTION/CCB/MAINTENANCE <i>A lot of the really good contributors are people that clean up other people's messes. For example merging bug reports that are duplicates, cleaning up code ...</i>
Axial code = CONTRIBUTION/CCB/MAINTENANCE <i>I think it is a good one because there is not many people that would volunteer to do these things necessarily. That's actually a really good way to get started because people like you a lot more if you do the things that nobody wants to do.</i>
CASE 3: Open code is not relevant, code is discarded
Open code = SOCIALIZATION/CONTENT/FIXED <i>Basically, absolutely not. Those projects have no sort of timetable for new contributors especially because the sort of high watermark for a project to make somebody go to the next stage of power in your project, say commit access, or maybe read access to the primary mailing list, or write access to the bug tracker is very subjective.</i>
CASE 4: Open code label and scope are modified (notions are related)
Open code = SOCIALIZATION/CONTEXT/INVESTITURE <i>If somebody comes in with a lot of skills, we love it. They get some instant reputation especially if they are known in other open source communities already. If somebody comes in without skills, we are not hostile but we are not afraid to tell them "You need to go and learn this, come back and you will be able to contribute". Until you learn these things, you will not be able to contribute in a meaningful way.</i>
Axial code = SOCIALIZATION/COMMUNITY/SUPPORTIVENESS <i>That depends a lot on the community members. In GNOME we are very supportive. We are trying to at least support in spirit. I think GNOME on the whole is supportive.</i>
CASE 5: Several open codes are combined into one axial code
Open code = CONTRIBUTION/CCB/BEING NICE <i>When I say nice here, I mostly mean treats people in accordance with the community expectations for how people are to be treated</i>
Open code = CONTRIBUTION/CCB/CONSCIENCIOUSNESS <i>We have a lot of unwritten technical rules, how things are supposed to work that would take years to pick up. But the social rules are much easier to pick up. You learn them from the most active or respected people posting on the lists.</i>
Open code = CONTRIBUTION/CCB/CODE OF CONDUCT COMPLIANCE <i>To be a good contributor, you have to make contributions and comply with the code of conduct.</i>
Axial code = CONTRIBUTION/CCB/CULTURE COMPLIANCE <i>Because there are always roadblocks and hurdles, you are walking into a culture that you don't know where people know each other, it takes some effort to get into that. I think it takes a fair amount of effort on both sides.</i>

The use of qualitative data analysis software allowed the researcher to generate reports that summarized the coded data per code and per participant. The reports were automatically generated by the software and provided a visual display of the coded data that helped in revising the preliminary codes as well as identifying emerging ones. The reports generated by NVivo 9 can be seen as variations of the data matrix data displays recommended by Miles and Huberman (1994) when analyzing data after axial coding (see Appendix B for an example of display that was used during the analysis of the data). The reports contained hyperlinks for each quoted piece of text allowing easy navigation between the report and the original transcripts.

4.3.3 Conclusion drawing and verification

The final stage of Miles and Huberman's method involves the verification of the conclusions that were drawn in terms of validity as well as an examination of the overall quality of the undertaken qualitative study. Qualitative researchers have used the umbrella term of 'trustworthiness' to encompass both notions (Creswell & Miller, 2000; Guba, 1981; Krefting, 1991). Often, researchers have raised concern about the inappropriate use of evaluation criteria that pertain more to quantitative research rather than qualitative research (Krefting, 1991). To answer the need to rely on criteria that are specific to qualitative research, Guba (1981) developed a model to assess the trustworthiness of qualitative studies. This research relied on Guba's model.

Truth value relates to answering the question: "How can one establish confidence in the "truth" of the findings of a particular inquiry for the subjects with which and the context in which the inquiry was carried out?" (Guba, 1981, p. 79). In other words, evaluating truth value establishes the degree of confidence associated with the truth of the findings based on the research design, informants, and context (Lincoln & Guba, 1985). Lincoln and Guba (1985) emphasize the use of member checks as "the most crucial technique for establishing credibility" (p. 314). Performing member checks involves communicating data and interpretation to the participants and asking them to confirm (or not) the credibility of the information. Once data analysis was completed, interview respondents as well as experienced FLOSS contributors were informally consulted to assess the confidence level associated with the quality of the results. They all strongly acknowledged the pertinence of the results for both the socialization framework and the CCB aspects that were identified.

Applicability refers to determining the degree to which the findings of the inquiry have applicability in other contexts or with other subjects (Guba, 1981). This was assessed when the researcher contacted experienced FLOSS contributors who did not take part to the interviews and who were contributors to FLOSS communities that were not represented such as Perl, LibreOffice, or Mozilla. A lot of support and encouragements were provided to the researcher.

Consistency of the data relates to the notion of stability from quantitative research and pertains to whether the findings would be consistent if the inquiry were replicated with the same subjects or in a related context (Krefting, 1991). Condensing samples during both the open coding and axial coding phases were regularly sent to the supervising staff to ensure the quality of the coding procedure. In addition, the collected data was entirely recoded manually at the end of axial coding to assess the stability

of the results. Overall, the outcome was highly satisfactory as with a few exceptions, the results were nearly identical.

Finally, *neutrality* is the final aspect to be checked in Guba's model. It relates to the freedom from bias in the research procedures as well as the results (Guba, 1981). The neutrality of a qualitative research is usually guaranteed by the implemented methodology (Krefting, 1991). Guba (1981) did not view neutrality in terms of a researcher's objectivity but rather as the interpretational confirmability of the data and the results. As a consequence, the author recommends the use of an audit strategy that consists of involving an external auditor from the beginning of the research in order to follow through the sequence of events and the decisions that were made along the process. The supervising staff played that role as the researcher constantly informed them of the course of actions that was undertaken. Besides, they were also consulted for each decision that had to be made throughout the progression of the research.

4.4 INTERVIEW PARTICIPANTS' CHARACTERISTICS

The interviewees included 9 men and 2 women. In terms of educational qualifications, one respondent had a postsecondary certificate, 3 had studied at the bachelor's degree level (one person did not complete the degree), 5 individuals had a master's degree, one person had a PhD degree and another was in the process of completing a PhD degree. Finally, one interviewee did not mention his or her educational background. The experience of the interviewees as FLOSS contributors ranged from several years to more than 15 years.

Seven out of the 11 interviewees had some direct experience in terms of newcomer management within the communities in which they had been involved. All seven mentioned having mentored community newcomers to some extent and 6 of them had taken part in organizing and managing community programs dedicated to attracting and engaging newcomers. Five participants contributed to FLOSS communities as project managers and 2 individuals had experience as FLOSS community managers (which was their main occupation).

The interviewees contributed to a wide array of large and complex FLOSS projects: Debian, Gentoo, Gimp, Gnome, KDE, Maemo, Mahara, Meebo, OpenHatch, Redrap, Ubuntu, Wordpress, and Yorba. The breadth of experience of the interviewees lends credence to the generalizability of the findings from phase 1. The ethical permissions for the interviews stipulated that the thesis and any of the resulting publications would not provide any information that could allow the identification of particular respondents. As a consequence, the interviewees are not explicitly associated with their respective projects in this thesis.

4.4.1 Interview procedure

Each interview lasted between 60 to 90 minutes and was conducted by the researcher. Ethical approval was obtained before the interviews were conducted. The interview protocol consisted of the following (see the Appendix A for details about the HEC application and the full interview protocol):

Table 4: Interviewees' characteristics

Participant	Gender	Education	Occupation	Nb FLOSS projects	Contribution types	FLOSS experience
1A	M	postsecondary certificate	school principal	2	non-technical contributor, user	several years
1B	M	master's	PhD candidate	2	non-technical contributor, user	5 years
1C	M	bachelor's	senior software developer	4	project manager, maintainer, in charge of newcomer programme, social event participant	9 years
1D	M	bachelor's	IT security analyst	1	technical contributor, user	more than 15 years
2A	M	master's	not-for-profit org. administrator	2	maintainer, technical contributor, social event organizer and participant, newcomer management, mentor	13 years
2B	M	unfinished bachelor's	web designer	1	former project manager, project outreach	4 years
2C	M	master's	freelance FLOSS community consultant	5	technical contributor, board member, project manager, release manager, community manager, mentor, social event organizer and participant	13 years
2D	F	master's	community manager	1	community manager, social event participant, mentoring programme administrator, technical contributor	8 years
2E	F	master's	software engineer	1	paid project contributor, mentoring programme administrator	7 years
2F	M	not mentioned	freelance documentation consultant	1	technical and non-technical contributor, project manager, board member	10 years
2G	M	PhD	IT industry analyst	3	project manager, technical contributor, mentor, project outreach, desktop manager, mentoring programme administrator, social event participant	10 years

1. Brief presentation of the research project.
2. Background information about the interviewee (educational background, occupation, FLOSS experience).
3. Interviewee's newcomer experience(s).
4. Reflection about the various aspects that characterize one's newcomer experience in the FLOSS context.
5. Exploration of the applicability and relevance of Van Maanen and Schein's theoretical view of socialization.
6. Interviewee's contributions in FLOSS projects.
7. Reflection about what a 'good' contributor is and assessment of the relevance of the conceptualization of contribution behaviour introduced in this research project (assignment performance and citizenship behaviour).
8. Reflection about the meaning of citizenship behaviour in the FLOSS context and its different manifestations.
9. Applicability and relevance of Organ's theoretical view of organizational citizenship behaviour.

In addition to the questions present in the interview protocol, the researcher was able to raise specific questions, informed by the literature review, that related to important emerging topics during the interviews. Once all the interview data were transcribed (using Express Scribe Transcription Software¹, a free transcribing software that allowed easy manipulation of the interview sound recordings).

The following subsections present the findings of the data analysis phase. First, an explanation is provided as to how the overall approach towards conceptualizing contributor behaviour was validated, followed by a presentation of the identified community citizenship behaviours. Then, another subsection provides a thorough overview of the six socialization factors that were identified. The factors are then categorized and integrated into a framework.

4.5 RESULTS

The following subsections present the findings of the data analysis phase. First, an explanation is provided as to how the overall approach towards conceptualizing contributor behaviour was validated, followed by a presentation of the identified community citizenship behaviours. Then, another subsection provides a thorough overview of the six socialization factors that were identified. The factors are then categorized and integrated into a framework.

4.5.1 *Relevance of performance operationalization*

The analysis of the interview data confirmed the relevance of the conceptualization of contributor behaviour introduced in this thesis project. The interviewees were first

¹ <http://www.nch.com.au/scribe/index.html>

asked to reflect about what makes a ‘good’ contributor in a FLOSS community. A majority of them clearly identified two complementary aspects. First, a person is good for the community based on his or her direct contributions to the project (technical or non-technical) by carrying out project tasks. Second, the interviewees highlighted another set of behaviours, independent of project-related tasks, that constitute a social ability to contribute to the well-being of communities:

“There are two things where somebody might be good. If you just have a really really top notch programmer and they are very good for the project [...] The other one would be people that are positive voices in the community, people who are easy to get along with, people who help new contributors, people who go to conferences representing your community.”

Following this question, the interviewees were confronted with the term *citizenship* by being asked to reflect on the assumption that ‘a good contributor is someone who gets the work done as well as one who is a good citizen to the community’. Overall, the interviewees strongly acknowledged the relevance of the statement as well as the importance of the notion of citizenship:

“A good contributor is one who gets the work done in terms of that product, and then there is some extra side thing they can do in order to be qualified a good citizen.”

“I think that being a good citizen is certainly variable, and it’s a lot of what makes the product of it successful. I usually consider success in a free software community is things like how big it is, how effective it is, creating a high quality of code and documentation products ... Good citizenship is by in large essential to make the product achieve that”.

4.5.2 Findings about FLOSS community citizenship behaviours

The literature review chapter discussed the relevance of using Williams and Anderson (1991)’s bi-dimensional view of citizenship behaviours in terms of those directed towards other individuals (also called OCB-I) and those directed towards the benefit of the organization itself (or OCB-O). This research extrapolated the notions of OCB-I and OCB-O to respectively, CCB-I (or community citizenship behaviours oriented towards the benefits of other individuals), and CCB-P (or community citizenship behaviours oriented towards the benefits of the project and its community). During the interviews, the researcher questioned the interviewees about each of Organ’s initial OCB dimensions (altruism, courtesy, conscientiousness, sportsmanship, and civic virtue) as a starting point and then probed the participants on other important FLOSS-related citizenship behaviours. The following subsections introduce the list of citizenship behaviours that were identified as the outcomes of the data analysis. Each behaviour is also justified in light of the extant OCB literature.

4.5.2.1 Instances of CCB-I

The analysis of the interview data demonstrated the importance of five essential CCB-I behaviours: helping and assisting other community members, helping to prevent project problems to occur for other members, treating others with respect, maintaining a positive attitude, and facilitating member conflicts. The following subsections

present each of the identified behaviours and justify them in light of the OCB literature. The results are summarized in Table 5.

Table 5: Interview findings - Instances of CCB-I behaviours

CCB-I behaviours	OCB Literature supporting citizenship nature of behaviour
Helping and assisiting other community members	altruism (Organ, 1988, 1990a; Smith, Organ & Near, 1983), interpersonal helping (Graham, 1991; Moorman & Blakely, 1995), helping coworkers (George & Brief, 1992; George & Jones, 1997)
Helping to prevent project problems to occur for other members	courtesy (Organ, 1988, 1990a), interpersonal conscientiousness (Coleman & Borman, 2000)
Treating others with respect	courtesy (Alge, Ballinger, Subrahmaniam & Oakley, 2006; Niehoff & Moorman, 1993)
Maintaining a positive attitude	helping coworkers (George & Brief, 1992; George & Jones, 1997), social participation (Van Dyne, Graham & Dienesch, 1994)
Facilitating member conflicts	peacekeeping (MacKenzie, Podsakoff & Paine, 1999; Organ, 1988, 1990a)

ASSISTING OTHER PROJECT MEMBERS The idea of *helping and assisting other project members* was found to be particularly relevant in the FLOSS context as one participant pointed out that helping is “the entire foundation of a community”. Helping other members appeared as an essential behaviour in order to qualify an individual for being an actual community member:

“It is important that contributors do help each other and act as a big family.”

“I don’t think it is possible to become a community member without having some kind of empathy for the other people in that community and be helping them.”

The OCB literature has often emphasized the importance of altruism in organizational contexts (Morrison, 1994; Organ, 1988; P. M. Podsakoff, Scott, Moorman & Fetter, 1990; Smith et al., 1983). Nonetheless, some interviewees raised some doubt about the altruistic nature of helping behaviours in FLOSS communities. For instance, a respondent qualified such behaviours as “radically not altruistic” whereas another respondent pointed out that helping is important but members “don’t need to have altruistic motivations” when doing so. As a result, the identified subdimension of CCB-I was more in line with the related concept introduced by Moorman and Blakely (1995) of *interpersonal helping*. This term particularly focuses on helping coworkers in their jobs when such help is needed (p. 130). The notion appears more appropriate in the context of this study as it de-emphasizes the nature of the motivation behind the actions and highlights the human aspect of help exchanges. In FLOSS projects, it can be argued that any kind of project contribution could be to some extent seen as

some form of help. As a participant said: “Pretty much everything is help”. Interpersonal helping has a narrower scope. Writing a bug report and submitting it to the bug tracker can be seen as a way of “helping a project” but it is not an instance of interpersonal helping as it does not involve an interaction between two individuals in which one person expresses a need for help and another person provides some assistance to the first person.

The interviewees highlighted various types of interpersonal helping behaviours such as providing assistance to people who have technical or non-technical problems (who could be either newcomers or experienced members) but also redirecting individuals who expressed a need for help to places or people where they can be provided answers:

“If people come to you and they need help, you should not turn them away. At the very least you should direct them in the right direction, like go and ask your question to this place.”

BEING COURTEOUS In OCB research, courtesy behaviours involved the attitudes and gestures to help prevent work problems for others as introduced by [Organ \(1988\)](#). When reflecting on the notion of FLOSS citizenship, a majority of interviewees confirmed the particular relevance of courtesy behaviours when characterizing FLOSS community citizenship:

“You do have to be self-aware of what you are going to do. Because everybody can change everything, you need to make sure that what you are changing, you have to think about it. You cannot just change some major stuff before thinking before.”

“Especially when you have something with a lot of dependencies like a Linux distribution and multiple layers of packages depending on others, you don’t want to make a change to a library then break everything that uses the library. You have to be thinking through what is going to be the effect of this.”

The analysis revealed that the term considerateness was commonly used in FLOSS communities to refer this notion (rather than courtesy). Several interviewees mentioned that FLOSS communities are familiar with the term considerateness which they see as an important trait and which is clearly emphasized in most communities’ code of conduct (e.g. Ubuntu², Joomla³, or KDE⁴).

BEING NICE The analysis of the interview data revealed that nearly all interviewees emphasized the importance of being “nice”, “respectful”, and “friendly” when interacting with other community members. Such behaviours were part of the community behavioural expectations that members are expected to comply with:

“When I say nice here, I mostly mean treats people in accordance with the community expectations for how people are to be treated”.

² <http://www.ubuntu.com/project/about-ubuntu/conduct>

³ <http://www.joomla.org/about-joomla/the-project/code-of-conduct.html>

⁴ <http://www.kde.org/code-of-conduct>

Nearly all interviewees highlighted that such behaviours are indeed clearly specified in most communities' codes of conduct. As some interviewees indicated, compliance to the community behavioural norms and rules is directly related to "maintaining the culture of the project". One respondent pointed out: "[technical issues] are easy to fix but community is hard to fix".

Being nice can be seen as a mechanism that contributes to the growth of communities; a participant noted that "being nice when people are making mistakes means that they are more likely to correct those mistakes and stick around." A respondent also argued that not 'being nice' has a negative impact on the overall community:

"If you have somebody, every time, when you say 'open source' he would say 'No, it's free software'. Every time you say Linux, they say 'No, it's GNU/Linux'. This is continuous belligerent behaviour. Then, I would say that's being a bad community citizen because you are making the general ambiance of the project less agreeable."

The interview data suggested that 'being nice' was used as a generic term to encompass two main characteristics when interacting with others. The interviewees highlighted the importance of maintaining a positive attitude as well as treating other members with respect:

"You don't have to smile at people to be nice, you just have to be respectful and effective."

"I think it is important that to be a good citizen that you get on well with the community, that you get on well with the other people."

Showing respect towards other members has been identified as a form of courtesy in the OCB literature. For instance, K. Lee and Allen (2002) included this notion in their OCB-I construct by asking interviewees about their "genuine concern and courtesy toward coworkers". Similarly, the idea of respecting individuals raises the issue of considering the rights of others. Niehoff and Moorman (1993) and later Alge et al. (2006) measured courtesy using the measurement item: "does not abuse the rights of others", which clearly underlies the idea of respecting and treating well others.

Maintaining a positive attitude when interacting with others is consistent with Van Dyne et al. (1994)'s form of OCB-I entitled 'social participation'. Those authors assessed the extent to which one engages in positive communications as part of social participation. Organ's early works (1988; 1990a) identified another overlapping interpersonal factor entitled 'cheerleading', defined as "the words and gestures of encouragement and reinforcement of coworkers' accomplishments and professional development" (p. 96). George and Brief (1992) introduced a help-related category of OCB entitled *helping coworkers* that encompasses "deliberate acts that improve morale" echoing the idea of behaving and interacting in a positive manner.

MODERATING INTERPERSONAL CONFLICTS Even though there is a strong expectation within FLOSS communities that members be nice, friendly, and respectful towards each other, the interviewees clearly indicated that this does not happen all the time. It was found that the severity of a participant's negative behaviour or attitude is seen differently by different communities. Some of the communities mentioned during

the interviews were said to adopt a rather drastic approach, strictly expelling members from projects for bad behaviour, whereas other communities were more moderate, assessing the overall value provided by a member in spite of bad behaviours and attitudes:

“There might be geniuses and contribute great code, but the number of people that are scared away because of them has always been greater than what they contribute to the project.”

“People would just be toxic to the community but they would just write great code and it kind of sucks. The community has to tell them to clean up their act or go away.”

The analysis of the interview data indicated that dealing with negative behaviours and moderating interpersonal conflicts can be seen as instances of citizenship behaviours. A respondent argued that one way to address individuals’ negative behaviours is to “put a layer around them” in which other project members help in smoothing interpersonal conflicts:

“I will now make sure I watch everything he posts. If he is being abusive, I’ll immediately reply and say that’s just #NAME#. This is what we think in general. And they will put in the extra work to cover up . . . to sort of smooth over the damage, to fix the damage he is creating over here.”

The moderation of interpersonal conflicts has been seen as an instance of citizenship behaviours in the OCB literature (MacKenzie et al., 1999; Organ, 1988, 1990a). The term ‘*peacemaking*’ was introduced by Organ and defined as “actions that help to prevent, resolve or mitigate unconstructive interpersonal conflict” (Organ, 1990a).

4.5.2.2 *Instances of CCB-P*

Seven CCB-P behaviours were identified from the analysis of the interview data (see Table 6 for a summary of the findings). They are: contributing with the best interest for the community, complying to the community behavioural rules, keeping the community informed about one’s work and progress, keeping project resources to a good standard, exercising civic responsibility, the responsible involvement in project decisions, and promoting the project to potential contributors and to the outside world. Each of the CCB-P sub-categories listed below are compared with overlapping notions from the OCB literature.

CONTRIBUTING WITH THE BEST INTEREST OF THE COMMUNITY IN MIND When reflecting about the overall reason for an individual to be considered a valuable contributor within a FLOSS community, a majority of interviewees clearly indicated that the ideal member provides contributions in accordance with the shared goals and direction of a project. As an interviewee pointed out, a contributor should “preferably drive the direction of the project in some meaningful way”.

In other words, communities value individuals who tend to act and behave beyond their self-interest, rather in the best interests of the community. A number of interviewees emphasized the importance for communities that members clearly know the project goals and contribute in a way that allows the project to move towards reaching those goals:

Table 6: Interview findings - Instances of CCB-P behaviours

CCB-P behaviours	OCB Literature supporting citizenship nature of behaviour
Contributing with the best interest for the community	loyalty (Graham, 1991), endorsing, supporting, and defending organizational objectives (Borman & Motowidlo, 1997), spreading goodwill (George & Brief, 1992; George & Jones, 1997), loyal boosterism (Moorman & Blakely, 1995)
Complying to the community behavioural rules	conscientiousness (X.-P. Chen, Hui & Sego, 1998; Niehoff & Moorman, 1993; Organ, 1988), OCB-O (Williams & Anderson, 1991)
Keeping the community informed about one's work and progress	conscientiousness (Organ, 1988), following organizational rules and procedures (Borman & Motowidlo, 1997), obedience (Van Dyne et al., 1994), generalized compliance (Konovsky & Organ, 1996; Smith et al., 1983), OCB-O (Williams & Anderson, 1991)
Keeping project resources to a good standard	protecting company resources (Farh et al., 1997), OCB-O (J. Yang, Mossholder & Peng, 2007), obedience (Van Dyne et al., 1994)
Exercising civic responsibility	civic virtue (Coyle-Shapiro, 2004; Graham, 1986; MacKenzie et al., 1991; MacKenzie, Podsakoff & Fetter, 1993; Niehoff & Moorman, 1993; Organ, 1988; Organ & Ryan, 1995; P. M. Podsakoff et al., 1990; Robinson & Morrison, 1995)
Responsible involvement in project decisions	making constructive suggestions (George & Brief, 1992; George & Jones, 1997), civic virtue (P. M. Podsakoff, Ahearne & MacKenzie, 1997; Vanyperen, Berg & Willering, 1999), identification with the organization (Farh et al., 1997), OCB-O (J. Yang et al., 2007)
Promoting the project to potential contributors and to the outside world	organizational loyalty (Farh et al., 1997; Graham, 1991), loyal boosting (Moorman & Blakely, 1995), spreading goodwill (George & Brief, 1992; George & Jones, 1997)

"I think that a good contributor is somebody who can come in and not just do stuff in the project but also do stuff in that context of those shared goals, culture, and priorities."

"First, they have to understand what the community wants and you have to be backing the overall goal where there is one. And where there isn't an explicit one, you have to be backing the current implicit goal of the community otherwise your contribution is not a good one."

The definition of which closely matches the concept developed above. The author defined loyalty as "the expansion of individual welfare functions to include the interests

of others, the state as a whole, and the values it embodies" (p. 254). This is similar to Borman and Motowidlo's (1997) definition of "endorsing, supporting, and defending organizational objectives," to George and Jones's (1997) "spreading goodwill," and also to Moorman and Blakely's (1995) "loyal boosterism."

COMPLYING WITH THE COMMUNITY BEHAVIOURAL RULES The interviewees were questioned about the relevance of the notion of conscientiousness. Since the beginnings of the OCB tradition, conscientiousness has been considered an essential component of OCB (MacKenzie et al., 1993; Organ, 1988; Smith et al., 1983). In a study involving 422 employees and their supervisors in the banking sector, Smith et al. (1983) explored the nature of OCB and identified a type of behaviours which they termed 'generalized compliance' and qualified as "a more impersonal form of conscientiousness that does not provide immediate aid to any one specific person, but rather is indirectly helpful to others involved in the system" (p. 657). Later, Organ (1988) defined conscientiousness as a discretionary behaviour that goes well beyond the minimum role requirements of the organization. Typically, OCB studies captured conscientiousness in terms of working extra-long days, returning phone calls from the home office promptly, never bending the rules (MacKenzie et al., 1993), being punctual, not taking extra breaks (Smith et al., 1983), or else always following company rules (Konovsky & Organ, 1996).

A lot of support was also provided by the participants for the notion of *conscientiousness* which was qualified as "extremely important", "very important", or "critical" by most interviewees.

Following rules and regulations has been seen as an important part of conscientiousness in OCB research (X.-P. Chen et al., 1998; Niehoff & Moorman, 1993; Organ, 1988). The interviewees were questioned about the importance of such behaviour in the FLOSS context. In spite of acknowledging that it helps in "keep[ing] everything going smooth", several interviewees expressed some reserve:

"The rules cannot count for all the little corners ... I think you have to tell people to be smarter with emotional intelligence point of view, than just strictly following the rules."

In this research, the notion of rule compliance was taken into consideration by considering the extent to which FLOSS communities expect their members to behave in accordance with the behavioural expectations that are transcriptions of the project culture. For instance, 'being nice' is not encouraged because it is a moral virtue, rather because 'being nice' is a way to ensure that newcomers are attracted to a project, and to reduce project turnover resulting from individuals having been improperly treated by other members. A respondent provided the following reflection about the reasons for why members should be 'nice':

"Although you can even take a very utilitarian perspective and declare that the reason you are nice to people in FLOSS projects is because on our products is a nice community, and we create a nice community, in order to support the other product which is the code and documentation, et cetera."

When reflecting on the reasons behind the need for members to 'be nice', interviewees pointed out that it was an antidote that could combat negative behaviours that

harm FLOSS communities. Several interviewees highlighted the 'toxic' nature of such behaviours. They highlighted that "toxic behaviours" are a real threat to the survival and success of communities which usually expel members who repeatedly behave in an improper manner no matter their competency for a project:

"You can be turning people away by your attitude. People would just be toxic to the community but they would just write great code and it kind of sucks. The community has to tell them to clean up their act or go away".

A number of FLOSS projects have realized the necessity of making such behavioural expectations highly visible to project members, and strongly encourage members to follow the stipulated rules during project-related interactions. The use of a community code of conduct is becoming increasingly common among FLOSS communities. A code of conduct can be seen as the embodiment of a community's shared goals and values in a set of behavioural rules as exemplified in the following excerpts of the GNOME⁵ code of conduct:

GNOME creates software for a better world. We achieve this by behaving well towards each other [...] Therefore this document suggests what we consider ideal behaviour, so you know what to expect when getting involved in GNOME. This is who we are and what we want to be.

The interview respondents emphasized the need for members to follow the behavioural rules of communities whether they are explicitly specified in project resources such as codes of conduct or else implicitly acknowledged among community members:

"We have a code of conduct that actually lays down some of our community rules. It is very important especially for long-term community members who are the role models that you comply to our community standards."

"To be a good contributor, you have to make contributions and comply with the code of conduct."

The idea of adhering to the behavioural rules of a community echoes some of the aspects that OCB researchers have attributed to conscientiousness, compliance, and OCB-O. For example, one of the items used by Williams and Anderson (1991) to assess OCB-O was "Adheres to informal rules devised to maintain order", suggesting similar behavioural rules.

KEEPING THE COMMUNITY INFORMED ABOUT ONE'S WORK AND PROGRESS When reflecting on how conscientiousness can be expressed by a member when contributing to a project, the interviewees stressed an important aspect that pertains to the nature itself of FLOSS communities. Project management is a delicate task in a FLOSS project due to the overall lack of visibility of task advancement and completion. This aspect is significantly different from traditional organizational settings and is an important characteristic of FLOSS communities. Most participants emphasized the idea that project managers need to regularly inform the community about the advancements of individuals' assignments. They also indicated that it was extremely important that contributors let the community know when they are not able to complete given or self-assigned tasks as the community will then be in a position to find a replacement (or else some other member can volunteer to carry on with the tasks):

⁵ <https://live.gnome.org/CodeOfConduct>

“I think it is vitally important to communicate what you will and will not do, and when.”

“If you are saying you’ll do something by January 31 and then you don’t, this is important but it is even more important to say that you can’t and let people know about that.”

The findings echo the notion of conscientiousness investigated in influential OCB works. Meeting deadlines was seen by Organ (1988) as a manifestation of conscientiousness, while Borman and Motowidlo (1997) regarded such behaviour as part of following organizational rules and procedures, and Van Dyne et al. (1994) as part of obedience. Smith et al. (1983) and Konovsky and Organ (1996) captured an aspect of *generalized compliance* with the item: “Gives advance notice if unable to come to work”. Williams and Anderson (1991) used the same item in their OCB-O measure.

KEEPING PROJECT RESOURCES TO A GOOD STANDARD Some interviewees emphasized that FLOSS communities value members who voluntarily undertake resource maintenance tasks. Such behaviour allows the maintenance of the project resources to a good standard thus greatly benefitting the project as a whole:

“I have great admiration for community members who find an interest in thankless tasks, and this is very important.”

“A lot of the really good contributors are people that clean up other people’s messes. For example merging bug reports that are duplicates, cleaning up code ...”

The results are consistent with past OCB research. Farh et al. (1997) developed a measure of OCB adopted to the Chinese context and identified *Protecting company resources* as a separate form of OCB. Z. Yang and Jiang (2007) used Farh et al. (1997)’s items into their OCB-O measure. Van Dyne et al. (1994) introduced the notion of obedience as a form of OCB that encompass the maintenance of company resources.

EXERCISING CIVIC RESPONSIBILITY The importance of exercising civic responsibility was highlighted by the participants. The first behaviour identified by interviewees that reflects civic responsibility in a FLOSS community, is a conscious effort to keep up with the latest news and advancements of a project:

“You don’t necessarily need to know every detail happening in every team. But it is important to know kind of the big picture, the big announcements, how things are going, how things are changing.”

“It is important to know what is going on and to actually make an informal effort there as well.”

These findings are in line with the civic virtue dimension of OCB (Graham, 1986; Niehoff & Moorman, 1993; Organ, 1988; Organ & Ryan, 1995). The OCB tradition has acknowledged the importance of keeping up with an organization’s news and changes as an integral part of civic virtue. For instance, MacKenzie et al. (1993) measured civic virtue using 3 items. “keeps up with developments in the company” was one of them. MacKenzie et al. (1991) assessed civic virtue with a three-item measure, two of which were: “Keeps up with developments in the agency/company X” and “Reads and keeps up with the agency/company announcements, messages, memos,

etc.” Similar items were also used in other related OCB works (Alge et al., 2006; Netemeyer, Boles, McKee, McMurrian & McKee, 1997; Niehoff & Moorman, 1993).

The interviewees also confirmed the importance of attending and partaking in community-related social events. This aspect has been conceptualized by OCB research as a second essential form of civic virtue (Konovsky & Organ, 1996; P. M. Podsakoff et al., 1997, 1990). Such behaviour was found to be an important part of a FLOSS community member’s sense of civic responsibility:

“If it is possible then you should be going to conferences. Yes, I think it is good for people to be active in the community and bringing their own something to the community.”

RESPONSIBLE INVOLVEMENT IN PROJECT DECISIONS The appropriate involvement in project discussions and decisions was found to be another area of citizenship behaviour in the FLOSS context. Even though the interviewees agreed that members should ideally take part in project decisions, it was clearly indicated that such involvement needed to be “effective” and “constructive”:

“If the person is totally irrelevant to the decision being made, then no we do not care if they say anything.”

“I think it depends on how much credibility the person expressing an opinion has. Some people have opinions about everything but they are not contributing themselves.”

Moreover, some participants added that one is a good community citizen when one is also able to accept decisions no matter whether they agree or disagree with them and act accordingly:

“It is the person who, when there is a group decision he does not agree with, will say “I don’t agree with it, here is why” but he will let it drop afterwards.”

“I think that people do need to voice their opinion and be heard, but sometimes you have to let things go.”

George and Brief (1992) and George and Jones (1997) introduced the notion of *making constructive suggestions* as a separate dimension of OCB (along with the dimensions helping co-workers, protecting the organization, developing oneself, and spreading goodwill). Other authors have considered the behaviour as a manifestation of civic virtue (P. M. Podsakoff et al., 1997; Vanyperen et al., 1999) by seeing it as part of one’s responsible and constructive involvement in the life of one’s organization, but also identification with the organization (Farh et al., 1997), and OCB-O (J. Yang et al., 2007).

PROMOTING THE PROJECT TO POTENTIAL CONTRIBUTORS AND TO THE OUTSIDE WORLD Related aspects of “improving image outside of the project”, “advocating a project outside”, “raising awareness”. or “representing the values” of a project were emphasized by most interviewees. Some participants pointed out that:

“If you are a true passionate citizen then naturally you will be talking about the project in good terms and spread your passion to other people, promoting it to others.”.

(about FLOSS community citizens) “Somebody who represents the project well to new community members and outside of the community, to tell the project especially to the commercial world because you often to deal with organizations who want to make money as well.”

The promotion of a FLOSS project echoes several citizenship dimensions from the OCB literature. Some OCB researchers such as Farh et al. (1997) and Graham (1991) introduced the notion of organizational loyalty as an important component of OCB. P. M. Podsakoff et al. (2000) characterized organizational loyalty as promoting the organization to outsiders, protecting and defending it against external threats, and remaining committed to it even under adverse conditions. Some research by Moorman and Blakely (1995) investigated the notion of organizational loyalty and indicated that this dimension is distinct from several other forms of citizenship behaviours. They then introduced the notion of *loyal boosting*, defined as “the promotion of the organizational image to outsiders”. The concept of *spreading goodwill* from George and Brief (1992) and George and Jones (1997) relates to the means by which organizational members voluntarily contribute to an organization’s effectiveness by representing their organizations to wider communities.

4.5.3 Interview findings about socialization in FLOSS communities

The second objective of the qualitative phase of this research project was to help derive the key characteristics of newcomer socialization in the FLOSS community context. In addition to asking interviewees to reflect about the specificities and essential aspects of one’s FLOSS community newcomer experience, Van Maanen and Schein (1979)’s conceptualization of socialization was used as a theoretical lens to shed some light on the FLOSS newcomer experience phenomenon.

Chapter 2 justified the overall applicability of Van Maanen and Schein (1979)’s socialization dimensions in the context of large FLOSS projects. For instance, Kraut and Resnick (2011) observed that some FLOSS and online communities do provide elements of institutionalized socialization, and then hypothesized that formal, sequential and collective socialization tactics would play an important role on newcomer socialization.

The overall motivation behind such strategy was not to rigorously apply Van Maanen and Schein (1979)’s socialization concepts to the FLOSS context, as they were shown to have a moderate degree of direct applicability. It was rather to reflect on some of the theoretical considerations of Van Maanen and Schein (1979)’s works that appear to be essential components of a person’s FLOSS community newcomer socialization. For instance, the notion of formal tactics from Van Maanen and Schein (1979) suggests the idea of segregating newcomers from regular employees, which led the researcher to investigate whether some aspects of a newcomer’s experience happened in isolation from other experienced members. Similarly, Van Maanen and Schein (1979)’s serial tactics refer to the use of experienced members accompanying newcomers during their socialization phase. As a result, the researcher explored the relevance of related notions such as the importance of mentoring-like relationships in the FLOSS community context.

The following subsections present the result of the qualitative data analysis which allowed the identification of six distinct aspects that characterize a newcomer's experience in a FLOSS community: *task segregation*, *interaction intensity*, *task purposefulness*, *joining structuredness*, *mentoring*, and *supportiveness*. Each of these factors is a separate facet of the experience through which an individual had to go to become a project contributor. The degree of overlap between the identified socialization variables and Van Maanen and Schein (1979)'s dimensions vary significantly from one socialization variable to another.

4.5.3.1 *Task segregation*

Formal tactics in the organizational context refer to the physical segregation of newcomers from regular employees in order for newcomers to go through a series of experiences that have been designed by the organization. The analysis of the interview data revealed that some degree of segregation may exist for FLOSS community newcomers in the nature of the project tasks they may undertake. The degree to which a newcomer performs tasks that are specifically designed and tailored for newcomers appeared as a key characteristic and was named: *task segregation*.

In line with the view adopted in this research that sees FLOSS socialization as triggered and shaped by newcomers themselves (justifying the use of the term socialization experience instead of socialization tactics), newcomers in FLOSS communities may be provided with an array of tasks dedicated to newcomers that can help them learn about how to become a project contributor. Newcomers are normally free to undertake them or not.

Some interview participants indicated that certain communities rely on the use of tutorials or workshops that are designed to help newcomers learn about a project:

"Yes, I would say that one kind of training is tutorials for new contributors ... One thing that we do is that we run weekend workshops that teach students how to contribute to open source".

It was found that some FLOSS communities requested that their newcomers take quizzes or tests in order to be qualified as community members:

"The quizzes are here to teach them: Here are the things that you absolutely have to know out of all of those places. We are going to ask you the questions with the intent that you find the answer and learn how it is supposed to work whereas of having the intent that you already know the answer."

The analysis also revealed that it is a rather common practice for FLOSS projects to tag tasks that are particularly suitable for newcomers. Those tasks are then used as means for newcomers to gain knowledge about a project and its practices:

"There are junior jobs that people can look at. There are really simple bugs, really simple tasks."

In addition, in spite of the availability of such tasks, newcomers are usually free to move straight into regular project contributions and skip those tailored for newcomers:

“It really depends on the newcomer. There is some people that jump right in, they don’t even read anything, do not ask questions and they do everything wrong according to the community sort of habits and norms”.

As a result, *task segregation* (see Table 7 for definition) was found to vary significantly per FLOSS project and per newcomer from a high degree of task segregation to no segregation at all. Some newcomers may dedicate a lot of time and effort in undertaking those special tasks that community have designed to help newcomers learn about a project and join a community, whereas other newcomers may jump right in and start contributing.

Table 7: From *formal tactics* to *task segregation*: definitions

Construct name and definition	
<i>Van Maanen and Schein (1978) - Formal tactics</i>	Those processes in which a newcomer is more or less segregated from regular organizational members while being put through a set of experiences tailored explicitly for the newcomer.
<i>Task segregation</i>	The degree to which a newcomer has performed tasks that are specifically tailored for newcomers.

4.5.3.2 Interaction intensity

Some organizations have traditionally relied on a group approach when dealing with new recruits in an attempt to produce standardized responses to situations (Jones, 1986). In such practices, newcomers are collectively involved in a set of experiences and have to face a “boundary passage” (Van Maanen & Schein, 1979).

The researcher explored the idea of collective newcomer experience in FLOSS communities, but the interviewees did not acknowledge the existence of such group practices. A respondent justified the absence of collective tactics in FLOSS communities by the virtual and asynchronous nature of interactions:

“I don’t think so because they show up at random times. I don’t think it is happening much”.

Other participants expressed their reluctance towards forming groups of newcomers as they did not believe in its effectiveness:

“I don’t really know if having a bunch of beginners in the same place is going to help much because what are we going to tell each other apart for what we already know?”

“It is good to have these groups but at the same time I feel they are keeping a lot back. Not pushing people enough, you should go to the real world and do your work”.

The analysis of the interview data thus revealed that collective tactics from a strict point of view are quite rare in all the FLOSS communities the interviewees had been involved in. However, the experience of a newcomer may still involve a collective element since individuals can seek out other newcomers of their own volition when

learning about how to contribute to a project. Whereas a FLOSS community newcomer may become a project contributor as the result of a solitary experience during which the person has learned from the various project resources and/or by direct experimentation, others may rely on other members through the active use of communication channels such as forums, mailing lists, or IRC channels. As a result, the degree or intensity of interaction with other contributors appeared as an important aspect.

Even though FLOSS projects were not found to rely on newcomer initiatives in which individuals would be collectively induced in the community, the analysis of the interview data revealed that it was rather common for FLOSS communities to dedicate specific communication channels to assist newcomers and encourage them to interact together:

“There is also a full beginner forum, for beginners, staffed by moderators who are tolerant enough to answer the dumb questions. That’s probably the most formal beginner stuff in [project name].”

Other participants who had been involved with various mentoring programmes acknowledged that instances of newcomer-to-newcomer interactions could happen within formal mentoring programmes:

“It absolutely happens with the Google Summer of Code and the [programme name].”

Finally, a respondent pointed out that in very large communities such interactions may not be solely virtual as newcomers may take part to social events that encourage the involvement of newcomers:

“Certainly release parties, when a new version of [project name] comes out, they try to have actual parties in person at bars or restaurants and attract people who know nothing about [project name] to come along and chat.”

In this research, the term *interaction intensity* (see Table 8 for definition) was chosen in order to encompass the degree to which a newcomer is actively involved with other newcomers and community members while learning about how to become a project contributor. Based on the findings, the notion of *interaction intensity* defined in this research encompasses any type of interactions whether they are Internet-enabled or face-to-face, and whether they happen on general communication channels of a project or on channels specifically dedicated to newcomers.

Table 8: From *collective tactics* to *interaction intensity*: definitions

Construct name and definition	
<i>Van Maanen and Schein (1978) - Collective tactics</i>	The tactic of taking a group of recruits who are facing a given boundary passage and putting them through a common set of experiences together.
<i>Interaction intensity</i>	the degree to which a newcomer is actively <i>engaged</i> with other newcomers and community members while learning how to become a project contributor.

4.5.3.3 *Structured sequence of tasks and structured joining processes*

When investigating the relevance and applicability of both sequential and fixed tactics (see Chapter 2 for a definition of both types of socialization tactic) in the FLOSS context, it clearly appeared that FLOSS socialization significantly differs from organizational socialization along both aspects.

Both types of socialization tactics, sequential and fixed, refer to the sequence of activities or experiences a newcomer has to go through within an organization. Sequential tactics assesses the existence of some explicit information about such sequence whereas fixed tactics characterize whether a newcomer is provided precise knowledge of the timetable associated with completing each stage in this sequence of activities (Jones, 1986).

According to the traditional socialization literature, sequential (versus random) and fixed (versus variable) deal with the content of the information provided to new recruits via socialization (Jones, 1986). The nature of such content in FLOSS communities was found to be quite distinctive in terms of both the tasks performed by newcomers and the process newcomers have to follow when joining.

An important result that stemmed from the analysis of the interviews and that made FLOSS socialization unique compared to traditional socialization is that there is no time constraint associated with the socialization experience of newcomers as individuals' degree of involvement can notably vary and as status promotion tends to be subjectively triggered (instead of arisen after a given period of time):

“Basically, absolutely not. Those projects have no sort of timetable for new contributors especially because the sort of high watermark for a project to make somebody go to the next stage of power in your project, say commit access, or maybe read access to the primary mailing list, or write access to the bug tracker is very subjective.”

“So in general, in any project, there is never a fixed timetable. I think I have never seen one.”

As a result, the idea that newcomers are provided time indications about when each stage of a given sequence of activities shall be performed could not be extended to the FLOSS context. However, while reflecting on the potential existence of some degree of structure in the tasks undertaken by newcomers, questions about the nature of such structure were raised during the interviews. The analysis of the interview data suggested two separate aspects in which some degree of structure can be observed during a newcomer's socialization experience.

First, there can exist a logical pattern in the sequence of project tasks performed by a newcomer, the purpose of which is to help the newcomer to learn how to contribute to a project. Second, some communities rely on a highly-structured joining process which requires their newcomers to follow a series of steps in order to become a contributor. The terms *task purposefulness* and *joining structuredness* were derived by the researcher to capture these concepts.

TASK PURPOSEFULNESS *Task purposefulness* refers to the degree to which the sequence of project tasks performed by a newcomer is not randomly chosen (from a learning perspective) but rather obeys a logic which aims at helping a newcomer to learn how to contribute to a project. The tasks may be entirely self-selected, recommended or even assigned by the community to the newcomer. The logical connection between tasks may pertain to the difficulty level in which case the complexity of the performed tasks gradually increases:

“For example, there is a list of potential project where you can start, categorized by difficulty level, by prerequisite you have to be coming with, with mentors available to talk to about them.”

“Basically, you start with some bugs that are small that are agreed on that you can work with and also connect with the people who can help you with that bug.”

It was also found that the logical connection among the set of tasks being undertaken may also pertain to project scope, as a participant noted:

“People don’t have to be filing bugs before a year in [project name] before they start working on fixing a bug. They can jump in and learn the technology for a particular project, look at its code and fix a small bug. You can find things that are more narrow in scope, yes.”

Whether a newcomer is assigned tasks by an experienced contributor or the person simply follows the recommendations provided by the community, the interviewees noted that handling bug reports such as bug triaging and reporting are usually the type of activity that communities encourage their newcomers to perform for those who want to be technical contributors. Once a person has demonstrated a thorough enough knowledge about how bugs are handled within a project, a newcomer is then usually encouraged to submit a few simple bug fixes (or patches), followed by contributions of larger scope such as submitting new features:

“Your ability to submit a good bug report is your understanding of the computer. Once you’ve shown that, you can submit a good patch which is leaved on the service of it because you’ve shown that you understand the computer so the patch is probably valid. And then the next step in is a patch for a non-bug, a feature patch. It probably wouldn’t be looked as a feature unless you’ve already shown that you don’t break code when you submitting fixes.”

JOINING (PROCESS) STRUCTUREDNESS *Joining structuredness* does not characterize the actual project contributions undertaken by a newcomer when learning how to become an official contributor. Instead, it refers to the degree to which a newcomer has to follow a formal procedure to join a project and the set of steps that a community expects its new members to follow.

The community process that one has to follow to officially become a project contributor was found to vary significantly. In FLOSS projects, it is common to acknowledge that a person has officially become a technical project member when commit rights are given to this person. One ‘commits’ code when submitting a change to the source

code repository, which then is integrated into the software. Certain FLOSS communities have adopted a simplistic approach in which newcomers become official project members once an experienced member judges that a newcomer has demonstrated solid skills and contributed enough to the project:

“So it is very easy to get commit access to our repository for example. If you have two or three merge requests then someone will likely say ‘you should get a commit account, we will advocate that for you’ then you have it.”

“It is almost always by gut feeling, if you are sensitive to this person, you will do something for him.”

Other communities relies on a much more formal approach in which all newcomers have to go through a structured joining procedure in order to be granted the status of official contributor:

“You can’t become a developer before going through this process. You can till put a patch on Bugzilla, you can still post to the mailing list. Your ability to commit ode yourself is not existent.”

The nature of the joining procedure as well as the set of steps that have to be followed were found to vary from one community to another. The interviewees identified a range of mandatory joining activities such as sponsorship mechanisms or quizzes that a newcomer needs to pass before being able to be granted membership status:

“Step 1, you create some change to a package, by maybe creating a new package or by submitting a patch for a bug on the packages. Sometimes, you can create some artefact that is a contribution to [project name] without having being put into [project name] yet. That’s step one. Step 2 is you contact somebody ... who has upload access to the archives and who adds your contribution to [project name]. There is different ways of this to happen. As you do that second step more and more, that person says ‘You should become a developer in [project name]’. Then step 3, you go through this long process at the end of which you are a [project name] developer”.

One way to justify the reliance on a highly-structured joining processes is that it allows a finer filtering of potential contributors, and helps retain those whose profiles match a community’s expectations, as a participant pointed out:

“With [project name], it’s a little bit harder now to get in. The test has become a bit harder. But, at the same time, it’s harder for people to slip through the cracks.”

In conclusion, *task purposefulness* and *joining structuredness* (see Table 9) both emerged as important aspects of a newcomer’s socialization experience. They were both found to vary significantly depending on the choices and actions of a newcomer or depending on communities themselves.

4.5.3.4 *Mentoring*

The term *mentor* originates from the Greek mythology, where it was used to describe a “relationship between a younger adult and an older, more experienced adult [who]

Table 9: *Task purposefulness and joining structuredness: definitions*

Construct name and definition	
<i>Van Maanen and Schein (1978) - Sequential tactics</i>	The degree to which the organization or occupation specifies a given sequence of discrete and identifiable steps leading to the-target role.
<i>Van Maanen and Schein (1978) - Fixed tactics</i>	The degree to which the steps involved in a socialization process have a timetable associated with them that is both adhered to by the organization and communicated to the recruit.
<i>Task purposefulness</i>	the degree to which the sequence of tasks performed by a newcomer has been purposefully selected by the newcomer in order to learn how to become a contributor.
<i>Joining (process) structuredness</i>	the degree to which a newcomer has to adhere to a step-by-stepwise joining process in order to become a community contributor.

helps the younger individual learn to navigate the adult world and the world of work” (Kram & Isabella, 1985, p. 2).

Traditionally, organizations have relied on serial tactics in which a dedicated experienced employee is in charge of ‘grooming’ a newcomer (Jones, 1986; Van Maanen & Schein, 1979). When asking interview respondents to reflect on the relevance of experienced members supporting newcomers during the socialization process, all the interview participants noted the rather common practice of mentoring within FLOSS communities. As mentioned in the literature review section, both formal and informal mentoring do happen frequently in FLOSS communities. The success of the Google Summer of Code (GSoC) program is probably the best known initiative exemplifying the relevance and rapid spread of formal mentoring in FLOSS communities. The participants also highlighted the occurrence of informal mentoring. Seven of the 11 interview participants indicated that they had been mentored (informally or formally) by one or several experienced community members.

Informal mentoring was found to be quite widespread. It is characterized as a relationship between a newcomer and an experienced member that develops over time while a newcomer is learning the ropes of a project:

“Yes, there is a lot of you see somebody on the forums or the mailing lists who is struggling, and after a couple of interchanges of email, you end up exchanging personal messages like in IRC, Twitter, or Google Chat. From there, you develop a little relationship with somebody.”

In parallel, the participants acknowledged the use of formal mentoring in FLOSS communities. Nearly all the communities in which the interviewees had contributed, have been involved in the Google Summer of Code program for a number of years. The popularity of GSoC has rapidly increased since the beginning of the program in 2005. Whereas the first instance of GSoC gathered 41 FLOSS organizations (for about 200 students), the 2012 GSoC program involved 180 FLOSS organizations (with more than

1,200 students)⁶. Some communities (including those involved in the GSoC) also organize their own mentoring programs or else include mentoring as part of the process when joining a community:

“You cannot get a new developer bug filed if you do not have a sponsor or a mentor. Because the mentor is the person who files the bug for you.”

“When we file a new developer bug, there is always at least one mentor for that bug. The person who says ‘if this new recruit is going to be working on X then I already work on X so I am bringing them on to teach them exactly how X works and how it is maintained”.

As a result, the analysis of the interview transcripts highlighted the pertinence of *mentoring* (see Table 10) within FLOSS communities when dealing with newcomers.

Table 10: From *serial tactics* to *mentoring*: definitions

Construct name and definition	
<i>Van Maanen and Schein (1978) - Serial tactics</i>	A process in which experienced members of the organization groom newcomers who are about to assume similar kinds of positions in the organization.
<i>Mentoring</i>	The degree to which a newcomer has been taken under the wing of one or more experienced members while learning how to become a project contributor.

4.5.3.5 Supportiveness

According to [Van Maanen and Schein \(1979\)](#), investiture is an important aspect of newcomer experience. Investiture concerns the confirmation or disconfirmation of a newcomer identity when entering an organization. [Jones \(1986\)](#) noted that it refers to the degree of positive or negative social support that newcomers receive from experienced organizational members. As discussed below, investiture as such was not found to be significant in FLOSS communities. Instead, the community support provided during one’s newcomer experience was found to be an essential characteristic of FLOSS socialization.

Most interviewees emphasized the meritocratic nature of FLOSS communities and its influence on interactions among members. As an illustration, the Apache project officially claims that meritocracy is the basic principle that drives the community and describes it as “the more you do the more you are allowed to do. As a person acquires merit, his or her stature in the community grows, and (to a certain extent) the weight given to his or her opinions” (Apache Software Foundation, [2012](#)).

A majority of interview participants explained that the past identity of newcomers in terms of skills or experience is mostly disregarded in FLOSS communities as some-

⁶ See <http://code.google.com/soc/> for more demographic information about the Google Summer of Code program.

one's value is solely evaluated based on what individuals have actually contributed to a project:

"People rarely bring it up. Especially as a beginner when you start contributing there is a big share of anonymity where you are basically as good as your patches are."

"Most communities basically disregard them. Because in a sense, nobody cares about your qualifications, they care about the way you interact with people, that proves that you have knowledge."

As a consequence, it appeared that the idea of identity confirmation was not found to be relevant in the particular context of FLOSS communities. Despite the lack of evidence in showing the importance of identity confirmation/disconfirmation in FLOSS communities, the participants highlighted that the degree of positive support that is provided to newcomers varies significantly and played an important role in the successful/unsuccessful integration of potential new contributors:

"Some communities are supportive and other communities are not."

"I think it varies a lot. There are communities that are a little bit friendlier to newcomers. I think it depends, it really depends on who is a member, what sort of members you can expect in your community."

In addition, within a given FLOSS community, participants suggested that the degree of supportiveness perceived by a given newcomer differs significantly:

"You don't know. It will depend. The [project name] community is big. It's several hundred people. In certain projects, you will have a really outstanding experience. In some others, people may go away saying 'That guy is an asshole'."

"Some people are supportive in communities and some people somehow don't receive the support from supportive communities."

A number of interviewees pointed out that the notion of identity building in a meritocratic social structure is quite distinctive from traditional organizational entities. A newcomer's identity is shaped from scratch and based on the value of the person's recent contributions. Nonetheless, identity building was found to be influenced by the degree of support that is provided to newcomers by community members. Most interview participants insisted on the importance of being supportive to newcomers as a means to attract more new members and hence ensure the success of a project:

"I think [project name] on the whole is supportive. We at least consciously know that newcomers are vital to the continued success of the project [...] You have to keep bringing in new talents. "

In summary, the notion of *supportiveness* (see Table 11) was identified as a key aspect of a FLOSS community newcomer's experience.

4.6 A FLOSS SOCIALIZATION MODEL

The previous subsection presented the findings that were derived from the analysis of the collected interview data. Six key socialization factors were identified: *task segregation*, *task purposefulness*, *interaction intensity*, *mentoring*, *joining structuredness*, and

Table 11: From *investiture tactics* to *supportiveness*: definitions

Construct name and definition	
<i>Van Maanen and Schein (1978) - Investiture tactics</i>	The degree to which a socialization process is constructed to either confirm or disconfirm the entering identity of the recruit.
<i>Jones (1986) - Investiture tactics</i>	Investiture versus divestiture concerns the degree to which newcomers receive positive or negative social support after entry from experienced organizational members.
<i>Supportiveness</i>	The degree to which a newcomer has perceived a community to be supportive while learning how to become a project contributor.

supportiveness.

Task segregation relates to the nature of tasks performed by a newcomer (in terms of whether they are tailored or not for newcomers) whereas *task purposefulness* is about the existence of logical links among the performed tasks with the overall intention to learn about a project. As a result, both constructs address complementary aspects that characterize the tasks that FLOSS community newcomers undertake during their socialization phase. *Task segregation* and *task purposefulness* then characterize the interaction between a newcomer and the software project artifact.

Mentoring and *interaction intensity* characterize forms of social interaction during a newcomer's FLOSS community socialization experience, as they both deal with interactions between the newcomer and a mentor or other community members. The two constructs are thus two aspects that characterize a newcomer's experience in terms of the interaction between a newcomer and other members of a FLOSS community.

Finally, *supportiveness* and *joining structuredness* pertain to the community dimension of newcomer socialization. *Supportiveness* refers to the degree to which a newcomer perceives a community to be supportive. Meanwhile, *joining structuredness* concerns the extent to which a community expects a newcomer to go through a step-by-step joining process in order to become a community member. It is important to bear in mind that both factors are newcomer perceptions, as opposed to 'facts.' Nonetheless, both types of perception are about the interaction between a newcomer and the community during the socialization phase.

In conclusion, the six socialization factors identified in this research can be integrated in a socialization experience framework as presented as in Figure 17. The proposed sub-categorization into task, individuals, and community factors provides a more encompassing view of FLOSS newcomer socialization experience. It identifies three core aspects of socialization experience and forms the basis for the quantitative investigation that follows.

Figure 17: FLOSS socialization experience - Model
SOCIALIZATION EXPERIENCE

4.7 CHAPTER SUMMARY

The chapter starts with a description of the procedures used to collect the interview data. Some demographic information about the interviewees is then presented. The results of the interview data analysis are then introduced. First, the key instances of citizenship behaviours are described and categorized in terms of CCB-I (citizenship behaviours directed towards community members), and CCB-P (citizenship behaviours directed towards the project and its community). The extant literature is used to justify the relevance of each type of behaviour. The chapter ends with the identification of six socialization factors that were derived from the interview findings, and their integration in a FLOSS socialization framework. The next chapter introduces the conceptual models that were derived from the combination of the results of the literature review and interview data analysis.

CONCEPTUAL MODELS AND RESEARCH HYPOTHESES

Chapter 2 reviewed the literature relevant to this study to gain a firm understanding of the phenomenon under study. This chapter introduces additional literature to develop the two conceptual models that were tested further in this research project. Based on the results of the literature review and the outcome of the qualitative phase of this research, this chapter introduces the theoretical constructs that were investigated in this research as well as the research hypotheses that were derived. Finally, the use of partial least squares as the structural equation modeling technique is explained.

5.1 CONCEPTUAL DEFINITIONS

The conceptual definitions used in this research are summarized in Table 12. This research conceptualizes FLOSS socialization experience as *the process by which newcomers make the transition from being community outsiders to being insiders*. Chapter 4 presented the six socialization experience factors that were identified during the first phase of the methodology.

This research combines two overall approaches from the socialization tradition. Following Van Maanen and Schein (1979), newcomers in FLOSS communities are knowledge and experience recipients as FLOSS projects relies on the use of various initiatives to facilitate their integration (socialization tactics tradition). As well, the newcomer proactivity tradition argues that FLOSS newcomers are also active participants who shape their own socialization experience (Gruman et al., 2006; Morrison, 1993a, 1993b)

This research also relies on Katz' view of in-role and extra-role performance (1964; 1966) in operationalizing contributor performance as *task performance* and *community citizenship behaviour*.

Task performance is defined as *the extent to which an individual carries out assigned tasks by meeting some minimal level of performance defined by the FLOSS community to which he/she belongs*. Community citizenship behaviours are defined as *behaviours that are discretionary, not directly or explicitly recognized by the formal reward system of the community and that in the aggregate promote the effective functioning of the community* (inspired from Organ, 1988). This research operationalizes Community Citizenship Behaviours (or CCB) following Williams and Anderson's works in OCB wherein they distinguish citizenship behaviours targetted towards individuals and those targetting the organization.

This project introduces the notion of CCB-I (community citizenship behaviours targetted towards individuals) and CCB-P (community citizenship behaviours targetted towards the project and its community). Chapter 4 presented the results from the qualitative phase of this research in which important instances of CCB-I and CCB-P were identified. The CCB construct used in this research is then a second-order construct

whose dimensions are CCB-I and CCB-P. This approach is consistent with the OCB research tradition which relied on a similar operationalization of the OCB construct (N. P. Podsakoff et al., 2009).

5.2 DUAL INVESTIGATION AND CONCEPTUAL MODELS

The qualitative component of this research identified six latent variables that characterize one's newcomer experience in a FLOSS community. Even though some aspects of the identified variables were found to be theoretically-grounded, the FLOSS-tailored socialization framework introduced in this research is overall novel. It was decided to test two alternative models and to determine which one was the best predictor of individual performance. Such decision was motivated by the fact that the effects of the differences between the two alternative FLOSS-specific models could not be hypothesized from previous literature.

In the first stage, a conceptual model was derived by linking all socialization factors to the individual performance constructs (*task performance* and *CCB*). The direct effect of all six factors on the performance measures was assessed before considering other types of relationships such as mediating or moderating effects. This approach is consistent with the socialization literature, wherein the bulk of research works investigated the direct impact of socialization on performance (see Chapter 2).

The second stage of the investigation consisted of refining the first conceptual model by introducing the mediating effect of important proximal socialization factors that particularly suit the FLOSS context: *social identification* and *social integration*. Indeed, the review of the literature identified these two particular proximal factors as potential candidates for mediating the relationship between socialization and individual performance. The following subsections introduce and justify the development of the two conceptual models and their derived hypotheses.

5.3 FIRST RESEARCH MODEL

Inspired by the early socialization literature tradition, the first research model hypothesizes that factors characterizing FLOSS newcomer socialization experience have an influence on FLOSS community contributor performance. Chapter 4 presented the results that were derived from the first data collection phase of this project (through interviews with FLOSS contributors). Six socialization factors were then introduced: *task segregation*, *task purposefulness*, *interaction intensity*, *mentoring*, *supportiveness*, and *joining structuredness* characterizing task-related, individuals-related, or community-related aspects of one's FLOSS community newcomer experience. The literature review chapter presented findings of the socialization literature that investigated the direct impact of socialization on individual performance (e.g. D. G. Allen, 2006; Ashforth & Saks, 1996; Gruman et al., 2006; Saks & Ashforth, 1997b), findings which justify the underlying structure of the conceptual model (see Figure 18). The model hypothesizes a direct relationship between the six socialization factors and the two performance-related dependent variables: *task performance* and *CCB*.

Table 12: Summary of conceptual definitions

Construct name	Definition
TASK SEGREGATION	<i>the degree to which a newcomer has performed tasks that are specifically tailored for newcomers.</i>
TASK PURPOSEFULNESS	<i>the degree to which the sequence of tasks performed by a newcomer has been purposefully selected by the newcomer in order to learn how to become a contributor.</i>
INTERACTION INTENSITY	<i>the degree to which a newcomer is actively engaged with other newcomers and community members while learning how to become a project contributor.</i>
MENTORING	<i>the degree to which a newcomer has been taken under the wing of one or more experienced members while learning how to become a project contributor.</i>
JOINING STRUCTUREDNESS	<i>the degree to which a newcomer has to adhere to a step-by-stepwise joining process in order to become a community contributor.</i>
SUPPORTIVENESS	<i>the degree to which a newcomer has perceived a community to be supportive while learning how to become a project contributor.</i>
SOCIAL IDENTIFICATION	<i>the extent to which one identifies with a FLOSS community, leading to viewing the community's successes and failures as one's own.</i>
SOCIAL INTEGRATION	<i>the extent to which an individual perceives himself/herself to be trusted and accepted by the other FLOSS community contributors.</i>
TASK PERFORMANCE	<i>the extent to which an individual carries out assigned tasks by meeting some minimal level of performance defined by the FLOSS community to which he/she belongs.</i>
COMMUNITY CITIZENSHIP BEHAVIOUR	<i>behaviours that are discretionary, not directly or explicitly recognized by the formal reward system of the community and that in the aggregate promote the effective functioning of the community.</i>

In order to help newcomers learn how to contribute, FLOSS communities often tailor certain tasks specifically for newcomers or at least tag project tasks that newcomers are able to perform. In the GNOME community for instance, certain bug reports are marked with the "gnome-love" keyword and are recommended for project newcomers. Other examples include the Debian project having a set of manuals for new contributors¹ or the KDE project providing dedicated tutorials². FLOSS projects expect that such tasks specifically designed for newcomers will help new contributors learn about the project and thus perform better. As a result, the following hypotheses were derived:

Research Hypothesis 1a (H_{1a}): *Task segregation will be positively associated with task performance.*

Research Hypothesis 1b (H_{1b}): *Task segregation will be positively associated with community citizenship behaviour.*

¹ <http://www.debian.org/doc/#manuals>

² <http://techbase.kde.org/Development/Tutorials>

Figure 18: First conceptual model

FLOSS project newcomers are free to perform tasks of their own choice. They are free to get started with submitted a patch which is expected to be handled by experienced contributors, or write an entire section in the documentation about a particularly complex module of the project. However, a newcomer may decide to purposefully perform a set of tasks that would help him or her learn about a project, gradually increasing the complexity level or else undertaking tasks from complementary parts of a software project. In addition, FLOSS projects often recommend that newcomers perform tasks in a certain order which is judged to facilitate project learning. For instance, the Mozilla developer platform provides a precise set of steps to follow in order to learn how to contribute to the Mozilla codebase³:

- Step 1 - Build application (Firefox, Thunderbird or another application)
- Step 2 - Understand how contributing to Mozilla works
- Step 3 - Find something to work on (Fix your pet peeve / Find a bug we've identified as being good for newcomers)
- Step 4 - Fix the bug
- Step 5 - Get your code reviewed
- Step 5b - Follow it up
- Step 6 - Respond to the review
- Step 7 - Actually get the code into the tree
- Step 8 - Repeat

As a consequence, structuring tasks in a way that optimizes learning seems to be expected to increase newcomer performance. Whether newcomers choose their own

³ <https://developer.mozilla.org/en-US/docs/Introduction>

sequence of tasks to perform or whether they follow a sequence of tasks that a FLOSS project recommends, *task purposefulness* is then hypothetically associated with higher performance levels. The following hypotheses were generated:

Research Hypothesis 2a (H_{2a}): *Task purposefulness will be positively associated with task performance.*

Research Hypothesis 2b (H_{2b}): *Task purposefulness will be positively associated with community citizenship behaviour.*

In their extensive review of the socialization literature applied to the domain of virtual communities, Kraut, Burke and Riedl (2011) concluded that by using collective socialization tactics, new members are likely to learn better how to behave in a community, and more importantly contribute more to the community. Social interaction with network ties within online communities were found to have a positive influence on the amount of knowledge being shared (Chiu et al., 2006). Group research also demonstrated that collaborative learning has a positive impact on individual performance (Brodbeck & Greitemeyer, 2000; Olivera & Straus, 2004). Such findings are in line with the notion of positive group-to-individual transfer (G-I transfer) from Laughlin and Sweeney (1977), wherein individual performance is enhanced through the participation in a collaborative group. In the FLOSS context, interpersonal relationships were found to have an impact on contributor involvement, which in turns affects individual performance (Xu et al., 2009). Korte (2010) emphasized the importance of social interaction during socialization by pointing out that, “the initial interactions between newcomers and their work groups are extremely important, because they establish the quality of the relationship, which in turn affects the newcomers’ attitudes, satisfaction, and performance on the job” (p. 30). Consequently, the following hypotheses were proposed:

Research Hypothesis 3a (H_{3a}): *Interaction intensity will be positively associated with task performance.*

Research Hypothesis 3b (H_{3b}): *Interaction intensity will be positively associated with community citizenship behaviour.*

Research on mentoring in organizational settings has a long tradition (Higgins & Kram, 2001; Kram & Isabella, 1985). Pioneering works about mentoring such as Levinson, Darrow, Klein, Levinson and McKee (1978) or Kram (1983) emphasized the powerful influence of mentoring on newcomer success in organizational environments. Individual performance has been one of the investigated outcomes of mentoring (Kammeyer-Mueller & Judge, 2008). A meta-analysis about the impact of mentoring on performance revealed that performance has been measured from various perspectives such as objective measures of performance, perceived competence of the protege, or the mentor’s appraisal of the quantity of work provided by the protege (Kammeyer-Mueller & Judge, 2008). The following hypotheses are proposed:

Research Hypothesis 4a (H_{4a}): *Mentoring will be positively associated with task performance.*

Research Hypothesis 4b (H_{4b}): *Mentoring will be positively associated with community citizenship behaviour.*

Some FLOSS projects such as Debian rely on a very formal joining process that consists of having newcomers follow a detailed step-by-step process at the end of which they become official contributors. The findings provided by the qualitative phase of this research thus allowed to identify *joining structuredness* defined as *the degree to which the steps involved to in becoming a contributor follow a structured joining process that is both adhered to by the community and communicated to the newcomer*, as an important socialization factor. Such socialization practices can be seen as a community socialization ritual which aims at gradually familiarizing newcomers to the project culture. Because such practices are intended to increase the effectiveness of newcomer socialization (Islam & Zyphur, 2009), it is hypothesized that:

Research Hypothesis 5a (H_{5a}): *Joining structuredness will be positively associated with task performance.*

Research Hypothesis 5b (H_{5b}): *Joining structuredness will be positively associated with community citizenship behaviour.*

Organizational support has been examined in the context of organizational socialization research (J. Chen & Eldridge, 2011). In their study that investigated the role of perceived organizational support and leader-member exchange on newcomer adjustment, J. Chen and Eldridge (2011) concluded that perceived organizational support may have an influence on turnover intention (socialization distal outcome). Research has also suggested that perceived organizational support has an influence on various work outcomes, such as job satisfaction, positive mood, affective commitment, and lessened withdrawal behaviour (Rhoades & Eisenberger, 2002). Meanwhile, organizational support was also found to have a positive impact on group effectiveness in the context of Wikipedia article production (Carillo & Okoli, 2011). Consequently, this research tested the following hypotheses:

Research Hypothesis 6a (H_{6a}): *Supportiveness will be positively associated with task performance.*

Research Hypothesis 6b (H_{6b}): *Supportiveness will be positively associated with community citizenship behaviour.*

5.4 SECOND CONCEPTUAL RESEARCH MODEL

5.4.1 Overall description

In addition to emphasizing the impact of socialization factors on distal outcomes, the review of the socialization literature also showed the importance of proximal socialization factors in explaining distal outcomes. This suggests that in the FLOSS community context, there may be key socialization proximal factors that mediate the relationship between socialization factors and a distal outcome such as individual performance.

The second conceptual model (see Figure 19) is introduced in this subsection. The model includes the mediating effect of two socialization variables that can be seen as

Figure 19: Second conceptual model

proximal socialization outcomes: *social identification* and *social integration*. The mediating variables are defined and the generated research hypotheses are then justified in light of the literature. The conceptual model integrates the overall framework adopted in socialization research that sees factors characterizing FLOSS newcomer experience affecting proximal socialization outcomes which in turn impact a socialization distal outcome: contributor performance. The model overall hypothesizes that task-related socialization experience factors have a direct impact on contributor performance. In addition, the four other socialization factors that have a social component: *mentoring*, *interaction intensity*, *supportiveness*, and *joining structuredness* have a positive influence on the two proximal outcomes: *social identification* and *social integration*, which in turn have an impact on both *task performance* and CCB.

5.4.2 Social identification

According to social identity theory (Tajfel & Turner, 1985), individuals classify themselves into social categories such as religious affiliation or organizational membership. Ashforth and Mael (1989) re-investigated the notion of organizational identification in light of social identity theory and applied their work in the domain of socialization, role conflict, and inter-group relations. In this research social identification is defined as *the extent to which one identifies with a FLOSS community, leading to viewing the community's successes and failures as one's own* (Mael & Ashforth, 1992).

According to the organizational socialization literature, social identification plays an important role in capturing how organizational newcomers are concerned with building a situational definition (Ashforth & Mael, 1989). Van Maanen and Schein (1979) argued that conceptions of the self are learned by interpreting the responses of others in situated social interactions. Furthermore, social identification was identified as one of the influential proximal outcomes of organizational socialization (Saks & Ashforth, 1997b), having an impact on distal outcomes such as job performance or OCB.

The literature has highlighted some overlap between the affective dimension of social identity and the notion of affective commitment (Bergami & Bagozzi, 2000; Ellemers, Kortekaas & Ouwerkerk, 1999). N. J. Allen and Meyer (1996, p. 253)'s definition of affective commitment, "identification with, involvement in, and emotional attachment to the organization," confirms the similarities of some of the aspects of the two notions.

Social identity was found to influence individual productivity within groups (Worchel, Rothgerber, Day, Hart & Butemeyer, 1998). The impact of social identity on individual participation has also been investigated in the context of online communities. Social identity was found to have a direct impact on overall online community participation (Liu & Iyer, 2007) or else through the mediation of participation desires and "we-intentions" (Bagozzi, Dholakia & Pearo, 2007; Dholakia, Bagozzi & Pearo, 2004). We-intention was defined as the situation in which an agent intends to perform a certain action jointly with others. In applying both common identity and bond theories to provide overall recommendations to design online communities, Kraut and Kiesler (2007) highlighted the importance of social categorization and identity-based attachment in generating active participation, high conformation to group norms, robustness to turnover among other behavioural variables.

In FLOSS research, social identity was also found to influence the behaviour of community members. Linux usergroup members' perceptions of social identity were found to influence we-intentions to participate in the usergroups which in turn was found to have an impact on participation (Bagozzi & Dholakia, 2006). Social identity was also found to have a direct and indirect impact through the mediation of effort intensity on contributor task performance (Ke & Zhang, 2009). A study empirically verified how adherence to components of the FLOSS ideology impacts effectiveness in FLOSS teams (Stewart & Gosain, 2006). Similar results identified a category of FLOSS community participants (called 'hobbyists' by the author) who are critical to the long-term viability of FLOSS projects and communities (Shah, 2006). These individuals were found to take on tasks that might otherwise go undone and work to maintain the simplicity and modularity of project code.

The notion of social identification appears particularly relevant in the context of large and complex FLOSS projects. First, the building of a situational definition for newcomers seems to be particularly challenging in projects that are highly modular and that require special efforts to understand how one can contribute to such complex IT artifacts. Besides, large FLOSS projects are characterized by large communities of contributors interacting online. It thus echoes the importance of social identity building in online community participation that was found in past research. Finally, large and complex FLOSS projects usually have shared values and norms embodied into a code of conduct that shall be followed by all contributors. This aspect indicates that Stewart and Gosain's (2006) results about the link between the adherence to components of the FLOSS ideology and team effectiveness particularly apply to the case of large and complex FLOSS projects.

Thus, the following hypothesis was generated:

Research Hypothesis 7a (H_{7a}): Social identification will be positively associated with task performance.

Studies from the socialization literature have demonstrated the significant impact of organizational identification on OCB (Ge, Su & Zhou, 2010). Gorden, Anderson and Bruning (1992) empirically demonstrated a positive relationship between employees' perceptions of the commitment that exists in their relationship with their employing organization and citizenship behaviour. In a similar manner, organizational identification and internalization were found to be positively related to prosocial behaviors (C. O'Reilly & Chatman, 1986).

FLOSS research has also highlighted the presence of a strong sense of community identification and adherence to norms of behaviour within FLOSS communities and projects (Lakhani & Wolf, 2007). Helping and knowledge sharing behaviours within FLOSS communities have been examined by various researchers (Endres et al., 2007; Kuk, 2006; Sowe et al., 2008). A subset of the motivational studies identified intrinsic motivations such as helping, altruism, sharing as influencing contributors to join FLOSS communities and projects (Ghosh, Glott, Krieger & Robles, 2002; Hars & Ou, 2002). Nonetheless, the literature review did not uncover studies which empirically tested the relationship between social identification and community citizenship behaviour. This research project strives to address this research gap, drawing insights from results provided by both the socialization and online communities literature. The following hypothesis was therefore proposed:

Research Hypothesis 7b (H_{7b}): Social identification will be positively associated with community citizenship behaviour.

5.4.3 *Relationship between socialization experience and social identification*

Early research about socialization has asserted that members of organizations incorporate the meaning of the organization into their self-concept (Ashforth & Mael, 1989; Feldman, 1976; Van Maanen & Schein, 1979). The socialization tactics literature has provided a wealth of evidence about the influence of institutionalized tactics on organizational identification (Ashforth, 1998; Ashforth & Saks, 1996) as well on identity-related constructs such as person-organization fit perceptions (Cable & Parsons, 2001; Kim et al., 2005; Saks & Gruman, 2011), or organizational commitment (Ashforth & Saks, 1996; Cooper-Thomas & Anderson, 2002; Gruman et al., 2006; Jones, 1986). Even though person-organization fit and organizational commitment are distinct constructs, an investigation of the employed conceptual and operational definitions revealed some theoretical aspects which overlap with social identification.

In their seminal work about mentoring relationships in organizational settings, Kram and Isabella (1985) identified psychosocial support as one of the two main functions played by a mentor. The authors explained that psychosocial support concerns the interpersonal aspects of the mentor-mentee relationship that enhance an individual's sense of competence and identification to an organization. It can then be argued that a mentor-mentee relationship may have an influence on the degree to which a new-comer relates to an organization or a FLOSS community and identifies with it.

A study involving both MBA graduates and engineering professionals used Van Maanen and Schein (1979)'s original model and revealed that serial tactics have a significant positive effect on organizational identification, whereas collective tactics and fixed tactics had a significant influence through the moderating effect of work experience (Yi & Uen, 2006). Serial tactics were also shown to have a significant impact on the congruence between organizational and individual values in the context of sales-force socialization (Grant & Bush, 1996). A previous study by Chatman (1991) found that time spent with a mentor predicted both fit and changes in fit after one year.

Allen and Meyer (1990a) verified the significance of the positive relationship between investiture tactics and organizational commitment; the notion of commitment tapping into the affective dimension of social identity. Person-organization fit and organizational commitment were both found to be affected by investiture tactics (Cooper-Thomas, van Vianen & Anderson, 2004). Simosi (2010) demonstrated the role of investiture-divestiture tactics as a moderator in the relationship between newcomers' task-related information acquisition and organizational affective commitment.

In sum, the literature suggests that mentoring relationships facilitate the emotional attachment of an individual to the group, suggesting that identification with a FLOSS community by a newcomer would be positively influenced by mentoring practices and support from the embedding social entity. Hypotheses were then derived accordingly:

Research Hypothesis 10a (H_{10a}): Mentoring will be positively associated with social identification.

Research Hypothesis 12a (H_{12a}): Supportiveness will be positively associated with social identification.

Sensemaking (including information and feedback seeking) has a long tradition in socialization research (Ashford & Black, 1996). Early works demonstrated that information-seeking efforts towards co-workers were linked to newcomers' job satisfaction, retention, and commitment (Louis et al., 1983). Saks and Ashforth (1997a) argued that "uncertainty is reduced through information provided via various communication channels, notably social interactions with superiors and peers". Information acquisition from supervisors and coworkers was shown to positively affect attitudinal socialization outcomes such as satisfaction, commitment, and feelings of adjustment (Ostroff & Kozlowski, 1992). A meta-analysis of 70 socialization studies found that information seeking had an impact on organizational commitment (Bauer et al., 2007). Robert Kraut and colleagues' works about socialization in virtual communities concludes that by using collective socialization tactics, new members are more likely to become more committed to the community (Kraut, Burke & Riedl, 2011). As a consequence, the following hypothesis was derived:

Research Hypothesis 2b (H_{9a}): Interaction intensity will be positively associated with social identification.

Trice and Beyer (1984) were among the first scholars to study rites and ceremonies in the organizational context and investigated their influence on the notions of commitment and social identity. Early studies from Van Maanen (1973, 1975) showed that the ritualized socialization of policemen engendered a stronger commitment from the

recruits. It then appears that a formal and structured joining process with FLOSS communities tend to develop a higher sense of identification to the community. Kraut, Riedl and Burke (2011) reflected on the various means to keep newcomers engaged in online communities and pointed out that “barriers and initiation rituals that cause newcomers to suffer a little before joining a group should increase their eventual commitment” (p. 22). The authors justified their point by referring to Festinger’s (1957) theory of cognitive dissonance. The theory explains why newcomers are more attached to groups when they have to go through a challenging initiation process to join them. The hypothesis that follows is then proposed:

Research Hypothesis 2c (H_{11a}): Joining structuredness will be positively associated with social identification.

5.4.4 *Social integration*

One of the critical challenges for newcomers is to become successfully integrated into their work group (Feldman, 1976; C. D. Fisher, 1986; Louis, 1980). The socialization literature has introduced the notion of social integration as one of the important proximal outcomes of employee socialization. Bauer, Morrison and Callister (1998) defined social integration as the newcomer’s integration into his or her new work group and Feldman (1981) stated that “social integration concerns the status of newcomers being trusted and accepted by group members”. Early works on social integration also recognized the multifaceted nature of the phenomenon that reflects attraction to the group, satisfaction with other group members, and social interaction with members of the group (1978; 1989).

von Hippel and Von Krogh (2003) developed the view that FLOSS is a compound development model that comprises both elements of the private investment and the collective action models. The authors emphasized the importance of studying the notion of social integration in the FLOSS context:

Social integration may not prevent withdrawal from the project through punishment, but perhaps through the individual utility derived from a social category such as a core-developer status (p. 218).

In other words, social integration is thus seen as a means for FLOSS communities to ensure that members sustain their participation in a project. Drawing insights from Bauer et al. (2007)’s work, Kraut, Riedl and Burke (2011) also suggest the importance of social acceptance when dealing with online community newcomers to influence their performance level. It may therefore be hypothesized that social integration in FLOSS communities plays a role in influencing contributor performance.

FLOSS newcomers have been shown to rely on various strategies to facilitate their acceptance with a community. For instance, Bergquist and Ljungberg (2001) showed that humility was used by newbies through various narrative norms (such as using the headline ‘newbie’ in messages) to get acceptance from the community. Gaining reputation through the demonstration of skills can be seen as another such mechanism as it increases the trust from other community members (Stewart & Gosain, 2006) thus affecting the feeling that one is accepted by the community.

This aspect seems particularly pertinent in large and complex FLOSS projects since they can be seen as instances of commons-based peer production. It has been shown that in such entities, social cues such as status and reputation are key mechanisms that govern motivation and participation (Benkler, 2002, 2006). Building on the results reviewed in the previous paragraph, the role of social integration thus seems critical in predicting the behaviour and performance of newcomers in large and complex FLOSS projects.

In this research, the definition of social integration was adapted from Feldman (1981); a definition which has been consistently used in organizational behaviour research (J. Chen & Eldridge, 2011; Kammeyer-Mueller & Wanberg, 2003; Wanberg & Kammeyer-Mueller, 2000). Social integration is defined as *the extent to which an individual perceives himself/herself to be trusted and accepted by the other FLOSS community contributors*.

5.4.4.1 *Influence of social integration on performance*

The socialization literature has investigated the influence of social integration on various distal outcomes, Bauer et al. (2007) found a positive influence of social integration on five distal socialization outcomes: performance, job satisfaction, organizational commitment, intention to remain and turnover. J. Chen and Eldridge (2011) demonstrated the negative impact of social integration on turnover intention. In a 4-wave longitudinal study, a positive direct relationship was found between group integration and organizational commitment (Kammeyer-Mueller & Wanberg, 2003). C. O'Reilly et al. (1989) found group social integration to be significantly and negatively related to group turnover rates.

An important part of social integration is about the development of relationships with co-workers (Morrison, 1993a). This echoes an important stream of research in organizational behaviour, as well as virtual communities and FLOSS research, that has relied on social capital theory. Social capital refers to the resources embedded within networks of human relationships (Nahapiet & Ghoshal, 1998). The concept of social capital encompasses various social many aspects, of a social context which Nahapiet and Ghoshal (1998) categorized in terms of the structural, the cognitive, and the relational dimensions. In socialization research, a number of research projects have studied the influence of socialization tactics on the social integration of newcomers which bear a strong (Morrison, 2002a) or moderate (Gruman et al., 2006; Morrison, 1993a) relationship with aspects of social capital theory. The FLOSS literature have recently begun to recognize the relevance and potential of social capital theory in understanding individual contribution behaviour (Oh & Jeon, 2007; Wasko & Faraj, 2005).

An investigation of the importance of the structural dimension of social capital in FLOSS project success revealed the importance of both connections through contributors who bring their knowledge to the project, and the connections through members who transfer their knowledge to other projects (Méndez-Durón & García, 2008). Similarly, the number of direct and indirect ties has been shown to positively affect a FLOSS project's success (Singh, Tan & Mookerjee, 2011). Within the Wikipedia open content encyclopedia community, direct and indirect ties were found to have a significant positive effect on increasing participants' recognition-based performance (Okoli

& Oh, 2007). A study about the influence of all three dimensions of social capital found that the structural dimension was the strongest predictor of FLOSS expertise integration seen as the coordinated application of individual expertise in accomplishing project tasks (Chou & He, 2010). Furthermore, both direct and indirect ties were found to positively influence the productivity of FLOSS teams in Sourceforge projects (Tan, Mookerjee & Singh, 2007). Interpersonal relationships were found that have an impact on contributor involvement which in turns affects individual performance (Xu et al., 2009).

The above findings overall suggest that social integration may have an influence on both project-level and individual-level performance. Thus, the following hypotheses were derived:

Research Hypothesis 8a (H8a): Social integration will be positively associated with task performance.

Research Hypothesis 8b (H8b): Social integration will be positively associated with community citizenship behaviour.

5.4.5 *Relationship between socialization experience and social integration*

The socialization literature has investigated the influence of various aspects of one's socialization on the integration of newcomers within organizations. Institutionalized socialization tactics were found to have a positive effect on social integration (Gruman et al., 2006). Also, coworker socialization influence was found to be positively associated with work group integration (Kammeyer-Mueller & Wanberg, 2003) suggesting the importance of both context and social socialization tactics in engendering the development of relationships between a newcomer and other members. The related concept of newcomer embeddedness was also investigated by some socialization researchers. Results revealed that socialization tactics enable organizations to actively embed newcomers (D. G. Allen, 2006).

Studying mentor-mentee relationships, Kram and Isabella (1985) identified the relevance of career-related support, which characterizes how a mentee may benefit from the experience and influence of a mentor. Career-related support functions include actions from the mentor related to the mentee's human capital enhancement providing opportunities to create links to important individuals in the organization (Kammeyer-Mueller & Judge, 2008). Aspects of career-related support include sponsorship, exposure and visibility (T. D. Allen, Eby, Poteet, Lentz & Lima, 2004). Within the FLOSS community context, a mentor may thus help a newcomer become integrated in the community by introducing the person to other community members, including influential ones. This aspect of mentoring relationship was clearly emphasized by the interview respondents during the qualitative phase of this research. A meta-analysis of the socialization tactics research also found serial tactics to be significantly correlated with social acceptance (Bauer et al., 2007). The following hypothesis is therefore proposed:

Research Hypothesis 9ab (H9ab): Mentoring will be positively associated with social integration.

A number of research projects have criticized the tendency of the traditional socialization literature to portray newcomers as passive actors of the socialization process, and have focused on the proactive role played by newcomers in gaining information related to learning and adjustment within organizations (Gruman et al., 2006; Kim et al., 2005; Morrison, 1993a, 1993b). In their work on feedback-seeking behaviour, Ashford and Cummings (1983) identified inquiry as one of the primary modes of information-seeking behaviour (the other mode being monitoring). Inquiry entails information-seeking behaviours in which newcomers directly ask other people for information. Newcomers may solicit information from various types of individuals such as direct or indirect supervisors, experienced co-workers, other newcomers, or subordinates (C. D. Fisher, 1986; Miller & Jablin, 1991; Morrison, 1993b).

Some empirical evidence was found in the literature to suggest a relationship between the extent to which a newcomer interacts with individuals and the person's perception towards his or her integration within a group or an organization. For instance, a study investigated the influence of socialization tactics on newcomer turnover by embedding newcomers into the organization (D. G. Allen, 2006). Studying a large financial services organization, the author found that collective tactics were positively related to on-the-job embeddedness. Morrison (1993b) showed that information seeking and feedback seeking were related to social integration in a sample of newly recruited staff accountants. In contrast, conflicting results were provided by Wanberg and Kammeyer-Mueller (2000) who concluded that neither of the sensemaking behaviours (information seeking and feedback seeking) was related to social integration.

The delineation between the inquiry mode of information-seeking versus monitoring (Ashford & Cummings, 1983) is particularly important in the FLOSS context. In FLOSS communities, newcomers have various communication means to interact with community members in order to ask for socialization-related information, such as general or newcomer-dedicated IRC channels, mailing lists, forums, or even direct emails. However, rather than directly interacting with other project contributors, newcomers may prefer to observe current and past member interactions, since mailing lists, forums, or IRC channels are normally open to anyone. In addition, past messages posted in mailing lists can be consulted as they are usually freely available on the project platform, as well as the past logs of the various IRC channels. The hypothesis below is derived:

Research Hypothesis 10ab (H_{10ab}): Interaction intensity will be positively associated with social integration.

Employees who perceive the support from their organization are more likely to feel as valued and trusted organizational members (Stamper & Masterson, 2002). Meanwhile, perceived organizational support was found to increase communication and cooperation among employees (Erdogan, Kraimer & Liden, 2004). The socialization literature provided some empirical evidence as investiture tactics were found to have a positive effect on on-the-job embeddedness (D. G. Allen, 2006) and social acceptance (Bauer et al., 2007) suggesting that the positive support provided by an organization may play a role in increasing the integration of a newcomer.

However, both in the organizational behaviour and FLOSS literatures, the evidence concerning the influence of organizational/community support on social integration, whether it is within a socialization context or not, is limited. For instance, a study investigated the relationship between perceived organizational support on social integration from a sample of 167 newcomers in a leading insurance organization in China, and failed to find a statistically significant relationship (J. Chen & Eldridge, 2011). In the context of the Wikipedia encyclopedia, Carillo and Okoli (2011) investigated, among other group factors, the effect of organizational support on member support (conceptualized as the activities related to the ways individuals are embedded within a group through relations among group members). The authors analyzed data drawn from a sample of 10,000 encyclopedia articles and did not find any significant relationship. This research aims to provide additional insights to clarify the above mentioned conflicting results by studying the following hypothesis:

Research Hypothesis 11ab (H_{11ab}): Supportiveness will be positively associated with social integration.

Actions in organizations and other social entities have often a dual significance (Peter, 1981). First, actions have a tangible characteristic that pertains to the instrumental nature of actions: to gain profits, to get a promotion, or reach other calculated goals (Islam & Zyphur, 2009; Trice, Belasco & Alutto, 1969). Second, actions often have a symbolic nature that transcribe the culture of a social entity and that has an impact on individuals' beliefs, emotions, or identity (Islam & Zyphur, 2009). The symbolic character of actions may also play a role in helping individuals become integrated into larger social entities (Trice et al., 1969).

Trice and Beyer (1984) integrated early works about organizational culture and developed a typology of rites and ceremonials. Among the six categories of rite, *rites of integration* aims at binding members together into a larger social group by lessening the social distance among individuals. The authors provided examples such as office Christmas parties, which bring together employees from all levels of the hierarchy, and events mixing students and faculty members.

Drawing from earlier insights presented in this chapter, the reliance of formal joining processes in FLOSS communities can be seen as instances of rites within a social entity. For instance, sponsorship mechanisms are rather common practices in FLOSS projects. One of the steps of the Debian New Member Process⁴ is that new contributors who want to upload a package onto the Debian repository must find a sponsor (an official Debian maintainer) to upload their packages for them. The sponsoring process is described on the project platform⁵ and indicates three possible ways to be sponsored: *join a packaging team*, *ask the debian-mentors mailing list*, or *talk directly to people willing to sponsor your package*. It can be argued that such practice is a way to bridge the gap between community newcomers and experienced members, and constitute specific instances of rites of integration. The sponsorship mechanism implemented in Debian can be seen as a means of pairing a newcomer with one or several experienced member(s) who both have matching areas of interest and expertise, as is indicated in the sponsoring process page:

⁴ <http://www.debian.org/devel/join/newmaint>

⁵ <http://mentors.debian.net/sponsors>

Contact individuals only if your package is compatible to their respective requirements and matches their area of interest. To tell apart sponsors who are interested in your package from those who are not, we asked developers to formulate their own sponsor traits.

As a result, this suggests the hypothesis that one of the goals of formal joining processes in FLOSS communities is to facilitate the integration of recruits by connecting them to other community members. One's sense of being socially integrated in a FLOSS community should be affected when taking part in a joining process that involves the interaction with other community members (and especially experienced ones). On the other hand, when no such process is required, a newcomer may not interact with any members and may then feel more isolated from the rest of the community. Based on this rationale, this research poses the hypothesis:

Research Hypothesis 12ab (H_{12ab}): Joining structuredness will be positively associated with social integration.

5.5 CHAPTER SUMMARY

This chapter introduces some of the conceptual definitions used when deriving the conceptual models and justifies the use of two conceptual models. Then, the chapter describes the development of the research models informed by the integration of the literature review and the results of the qualitative phase of this research. The derived research hypotheses that were used to validate this research are also presented.

METHODOLOGY

This research project relied on a three-phase mixed-methods methodology. The first phase involved the collection of interview data to explore the newcomer experience and citizenship phenomena, which led to the design of two alternative conceptual models in line with the findings from the review of the literature. The second phase consisted of the design and validation of a questionnaire instrument to test the models. The final phase included the collection of data from FLOSS community members to quantitatively test the research models (see Figure 20 for an overview of the implemented methodology). An online survey was then conducted.

Figure 20: Methodology overview

6.1 METHODOLOGICAL CHOICES

6.1.1 Survey method

The choice of a survey method to test the conceptual models designed during the first phase of the methodology can be justified using Pinsonneault and Kraemer (1993)'s criteria about the appropriateness of survey research when investigating a research

problem:

The central questions of interest about the phenomena are "what is happening?", and "how and why is it happening?": This research project investigates the socialization factors that have an impact on the behaviour of contributors. This study is then trying to understand the *hows* and *whys* of successful FLOSS newcomer socialization from a FLOSS community perspective.

The phenomena of interest must be studied in its natural setting: FLOSS communities are large complex entities which cannot be reproduced within an artificial setting (e.g., in a laboratory). The literature review chapter has provided an overview of various internal mechanisms that govern FLOSS communities. Each community is for instance characterized by a set of shared norms and beliefs that have matured throughout the history of the community and that influence members' actions and behaviours. The temporal evolution of a FLOSS community is a unique feature that cannot be simulated in an artificial research setting such as a laboratory.

Control of the independent and dependent variables is not possible or not desirable: FLOSS communities rely on a wide array of options and initiatives to facilitate the integration of new recruits. In addition, newcomers in FLOSS communities have the freedom to shape their own experiences by utilizing certain of the provided resources, and by deciding upon the amount and type of interactions with other community members with which they will involve themselves. Each individual's socialization experience is then a unique combination of various technical and social interactions that renders the control of socialization variables simply impossible. Forcing control on such aspects would denature the socialization phenomenon from its inherent complexity, thus compromising the validity of this study.

The phenomena of interest occur in current time or the recent past: This research aims to better understand the current performance of individuals within a FLOSS community by examining the experiences of these individuals when they were community newcomers. The target population of this study is community members who joined FLOSS communities relatively recently. As a result, the socialization factors that are captured in this study pertain to facts that occurred in the recent past.

In conclusion, there was little doubt about the choice of the most appropriate method to test the socialization models introduced in this project. In addition to the very context-specific nature of one's interactions within a FLOSS community, the temporal nature of the socialization phenomenon justified the appropriateness of using survey research to investigate the research questions and test the conceptual models.

6.1.2 *Online survey versus regular survey*

In this research, a web-based survey approach was chosen to collect data. Over the last 30 years, technology has transformed how survey research is conducted, with the advent of the first e-mail surveys in the 1980s and the first web-based surveys in the 1990s (Evans & Mathur, 2005).

Web-surveys (or e-surveys) have been shown to offer unique advantages such as dramatically reducing the cost of conducting surveys, the rapid collection of survey data, and the possibility of transmitting a rich array of cues such as the use of sound and graphics (Simsek & Veiga, 2001). However, web-surveys have specific challenges such as a high risk of low response rate and non-response bias, as well as the difficulty of controlling the quality of the sampling frame (Dillman, 2009).

In this research, the Internet medium was seen as the most appropriate channel for conducting the survey for two essential reasons. First, web surveys are seen as the most effective way to collect data when the population of interest is largely made up of members of an online community (Parasuraman & Zinkhan, 2002). FLOSS communities are in essence social entities comprised of geographically distributed individuals that interact and contribute via the Internet (Subramanyam & Xia, 2008). Second, the nature of this research project and in particular the complexity of the designed research models (see Chapter 4) require a large sample in order to be able to successfully examine all the hypothesized relationships. The Internet is an ideal channel to use when conducting surveys requiring a large number of respondents (Evans & Mathur, 2005).

6.2 1ST PHASE

Informed by the findings of the literature review earlier performed in this research, the first phase involved the collection of qualitative data to investigate the socialization phenomenon in the FLOSS community context as well as the notion of FLOSS community citizenship behaviour. The results were then used to answer two of the subsidiary research questions: identify the key factors that characterize the socialization experience of FLOSS newcomers, and identify the important types of FLOSS community citizenship behaviours. As a result, interviews were conducted with FLOSS practitioners covering various roles and from different FLOSS communities. The interview data collection, analysis procedure, and findings are presented in the next chapter.

The findings from the analysis of the interview data were then used along with the findings from the literature review to derive two alternative conceptual models and their associated research hypotheses. Both models were tested and their explanatory power compared so as to identify the most appropriate model. Further justification concerning the development of the two conceptual models is provided in Chapter 5.

6.3 2ND PHASE

The primary goal of the second phase was to design and validate a survey questionnaire instrument to test the two theoretical models. Measurement items for the socialization constructs were derived from the interview findings, while the items used to measure the other constructs were identified during the review of the literature and later tailored and extended to the FLOSS context.

Survey research is widespread in IS research and is the most common method of collecting primary quantitative data in the field (Pinsonneault & Kraemer, 1993). This method has the benefit of being able to generate valid and easily interpretable data

Figure 21: Instrument development process

based on the use of a formalized set of questions that provides information about respondents' habits, perceptions, and experiences (Malhotra, 2004). On the other hand, the inadequacy of construct measurement in survey research has been a recurring issue in fields such as organizational behaviour (Hinkin, 1998), marketing (Churchill, 1979), and information systems (Gefen, Straub & Boudreau, 2000; Moore & Benbasat, 1991).

As Malhotra (2004) highlights, there is no scientific principle that guarantees the design of an optimal or ideal questionnaire: "Questionnaire design is as much an art as it is a science" (p. 176). However, authors have provided an array of guidelines to assist social scientists in the process of questionnaire instrument development to help them avoid common errors (e.g. Churchill, 1979; Gefen et al., 2000; Hinkin, 1998; Moore & Benbasat, 1991; Straub, Boudreau & Gefen, 2004). For instance, Moore and Benbasat (1991) point out that the proper operationalization of constructs through measures with high degrees of validity and reliability is a sine qua non condition for cumulatively contributing to the IS body of knowledge. Content validity, construct validity, and also reliability have been cited as among the most important evaluation criteria when developing survey instruments (Gefen et al., 2000). A compilation of the guidelines and advices provided by the researchers mentioned above helped in elaborating the instrument development process that was used in this research. Figure 21 describes the sequence of steps that was followed.

It has been suggested that the first step in the instrument development process should be the delimitation of the domain of each construct, as this allows researchers to specify content boundaries for each construct, and allows social scientists to know with precision what is included and what is excluded from each construct domain (Churchill, 1979; Hinkin, 1998; Moore & Benbasat, 1991). In this research, conceptual

construct definitions were generated from the findings derived from the literature review and the qualitative phase of this research. The next step of the instrument development process involved the creation of a pool of measurement items through the collection of items from the extant literature and the generation of newly-designed items from the results of the qualitative data analysis phase.

6.3.1 *Item generation*

The literature was reviewed to search for existing measurement items and scales that had been previously validated in past research works (Moore & Benbasat, 1991). Moore and Benbasat suggest that a researcher who is investigating the potential use of a measurement scale should then consider its reliability reported by studies that used such measure and should also reflect on the applicability of the scale to the current research context. In case no scale satisfying both criteria is available in the academic literature, newly-designed items relevant to the context of the research must then be generated by the researcher.

When generating new measurement items in this research, Hinkin's (1998) recommendations were followed. Hinkin (1998) suggests that statements shall be as simple and as short as possible, the chosen words and expressions shall be familiar to the respondents, and items should address a single issue (i.e., not be double-barreled). Additionally, the author indicates two possible approaches in creating items: the deductive approach in which items are derived from the construct domain, and the inductive approach where items are generated by respondents and then categorized into constructs based on the analysis of the researcher. In addition, special attention should be provided in generating sufficient items that ensured the coverage of the domain of each construct (Straub et al., 2004).

6.3.1.1 *Measurement of the socialization constructs*

This research investigated the FLOSS newcomer socialization phenomenon in a prior qualitative data collection and analysis phase which allowed the identification of key factors. The literature review and the conceptual model development chapters justified the applicability of Van Maanen and Schein (1979)'s approach in conceptualizing socialization by highlighting that large FLOSS communities share, to some extent, some structural and managerial similarities with organizations.

Chapter 4 introduced the six socialization variables which were identified through the analysis of the collected qualitative data. As explained in the previous chapter, some of the socialization variables manifest some overlap with some of the original socialization constructs from Van Maanen and Schein (1979). Consequently, the measurement scales that were developed to capture Van Maanen and Schein (1979)'s constructs, such as the work of Jones (1986), were used to the extent possible when generating items for certain of the socialization variables (see Appendix C for Jones' measurement scales). For instance, formal tactics deal with the segregation of new employees who go through experiences tailored for newcomers. Task segregation shares some similarities with formal tactics since, even though the physical segregation (even virtually) is not relevant in FLOSS communities, newcomers may face some degree of segregation in relation to the tasks they may perform, thus engendering experiences

designed specifically for newcomers.

The use of items that belong to existing and already validated scales lowers the potential for measurement errors (Churchill, 1979). Jones' measurement scales were examined for each of the socialization constructs in search for items that address overlapping aspects of the socialization variables of this research. The selected items were used as a starting point to generate the FLOSS-tailored final items that were added to the initial pool of items. For instance, Jones used the item "I have been through a set of training experiences which are specifically designed to give newcomers a thorough knowledge of job related skills" to reflect formal tactics. This captures an aspect of task segregation, so the item "I have undertaken activities which are intended to turn new community members into project contributors" was added to the initial pool of items. The same operation was repeated for all socialization constructs and items from Van Maanen and Schein (1979) and Jones (1986). The degree of overlap between Van Maanen and Schein (1979)'s socialization constructs and those derived in this research varies greatly from construct to construct. As explained in the previous chapter, fixed tactics for instance, is foreign in a FLOSS context since the notion of time constraint or deadline is mostly irrelevant. In such a case, the measurement scale of fixed tactics was ignored.

Moreover, the researcher generated new items based on the domain and conceptual definition of each construct. Additionally, an IS senior lecturer, expert in online communities, was consulted to help derive additional items during a 2-hour session. A deductive approach (Hinkin, 1998) towards scale development (as opposed to an inductive approach) was adopted to create the items. As Hinkin (1998) indicates, the development of a scale using the deductive approach assumes that the theoretical foundation of a construct is sufficient for the researcher to generate items. The construct definition is then used as a guide for generating items. After consultation with the supervising staff and several rounds of revision, it was decided to initially develop 8 measurement items for each socialization construct. Due to the novelty and lack of validation of the socialization construct measures used in this research, it was anticipated that there was a chance that a number of items would exhibit weak psychometric properties during the expert panel, Q-sorting, pretest, and pilot test phases. The researcher targeted an ideal number of 4 to 5 items remaining in the final version of the survey instrument. The initial pool for all socialization constructs contained 48 items and is presented in Tables 13, 14, and 15.

6.3.1.2 *Measurement of task performance*

Task performance has been measured in FLOSS research using various perspectives. A review of the studies that captured performance in FLOSS projects identified the following approaches. Some authors generated task completion scores based on the computation of the percentage of completed tasks (Chou & He, 2010; Daniel, Agarwal & Stewart, 2006); others derived project activity and development status scores (Hahn et al., 2008) while others captured rank promotion, defines as the progression through a community's hierarchy based on performance rankings (Roberts et al., 2006). All the reviewed measures were judged to be inadequate in the context of this study as they are all software project-centric and they omit non-technical participation which is essential in FLOSS communities. In other words, the scope of the measures

Table 13: Pool of items before expert panel and card sorting procedure -1

Task Segregation	
<i>the degree to which a newcomer has performed tasks that are specifically tailored for newcomers.</i>	
SE1	I have performed tasks that were specifically designed for newcomers.
SE2	I have undertaken activities which are intended to turn new community members into project contributors.
SE3	Much of my project-related knowledge has been acquired while working on general project assignments (R).
SE4	I have been involved in tasks which I was clearly aware were appropriate for newcomers.
SE5	I have carried out assignments designed by the community to help newcomers become project contributors.
SE6	I have been through a set of tasks that that I was advised to do as a newcomer
SE7	I have learned how to become a project contributor by carrying out assignments that any community member could do (R).
SE8	My contributions to this community have included tasks that were particularly suitable for newcomers.
Interaction Intensity	
<i>the degree to which a newcomer is actively involved with other newcomers and community members while learning how to become a project contributor.</i>	
IN1	I have worked with other newcomers and experienced members whose role included helping community newcomers.
IN2	I have been active in the community communication channels such as forums, mailing lists, or IRC channels.
IN3	Most of my learning has been carried out on my own, with no involvement from other newcomers and community members (R).
IN4	I have interacted with many other community members to gain project-related knowledge.
IN5	I have learned how to contribute to the project all by myself (R).
IN6	I have felt part of a group during my experience as a newcomer in this community.
IN7	Working with other people as a newcomer has contributed to my learning.
IN8	The way I have learned to become a contributor in this project has been highly collaborative.

Table 14: Pool of items before expert panel and card sorting procedure -2

Task Purposefulness	
<i>the degree to which the sequence of tasks performed by a newcomer follows a logical pattern designed to help the newcomer to become a contributor.</i>	
TP ₁	The tasks that I have performed to gain an overall understanding of the project follow a logical pattern.
TP ₂	The tasks I have carried out have been specifically designed to help me understand how to contribute to the project.
TP ₃	Each task that I have performed has expanded and built upon the knowledge gained from the preceding tasks.
TP ₄	To become a contributor, I have undertaken assignments with gradually increasing complexity.
TP ₅	To become a contributor, I have performed tasks whose scope has gradually increased.
TP ₆	The tasks I have undertaken have helped me learn different aspects of the project.
TP ₇	There is cohesion in the sequence of tasks I have carried out while learning how to become a contributor.
TP ₈	The tasks I have performed to gain an understanding of the project have not been logically connected (R).
Joining Structuredness	
<i>the degree to which a newcomer is actively involved with other newcomers and community members while learning how to become a project contributor.</i>	
JS ₁	There was a clear set of steps I needed to follow in order to join this community.
JS ₂	This community does not require newcomers to follow a specific joining process to become a project contributor (R).
JS ₃	I had little information about the procedure that I had to follow to become an official project contributor (R).
JS ₄	The community has clearly communicated the course of action I had to pursue to become a recognised project member.
JS ₅	There was a formal course of action that I had to follow to join this community.
JS ₆	In this community, the steps one has to go through to become a project contributor are well documented.
JS ₇	The process of joining this community has been clearly structured in a way that I could understand.
JS ₈	The process of joining this community does not follow a defined pathway (R).

Table 15: Pool of items before expert panel and card sorting procedure -3

Mentoring	
<i>the degree to which a newcomer has been taken under the wing of one or more experienced members while learning how to become a project contributor.</i>	
ME1	One experienced member [or more] provided me with support and feedback when I was new to this community.
ME2	One experienced member [or more] took me under his/her wing while I was learning how to contribute to this project.
ME3	I became a project contributor thanks to the help of one [or more] experienced member [or more] who has been like a mentor to me.
ME4	One experienced contributor [or more] has been instrumental in helping me learn how to become a contributor to this project.
ME5	One or more experienced member(s) has always been there when I needed help.
ME6	I could not have become a contributor to this project without the guidance of one particular person [or more] in this community.
ME7	I do not feel any individual member supported me more than others in becoming a contributor to this project. (R)
ME8	I have been mentored by one experienced contributor [or more].
Supportiveness	
<i>the degree to which a newcomer has perceived a community to be supportive while learning how to become a project contributor.</i>	
SU1	Most community members that I have been in contact with have gone out of their way to help me become a contributor.
SU2	I have been made to feel that my contributions are valuable in this project.
SU3	This community is very supportive towards its newcomers.
SU4	The community has been there for me when I needed help.
SU5	This community cares a lot about its newcomers.
SU6	In this community, newcomers feel that they are part of the family.
SU7	I have felt valued by the community even though I was a newcomer.
SU8	I have felt cared for by the wider community while I was learning to become a project contributor.

was narrower (technical contributions to a software project) than the one adopted in this research (contributions to the overall community that encompasses the software project).

A review of the performance measures used in the socialization and overall organizational behaviour literatures identified several approaches that can be categorized into two broad types of measures. Some researchers used performance measures that are closely tailored to the context under investigation. For example, MacKenzie et al. (1991) used company records to compute a measure of salespersons' performance; Motowidlo and Van Scotter (1994) measured task performance using U.S. Air Force mechanic's supervisor rating of critical incident scores specifically developed for vehicle mechanics; Geller and Bamberger (2012) used archival data to compute the average number of calls handled per hour in the past month when studying customer contact employees. Such approaches provide accurate performance measures but their potential to be reused by other researchers is limited to contexts that are similar to those in which the measures were developed. Such measures were then not applicable to the FLOSS context.

The second type of measures adopts an overall organizational perspective when assessing employee task performance. This research has been influenced by the early works from Katz (1964) who distinguishes in-role from extra-role behaviours. With this in mind, various measures of in-role performance were derived and validated. In a widely cited study, Williams and Anderson (1991) measured in-role behaviour in terms of employees' adherence to and completion of formal job duties (Katz, 1964). They derived some of their measurement items from C. O'Reilly and Chatman (1986) (see Table 16) and reported an initial Cronbach's alpha coefficient of 0.91. Responses were given on a 5-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Table 16: Task performance measurement items

MEASUREMENT ITEMS FROM WILLIAMS AND ANDERSON (1991)	
1	Adequately completes assigned duties.
2	Fulfills responsibilities specified in job description.
3	Meets formal performance requirements of the job.
4	Performs tasks that are expected of him/her.
5	Neglect aspects of the job he/she is obligated to perform. (R)
6	Fails to perform essential duties.
7	Engages in activities that will directly affect his/her performance evaluation.
INITIAL MEASUREMENT ITEMS USED IN THIS STUDY	
PER1	I adequately complete my tasks and responsibilities.
PER2	I effectively fulfill my roles and responsibilities in this project.
PER3	My contributions meet or exceed the performance expectations from the community.
PER4	My work in this project meets the quality that the community expects from its members.
PER5	I carry out assignments to meet the standard that is expected by the community.
PER6	I get the job done when I work on a project assignment.
PER7	I consider myself to be a dependable contributor to this community.

The task performance measurement scale of Williams and Anderson (1991) demonstrated satisfactory reliabilities in a variety of organizational and professional contexts. In a study that explored the link between OCB and supervisor performance judgments, T. D. Allen and Rush (1998) reported a reliability coefficient $\alpha = 0.95$. Sparrowe and Liden (2001) investigated the influence of social networks on individual performance and used Williams and Anderson's measure of in-role performance (all 7 items), reporting a Cronbach's α coefficient of 0.94. A study by Turnley, Bolino, Lester and Bloodgood (2003) explored the relationship between psychological contract fulfilment and in-role performance using six of the seven items from Williams and Anderson ($\alpha = 0.93$) having omitted the item: "Engages in activities that will directly affect his/her performance evaluation". The measure was also used when studying the impact of job embeddedness on individual performance (T. Lee, Mitchell & Sablinski, 2004). The authors selected six of the seven items from the initial measure and added an extra item: "performs all tasks that are expected of him or her" ($\alpha = 0.92$). Another study developed and validated a multidimensional scale of supervisor leader-member exchange and investigated the influence of each of the dimensions on in-role behaviour (Greguras & Ford, 2006). The reliability coefficient for this study was $\alpha = 0.68$. Finally, Lynch, Eisenberger and Armeli (1999) investigated the moderating effect of perceived organizational support on the relationship of employees' fear of exploitation in exchange relationships and their in-role performance. The authors used the seven items from Williams and Anderson and added two items from Smith et al. (1983) - "This employee exhibits punctuality in arriving at his/her work station on time after breaks" and "This employee spends time in idle conversation (R)". The authors performed

two rounds of data collection and reported Cronbach's α of 0.87 and 0.88.

Williams and Anderson's measurement items were carefully analyzed to determine the extent to which each of them applied to the FLOSS community context. The first three items (see Table 16) were found to be relevant for this study. The interview data addressed the notion of task performance with terms that were semantically consistent with the 3 items. The item "Performs tasks that are expected of him/her" was reworded for this study, as within a FLOSS community tasks are mostly self-assigned. There is thus no such notion of members being expected, a priori, to perform particular tasks. Similarly, the item "Neglect aspects of the job he/she is obligated to perform" was dropped as for the same reason, in a FLOSS community there is no sense of a priori obligation to undertake particular project tasks. The same justification applied when dropping the item "Fails to perform essential duties" as again, the sense of duty is mostly irrelevant in a FLOSS community. Finally, "Engages in activities that will directly affect his/her performance evaluation" was discarded for the reason that FLOSS communities differ from traditional organizations as performance is not evaluated and rewarded in the same way and also not for the same purpose. Items PER4 and PER5 were derived from Williams and Anderson's conceptual definition and items for in-role performance. Finally, items PER6 and PER7 were created using an inductive approach by carefully considering the domain covered by the construct definition used in this project and insights obtained from the interview respondents. For instance, 'getting the job done' was a recurring expression used by respondents when addressing task performance in FLOSS projects.

6.3.1.3 *Measure of community citizenship behaviours*

Community citizenship behaviours address a type of performance that complements task performance. Whereas task performance concerns the fulfilment of project roles and direct contributions, CCB contributes to "the maintenance and enhancement of the social and psychological context that supports task performance" (Organ, 1997, p. 91). The qualitative phase of this research confirmed the relevance and appropriateness of the overall concept of citizenship behaviours in the context of FLOSS project participation. In addition, the first phase of this research project confirmed the relevance of the notions of CCB-I (or community citizenship behaviours directed towards project members) and CCB-P (or community citizenship behaviours directed towards the project and its community as a whole). While some of the traditional OCB dimensions were judged to be relevant to a FLOSS context and integrated into the item pool, the results of the first, qualitative stage of this research identified additional behaviours unique to the FLOSS context.

The generation of the measurement items was performed using a two-phase procedure. First, items from existing scales for each dimension were carefully analyzed in line with the findings of the qualitative phase of the project. Items were added to the item pool when they satisfied two conditions: they had to have demonstrated satisfactory validity and reliability in past studies, and they had to successfully apply to the particularities of the FLOSS context. The previous chapter introduced, for each of the identified CCB-I/CCB-P behaviour, the related OCB works and concepts that overlap with the various instances of CCB. The cited references were reviewed when analyzing the various existing measurement scales and items. Second, new items were

created for the CCB types that were unique to the FLOSS community context.

For instance, the analysis of the interview data revealed the importance of redirecting members who seek help to individuals who are in a position to help. The review of the existing items that were used in the literature to capture notions to interpersonal helping such as altruism (Organ, 1988, 1990a; Smith et al., 1983), interpersonal helping (Graham, 1991; Moorman & Blakely, 1995), or helping coworkers (George & Brief, 1992; George & Jones, 1997), did not identify any items that could be used. It was then decided to add the item “I redirect people who have a problem for which I do not know the answer to members who are in a position to help” to the initial pool of CCB-I items. Similarly, the measurement items from the helping coworkers (George & Brief, 1992; George & Jones, 1997) and social participation (Van Dyne et al., 1994) constructs were reviewed in search for items that would appropriately capture the idea of maintaining a positive attitude in the FLOSS context. The item “I maintain a positive attitude when interacting with other members” was created as no reviewed item was found to be satisfactory. The items that were added to the initial pool for both CCB-I and CCB-P are listed in Tables 17 and 18.

6.3.1.4 *Measure of social identification*

The notion of identification has been thoroughly investigated in a variety of organizational contexts (Ashforth & Saks, 1996; Bergami & Bagozzi, 2000; Mael & Ashforth, 1992) but also in FLOSS communities (Bagozzi & Dholakia, 2006; Gwebu & Wang, 2011; Stewart & Gosain, 2006). The extant literature was then reviewed to identify the measurement scales that would be appropriate for this study. Mael and Ashforth (1992)’s was found to be a satisfactory measure as it consistently demonstrated good reliability scores and a broad domain of applicability. Following Moore and Benbasat’s (1991) recommendations, the reliability of the scale was analyzed and the items were revised in an attempt to tailor them to the context of this research (see measurement scale in Table 19).

Mael and Ashforth (1992) developed a six-item measure of organizational identification which was validated in the context of alumni in an American educational institution. The authors reported an alpha coefficient of 0.87 (see measurement scale in Table 19). The authors indicated the applicability of the designed scale in organizational contexts. Their measure was reused in a number of studies and demonstrated strong psychometric properties.

Ashforth and Saks (1996) investigated the influence of socialization tactics, using van Maanen and Schein’s (1979) typology, on newcomer adjustment and collected data from graduates from a Canadian university. They reported an alpha coefficient of 0.80. Ashforth et al. (2007) studied how socialization processes (socialization tactics and proactive behavior) jointly affect socialization content and adjustment. The authors used Mael and Ashforth (1992)’s scale and found a Cronbach’s alpha coefficient of 0.85. Knippenberg and Schie (2000) investigated the importance of work-group identification relative to organizational identification using two separate samples. The study reported alpha coefficients of 0.82 and 0.83. Organizational identification was studied in the context of virtual workers by Wiesenfeld, Raghuram and Garud (2001). The study indicated an alpha coefficient of 0.86. Finally, Kreiner and Ashforth (2004)

Table 17: Pool of CCB-I items before expert panel and card sorting procedure -1

Community citizenship behaviour - Individuals			
<i>Community citizenship behaviours directed toward the benefit of other community members</i>			
ITEM	STATEMENT	CCB BEHAVIOUR	ORIGIN
CCB-I1	I lend a hand to members who express a need for help in project areas for which I am knowledgeable.	<i>assisting other project members</i>	Podsakoff et al. (1990)
CCB-I2	I go out of my way to help other community members with project-related problems.		Moorman and Blakely (1995)
CCB-I3	I help community members who seek support from other members.		Moorman and Blakely (1995)
CCB-I4	I redirect people who have a problem for which I do not know the answer to members who are in a position to help.		NEW (interviews)
CCB-I 5	I willingly assist newcomers in becoming familiar with the community.		Fahr et al. (1997)
CCB-I6	When conflict occurs between other members of this community, I try to act as a peacekeeper.	<i>facilitating interpersonal conflicts</i>	Podsakoff and Mackenzie (1994)
CCB-I7	I pay attention to how my actions and contributions will affect other members.	<i>being considerate</i>	Podsakoff et al. (1990)
CCB-I8	I consult people who might be affected by my actions or decisions.		Konovsky and Organ (1996)
CCB-I9	I try to avoid creating problems for other members when contributing to the project.		Podsakoff et al. (1990)
CCB-I10	I treat others with respect.	<i>treating others with respect</i>	NEW (interviews)
CCB-I11	I maintain a positive attitude when interacting with other members.	<i>maintaining a positive attitude</i>	NEW (interviews)

Table 18: Pool of CCB-P items before expert panel and card sorting procedure

Community citizenship behaviour - Project			
<i>Community citizenship behaviours directed toward the benefit of the project and its community.</i>			
ITEM	STATEMENT	CCB BEHAVIOUR	ORIGIN
CCB-P ₁	I think beyond myself when contributing to the project.	<i>Contributing with the best interest for the community</i>	NEW (interviews)
CCB-P ₂	I try to align my contributions with the shared goals of the project.		NEW (interviews)
CCB-P ₃	I comply with the behavioural expectations of this community.	<i>Complying to the community behavioural rules</i>	NEW (interviews)
CCB-P ₄	I behave in a way that is aligned with the shared values of this community.		NEW (interviews)
CCB-P ₅	I respect the behavioural rules of the community.		NEW (interviews)
CCB-P ₆	I keep the community informed about my work and its progress.	<i>Keeping the community informed about one's work and progress</i>	NEW (interviews)
CCB-P ₇	I let the community know if I am not able to complete a task on which I am working.		NEW (interviews)
CCB-P ₈	I voluntarily perform maintenance tasks to keep project resources to a good standard.	<i>Keeping project resources to a good standard</i>	Graham (1991) Organ (1990)
CCB-P ₉	I keep up with the latest news about the overall project as well as in areas in which I am involved.	<i>Exercising civic responsibility</i>	Netemeyer et al. (1997), Podsakoff et al. (1990)
CCB-P ₁₀	I attend community events such as conferences, hackfests, or sprints.		Konovsky and Organ (1996) Podsakoff et al. (1990)
CCB-P ₁₁	I express my opinion in a way that is constructive.	<i>Responsible involvement in project decisions</i>	NEW (interviews)
CCB-P ₁₂	I take part in project decisions in an effective way.		NEW (interviews)
CCB-P ₁₃	Once a decision has been made for which I either agree or disagree, I accept it and act accordingly.		NEW (interviews)
CCB-P ₁₄	I make an effort to represent the project to potential new members and community outsiders.	<i>Promoting the project to potential contributors and to the outside world</i>	Moorman and Blakely (1995), Koh (2004)
CCB-P ₁₅	I spend time talking about the project to my friends, colleagues, and family.		Koh (2004)
CCB-P ₁₆	I actively promote the image of the project's community.		Moorman and Blakely (1995)
CCB-P ₁₇	I encourage friends, colleagues, and family to use the software produced by the community.		Moorman and Blakely (1995)

Table 19: Organizational identification scale

Mael and Ashforth (1992)'s measure of Organizational identification [1 = Strongly agree; 5 = Strongly disagree]
When someone criticizes (name of school), it feels like a personal insult.
I am very interested in what others think about (name of school).
When I talk about this school, I usually say "we rather than 'they'".
This school's successes are my successes.
When someone praises this school, it feels like a personal compliment.
If a story in the media criticized the school, I would feel embarrassed.
Measure of Social identification used in this research
IDE1: When someone criticizes the community, it feels like a personal insult.
IDE2: I am very interested in what outsiders think about the community.
IDE3: When I talk about this community, I usually say 'we' rather than 'they'.
IDE4: This community's successes are my successes.
IDE5: When someone praises this community, it feels like a personal compliment.
IDE6: If a story in the media criticized the community, I would feel embarrassed.

introduced and tested an expanded model of identification and found an alpha coefficient of 0.86 when measuring *organizational identification* using Mael and Ashforth (1992)'s scale.

6.3.1.5 Measure of social integration

Both the review of the literature and the conceptual model development chapters emphasized the relevance of the notion of social integration in the overall FLOSS context and more particularly to FLOSS socialization.

In addition, there has been considerable evidence that suggests that social integration is a critical indicator of successful adjustment (Bauer et al., 2007; Kammeyer-Mueller & Wanberg, 2003; Morrison, 1993a; Wanberg & Kammeyer-Mueller, 2000). In this research, the definition of *social integration* was adapted from Feldman (1981): *the extent to which an individual perceives himself/herself to be trusted and accepted by the other FLOSS community contributors*.

When measuring *social integration*, the OB literature has mainly relied on the works from Morrison (1993a), Wanberg and Kammeyer-Mueller (2000), and Kammeyer-Mueller and Wanberg (2003). In this research, *social integration* was measured through an adaptation of the four-item scale from Wanberg and Kammeyer-Mueller (2000) which items were derived from Morrison (1993a) and Price and Mueller (1986). In their study, the authors reported a satisfactory reliability score (Cronbach's $\alpha = 0.84$). The initial and the modified items can be found in Table 20. This measure is an adaptation of Morri-

Table 20: Social integration scale

**Wanberg and Kammeyer-Mueller (2000)'s measure of Social Integration
[7-point Likert scale]**

The people I work with respect me
 My coworkers seem to accept me as one of them
 I get along with the people I work with very well
 I feel comfortable around my coworkers

Measure of Social integration used in this research

INT1: In this community, members seem to accept me as one of them.
 INT2: I get along with the project members I am in touch with very well.
 INT3: I feel comfortable when interacting with other community members.
 INT4: The community members I interact with respect me.

son's (1993) social integration scale and a careful analysis of each of the items allowed to conclude that the scale was the most appropriate in the FLOSS context.

6.3.2 Expert panel

The design and validation of content-valid measurements is a difficult task as researchers need to generate measurement items from a content universe that is theoretically infinite (Straub, 1989; Straub et al., 2004). Drawing insights from Cronbach (1971), Straub (1989) encourages researchers to conduct several rounds of instrument pretesting with groups of expert judges in order to establish the content validity of the measurement scales. Content validity refers to the degree to which measurement items reflect the content universe to which the findings will be generalized (Cronbach, 1971). Boudreau, Gefen and Straub (2001) reviewed the IS literature between 1997 and 2000 to investigate how well instruments had been validated in the IS field. The authors concluded that for studies that used newly designed instruments, only about 25% of the reviewed articles had assessed content validity (whereas 20% of the studies that relied on existing instruments tested content validity).

A panel of four experts was utilized in this research to assist in the questionnaire development. The expert panel was composed of an IS senior lecturer, an IS professor, an IS associate professor specialized in FLOSS research, and a PhD candidate with expertise in FLOSS research and several years of contributions to FLOSS projects. The purpose of the procedure was to verify whether the constructs were clearly defined and that each measurement scale captured the entire domain of its associated construct.

Each expert received an email containing detailed instructions and a set of tables containing the construct names, definitions, and measurement items. Feedback was discussed during a 30-minute to one-hour face-to-face or online video session. The experts were asked to provide feedback along the following set of questions:

1. Is each aspect and its associated definition clear and well-worded?

2. Does each statement capture the aspect it is supposed to capture?
3. Are there any statements less important than the other statements within a category?
4. Is there anything important that is missing in the set of statements that capture each aspect?
5. Overall, does the proposed grouping of the statements seem logical to you?

6.3.2.1 General feedback

One expert raised a concern about the use of reverse-coded items as he argued they often end up being dropped during the various stages of instrument validation. The expert with experience as a FLOSS contributor recommended avoiding the use of the word *assignment* as it gives connotations that tasks are being assigned to contributors whereas it is seldom the case in FLOSS projects. The expert suggested to use other words such as *job*, *task*, or *work*. Both suggestions were taken into consideration when refining the items. Reverse-coded items were analyzed with extra attention, while the word *assignment* was replaced in all the items that contained it.

6.3.2.2 Expert feedback about the socialization construct items

The experts were asked to carefully consider the derived items for each socialization construct. Their feedback as well as the actions that were taken to address the issues they raised are presented in the following subsections.

TASK SEGREGATION One expert argued that some of the task segregation items provided a timeframe whereas others did not have any temporal indication. They further indicated that even if the timeframe is clearly stated in an introductory paragraph of the survey, respondents will still have a tendency to forget the timeframe when completing the socialization section of the survey instrument. After consultation with the supervising staff, it was decided that indicating a timeframe for each of the items would not be a satisfactory solution as this could increase the overall length of the items considerably. This would contradict Hinkin (1998)'s recommendations to use items as simple and short as possible. In turn, this may also affect the overall survey completion rate and then lower the number of respondents. A compromise solution was adopted: a short paragraph was added before the socialization variable items making clear that the items are concerned with the time when the person was a newcomer in the community.

Another issue was raised about the items SE3 and SE7, both of which refer to project tasks that are usually performed by experienced members. Two experts pointed out that the wording was unclear and could tap into other constructs. The wording of SE3 was refined following the recommendations of the experts. SE7 was dropped because experts noted that the two items were semantically very close.

One expert suggested replacing the term *appropriate* with *designed* in SE4 in order to make the item clearer. In addition, *assignments* was replaced with *tasks* in SE5. The word *advised* used in SE6 was found to be ambiguous as it implied a third party. Another expert recommended adding (*such as tutorials*) in the item in order to make

Table 21: Refined *task segregation* items - Expert panel

ITEM	ACTION	STATEMENT
SE3	REPHRASED	Much of my project-related knowledge has been acquired while working on tasks <i>that are normally carried out by experienced members</i> (R).
SE4	REPHRASED	I have been involved in tasks which I was clearly aware were <i>designed</i> for newcomers.
SE5	REPHRASED	I have carried out <i>tasks</i> designed by the community to help newcomers become project contributors.
SE6	REPHRASED	I have been through a set of tasks (<i>such as tutorials</i>) that the community recommends its newcomers perform.
SE7	DROPPED	

the meaning of the item more specific by adding a concrete example, tutorials, in the item wording (see Table 21 for the refined items).

INTERACTION INTENSITY IN1 was argued to be double-barreled and ambiguous; the item was dropped. An expert recommended making item IN2 more specific, by indicating the overall purpose of using the communication channels. The expert noted that a newcomer can be active in communication channels for reasons that may not be related to learning and becoming a project member. IN3 and IN5 were flagged as one expert was doubtful about the two reverse-coded items. *Interacted* was replaced with *engaged* in IN4 following a suggestion from one of the experts. IN6 was judged to be capturing aspects of social identification (being part of a group) while interaction intensity is behavioural in nature. The item was dropped. An expert raised concern about IN7 and IN8 which he saw as potentially correlating with supportiveness. The items were kept in the measurement scale of the construct, but were flagged in order to carry further investigation during the later instrument validation stages. The newly refined items are presented in Table 22.

TASK PURPOSEFULNESS One of the experts indicated that TP1 was unclear as it was not properly capturing the existence of links within tasks, the purpose of which is to learn about a project. The item was reformulated according to the expert's suggestions. TP2 and TP6 were dropped because the items were found to be both more about actual tasks performed by a newcomer rather than the links among them. As a result, the items were seen as more related to *task segregation* rather than *task purposefulness*. *Each task* was replaced with *Almost every task* in TP3 for the reason that if a member has performed a well-structured sequence of tasks in an attempt to learn, with the exception of one single task, the person should still score high in *task purposefulness*. TP4 was slightly reworded in order to make more explicit that the tasks performed by a newcomer are chosen with learning purposes in mind. TP5 was dropped because the word *scope* in the context of a FLOSS project appeared to be vague and ambiguous. TP7 was also refined following suggestions from some of the experts to make the item clearer. Two experts recommended modifying TP8 so that it was not reverse-coded and was more explicitly focused on the learning purpose when choosing tasks. Finally, a

Table 22: Refined *interaction intensity* items - Expert panel

ITEM	ACTION	STATEMENT
IN ₁	DROPPED	
IN ₂	REPHRASED	I have been active in the project communication channels (such as forums, mailing lists, or IRC channels) <i>to help me learn about the project</i> .
IN ₃	FLAGGED	Most of my learning has been carried out on my own, with no involvement from other newcomers or experienced community members (R).
IN ₄	REPHRASED	I have been <i>engaged</i> with many other community members to gain project-related knowledge.
IN ₅	FLAGGED	I have learned how to contribute to the project all by myself (R).
IN ₆	DROPPED	
IN ₇	FLAGGED	Working with other people as a newcomer has contributed to my learning.
IN ₈	FLAGGED	The way I have learned to become a contributor in this project has been highly collaborative.

compilation of the experts' suggestions was used to generate an extra item, TP₉. Table 23 presents the refined items.

Table 23: Refined *task purposefulness* items - Expert panel

ITEM	ACTION	STATEMENT
TP ₁	REPHRASED	<i>The pattern of the sequence of tasks that I have performed helped me gain an understanding of the project.</i>
TP ₂	DROPPED	
TP ₃	REPHRASED	<i>Almost every task that I have performed has expanded and built upon the knowledge gained from the preceding tasks.</i>
TP ₄	REPHRASED	<i>To learn how to become a contributor, I have undertaken jobs with gradually increasing complexity.</i>
TP ₅	DROPPED	
TP ₆	DROPPED	
TP ₇	REPHRASED	<i>There is cohesion in the sequence of tasks I have carried out while learning how to become a contributor.</i>
TP ₈	REPHRASED	<i>The logical connection among the tasks I have performed helped me gain an understanding of the project.</i>
TP ₉	NEW	<i>Most of the tasks that I have undertaken were chosen in a way that helped me learn about how to become a project contributor.</i>

JOINING STRUCTUREDNESS JS₂ was found to be confusing for an expert who gave the example of a community having a formal joining procedure which was not fol-

lowed by a newcomer. The item was reworded to make sure that it specifically focused on the actual experience of the newcomer. Two experts indicated that JS3 and JS4 were about communication and not newcomer experience. The items were thus dropped. JS6 and JS7 were reworded to omit any references or connotations to community communication and to focus solely on newcomer experience. Some experts recommended rephrasing JS8 so that it was not reverse-coded; the item was modified accordingly. The new item statements are presented in Table 24.

Table 24: Refined *joining structuredness* items - Expert panel

ITEM	ACTION	STATEMENT
JS2	REPHRASED	The community did not require me to follow any <i>step-by-step</i> joining process to become a project contributor (R).
JS3	DROPPED	
JS4	DROPPED	
JS6	REPHRASED	<i>In this community, I had to go through several steps in order to become a project contributor.</i>
JS7	REPHRASED	<i>The process I had to follow to join this community has been clearly structured.</i>
JS8	REPHRASED	<i>The process I undertook to join this community has clearly followed a defined pathway.</i>

MENTORING Several experts argued that the use of *One experienced member [or more]* in the mentoring items was misleading. They recommended using instead *One or more experienced member(s)*. An expert indicated that ME3 was double-barreled. The item was thus simplified. Some concern was raised by an expert about the use of the expression *have always been there* in ME5. It was perceived to be too vague and had the potential to be interpreted in various ways. The expression was replaced with *have always been responsive*. ME7 was dropped because it was seen by an expert as being strongly correlated to *supportiveness*. Table 25 summarizes the newly modified items used to capture *mentoring*.

Table 25: Refined *mentoring* items - Expert panel

ITEM	ACTION	STATEMENT
ME1	REPHRASED	<i>One or more experienced member(s) provided me with support and feedback when I was new to this community.</i>
ME2	REPHRASED	<i>One or more experienced member(s) took me under his/her wing while I was learning how to contribute to this project.</i>
ME3	REPHRASED	<i>I became a project contributor thanks to the mentoring of one or more experienced members.</i>
ME4	REPHRASED	<i>One or more experienced contributor(s) has been instrumental in helping me learn how to become a contributor to this project.</i>
ME5	REPHRASED	<i>One or more experienced member(s) have always been responsive when I needed help.</i>
ME6	REPHRASED	<i>I could not have become a contributor to this project without the guidance of one particular person [or more] in this community.</i>
ME7	DROPPED	
ME8	REPHRASED	<i>I have been mentored by one or more experienced contributor(s).</i>

SUPPORTIVENESS For the same reason that was raised about the use of the expression *being there* in ME5, SU4 was discarded. SU5 and SU6 were slightly reworded as they addressed newcomers in general rather than the particular person who is reading the item. An expert argued that a community may be overall supportive towards its newcomers while not being supportive toward a particular individual. This could have led to measurement issues. The modifications of the items used to capture supportiveness are presented in Table 26.

Table 26: Refined supportiveness items - Expert panel

ITEM	ACTION	STATEMENT
SU4	DROPPED	
SU5	REPHRASED	<i>This community cared a lot about me when I was a newcomer.</i>
SU6	REPHRASED	<i>In this project, I was made to feel part of the community when I was a newcomer.</i>

6.3.2.3 Experts' feedback about the performance constructs

No major concerns were raised about the *task performance* and CCB constructs. The construct definitions and items were judged to be clear and appropriate.

The expert who was an experienced FLOSS contributor suggested adding an additional item in the survey questionnaire to control for the extent to which it is easy to tell if a task has been adequately performed. It was pointed out that there could be some

Table 27: Refined Task Performance items - Expert panel

ITEM	ACTION	STATEMENT
PER ₁	REPHRASED	I adequately complete my project tasks in this community.
PER ₂	REPHRASED	I effectively fulfill my responsibilities in this project.
PER ₃	REPHRASED	My contributions meet the performance expectations from the community.

variation from one FLOSS community to another. Consequently the item “It is easy to understand if the job that I have been doing is adequate in this community” was added to the survey instrument to control for this issue.

Two experts noted that the nature of the items may lead to a social desirability bias as people have a tendency to think they are above average. One of the two experts suggested rephrasing the items in a way that they would reflect the beliefs of what others think of the concerned person (e.g. community members feel that I adequately complete my tasks and responsibilities). After careful consideration, the proposed suggestion was judged inappropriate, due to the overall lack of hierarchical structure in FLOSS projects as well as the fact that individuals are mostly self-guided. Only members who are highly involved or visible may be in a position to be aware of how others perceive their performance. As a result, adopting such approach would have triggered another form of bias related to how ‘connected’ one is within a project and its community. However, it was agreed with the supervising staff that social desirability was a serious threat. Each of the flagged items were carefully considered and rephrased in an attempt to minimize the risk of such bias. Moreover, it was also decided that a social desirability bias could be potentially decreased by adding a statement prior to the items related to individual performance emphasizing that the survey was anonymous. In addition, the statement would make clear that providing accurate answers was crucial in order for this research to effectively contribute to FLOSS communities.

An expert recommended simplifying the CCB definition as the expression *in the aggregate* was seen as being redundant. After consultation with the supervising staff, it was decided to modify the definition accordingly.

6.3.2.4 Feedback about task performance items

Overall, the experts indicated that the items effectively covered the domain of the construct. A few minor issues were raised by some of the experts.

PER₁, “*I adequately complete my tasks and responsibilities*” was double-barreled. The word *responsibilities* was then dropped from the item. A similar issue was raised about PER₂, “*I effectively fulfill my roles and responsibilities in this project*”. Again, it was decided to remove the term *responsibilities* from the item. Finally, PER₃, “*My contributions meet or exceed the performance expectations from the community*” appeared to be capturing two separate aspects (*meet* and *exceed*). The term *exceed* was removed from the statement. The changes are presented in Table 27.

Table 28: Refined CCB-I items - Expert panel

ITEM	ACTION	STATEMENT
CCB-I2	FLAGGED	I go out of my way to help other community members with project-related problems.
CCB-I3	FLAGGED	I help community members who seek support from other members.
CCB-I5	REPHRASED	I assist newcomers in becoming familiar with the community.
CCB-I6	REPHRASED	When conflict occurs between other members of this community, I try to act as a <i>moderator</i> .
CCB-I8	REPHRASED	I consult people who might be affected by my actions or decisions.
CCB-I10	REPHRASED	<i>Within the community</i> , I treat other community members with respect.

6.3.2.5 Feedback about CCB-I items

Two experts argued that there was a lot of overlap between CCB-I2 and CCB-I3. It was decided to keep both items during the instrument validation phase and to eventually retain the item that provided better validation.

One of the participants recommended removing the term *willingly* from CCB-I5. The item was modified accordingly. The term *moderator* was judged to be more appropriate than *peacekeeper* in CCB-I6. The item was slightly rephrased. CCB-I7 was signalled by an expert as having some potential social desirability bias. The item was eventually retained as it was. The researcher decided that participants of the pretest would be asked to give some feedback about the item and its potential bias.

A member of the expert panel recommended removing “*I try to*” from CCB-I9. Nonetheless, the researcher and the supervising staff decided to retain the phrase, after concluding that the use of *I try to* emphasized emotional state rather than the actual behaviour, which was more appropriate when capturing this particular instance of citizenship behaviours.

Finally, one expert suggested adding “*Within the community*” at the beginning of CCB-I10, arguing that the item may potentially be misunderstood otherwise. Some respondents may not relate the behaviour to people that belong to the project but rather to individuals in general. The item was reworded accordingly. The undertaken actions for the CCB-I items are presented in Table 28.

6.3.2.6 Feedback about CCB-P items

An expert indicated that the item CCB-P1 had some minor social desirability bias. The researcher and the supervising staff decided to rephrase the item to “*I think about the whole community when contributing to the project*” to reduce any potential bias. It was suggested by another expert that the item CCB-P7 “*I let the community know if I am not able to complete a task on which I am working*” may be found to correlate with task

performance. The item was thus flagged for later investigation.

The terms *maintenance tasks* in CCB-P8 were judged by an expert to be too code-oriented since projects involve both technical and non-technical contributors. It was decided to use *tedious tasks* instead. CCB-P10 was debated with an expert and later with the supervising staff as it was pointed out that a member may be unable to attend community events because of geographical constraints. It was then decided to slightly extend the scope of the item by reformulating it to “*I contribute to community events (such as conferences, hackfests, or sprints)*”. This phrasing applied to members who do attend events as well as those that make an effort to contribute to the organization of the events.

Regarding CCB-P11 and CCB-P12, two experts indicated that there was a risk of social desirability for both items. One of the two experts explained that it was difficult for a respondent to answer such questions as a person who is not constructive still has a tendency to think that he or she is *constructive* or *effective*. After consultation with the supervising staff, it was decided to make both items more specific. CCB-P11 was then rephrased to “*When I express my opinion in the community, I make a conscious effort to be constructive*” and CCB-P12 to “*When I take part in a project decision, I make a conscious effort to be effective*”.

Two experts raised the issue that CCB-P14 could be simplified as it could be seen as double-barreled. Following the experts’ recommendations, the item was modified accordingly: “*I promote the project to potential users*”. A related point was made by two experts for CCB-P15. In addition, it was noted that in the case where the software produced by a given project is highly technical or specific to a narrow domain of application (the example of an advanced statistical package was given), talking to friends and family may not be that important for the community. It was decided to change the item to “*I spend time talking about the project to project outsiders*”. An identical remark and course of action were taken about CCB-P17 which was modified to: “*I encourage potential users to use the software produced by the community*”. Finally, a participant on the expert panel indicated that the focus on community image in CCB-P16 was unnecessary. He recommended instead rephrasing the item in order to focus on the achievements of the community. The item was modified accordingly. All item refinements are summarized in Table 29.

6.3.3 Card sorting

Following the expert panel evaluation, two rounds of card sorting were performed. According to Moore and Benbasat (1991), the objective of this stage of the instrument validation process has two overall purposes. First, it is a means to assess the construct validity of the scales that are under development; and second, it serves to identify items that are still unclear or ambiguous. Card sorting consists of asking judges to sort a set of index cards containing item statements into categories, independently from the other judges. If an item is consistently placed within the same category, then the item is considered to demonstrate convergent validity with the construct related to the category, as well as discriminant validity with respect to the other constructs. Davis (1989) then Moore and Benbasat (1991) have contributed to making card sorting

Table 29: Refined CCB-P items - Expert panel

ITEM	ACTION	STATEMENT
CCB-P ₁	REPHRASED	<i>I think about the whole community when contributing to the project.</i>
CCB-P ₇	FLAGGED	I let the community know if I am not able to complete a task on which I am working.
CCB-P ₈	REPHRASED	I voluntarily perform <i>tedious</i> tasks to keep project resources to a good standard.
CCB-P ₁₀	REPHRASED	<i>I contribute to community events (such as conferences, hackfests, or sprints)</i>
CCB-P ₁₁	REPHRASED	<i>When I express my opinion in the community, I make a conscious effort to be constructive.</i>
CCB-P ₁₂	REPHRASED	<i>When I take part in a project decision, I make a conscious effort to be effective.</i>
CCB-P ₁₄	REPHRASED	I promote the project to <i>potential users</i> .
CCB-P ₁₅	REPHRASED	I spend time talking about the project to <i>project outsiders</i> .
CCB-P ₁₆	REPHRASED	I actively promote <i>the achievements of the project's community</i> .
CCB-P ₁₇	REPHRASED	I encourage <i>potential users</i> to use the software produced by the community.

a popular instrument validation technique in the IS field.

Card sorting procedures have two variants: open and closed. The construct labels (the actual categories) and definitions are provided in a closed card sorting procedure. In an open (or exploratory) card sorting procedure, judges do not know what the underlying constructs are and are asked to create constructs themselves, and to provide their own category labels and definitions for the constructs they create. The latter technique was introduced by Moore and Benbasat (1991) who saw such an approach as an extension of Davis (1989)'s method of asking judges to rank how well items fit into pre-defined construct definitions. Moore and Benbasat argued that open card sorting reduces the risk of "interpretational confounding" defined by Burt (1976) as "the assignment of empirical meaning to an unobserved variable [e.g. factor] other than the meaning assigned to it by an individual prior to estimating unknown parameters" (p. 4). If judges provide category definitions that match the overall intent of each of the scales, it is then a further indication of construct validity.

Following the guidelines provided by Moore and Benbasat (1991), a two-round card sorting procedure, open then closed, was organized with two different sets of judges. The card sorting exercises started once the approval from the Human Ethics Committee was granted (see Appendix A). Each round involved four judges which were selected using a convenient sampling approach. The participation criterion for judges was to have some degree of understanding about FLOSS communities and FLOSS development. A protocol containing a brief description of the project and a set of detailed instructions (see Appendix D) was sent to each judge by email at least two days before the card sorting exercise was conducted. A hard-copy of the protocol was also

provided to the judges at the beginning of the session. The protocol for both rounds was reviewed by two IS senior lecturers to ensure that instructions were unambiguous and easy to follow. Several modifications were performed on the final protocol documents to address some of the issues of mentioned in the provided feedback.

6.3.3.1 Round 1

It was decided to only include the items that were derived for the measurement scales of the six socialization constructs. The reason to temporarily exclude the items from the other constructs was twofold. First, open card sorting procedures are usually intellectually demanding for judges and it was important to avoid cognitive overburden. Second, since the socialization constructs were newly designed and had thus no prior empirical validation, it was most important to assess the construct validity of these six constructs specifically. Furthermore, inclusion of the items measuring social identification, task performance and citizenship behaviours may have reduced the effectiveness of identifying potential convergent and discriminant validity issues of the socialization constructs.

Participants consisted of a male IS Senior Lecturer and a female PhD candidate that had a good understanding of FLOSS and software development, and two male FLOSS contributors (one very experienced and the other with several years of experience). For both the IS Senior Lecturer and the PhD candidate, a short discussion preceded the explanation about the instructions to ensure that they both had a good understanding about FLOSS in general as well as about FLOSS communities and FLOSS development. Participants were chosen in a way that maximized the range of ages, backgrounds and expertise to ensure a variety of perceptions in the analysis. Each of the 39 socialization items was printed on an index card that had a random number printed on the back (to help identify items and compute the inter-rater scores).

Following [Moore and Benbasat \(1991\)](#), the judges participated in a practice card sorting exercise during which 11 cards containing statements about perceptions towards driving were provided (see [Appendix D](#)). Six items were related to perceived driving ability, four items to perceived risks, one item was ambiguous, and a last item 'unfitting'. The practice round helped to ensure that the judges felt confident about how to sort index cards according to their underlying idea. Three of the four judges performed the practice exercise within a minute without any error. The fourth judge initially assigned the cards in a different way, and asked that the instructions be re-explained. The researcher made sure that the judge then clearly understood the nature of the exercise.

Judges were asked to perform the following tasks (as indicated in the provided instructions):

1. Sort cards into categories according to the primary underlying idea that each statement reflects.
2. Write a label and a definition for each of the identified categories.
3. Identify potential ambiguous, repeated and/or indeterminate items.

Once the practice exercise was completed, the 39 cards were presented to the judges. For each judge, the stack of index cards had been thoroughly shuffled to ensure that there was no pattern in the sequence of cards. The instructions made clear that judges were free to use whatever number of categories they thought was necessary, that each category may contain any number of cards, and that judges could change their mind as many times as needed. A stack of blank yellow cards was also provided to them in order to write the category labels and definitions. Each stack of yellow cards contained more than 20 cards so as not to give any numerical indication about the targeted number of categories. The four judges finished the exercise within 20 to 25 minutes. None of the judges requested any clarification during the exercise.

RESULTS The construct labels provided by the judges are presented in Table 30. Both *supportiveness* and *joining structuredness* were clearly identified by the four judges whose labels were semantically close to the target constructs. The provided definitions also matched closely the targeted theoretical constructs, providing satisfactory evidence of construct validity for the two measurement scales.

Table 30: First sorting round: Individual judge's construct labels

CONSTRUCTS	JUDGES			
	<i>Judge A</i>	<i>Judge B</i>	<i>Judge C</i>	<i>Judge D</i>
<i>Task segregation</i>	Tasks	Ability progression	Work	Structured learning pathway
<i>Task purposefulness</i>				
<i>Interaction intensity</i>	Mentoring	Help from community	Relationships	Mentoring
<i>Mentoring</i>				
<i>Joining structuredness</i>	Process of joining	Newcomer process	Process	Joining the community
<i>Supportiveness</i>	Acceptance	Feeling valued	Community support	Feeling part of community

The four judges did not distinguish *task segregation* from *task purposefulness* as they all included the two constructs in a higher-level category related to tasks performed by newcomers. For analysis purposes, the two constructs were temporarily grouped together, and a decision to refine the constructs after the results of the closed card sorting procedure postponed until later.

Two judges identified *mentoring* as a separate category whereas the other two judges merged *mentoring* and *interaction intensity* in a higher-level category capturing community assistance. Again for analysis purposes, both constructs were merged together and a further investigation was performed during the second round of card sorting. It was interesting to note that the judges who had a *mentoring* category, did not have any separate category for the *interaction intensity* items. This raised an issue about the appropriateness of the *interaction intensity* items which also appeared in the computation of the various agreement scores. It is interesting to note that the clubbing

of the task-related as well as the individuals-related constructs provided further evidence about the socialization framework derived during the qualitative phase of this research which identified three overall categories of socialization factors: task, individuals, and community.

Four categories were used to assess the overall degree of agreement among the four judges: tasks structuredness (the nature of the tasks and the nature of the links among them), community assistance, joining structuredness, and supportiveness.

There are two overall approaches when assessing the degree of agreement among a set of judges: inter-judge agreement scores and overall item placement scores. Inter-judge agreement scores compare the degree of agreement between each distinctive pair of judges and an average of those scores is used to have a global agreement measure. The level of agreement between two judges may be computed using Raw Percentage Agreement, Cohen's Kappa (Cohen, 1960), or Fixed Marginal Multi-rater Kappa (Siegel & Castellan, 1988). All three scores were computed (see Table 31 for the compiled results).

Table 31: Inter-judge agreement

PAIRED JUDGES	RAW AGREEMENT	COHEN'S KAPPA
1&2	0.641	0.535
1&3	0.744	0.647
1&4	0.821	0.757
2&3	0.821	0.759
2&4	0.692	0.603
3&4	0.795	0.727
<i>Average</i>	0.752	0.671
Fixed marginal multirater Kappa: 0.666 (Siegel and Castellan, 1988)		

There is no agreed threshold for Raw Percentage Agreement scores. However, Hinkin (1998) recommended using 75% as an acceptable minimum agreement level. Similarly, Cohen's Kappa coefficients, that measure the proportion of agreement between judges after removal of agreement attributed to chance, are usually considered acceptable when they are higher than 0.65 (Jarvenpaa, 1989; Moore & Benbasat, 1991). In this research, the Kappa nomenclature from Landis and Koch (1977) that describes the relative strength of agreement associated with Kappa statistics was used (see Table 32). Both Raw Percentage Agreement (0.752) and Cohen's Kappa scores (0.671) were slightly higher than the minimum recommended thresholds. The results were encouraging as they indicated some degree of confidence in the overall appropriateness of the socialization construct measurement scales.

Table 32: Landis and Koch's Kappa statistics nomenclature

KAPPA STATISTIC	STRENGTH OF AGREEMENT
< 0.00	<i>Poor</i>
$0.00-0.20$	<i>Slight</i>
$0.21-0.40$	<i>Fair</i>
$0.41-0.60$	<i>Moderate</i>
$0.61-0.80$	<i>Substantial</i>
$0.81-1.00$	<i>Almost Perfect</i>

Item Placement Ratio is an indicator of how many items are placed in the appropriate target categories by all the judges. The procedure was developed by Moore and Benbasat (1991) who recommended using the resulting matrix to identify potential problem areas. The overall frequency with which judges have assigned items to the targeted theoretical constructs can be seen as a measure of construct validity. High placement scores are also indicators for potentially good reliability scores. The item Placement Score matrix is presented in Table 33.

Table 33: Items Placement Score – Round 1

TARGETED CONSTRUCTS	ACTUAL CATEGORIES						
	SE + TP	IN + ME	SU	JS	N/A	TOTAL	%HITS
SE + TP	46			6		52	88.46%
IN + ME	4	38	7	2	1	52	73.08%
SU	1	1	26			28	92.86%
JS	1			19		20	95.00%
Total of items placement: 152			Hits: 129		Overall Hit Ratio: 0.849		

The overall ratio was satisfactory, as it was around 85%. Confirming the successful identification and labelling of both joining structuredness and supportiveness, the items of the two constructs were placed by the judges with a high success rate (respectively 95% and 93%). The six mis-placements with respect to the combined category *task segregation* and *task purposefulness* were all performed by the same judge. The feedback discussion and the analysis of the judge's placements revealed that the judge had combined the *joining structuredness* and *task segregation* items within a category called 'Newcomer process'. The judge explained that he had interpreted this category to include all initiatives (both tasks and joining) launched by the community that are designed for newcomers. It was concluded that the judge had considered a higher-level category which included both *task segregation* and *task purposefulness*. No action was taken regarding the items as the three other judges had clearly distinguished (no single mis-placement) the two constructs.

The analysis of the item placement score matrix revealed more serious issues with the *interaction intensity* items. Whereas the *mentoring* items were overall successfully placed (2 mis-placements out of 28), 50% of the *interaction intensity* items were assigned

to an incorrect category (12 out of 24). However, it should be noted that the two judges who were FLOSS contributors properly assigned the items. The mis-placements came from the two other judges (the IS Senior lecturer and the PhD candidate) who had some knowledge of FLOSS but no direct experience. Both those judges also identified *mentoring* as a category and placed the *interaction intensity* items in the task-related or supportiveness constructs. Therefore the mis-placements were not judged to be overly problematic.

Among the incriminated items, the participants of the expert panel had already raised some concern about items IN3 and IN5 because of their reverse-coded nature:

IN3: Most of my learning has been carried out on my own, with no involvement from other newcomers or experienced community members (R).

IN5: I have learned how to contribute to the project all by myself (R).

The researcher judged that this was enough evidence to modify the items so that they would not be reverse-coded anymore. Indeed, an absence of interaction with other community members can simultaneously be perceived as an absence of support from the community and no mentoring. The items were thus refined accordingly:

IN3: I have been involved with many other community members to help me learn how to contribute to the project.

IN5: Interacting with a lot of other project members has helped me understand how to become a project contributor.

IN2 and IN8 were flagged and were carefully analyzed during the latter stages of instrument validation. Three out of four judges placed IN7 in the appropriate category. No action was taken for that item.

6.3.3.2 Round 2

The second round of card sorting consisted of two distinct closed card sorting exercises. It was decided to separate the socialization constructs from the performance-related ones to optimize the validation of the socialization measurement scales. The two groups of constructs being overall semantically different, the separate treatment of each set of items was not problematic. As a matter of fact, the risk of discriminant validity issues between socialization and performance items was judged to be minimal. *Social identification* and *social integration* were excluded from the procedure.

The judges included a female IS Senior Lecturer and a male IS Associate Professor, and two male FLOSS contributors with several years of experience in FLOSS. A short discussion about FLOSS principles and practices with the IS Senior Lecturer and the IS Associate Professor preceded the exercise. This was to ensure that both judges had a sufficient level of understanding about FLOSS to satisfactorily perform the sorting exercises.

Each of the 39 socialization items was printed on an index card that had a random number printed on the back (to help identify items and compute the inter-rater scores), the name and definition of each construct was printed on a yellow index card. Simi-

larly, the measurement items of *task performance*, CCB-I, CCB-P (for a total of 35 items) were printed on index cards with a random number printed on the back. Labels and definitions were provided on three yellow index cards.

Judges were asked to perform the following tasks for each of the two exercises (as indicated in the provided instructions):

1. Read carefully each category name and definition and place each card in front of you.
2. Assign each statement to a category according to the primary underlying idea that each statement reflects.
3. Write a label and a definition for each of the identified categories.
4. Identify potential ambiguous, repeated and/or indeterminate items.

For each judge, the two stacks of index cards were shuffled. The four judges completed each exercise in about 15 minutes.

RESULTS - SOCIALIZATION CONSTRUCTS None of the judges identified any ambiguous or ill-worded items. All judges indicated that they had had no concerns about the procedure, which they found straightforward. The level of agreement among judges was also assessed using Raw Percentage Agreement, Cohen's Kappa, Fixed Marginal Multi-rater Kappa (see Table 34) and Items Placement Score (see Table 35).

Table 34: Inter -judge agreement – Socialization

PAIRED JUDGES	RAW AGREEMENT	COHEN'S KAPPA
1&2	0.846	0.815
1&3	0.872	0.846
2&3	0.821	0.784
1&4	0.923	0.908
2&4	0.846	0.814
3&4	0.846	0.816
<i>Average</i>	0.859	0.831
Fixed marginal multirater Kappa: 0.824 (Siegel and Castellan, 1988)		

Table 35: Items Placement Score – Round 2

TARGETED CONSTRUCTS	ACTUAL CATEGORIES							
	SE	TP	IN	ME	JS	SU	TOTAL	%HITS
SE	21	2			5		28	88.5%
TP	1	19			4		24	73.1%
IN			22		1	1	24	92.9%
ME				27		1	28	95.0%
JS					24		24	100%
SU			2			26	28	75.0%
Total of items placement: 156			Hits: 139		Overall Hit Ratio: 0.891			

The results were overall satisfactory with an overall hit ratio of about 89%, an average Cohen's Kappa agreement coefficient of 0.83 and a Fixed marginal multirater Kappa coefficient of 0.82 thus suggesting an *Almost Perfect* level of agreement, following Landis and Koch (1977). The biggest issue that arose from the judges concerned the *task purposefulness* items. The construct definition clearly specified that newcomers voluntarily choose the sequence of tasks to perform in an attempt to learn about the project. Even though judges mostly assigned the items to the right category, three of them indicated that all the *task purposefulness* items were not explicit enough. For instance, item TP1 “The pattern of the sequence of tasks that I have performed helped me gain an understanding of the project” is not specific enough about who chooses the tasks. One of the judges suggested rephrasing the item to: “The pattern of the sequence of tasks that I have chosen to perform helped me gain an understanding of the project”. All *task purposefulness* items were then rephrased accordingly (see Table 36).

Table 36: Task Purposefulness Items - Rephrasing

ID	ACTION	STATEMENT
TP1	REPHRASED	The pattern of the sequence of tasks that <i>I have chosen to undertake</i> helped me gain an understanding of the project.
TP3	REPHRASED	Almost every task that <i>I have chosen to perform</i> has expanded and built upon the knowledge gained from the preceding tasks.
TP4	REPHRASED	To learn how to become a contributor, <i>I have selected</i> jobs with gradually increasing complexity.
TP7	DROPPED	.
TP8	REPHRASED	The logical connection among the tasks <i>I have opted to perform</i> helped me gain an understanding of the project.
TP9	DROPPED	

The examination of the off-diagonal placement matrix above suggested that some items from the *task segregation* construct were assigned to *joining structuredness* and *task purposefulness*. Three of the seven misplacements were due to item SE2, whose word-

ing included strong connotations about sequence and process “*I have been through a set of tasks (such as tutorials) that the community recommends its newcomers perform*”. The item was dropped.

SE₂, “*I have undertaken activities which are intended to turn new community members into project contributors*”, was also misplaced. The misunderstanding seemed to be caused by the use of the term *activity*. The term indeed implies a notion of duration thus suggesting the idea of process. The term *task* was then used instead when rephrasing the item.

SE₃, “*Much of my project-related knowledge has been acquired while working on tasks that are normally carried out by experienced members*”, was misplaced by both IS academics but correctly assigned by the two FLOSS contributors. The item was flagged but kept in the scale.

The analysis of the placement results provided supportive evidence for the convergent and discriminant validities of *interaction intensity*, *joining structuredness*, *mentoring*, and *supportiveness*. SU₆ “*In this project, I was made to feel part of the community when I was a newcomer*”, was dropped since two judges misplaced the item into *interaction intensity*.

Some judges indicated that the items of the *mentoring* measurement scale were repetitive. The issue was taken into consideration and it was decided to retain six of the seven items. The decision to discard additional items to avoid repetition was postponed until after the pretest and pilot study. ME₅, “*One or more experienced members have always been responsive when I needed help*”, was discarded because judges indicated that the word *responsive* was not effectively capturing the idea of mentoring.

RESULTS - PERFORMANCE CONSTRUCTS The motive behind running a card sorting procedure involving the performance-related constructs was to test the construct validity of *task performance*, CCB-I, and CCB-P in the FLOSS context.

The *task performance* measurement items were derived from a measure that had demonstrated satisfactory results in the organizational behaviour literature (see section 6.3.1.2). Some of the items were refined, others were dropped because of their strong organizational connotation and/or irrelevance to the FLOSS context, and new items were generated by the researcher (see section 6.3.1.2). It was thus important to test the appropriateness of the modified and newly created items. Similarly, the items comprising both CCB-I and CCB-P were derived from acknowledged OCB-I and OCB-O subdimensions. Most items originated from existing measurements and were simply tailored to the FLOSS context. Several items were created by the researcher (see section 6.3.1.3). The card sorting procedure was mainly undertaken to ensure that the CCB items were clearly discriminated from the *task performance* items.

No study was found during the review of the literature that examined the OCB concept to the FLOSS community context. There was thus no empirical evidence about the dimensionality of FLOSS CCB. This research integrated the past OCB literature to introduce two CCB dimensions: CCB-I and CCB-P. The rationale behind the conceptualization was essentially theoretical. Because of the specificity of the FLOSS context,

there was thus a chance that throughout the various data collection phases undertaken in this research project, the dimensionality of CCB might change. It was decided that exploratory factor analyses of the CCB items run during both the pilot and final studies would help confirm (or not) the CCB dimensions introduced in this research. Therefore it was important not to discard CCB items that were mis-assigned during the card sorting procedure. Indeed, the consistent misplacement of a CCB item could provide evidence about the existence of additional dimensions.

The Items Placement Score matrix indicated a Hit Ratio of 85% whereas the Raw Percentage Agreement was found to be 0.72 and the Cohen's Kappa coefficient: 0.594 (see Tables 37 and 38). For the reasons mentioned above the result of the Items Placement Score table were considered to be more important than Raw Percentage Agreement and Cohen's Kappa.

Table 37: Inter -judge agreement – Performance

PAIRED JUDGES	RAW AGREEMENT	COHEN'S KAPPA
1&2	0.676	0.521
1&3	0.649	0.493
2&3	0.811	0.717
1&4	0.757	0.654
2&4	0.676	0.533
3&4	0.757	0.649
<i>Average</i>	0.721	0.594
Fixed marginal multirater Kappa: 0.633 (Siegel and Castellan, 1988)		

Table 38: Items Placement Score – Socialization

TARGETED CONSTRUCTS	ACTUAL CATEGORIES				
	PER	CCB-I	CCB-P	TOTAL	%HITS
PER	25		3	28	89.29%
CCB-I		40	4	44	90.91%
CCB-P	9	5	54	68	79.41%
Total of item placements: 140		Hits: 119		Overall Hit Ratio: 0.85	

The reason for the relative weakness of the two coefficients is twofold. First, the use of IS academics as judges proved to be satisfactory during the first exercise but demonstrated some limitations during the second exercise. Indeed, the perception of the distinction between in-role and extra-role behaviours was less clear for the two judges. For instance, the item *"I keep up with the latest news about the overall project as*

well as in areas in which I am involved" were assigned to *task performance* by both IS academics and (correctly) to CCB-P by the two judges who were FLOSS contributors. Identical results occurred for several other CCB-P items such as *"I voluntarily perform tedious tasks to keep project resources to a good standard"*. The general rule of thumb that was applied to retain or discard items that demonstrated some divergence was to keep the items that were correctly assigned by both judges with direct FLOSS experience.

Second, the inherent correlation between CCB-I and CCB-P naturally affected the results of the agreement coefficients. Nine of the 21 misplacements were between CCB-I and CCB-P. A careful consideration of each misplacement was made so as to determine whether misplacements occurred due to some innate correlation, or to the ambiguity or inappropriateness of an item.

As a result, the following actions were taken. CCB-I 6, *"When conflict occurs between other members of this community, I try to act as a moderator,"* was assigned twice to CCB-P. This was identified as an instance mis-assignment due to correlation. One may help solve an interpersonal conflict for the benefit of the conflicted individuals as well as the project. The OCB literature was very explicit in attributing such behaviour to OCB-I (MacKenzie et al., 1999; Organ, 1988, 1990a; P. M. Podsakoff et al., 1997; P. M. Podsakoff & MacKenzie, 1994). The item's wording was slightly modified to emphasize the interpersonal nature of the behaviour: *"I try to act as a moderator when a conflict occurs between other members"*. It was decided to introduce an extra item to capture peacekeeping behaviours for the later instrument validation stages. The item *"I try to act like a peacekeeper when other contributors have disagreements"* (CCB-I 12) was derived from a validated measurement scale (MacKenzie et al., 1999; P. M. Podsakoff et al., 1997; P. M. Podsakoff & MacKenzie, 1994). The other CCB-I items were unchanged as there were few misplacements.

CCB-P 5, 6, 11, 12, and 13 were misplaced by either one or two of the judges who were IS academics. For the reasons mentioned above, the items were retained. CCB-P 7, *"I let the community know if I am not able to complete a task on which I am working"* was dropped as the item was misplaced by the two FLOSS judges into *task performance*. CCB-P 3, *"I comply with the behavioural expectations of this community"* was also discarded as it was misplaced by two judges (one FLOSS judge and one IS academic judge), while the two other judges indicated that they could not decide whether the item belonged to CCB-I or CCB-P. The other items capturing the same aspect, CCB-P4 and CCB-P5, provided satisfactory results. Construct domain coverage was then not affected. As a result, 15 items out of 17 were retained for the next phase of the research project.

The decisions to retain or discard *task performance* items were stricter as it was essential to ensure that there were no convergent or discriminant validity issues with respect to the dependent variables. As a result, PER2, *"I effectively fulfill my responsibilities in this project"* was dropped because the term *responsibilities* was judged to be ambiguous in a FLOSS community context. PER7, *"I consider myself to be a dependable contributor to this community"* was also discarded, as the word *dependable* was found to be ambiguous by the judges.

6.3.4 *Survey design*

Once the measurement scales were validated through the expert panel procedure and the two rounds of card sorting, the online questionnaire was developed. Researchers using surveys for data gathering must be aware of the major threats to research validity, such as mistrust about survey uses, survey length, or privacy and security (Evans & Mathur, 2005). A critical challenge for social scientists conducting survey research is to maximize the response rate among sampled individuals in order to mitigate non-response bias (Luong & Rogelberg, 1998; Stanton & Rogelberg, 2001). Furthermore, online surveys have their own weaknesses. Internet surveys can be perceived as spam or junk mail. In addition, there can be technological variations in terms of the users' Internet connection or computer configurations, and the risk for incomplete survey submissions or multiple responses is increased (Evans & Mathur, 2005; Stanton & Rogelberg, 2001).

6.3.4.1 *Online survey using LimeSurvey*

LimeSurvey¹, a PHP-based FLOSS online survey application, was chosen for the design and creation of the survey, for two reasons. First, this user-friendly application allows the development of high quality online surveys by providing advanced features such as the use a rich text editor to design survey templates, the availability of a wide array of question types, or the possibility to run basic statistical and graphical analysis within the software. Second, LimeSurvey has been released under a FLOSS license (the GNU General Public License) - a judicious choice since the target population of the study is FLOSS community members and Limesurvey has been widely accepted (and used) within the FLOSS sphere. The online questionnaire was meticulously constructed and extra attention was paid to the survey layout and design (Simsek & Veiga, 2000).

6.3.4.2 *Survey structure*

The design of the online questionnaire was performed in accordance with well established online and general survey development principles (Dillman, 2009; Evans & Mathur, 2005; Malhotra, 2004). In particular, as suggested by Simsek and Veiga (2000), electronic questionnaires should make use of features that "minimize questionnaire completion time and maximize respondent convenience". The authors recommend the use "scrolling, jump screen, quitting, no automatic next, no keyboard responses, help screens, and a progress thermometer indicating completed percentage of the questionnaire" when possible (p. 108). A progress bar was added on top of each survey page to indicate current survey completion rate to respondents; hover effects were implemented to render navigation among lists of Likert-scale questions easier; and the use of survey variables was used to consistently display the name of the FLOSS project initially indicated by respondents in an effort to personalize the survey.

The initial online survey questionnaire was structured as follows:

1. Pre-survey

¹ <http://www.limesurvey.org>

- a) Brief research description and consent
 - b) Conditions to participate and screening question
2. Survey Questionnaire
- a) Part 1 - Questions related to role and participation of individual in FLOSS project and FLOSS in general
 - b) Part 2 – Questions related to newcomer experience
 - c) Part 3 – Questions related to current perceptions and contributions in FLOSS project
 - d) Part 4 – Demographics
3. Post Survey
- a) Thank you note
 - b) Link to public summary statistics (number of respondents per FLOSS project)

The pre-survey section contained a brief overview about the nature of the research project as well as a definition of FLOSS community in the context of this study. It also provided respondents information about participation consent in accordance to the Human Ethics Committee (HEC) requirements of the Victoria University of Wellington. Confidentiality, anonymity and voluntariness were particularly highlighted on the page. In the second page of the survey, conditions for participation were presented to viewers. Potential respondents qualified for participation when they satisfied the following conditions:

- They were a member of one of the listed FLOSS projects (they were asked to select the one to which they belonged or related the most to in case of multiple membership)
- They had joined the project within the last 2 years (to ensure that the effect of newcomer experience on contribution behavior was still memorable)
- They had been contributing to the project on a regular basis (on average once a month or more) during the last six months (in order to include only members that have a minimum level of activity in the project).

It was then specified that term 'contribution' was used in a broad sense that includes any sort of technical or non-technical action that contributes to the project (e.g. submitting a patch, editing a project documentation page, participating in a mailing list or an IRC channel, representing the project to outsiders, providing feedback to someone, redirecting a person with a problem to another person in the project...). Respondents were invited to click on the Next button only if they satisfied all conditions. Those who did not satisfy the conditions were politely informed that they could not take part in the survey.

The next page (Part 1) solicited project-related information that was not required for the analysis of this research but that would serve to provide detailed information about when and how the respondent joined the selected project. The questions were chosen in a way that would provide additional data that would enrich the interpretation of the primary data. Respondents were asked questions such as the time it took

them to become a member of the project, whether they took part in any formal mentoring programme or community event while joining the project as well as several questions about their prior FLOSS-related experience and knowledge.

Part 2 included all items from the socialization constructs. An introductory paragraph explained that all the questions pertained to the respondent's experience when he/she was in the phase of becoming a contributor for the project. The following statement concluded the survey section introduction: *Think back about the time from when you decided to join the project until you became a project member. Answer the items below by reflecting on what your experience was during that period.*

Part 3 started with a brief paragraph that emphasized that the remaining questions of the survey were concerned with current perceptions of, and contributions to the project. In addition, a sentence requested respondents to provide honest responses when answering questions about contributions and actions within the project (and repeated the point that the survey was anonymous). This was done to minimize as much as possible the risk of social desirability bias inherent to self-reported behavioural measures (R. J. Fisher & Katz, 2000). Social desirability bias is associated with the tendency for individuals to present a favourable image of themselves (van de Mortel, 2008) and has a long history in the social sciences in general and in psychology in particular (R. J. Fisher & Katz, 2000). An important strategy for controlling for social desirability biases has been shown to be maximizing subject anonymity (Sudman & Bradburn, 1974). Questions related to measurement items of *social identification*, *social integration*, *task assignment*, and CCB then followed in the survey section. Finally, Part 4 collected demographic information about the respondents (age, gender, country of residence, and occupation).

The post-survey page contained a thank you message, encouraged respondents to spread the word about the project, and provided a link to a web page which displayed some descriptive statistics about the survey. Respondents could click on the link to see the number of respondents who took part in the survey, per FLOSS project.

6.3.5 Pretest

In social sciences research, it is strongly recommended that survey questionnaires be pretested before they are administered in the field (Malhotra, 2004). Alreck and Settle (1995) define a pretest as a preliminary trial of some or all aspects of a measurement instrument in order to avoid unanticipated difficulties. Some confusion between pretest and pilot study has often arisen in past research (Boudreau et al., 2001). Moser (1958) provided the following explanation about the difference between the two types of test: "the pilot survey is the dress rehearsal, and like a theatrical dress rehearsal, it will be preceded by a series of preliminary tests and trials [i.e., the pretests]."

In general, a pretest involves the testing of a questionnaire using a small number of respondents, that usually ranges from 15 to 30 (Malhotra, 2004). The researcher then invites participants to provide feedback about all aspects of the questionnaire such as form and layout, wording, instructions, or question content/sequence/difficulty (Malhotra, 2004).

The first pretest involved 5 IS academics who had some knowledge in FLOSS or in some related domains (such as online communities) as well as 5 PhD students in Information Systems. The online survey was tested on various Web browsers on personal computers (Internet Explorer, Firefox, Google Chrome) as well as tablet computers. The researcher then tested the data collection and conversion procedures (into .csv, .xls files as well in SPSS files).

Overall, the respondents were highly satisfied with the design and layout of the survey. The feedback allowed the identification of various typographical errors. Also, several instructions and questions were slightly reworded for clarity purposes. For instance, there were several inconsistencies in the use of US or UK English words. It was eventually decided to use American English as it was expected that respondents would be more familiar with that spelling. In the introductory paragraph of Part 3, respondents were requested to provide 'accurate' answers. The adjective 'honest' was suggested by 3 participants. The second paragraph of the survey introduction requested potential respondents to let other people know about the survey. Several pretest participants suggested moving the paragraph to the end of the page. This was done accordingly.

In addition, two IS academic staff raised concern about six particular items which were perceived to be double-barreled. After consultation with the supervising staff, it was decided to refine the wording of some of the six identified items in order to make sure that each item captured a single aspect, thus avoiding inconsistencies in the responses. The modifications are presented in Table 39.

Fifteen experienced FLOSS contributors took part in a second round of pretesting. This phase of the pretest was particularly important to ensure that the content and wording of the survey would be well perceived by future respondents. Overall, the feedback provided was positive and enthusiastic. Some comments were raised about reducing the overall length of the first page that included the survey description and consent information. A few additional remarks concerned some minor changes in the wording of the question instructions. A participant also suggested that the version number of the license under which the survey software was released (GNU General Public License v3.0) be indicated.

6.3.6 *Pilot study*

A pilot study is a small scale rehearsal of a research survey that aims at testing the instrumentation before the final survey is administered to a larger sample (Alreck & Settle, 1995; Boudreau et al., 2001; Pinsonneault & Kraemer, 1993). Even though it is common for researchers to use small convenience samples when conducting pilot studies (Boudreau et al., 2001), the ideal pilot study dataset should come from the same target population used for the main study (Lewis, Templeton & Byrd, 2005; Pinsonneault & Kraemer, 1993). The analysis of pilot study data should help appraise and refine instrument items and scale measures through the deletion, addition, or modification of items (Lewis et al., 2005).

Table 39: Rephrasing after first round of pretest

ID	ACTION	STATEMENT
TP ₃	NONE	Almost every task that I have chosen to perform has expanded and built upon the knowledge gained from the preceding tasks.
CCB-I7	REPHRASED	<i>Previous:</i> I pay attention to how my actions and contributions will affect other members. <i>New:</i> I pay attention to how my actions will affect other members.
CCB-I8	REPHRASED	<i>Previous:</i> I consult people who might be affected by my actions or decisions. <i>New:</i> I consult people who might be affected by my actions.
CCB-P6	REPHRASED	<i>Previous:</i> I keep the community informed about my work and its progress. <i>New:</i> I keep the community informed about the progress of my work.
CCB-P9	REPHRASED	<i>Previous:</i> I keep up with the latest news about the overall project as well as in areas in which I am involved. <i>New:</i> I keep up with developments that are happening in the project areas I am involved in.
CCB-P13	REPHRASED	<i>Previous:</i> Once a decision has been made for which I either agree or disagree, I accept it and act accordingly. <i>New:</i> I abide by project decisions whether I agree with them or not.

6.3.6.1 Data collection

An application was submitted to the Human Ethics Committee (HEC) of the School of Information Management, Victoria University of Wellington (see Appendix A). After approval of the application, the pilot study was initiated.

Conducting the pilot study with a single FLOSS community risked achieving low variability in the scales as newcomer experience may not vary a lot within a given project. It was thus decided to include two FLOSS communities, Perl and LibreOffice, that were significantly different in terms of project history, practices, and culture.

The Perl community² has developed alongside the evolution of the Perl language that was initially developed by Larry Wall in 1987 as a general-purpose Unix scripting language that would help improving report processing. The Perl community contributes to a large Perl-written software repository, the Comprehensive Perl Archive Network (CPAN³) that has over 100,000 Perl modules written by more than 10,000 authors as of November 2012. Legal responsibilities are carried by the Perl Foundation, a non-profit organization which runs various events worldwide such as YAPCs (Yet Another Perl Conference) and workshops. The overall Perl community is organized

² <http://www.perl.org/community.html>

³ <http://www.cpan.org>

into sub-groups, or PerlMongers⁴ that involve contributors from the same area. As of November 2012, there were 257 Perl monger groups around the world.

LibreOffice⁵ is a FLOSS equivalent of the Microsoft office suite and that was derived in 2010 from another FLOSS office suite, OpenOffice (acquired in 2010 by Oracle Corporation). LibreOffice is officially developed by the Document Foundation that also runs the LibreOffice Conference each year. As opposed to the Perl community, LibreOffice is a rather young community that is particularly active in promoting the project to potential new contributors.

It was decided along with the supervising staff to target individuals who had joined one or the other of these communities within the last 2 years and who had been active during the last six months. The following conditions to participate were noted in the beginning of the survey:

- *You joined the community within the last 2 years.*
- *You have been contributing to the Perl [or LibreOffice] project on a regular basis: on average once a month or more during the last six months.*

Here the term 'contribution' is broad as it encompasses any sort of technical/non-technical action that contributes to the project (e.g. submitting a patch, editing a project documentation page, participating in a mailing list or an IRC channel, representing the project to outsiders, providing feedback to someone, redirecting a person with a problem to another person in the project...).

A separate online survey was administered to each of community. The Perl survey was available between 8 October and 22 October 2012. In addition to sending survey invitations on various project mailing lists, blog posts⁶ (see Figure 22) were written on various Perl-related resources, and the Perl Foundation promoted the survey on their website (see Figure 23) as well as during a Perl conference and a workshop.

The LibreOffice survey was live between 17 October to 29 October, 2012. Direct contact was established with the head of the Document Foundation as well with project leaders who helped spread out the word by sending messages on various mailing lists (see Figure 24) and discussion boards. Luckily, the LibreOffice conference was held while the survey was live (17–19 October, 2012). Several LibreOffice community members advertised the survey during conference sessions.

6.3.6.2 Sample characteristics

Before discussing the analysis of the pilot test data in detail, this subsection provides some general information about the profile of respondents. The datasets were initially screened for completion leading to 20 full responses (out of 41 submissions) for LibreOffice contributors, and 26 out of 83 responses for Perl. As a result, a total of 46 answers were usable for analysis.

⁴ <http://www.pm.org>

⁵ www.libreoffice.org

⁶ http://blogs.perl.org/users/kevin_carillo/2012/10/newcomer-experience-and-contributor-behavior-in-perl-and-other-foss-communities-survey.html

BETA Kevin Carillo
A blog about the Perl programming language

Newcomer experience and contributor behavior in Perl and other FOSS communities - Survey

By Kevin Carillo on October 4, 2012 1:40 AM under [CPAN](#), [citizenship](#), [community](#), [invitation](#), [newcomer](#), [research](#)

My name is Kevin Carillo. I am a PhD student currently living in Wellington (New Zealand) and I am doing some research on Free/Open Source Software communities.

If you have joined the Perl community within the last 2 years, I would like to kindly request your help. I am interested in hearing from people who are either technical or non-technical contributors, and who have had either positive or negative newcomer experiences.

The purpose of the research is to work out how newcomers to a FOSS community become valued sustainable contributors.

You can complete the survey via my [university's survey platform](#). For more detail on the project aims, see below.

What is the project about?

I am basically studying how the experience of a FOSS community newcomer has an influence on this person's actions and project contributions in the community.

Citizenship behaviors in FOSS communities

The first assumption that motivated this project is that attracting new members has become crucial for a large majority of FOSS communities but this is not a sufficient condition to ensure the success and prosperity of a project.

Suppose a community manages to get 50 new members per month and suppose a large number of them do not comply to the code of conduct, commit changes without considering the people or modules/components being affected by the commits, do not attend or contribute to any of the community events, and do not assist any other members when they seek for help ... Down the line, the health of the community is going to be affected and the future of the project seriously jeopardized.

So, yes ... it is important to attract newcomers but a community needs to make sure that a certain proportion of these newcomers become 'good' contributors from the community

About Kevin Carillo

I am a PhD student doing some research about free/open source software communities. I am interested in newcomer experience and the concept of citizenship in FOSS.

[More info »](#)

Search this blog

Figure 22: Survey invitation in Perl blog post

Perl Survey

9th October 2012
Mark Keating

You can help with a PhD students survey into Newcomer experience in open source.

If you have joined the Perl community within the last 2 years, we would like to invite you to complete this online survey. It is interested in hearing from people who are either technical or non-technical contributors, and who have had either positive or negative newcomer experiences.

The data you provide will help gain insights about the experience of newcomers within the Perl community. It will also be used to improve the survey before it is later administered to contributors from other FOSS projects.

Please consider filling this in and spreading the word about it. [The survey can be found here](#), and should take no longer than 20 minutes of your time to complete.

Figure 23: Message on Perl Foundation's website

Table 40: Demographics of the respondents - pilot test

		NUMBER OF RESPONDENTS	
		<i>LibreOffice</i>	<i>Perl</i>
AGE	20 or younger	1	0
	21- 25	1	4
	26- 30	4	6
	31- 35	2	5
	36- 40	1	3
	41- 45	2	1
	46- 50	2	0
	51 or older	7	1
EDUCATIONAL BACKGROUND	PhD	4	1
	Master's degree	2	3
	Postgraduate certificate or diploma	4	3
	Undergraduate degree	6	6
	Postsecondary certificate or diploma	1	0
	Secondary or high school graduate	1	5
	None	0	1
	Other	2	1
PREVIOUS FLOSS EXPERIENCE	No experience	9	12
	Less than six months	1	0
	Between 6 months and a year	1	3
	Between 1 year and 2 years	2	2
	Between 2 years and 3 years	1	1
	More than 3 years	6	8
NEWCOMER EXPERIENCE LENGTH	Less than 6 months	17	8
	Between 6 months and 1 year	2	5
	Between 1 year and 1.5 years	0	7
	Between 1.5 years and 2 years	1	1
	More than 2 years	0	5
TOTAL		20	26

Figure 24: LibreOffice mailing list invitation

Table 40 presents some of the important demographic information. Out of the 46 respondents, 37 were male and 2 female (the remaining seven people chose not to answer). The age of the LibreOffice respondents was rather evenly distributed between 20 and 50 with a peak of people being 50+, while nearly all Perl respondents were mostly between 20 and 35.

For both projects, educational background varied widely. Previous FLOSS experience indicated about 50% of each sample being individuals with no prior FLOSS experience and about a third of each sample consisting of individuals with 3+ years of experience. A majority of respondents indicated in both projects that they were either somewhat or very knowledgeable in FLOSS before they joined the project. In addition, around 50% of respondents in each project had not contributed to any other FLOSS project before joining.

The respondents had various roles within their respective projects, with development-related roles being the most common category: 28 out of 46. The next most common roles were project advocacy (5 respondents), translators (5 respondents), documentation writers (4 respondents), and user support (3 respondents).

Finally, the survey captured the time it took for respondents to become contributors once they had decided to join their respective project. While it took less than 6 months for nearly all the LibreOffice respondents, there was more variation for Perl contributors who spent from 6 months to more than 2 years to become contributors.

6.3.6.3 Validation strategy

Pilot studies are usually used at the final validation stage of a survey instrument before it is administered to its targeted population. Typically, both scale reliability and construct validity are assessed for each of the construct scales leading to the modification or deletion of some of the measurement items (Moore & Benbasat, 1991; Straub, 1989).

Decisions had to be taken about the nature of the tests and their sequence when validating the different parts of the conceptual models through the assessment of reli-

ability and construct validity. The following subsections present the various validation procedures that were followed.

RELIABILITY Construct reliability refers to the extent to which a set of measurement items are consistent in what they are intended to measure (Straub et al., 2004). Reliability is basically about the internal consistency of a measurement scale. Cronbach's alpha (Cronbach, 1970), one of Guttman's measures, is acknowledged as the standard measure of reliability (Straub et al., 2004), and was thus used in this research. The Cronbach's alpha coefficient reflects the extent to which the scores for all measurement items of a given construct have the same range and meaning (Straub et al., 2004). The accepted level of reliability varies across research projects as it depends on the nature of the research project (Moore & Benbasat, 1991). Nunnally argues that reliabilities of 0.50 to 0.60 are acceptable in early stages of research or in exploratory research (Nunnally, 1978) whereas .70 is usually used for confirmatory research (Moore & Benbasat, 1991; Straub et al., 2004). It was decided to use a reliability threshold of 0.70 in this research. Following Moore and Benbasat (1991)'s recommendations, the effects on Cronbach's alphas when items are deleted were also considered when refining each scale.

In addition to assessing Cronbach's alpha scores for each construct, item-item and item-scale correlations provide additional information about the extent to which items 'hang together' (Malhotra, 2004, p. 413). Item-item correlations help assess the extent to which all item of a given scale correlate with each other whereas item-scale (or item-to-total) correlations allow to assess whether each measurement item is consistent with the averaged behaviour of the other items of a given scale. Following Hair, Black, Babin and Anderson (2009)'s recommendations, item-to-total correlations were considered acceptable when they exceeded 0.50, whereas inter-item correlations within the same construct were compared to a 0.30 correlation threshold.

Before any final decision to discard a measurement item that did not satisfy one or several of the reliability tests mentioned above, it was carefully verified that the deletion would not affect the domain coverage of the respective constructs (thus ensuring content validity). As a result, the following rules were applied when assessing the reliability of each construct:

- Cronbach's alpha > 0.70
- Cronbach's alpha if item deleted > Cronbach's α if item is not deleted
- Item-total correlation > 0.50
- item-item correlation > 0.30

CONSTRUCT VALIDITY Construct validation is at the core of the scientific method in social sciences and relates to how well an instrument measures what it is supposed to measure (Straub et al., 2004). Construct validity is usually addressed by assessing both convergent and discriminant validity (Boudreau et al., 2001; Straub et al., 2004). These two components of construct validity verifies whether measurement items are homogeneous internally and sufficiently different from other measures.

Factorial validity is a way to assess both convergent and discriminant validity (Gefen & Straub, 2005). When factorial validity is judged to be acceptable, each measurement item correlates strongly with the construct or factor it belongs to, while correlating weakly with the other constructs/factors. Factorial validity is commonly evaluated through an Exploratory Factor Analysis (or EFA) (Gefen & Straub, 2005). Its main objectives are:

- to verify that the measurement items converge into the intended number of theoretical factors
- to ensure that each measurement item loads highly on only one factor
- that this one factor is common to the measurement items that supposedly relate to a single latent construct (Gefen & Straub, 2005).

The literature recommends various factor loading thresholds depending on the nature of the research projects (Straub et al., 2004). A common rule of thumb is that a measurement item shall load higher than .60 on its intended factor and not more than .40 in the other constructs (Gefen & Straub, 2005; Hair et al., 2009). However, increasing the factor loading threshold does not necessarily mean a more rigorous factor analysis. The adopted threshold should consider the nature of the research project itself. For example, a .60 loading threshold is usually used for confirmatory research whereas various respected researchers have indicated that acceptable thresholds for exploratory factor analysis in IS research may be 0.50 (Straub, 1989), 0.45 (Lewis, Snyder & Rainer, 1995), or even 0.35 (Lederer & Sethi, 1992). A high threshold provides greater within-factor correlation (thus improving reliability) but the opted threshold should be such that it includes as many items as possible (Lewis et al., 2005). Indeed, unassigned items explain a certain proportion of the covariance among all items that should be minimized as much as possible. Finally, the threshold should be chosen in a way that minimizes item loading on multiple factors as much as possible. Because of the overall exploratory nature of this research project, it was decided to adopt a factor loading and cross-loading threshold of .50 for factor loadings and .40 for cross-loadings, following (Straub et al., 2004) and Straub (1989)'s recommendations.

Before analyzing the results of an EFA, some tests must be computed to ensure the appropriateness of factor analysis for a particular dataset. The Bartlett test of sphericity is a means to assess the statistical significance of the correlations among at least some of the variables. It is recommended that the outcome of Bartlett's Test of Sphericity should be below the 0.05 significance level (Field, 2010; Hair et al., 2009). A second check relates to the adequacy of sample size in order to legitimately run an EFA. This is done through the Kaiser-Meyer-Olkin Measure (KMO) of Sampling Adequacy (Field, 2010; Hair et al., 2009). A KMO score of at least 0.5 is recommended in order for a factor analysis to be satisfactory.

There is variety of factor extraction models that can be categorized into two main types: common factor models or components models (Conway & Huffcutt, 2003; Costello & Osborne, 2005). By far, Principal Component Analysis (or PCA) is the most popular extraction method (Gefen & Straub, 2005; Straub et al., 2004) and is the default extraction method for most statistical packages such as SPSS or SAS. Among common factor models, maximum likelihood and principal axis factoring are

the two most popular methods (Conway & Huffcutt, 2003).

Common factor models and component models differ in terms of their overall purpose. Common factor models aim at understanding the latent variables that account for relationships among measurement items. PCA is a means to reduce the number of items by finding linear combinations that retain as much variance as possible from the original measures' variance. As a result, PCA is a rather 'mechanical' method which result does not help identifying and interpreting latent constructs (Conway & Huffcutt, 2003). In this phase of the research, EFAs are being run to validate the measurement instrument by identifying the items that properly load on their intended construct, and those that do not (according to accepted loading thresholds). There is no concern about interpreting the factors resulting from EFAs as measurement items are all assigned to a respective latent variable derived from the results of the literature review and the qualitative component of this research. Principal Component Analysis thus appeared as the most relevant extraction method.

When extracting factors, a PCA relies on one of the two types of rotation techniques: orthogonal or non-orthogonal rotation. The aim of rotation is to simplify and clarify the structure of the dataset (Costello & Osborne, 2005). The choice for an appropriate rotation technique in IS research has been based on general criteria such as simplicity (Sethi & King, 1991), interpretability (Lederer & Sethi, 1992) or amount of variance explained by the factors (Straub, 1989).

In this research, the approach used to opt for the most appropriate rotation technique was based on considerations related to the nature of the constructs themselves. When factors are theoretically speaking, expected to correlate then an oblique rotation is preferred (Fabrigar, Wegener, MacCallum & Strahan, 1999; Ford, MacCallum & Tait, 1986; Gorsuch, 1997; Lewis et al., 2005). Indeed, the listed authors point out that when factors are supposed to be correlated, an orthogonal rotation forces the factors to be orthogonal. It then produces an unrealistic factor solution with distorted loadings and a less simple data structure. In contrast, an oblique rotation provides, in such cases, a simpler structure that reflects more accurately the reality. Lewis et al. (2005) add that when factors have little or no correlation, then both orthogonal and oblique rotations will provide very close solutions with cross-factor loadings close to zero and high factor loadings. As a result, an oblique rotation technique is preferred in most cases. Orthogonal rotations techniques, such as VARIMAX, used to be recommended because of computing power constraints that do not hold anymore (Conway & Huffcutt, 2003).

In summary, this research consistently used PCA and oblique rotation (direct oblimin) for each of the EFAs that were performed. Oblique rotation was all the more relevant in this research considering that some of the socialization variables are expected to correlate. For instance, the two task-related variables, *task segregation* and *task purposefulness* as well as the two people-related factors, *mentoring* and *interaction intensity*, were expected correlate to some extent.

NORMALITY ASSESSMENT Even though PLS statistical computations are free from normality assumption constraints (Chin, 1998a), the dataset was tested for normality to provide an overall view of the collected data, as well as an understanding of how

newly derived measurement scales behaved. A common method for assessing normality is to calculate the skewness and kurtosis scores (Field, 2010). Kurtosis refers to the “peakedness” or the “flatness” (Hair et al., 2009, p. 71) of the distribution compared to the shape of a normal distribution. Skewness relates to how balanced a distribution is in terms of asymmetry. A distribution shifted to the left will provide a positive skewness score whereas right-skewed distributions have negative skewness scores. A rule of thumb is that a dataset is normally distributed if both the skewness and kurtosis scores are within the +2 to -2 range (Field, 2010; Hair et al., 2009; Tabachnick & Fidell, 2007).

All socialization variables provided satisfactory normality assessment scores with an average skewness score of .434 (within a -.731 to 1.245 range) and kurtosis of -.489 (-1.440 to 1.029 range). The results were encouraging since all the socialization-related items had been newly designed. *Social identification* items were also found to satisfy the normality constraints: average skewness/kurtosis of .555/-.379 within A -.826 to 1.239 range. *Social integration* had an average skewness/kurtosis of 1.296/1.821 with two items having kurtosis scores slightly higher than 2.

Task performance items had average skewness/kurtosis scores of 1.221/1.793. Three items scored slightly higher than 2. There was a lot more variation for the CCB measurement items. The CCB-I items had an average skewness score of 1.059 and a kurtosis of 1.421. CCB-I-6 raised concerns with a skewness score of 2.065 and kurtosis of 6.222. The item was flagged for further investigation. CCB-I-5, CCB-I-7, and CCB-I-12 had kurtosis scores ranging between 2 and 3 which can be still seen as acceptable (Field, 2010; Tabachnick & Fidell, 2007).

CCB-P items scored an average skewness/kurtosis of 1.180/1.644. CCB-P-9 had skewness/kurtosis scores of 2.283/5.967. The item was flagged for further investigation. CCB-P-12 had a kurtosis score of 4.594 and was also flagged, along with CCB-P-5, which had a kurtosis score of 3.552. CCB-P-1, CCB-P-4, and CCB-P-15 had kurtosis scores between 2 and 3. Apart from the mentioned items, all remaining items had normality assessment scores between -2 and +2. As mentioned, normal distributions of the measurement data are not required when using PLS for analysis. However, the choice to flag the items that displayed skewness and kurtosis scores well above the recommended threshold in terms of normality was simply used as a strategy to identify items that may be problematic and that may need to be modified or discarded.

PILOT DATA ANALYSIS STRATEGY Factorial validity must be examined separately for each distinct causal stage of a conceptual model (Straub et al., 2004). In other words, it is irrelevant to assess how measures may or may not crossload across different stages in a multi-stage conceptual model, as measures from two causally linked constructs are expected to correlate. Straub et al. (2004) even add that it is plausible for measurement items associated with a construct to correlate more highly with other causally related constructs than with itself (p. 24).

The conceptual models derived in this research have, respectively, two and three causal stages. In the first conceptual model, the six socialization variables are causally linked to two performance constructs: task performance and CCB. In the second model, the six socialization variables are causally related to two proximal socializa-

tion outcomes (social identification and social integration) which are in turn linked to the performance constructs. As a result, the pilot data was analyzed through a three-phase strategy in which each phase pertains to the validation of constructs from a distinct causal stage: socialization experience / proximal outcomes / performance outcomes.

The strategy followed to validate the constructs for each of the causal stages of the two conceptual models followed the same overall guidelines. Reliability was assessed through the tests and rules mentioned previously in this section, whereas construct validity was assessed through exploratory factor analyses. However, due to the specificities of each cluster of constructs as well as the limited sample size that jeopardized the appropriateness of conducting factor analyses, the detailed rules as well as the sequence of tests slightly differed from one cluster of constructs to the other. Each phase of the validation is presented in detail in the next three subsections.

6.3.6.4 *Validation of socialization experience constructs*

Because of the moderate size of the sample, careful attention had to be paid to run factor analyses only in cases when both Kaiser-Meyer-Olkin Measure (KMO) of Sampling Adequacy and Bartlett's Test of Sphericity were satisfying (Hair et al., 2009). Hair et al. (2009) recommend removing items that negatively affect the outcome of a factor analysis (thus significantly lowering the Sampling Adequacy test score). Following the authors' recommendations, it was decided to assess the reliability of the socialization constructs prior to the evaluation of construct validity through factor analysis. A measurement item that has a low correlation with other items from a same construct and that negatively affects the overall reliability of a construct scale, will not load well when running a factor analysis. The item could even load higher in other factors or even load in a separate factor by itself. The reliability tests listed in the previous subsection were then performed prior to factor analysis, so as to purify the measurement items.

RELIABILITY The reliability assessment of *task purposefulness*, *interaction intensity*, and *supportiveness* were fully satisfactory. The results of the reliability assessment of the socialization constructs are presented in Table 41. SE3 and SE6 from *task segregation* had item-total correlation scores below the 0.50 limit. An examination of the item-item correlation of the two items also revealed correlations below 0.30. Cronbach's alpha scores "if item is deleted" confirmed that both items did not perform well. They were thus dropped, resulting in a final Cronbach's alpha of 0.907 (an initial alpha of 0.765). JS2 was also dropped for identical reasons. JS6 had an item-total correlation score just below 0.50 but passed the other reliability tests, therefore the item was flagged but retained. ME6 had an item-total correlation score slightly above the 0.50 limit and had 3 item-item correlation scores just above 0.30. Since the remaining 6 items from *mentoring* were fully satisfactory, it was decided to drop ME6 since its omission did not affect the domain coverage of the construct.

EXPLORATORY FACTOR ANALYSIS After ensuring the reliability of each of the measurement scales, an EFA was run in order to assess both construct convergent and discriminant validity. The EFA computation was performed using PASW Statistics 18 (SPSS). Following the recommendations provided in the former sub-section, the EFA

Table 41: Reliability assessment - Socialization experience constructs

Task Segregation Cronbach's $\alpha = 0.765 / 0.907$ (new)			Task Purposefulness Cronbach's $\alpha = 0.914$		
Item	Item-Total Correlation	Cronbach's α if Item Deleted	Item	Item Item-Total Correlation	Cronbach's α if Item Deleted
SE1	.807	.655	TP1	.849	.873
SE2	.649	.701	TP3	.802	.890
SE3 ^r	-.440	.877	TP4	.746	.909
SE4	.795	.659	TP8	.827	.882
SE5	.719	.684			
SE6	.292	.772			
SE7	.686	.694			
Interaction Intensity Cronbach's $\alpha = 0.884$			Mentoring Cronbach's $\alpha = 0.892 / 0.900$ (new)		
Item	Item-Total Correlation	Cronbach's α if Item Deleted	Item	Item Item-Total Correlation	Cronbach's α if Item Deleted
IN2	.554	.886	ME1	.724	.874
IN3	.730	.859	ME2	.768	.866
IN4	.816	.844	ME3	.763	.868
IN5	.691	.865	ME4	.757	.868
IN7	.661	.870	ME5	.604	.886
IN8	.734	.858	ME6	.506	.900
			ME8	.747	.869
Joining Structuredness Cronbach's $\alpha = 0.776 / 0.866$ (new)			Supportiveness Cronbach's $\alpha = 0.947$		
JS1	.804	.667	SU1	.643	.957
JS2	-.042	.866	SU2	.910	.928
JS5	.719	.691	SU3	.855	.936
JS6	.437	.763	SU5	.941	.924
JS7	.631	.714	SU7	.805	.941
JS8	.708	.691	SU8	.893	.931

used PCA and oblique rotation (direct oblimin).

The EFA resulted in 6 factors with an eigenvalue greater than 1, confirming the appropriate factor structure of the data (the eigenvalues were 13.486, 3.797, 2.563, 1.855, 1.412, and 1.245). The sum of the squared loadings from the 6 extracted factors cumulatively explained 76.120% of the total variance. The KMO and Bartlett's Test Kaiser-Meyer-Olkin Measure of Sampling Adequacy were both satisfactory confirming the appropriateness of running factor analysis with the given dataset. The KMO measure was 0.629 whereas the Bartlett's test of sphericity provided support for the existence of correlations among the variables with a chi-square score of 1313.283, $df = 496$, and degree of significance of $p=0.000$ (Hair et al., 2009; Malhotra, 2004).

Only one item, ME5, loaded higher on a different factor than its targeted one, with a loading of .622 in the factor representing *supportiveness* (and a loading of .259 for its respective factor). The item was dropped. Due to the importance to reduce the overall length of the questionnaire in order to reduce the risk of low response rate inherent to FLOSS research, factors with more than 5 items were carefully considered to identify potential items that were not fully satisfactory.

IN3 and SU8 were dropped because they had the lowest loading in their respective factor and loaded close to the .4 threshold in at least another construct. Finally, IN7 had the second lowest loading (0.528) and loaded nearly up to the 0.40 limit in the mentoring-related factor. The item was retained but slightly rephrased: *Working with a lot of other community members as a newcomer has contributed to my learning*.

A new EFA was run, excluding the three items that were chosen for deletion. Six factors were extracted (KMO = 0.724, Bartlett's Test = 1087.198, $df = 406$, $p=0.000$). 76.46% of the total variance was explained by all six factors. All items had loadings higher than 0.50 in their respective factor and all cross-loadings below 0.40 except the non-modified version of IN7 with a cross-loading with Mentoring of 0.402, confirming the need to refine the item. Table 54 summarizes the findings of the final round of factor analysis.

6.3.6.5 Validation: Social identification and social integration

The procedure that was followed to validate both social identification and social integration was identical to the one implemented with the socialization constructs. The reliability of the constructs was assessed by looking at Cronbach's alphas, item-total correlation, and item-item correlation (see Table 43).

An EFA was run by including all items confirming the 2-factor structure of the constructs (1 factor for *social identification* and 1 factor for *social integration*). All items loaded highly in their intended factor (see Table 44).

Table 42: Results of EFA - Socialization constructs

CONSTRUCT/ITEM	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Task Segregation						
SE1						0.768
SE2						0.703
SE4						0.826
SE5						0.739
SE6						0.755
Task Purposefulness						
TP1			0.960			
TP3			0.672			
TP4			0.813			
TP8			0.792			
Interaction Intensity						
IN2				.713		
IN4				.853		
IN5				.562		
IN7				.527	.402	
IN8				.558		
Mentoring						
ME1					0.557	
ME2					0.734	
ME3					0.809	
ME4					0.867	
ME8					0.768	
Joining Structuredness						
JS1		.838				
JS5		.715				
JS6		.607				
JS7		.699				
JS8		.706				
Supportiveness						
SU1	.788					
SU2	.732					
SU3	.765					
SU5	.653					
SU7	.781					

Table 43: Reliability assessment - Social identification and social integration

Social Identification Cronbach's $\alpha = 0.863$			Social Integration Cronbach's $\alpha = 0.894$		
Item	Item-Total Correlation	Cronbach's α if Item Deleted	Item	Item-Total Correlation	Cronbach's α if Item Deleted
IDE ₁	.638	.842			
IDE ₂	.658	.836	INT ₁	.793	.841
IDE ₃	.641	.838	INT ₂	.687	.880
IDE ₄	.624	.841	INT ₃	.754	.867
IDE ₅	.686	.830	INT ₄	.841	.827
IDE ₆	.682	.830			

Table 44: Results of EFA - Social identification/integration

CONSTRUCT/ITEM	Factor 1	Factor 2
Social Identification		
IDE ₁		.789
IDE ₂		.903
IDE ₃		.573
IDE ₄		.545
IDE ₅		.755
IDE ₆		.745
Social Integration		
INT ₁	.810	
INT ₂	.850	
INT ₃	.782	
INT ₄	.967	

6.3.6.6 Validation of the CCB construct

Because of the particularities of the CCB construct, the approach adopted for validating the construct differed slightly from the previous pattern. From the results of the literature review and the qualitative component of this research, CCB was hypothetically conceptualized as a second-order construct whose dimensions are CCB-I, or community citizenship behaviours directed towards the benefit of individuals, and CCB-P, community citizenship behaviours directed towards the benefit of the project and its community. The analysis of the interview data along with results from the OCB literature helped derive 12 items to capture CCB-I and 15 items for CCB-P. However, because of the specificities of the FLOSS context, the measurement items had to be significantly changed to be tailored to the FLOSS context or created.

As the result, the validation of the CCB construct consisted of evaluating the dimensionality of the constructs in addition to assessing construct reliability. There had been no previous empirical examination of the nature of the CCB construct in the FLOSS context. There was thus a need to use the pilot data as a means to make a first assessment of the structural nature of the CCB construct before assessing the reliability of each dimension.

CCB DIMENSIONALITY: EXPLORATORY FACTOR ANALYSIS An initial analysis including all CCB items provided a KMO sample adequacy measure of 0.500. Following [Hair et al. \(2009\)](#)'s recommendations, it was decided to discard the items that would obviously not perform well during factor analysis. The overall item-item correlation matrix revealed that CCB-P6 and CCB-P8 had very low correlations with nearly all CCB items, and were therefore dropped.

The second round analysis indicated a KMO score of 0.612 and a successful Bartlett's Test (Chi-Square = 815.824, df = 300, $p = 0.000$). This justified the appropriateness of running EFAs to assess construct validity. Six factors with eigenvalues greater than 1.0 were generated. When running factor analysis, [Hair et al. \(2009\)](#) and [Field \(2010\)](#) recommend discarding the items that load on separate factors. This was the case for CCB-I9, CCB-I10, CCB-I11, CCB-P9, CCB-P10, CCB-P12, CCB-P13 which had high loadings in separate factors and very low loadings in any of the other factors.

The third round of EFA resulted in a clean 3-factor solution (See Table 45) with a KMO sample adequacy score of 0.758 and a satisfactory Bartlett's Test of Sphericity (Chi-Square = 572.484, df = 153, $p = 0.000$). The three factors were found to cumulatively explained 71.54% of the total variance. A semantic interpretation of the three factors is presented in the following subsection.

INTERPRETATION OF THE THREE CCB FACTORS The results of the EFA identified three factors thus contradicting the 2-dimension structure of CCB hypothesized in this research. One factor clearly pertained to the CCB-I component of CCB confirming the findings from the qualitative phase of the research project. The factor included the following items: CCB-I1, CCB-I2, CCB-I3, CCB-I4, CCB-I5, CCB-I8.

The second factor included items from both CCB-I and CCB-P: CCB-I7, CCB-I9, CCB-I10, CCB-I11, CCB-P2, CCB-P4, and CCB-P5 thus raising some concern. After consultation with various experienced FLOSS contributors, 2 IS academics, and the supervising staff, it was found that all the items shared the common idea of compliance to a project in terms of adherence to behavioural rules, norms, and objectives. The notion of compliance has a long tradition in OCB research ([C. O'Reilly & Chatman, 1986](#); [Organ & Ryan, 1995](#); [Smith et al., 1983](#)). Being considerate (CCB-I7 and CCB-I9) were derived from the OCB notion of courtesy ([Organ, 1988, 1990b](#)) which was traditionally seen as a subdimension of OCBI ([Lepine, Erez & Johnson, 2002](#); [N. P. Podsakoff, Whiting, Podsakoff & Mishra, 2011](#); [Williams & Anderson, 1991](#)). However, the analysis of the interview data revealed that such a trait is part of the behavioural requirements in FLOSS projects that is usually clearly stated in the project's code of conduct (see Chapter 4). Similarly, maintaining a positive attitude (CCB-I10) as well as treating other members with respect (CCB-I11) are also another important behavioural norms that are expected in a large majority of projects. Finally, CCB-P2,

Table 45: Results of EFA - CCB Dimensions

ITEM ID	Factor 1	Factor 2	Factor 3	ITEM STATEMENT
CCB-I1		.618		I lend a hand to members who express a need for help in project areas for which I am knowledgeable.
CCB-I2		.734		I go out of my way to help other community members with project-related problems.
CCB-I3		.856		I help community members who seek support from other members.
CCB-I4		.551		I redirect people who have a problem for which I do not know the answer to members who are in a position to help.
CCB-I5		.713		I assist newcomers in becoming familiar with the community.
CCB-I8		.472		I consult people who might be affected by my actions.
CCB-I7	.595			I pay attention to how my actions and contributions will affect other members.
CCB-I9	.895			I try to avoid creating problems for other members when contributing to the project.
CCB-I10	.806			I treat other community members with respect.
CCB-I11	.830			I maintain a positive attitude when interacting with other members.
CCB-P2	.785			I try to align my contributions with the shared goals of the project.
CCB-P4	.886			I behave in a way that is aligned with the shared values of this community.
CCB-P5	.832			I respect the behavioural rules of the community.
CCB-P1			.688	I think about the whole community when contributing to the project.
CCB-P14			.926	I promote the project to potential users.
CCB-P15			.781	I spend time talking about the project to project outsiders.
CCB-P16			.721	I actively promote the achievements of the project's community.
CCB-P17			.681	I encourage potential users to use the software produced by the community.

CCB-P4, and CCB-P5 reflected the idea that members comply with the shared values and goals of a project. This factor thus seems in line with the overall notion of CCB-P since all items reflected contributor behaviours that benefit the project by itself.

The third factor consisted of the following items: CCB-P1, CCB-P14, CCB-P15, CCB-P16, CCB-P17. A careful semantic analysis showed that all items related to behaviours that directly benefit the overall community as a whole. CCB-P1 captures the extent to

which an individual provides contributions in the best interest of the community echoing the notion of loyal boosterism (Moorman & Blakely, 1995). The remaining items, CCB-P14, CCB-P15, CCB-P16, and CCB-P17 were found to capture the same notion by addressing aspects related to the promotion and advocacy of the project to the outside world. All the items had some overlap with concepts from the OCB tradition such as organizational loyalty (Farh et al., 1997; Graham, 1991), loyal boosting (Moorman & Blakely, 1995), and spreading goodwill (George & Brief, 1992; George & Jones, 1997). In order to refer to this factor in the further analyses, the acronym CCB-C (C for community) is used.

This research compiled the past OCB literature and introduced the notion of CCB-P, community citizenship behaviours that benefit a project and its community. Drawing from the OCB literature, this research introduced the notion of CCB-P, community citizenship behaviours that benefit a project and its community. The main justification for such conceptualization was the extrapolation of the concept of OCB-O (Lepine et al., 2002; N. P. Podsakoff et al., 2011; Williams & Anderson, 1991) to the context of FLOSS communities. However, the factor analysis of the pilot study data provided some evidence about the potential distinction between citizenship behaviours directed towards the project and those towards the community itself. In order to further investigate these preliminary findings, the items of all three factors were retained for the full-scale survey. No final decision could be taken about the dimensionality of CCB at such an early stage since the results were derived from a sample of moderate size with respondents from only two FLOSS projects. The researcher thus opted for retaining all three potential dimensions and carrying the investigation through the analysis of the final data derived from a much larger sample and that covers a higher number of FLOSS communities.

RELIABILITY ASSESSMENT OF THE THREE CCB FACTORS To further investigate the three identified CCB dimensions, the reliability of each factor was assessed. The results (see Table 46) confirmed the reliability of all three dimensions, CCB-I, CCB-P, and CCB-C, with respective Cronbach's alpha scores of 0.837, 0.934, and 0.925. All item-item correlations within each dimension were above 0.30 while item-total correlations were all above the 0.50 threshold. The findings were encouraging as they provided complementary evidence about the existence of the three distinct CCB dimensions.

Table 46: Reliability assessment - CCB dimensions

CCB-I			CCB-P		
Cronbach's $\alpha = 0.837$			Cronbach's $\alpha = 0.934$		
Item	Item-Total Correlation	Cronbach's α if Item Deleted	Item	Item-Total Correlation	Cronbach's α if Item Deleted
CCB-I1	.647	.799	CCB-I7	.696	.930
CCB-I2	.580	.815	CCB-I9	.798	.920
CCB-I3	.558	.815	CCB-I10	.734	.927
CCB-I4	.693	.790	CCB-I11	.750	.925
CCB-I5	.668	.792	CCB-P2	.717	.927
CCB-I8	.525	.821	CCB-P4	.940	.905
			CCB-P5	.878	.912
CCB-C					
Cronbach's $\alpha = 0.925$					
CCB-P1	.698	.926			
CCB-P14	.873	.891			
CCB-P15	.862	.895			
CCB-P16	.781	.910			
CCB-P17	.811	.905			

6.3.6.7 Validation of task performance

The approach adopted to validate *task performance* was consistent with the one adopted when validating the socialization constructs, as well as *social identification/integration*. The results of the reliability assessment of the construct are presented in Table 47. As PER6 had several correlations lower than the .3 threshold as well as an item-total correlation of 0.403. The item was dropped. The other items satisfied all the reliability tests. After removal of PER6, the final Cronbach's alpha was found to be 0.875.

Table 47: Reliability assessment - Task performance

Task Performance		
Cronbach's $\alpha = 0.848 / 0.875$		
Item	Item-Total Correlation	Cronbach's α if Item Deleted
PER1	.766	.786
PER3	.763	.792
PER4	.742	.797
PER5	.655	.822
PER6	.403	.875

In order to assess convergent and discriminant validity, an EFA was conducted including the items of the three CCB dimensions identified in the previous subsection. Even though those dimensions were not fully validated, it was decided that the results of the EFA provided an acceptable level of support for the construct's validity.

The EFA was found to be statistically appropriate with a KMO sample adequacy score of 0.731 and Bartlett's Test of Sphericity with Chi-Square = 680.827, $df = 231$, and $p = 0.000$. The EFA resulted in four factors providing further evidence about the stability of the three CCB dimensions previously identified (see Table 48). All four *task performance* items had high loadings in the targeted factor (all above 0.50) and no cross-loading higher than 0.40 (one cross-loading was 0.331 whereas the others were all significantly below 0.30). Surprisingly, CCB-I7 did not load any more on its intended factor and had a loading of 0.484 with the factor representing *task performance*. Because the CCB dimensions had not been finalized at this stage of the research, the item was flagged for further investigation when running factor analyses with the final sample. Overall, the results of the EFA demonstrated a satisfactory level of convergent and discriminant validity for *task performance*.

6.3.7 Refinements before survey administration

After compilation of the results and analysis of the feedback provided during the pilot study, it was decided to perform several refinements on the survey instrument. The first page of the survey (consent information page) was refined (see Figure 49) based on the feedback of some of the respondents of the pilot study as well as discussions with various FLOSS team members. The layout was modified to make the page easier to read and more attractive, while the amount of content was reduced to make the page more user-friendly.

Some concern was raised regarding the use of the term 'membership' when referring to FLOSS project contributors. Some respondents pointed out that to contribute to a FLOSS project, there was no need to be given some sort of "membership card." Also, the notion of membership was found to be ambiguous in some projects such as Debian which awards the official status of Debian Developers (DD) to contributors who have been maintainers for at least 6 months and who have been through the Debian New Member Process, a formal process one has to go through to become an official Debian developer. The use of the term 'membership' was then unclear as it could be understood as addressing only official Debian Developers but not all contributors who had been active in the Debian project. After consultation with several FLOSS community leaders, it was decided to consistently refer to the notion of 'contribution' to avoid confusion and to use the terms 'established contributor' when there was need to specifically refer to a person that had become an active and regular project contributor.

A final decision was made regarding the cutoff participation duration for potential participants to be eligible for the survey. A round of discussions with experienced FLOSS contributors led to the conclusion that including individuals who joined a project only within the last 2 years was a constraint that would significantly reduce the size of the target population. The consulted FLOSS project leaders estimated the number of such new contributors in a large FLOSS community to be within the 20-50 range. The issue was first observed during the pilot study, which gathered a sample

Table 48: Results of EFA - CCB / Task Performance

ITEM ID	Factor 1	Factor 2	Factor 3	Factor 4
CCB-I1		.588		
CCB-I2		.744		
CCB-I3		.888		
CCB-I4		.518		
CCB-I5		.656		
CCB-I8		.436		
CCB-I7	.273		.484	.348
CCB-I9	.917			
CCB-I10	.720			
CCB-I11	.917			
CCB-P2	.602			
CCB-P4	.670			
CCB-P5	.730			
CCB-P1				-.702
CCB-P11				-.714
CCB-P9				-.647
CCB-P15				-.937
CCB-P10	.406			-.692
PER1			.771	
PER3			.896	
PER4			.871	
PER5			.538	

of only moderate size even though it involved two large successful FLOSS projects. After consultation with the supervising staff, it was decided that extending the cutoff duration from 2 to 3 years would be a reasonable decision which would not jeopardize the validity of the results. Adopting a 3-year cutoff standard was accepted as a reasonable choice that would help lower the risk of gathering too small a sample whose size would not permit effective statistical analysis, but which was not so long that respondents' recollections concerning the effect of their socialization phase had faded out entirely (and thus were not capturable through the survey instrument).

After integration of all the refinements, the final questionnaire was pretested with 5 experienced FLOSS practitioners who were asked to provide some feedback about the clarity of the instructions and item statements as well as overall wording, flow, and length following suggestions from [Simsek and Veiga \(2001\)](#). No concerns were raised and a new application was submitted to the Human Ethics Committee (HEC) of the School of Information Management, Victoria University of Wellington (see Ap-

Table 49: Modified layout of the consent information page

Newcomer experience and contributor behavior in FOSS communities - Survey

Thank you for accessing the questionnaire!

The goal of this research is to understand how a person's experience as a newcomer to a Free/Open Source Software (FOSS) community influences that person's behavior and contributions within that community. I am interested in hearing from people who are either technical or non-technical contributors, and who have had either positive or negative newcomer experiences.

About the survey

- This survey is **anonymous**, and no information that would identify you is being collected.
- Your participation is voluntary, and you are implying consent to participate by completing and submitting this online survey.
- You will not be able to withdraw data once you submitted the survey, but it will be possible to withdraw anytime while completing the survey by simply closing the web browser.
- It should take you around **20 minutes** to complete the survey.

Data and results

- The collected data will be released under a '**share-alike**' **Open Data Commons Open Database License** (ODbL) ensuring that the dataset can be shared (copied, used, and distributed to others), used to produce derived works, or adapted. Any data that may help identify participants will be removed before the dataset is released.
- The software used for this survey is issued under the **GNU General Public License** (GPL version 3).
- The results of this research may be deposited in the library's institutional repository or presented at conferences or published as articles in professional or academic journals. Only aggregate data will be used in any presentations or publications that result from this research.

Do not hesitate to let other people know about the study. I will post news about my progress with this research, and the results on my blog: <http://kevincarillo.org>.

If you have any questions, please contact me at Kevin.Carillo@vuw.ac.nz or my supervisors Sid Huff at Sid.Huff@vuw.ac.nz and Brenda Chawner at Brenda.Chawner@vuw.ac.nz.

Next >>

pendix A). Once the HEC committee approved the application, the full-scale survey was launched.

6.4 3RD PHASE

The main purpose of the third phase was to collect data from FLOSS community members to quantitatively test the research models. The final measurement questionnaire was administered to contributors from a selected set of FLOSS communities using a web survey. Subsequently, the scales were assessed for construct validity (including discriminant and convergent) and reliability, using commonly accepted techniques such as factor analysis and Cronbach's alpha (see section 6.3.6.3 for more explanations about the two statistical concepts).

6.4.1 Survey administration and data collection

Following the instrument pretest and the pilot test, the final version of the questionnaire was administered to FLOSS community members. The pretest, pilot test, and final survey administration were performed only once the Human Ethics Committee had been informed and had approved the conduct of this research project and each of its data collection components (see Appendix A).

Potential participants from the involved FLOSS communities were invited (via email, mailing list, or the project platform) to access a website where an online version of the questionnaire was posted. The beliefs of most FLOSS participants towards the 'free' or 'libre' nature of software (Stewart & Gosain, 2006) made it critical to use a FLOSS tool to design and administer the questionnaire to participants. The use of a proprietary

software might have generated a bias in the data by excluding participants who have strong beliefs in the FLOSS ideology. LimeSurvey⁷ is among the popular FLOSS survey applications that has received support from the FLOSS world. It was the chosen software to design the web surveys (pilot and full-scale) in this research.

The increase of interest from academia concerning the FLOSS phenomenon has caused FLOSS communities to be more and more overwhelmed by research solicitations. Establishing contacts with FLOSS communities and gaining their agreement to participate in the study had to be handled carefully to avoid low-response problems. The approach taken was to establish initial contact with leaders of the various FLOSS community selected for study, and to rely on the support of influential FLOSS experts to invite participants to take part to the survey.

As discussed earlier in the chapters, the population under investigation in this research comprised contributors to large FLOSS projects who had become contributors within the last 3 years. Because of the geographically distributed nature of FLOSS communities, the use of a web-based survey questionnaire (Dillman, Smyth & Christian, 2008) was the most suitable means for collecting quantitative data to test the theoretical models. Because of the relative complexity of the models introduced in Chapter 5 and the need to obtain results that are statistically significant, it was important to gather a sample of large size (Field, 2010; Hair et al., 2009; Hinkin, 1998). In the case of PLS analyses, Chin (1998b) recommends a minimum sample size of ten times the number of structural paths leading into a construct, giving a minimum size of 60 in the case of this research (for Model 2). Besides, a confirmatory factor analysis was used to assess construct validity and sample size of at least 200 is usually recommended (Hinkin, 1998). As a result, a sample of at least 200 responses was targeted.

All FLOSS practitioners who had been involved in this research up to that stage of the project, agreed on the necessity to target FLOSS projects in which leaders would indicate their support for the research by promoting the survey within their respective community. As noted previously, FLOSS practitioners are often reluctant to take part in research projects due to a profusion of solicitations combined with a tendency from researchers to not 'give back' their results to the FLOSS projects. About 25 large FLOSS projects were invited to participate through direct contact and emails with project leaders. 12 active large FLOSS projects agreed to take part in the research (see Table 50 for key characteristics of the projects).

⁷ <http://www.limesurvey.org>

Table 50: FLOSS project description

Project Id/Name	Creation	Project type	Organization	Commercial sponsor
1 DEBIAN (www.debian.org)	1993	Linux distribution	Debian Project	community-driven
2 FEDORA (fedoraproject.org)	2003	Linux distribution	Fedora Project	commercially-sponsored (Redhat Inc.)
3 FREEBSD (www.freebsd.org)	1993	Unix-like operating system	FreeBSD Project	community-driven
4 GENTOO (www.gentoo.org)	2002	Linux distribution	Gentoo Foundation	community-driven
5 GNOME (www.gnome.org)	1999	desktop environment	GNOME Foundation	community-driven
6 KDE (www.kde.org)	1996	desktop environment	KDE e.V.	community-driven
7 MOZILLA (www.mozilla.org)	1998	Web browser (Firefox), other software (Thunderbird, Firefox OS, Bugzilla...)	Mozilla Foundation	commercially-sponsored (Mozilla Corp.)
8 NETBSD (www.netbsd.org)	1993	Unix-like operating system	NetBSD Foundation	community-driven
9 OPENSUSE	2006	Linux distribution	openSUSE Foundation	commercially-sponsored (SUSE/Novell)
10 PYTHON (www.python.org)	1991	programming language (e.g. PyPy, CPython)	Python Software Foundation	community-driven
11 UBUNTU (www.ubuntu.com)	2004	Linux distribution	Ubuntu Foundation	commercially-sponsored (Canonical Ltd.)
12 WIKIMEDIA/ MEDIAWIKI (www.wikimedia.org www.mediawiki.org)	2002	wiki software	Wikimedia Foundation	community-driven

The researcher decided to create a personal blog containing a description of his research interests, some details about the PhD research, and a page with the survey invitation (see Figure 25). This strategy was suggested by several consulted FLOSS practitioners who were consulted throughout the research in an attempt to both create a common point of entry for all potential participants as well as to increase the researcher's online presence.

Each project contact provided advice on the best way to promote the survey to their community members. In some cases, the researcher was invited to send a survey invitation message and/or to post a message on the project's main website (see Figure 26). In some other projects, the leaders themselves promoted the survey in one or several mailing lists and/or posted messages on the project website. In addition, the researcher posted survey invitations on websites with high readership among FLOSS practitioners such as Linux.com, h-online.com, or lwn.net (see Figure 27). Finally, while the survey was being conducted, some individuals spontaneously decided

Figure 25: Survey invitation page in researcher's blog

to promote the survey on their own website or some other FLOSS-related online resource.

The survey was available online from 7 November —31 December 2012. It proved impossible to rigorously synchronize the promotion of the survey across all 12 participating FLOSS projects. This explains the rather long duration during which the survey remained open. For instance, certain online resources pertaining to a given project were updated only on a weekly basis. Similarly, in some cases, project leaders waited for the survey to be live before meeting and deciding upon a course of action to help promote the survey (sending messages to mailing lists, using social media...). As a consequence, the strategy adopted by the researcher was to allow a 2-week time-frame for each project after the survey had started being promoted within the project. Project leaders were asked to send reminders 1 week after the survey had been initially promoted.

In total, 1137 individuals accessed the online survey. 582 people (51%) filled in the first page of the survey, 406 (36%) the second page, and a total of 367 people completed the survey giving a completion rate of about 32%. Such numbers are aligned with the estimates provided in the literature about web surveys (Dillman, 2009; Holland, Smith, Stephen, Hasselback & Payne, 2010). A careful analysis of the dataset resulted in discarding some survey responses which displayed a questionable pattern. For instance, a participant answered 'strongly agree' to all Likert-scale questions. Fur-

[< Mentoring in GNOME](#)
[Reaching out Together >](#)

Newcomer Experience Survey

Did you first start contributing to GNOME after January 2010? Then you must still remember the confusion and the clarity, the friendly people and the trolls you encountered on your path. Please take [Kevin Carillo's](#) well-researched [survey](#) about your experience as a newcomer and how it shaped you as a contributor.

The collected data will be released under a 'share-alike' Open Data Commons Open Database License (ODbL) and the published materials about the research will be publicly available. The results of this research will be of great interest to everyone in Free and Open Source Software and your response would help ensure that we have a comprehensive set of data about newcomer experiences in GNOME. Thank you!

Tags: [newcomer-experience](#)

This entry was posted on Wednesday, November 7th, 2012 at 9:16 pm and is filed under [Newcomer Experience](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

Figure 26: Survey invitation in a project's 'planet'

[Home](#)
[News](#)
[Linux Community](#)
[Learn Linux](#)
[Directory](#)
[Jobs](#)
[Store](#)

[Home](#) > [News](#) > [Software](#) > [Applications](#) > **Survey: How Important is Newcomer Experience in Free, Open Source Software Projects?**

Survey: How Important is Newcomer Experience in Free, Open Source Software Projects?

Tuesday, 13 November 2012 10:02 | Kevin Carillo | Exclusive

J'aimé 3

Tweet 22

+1 12

Share

Submit

Editor's Note: This is a guest post by Kevin Carillo, a PhD student at the School of Information Management of Victoria University of Wellington in New Zealand.

Kevin Carillo is conducting a survey to find out how open source projects can attract and keep talent.

Free, open source software projects have relied on a wide array of strategies and procedures to attract new contributors. Retaining newcomers and having them become valued sustainable contributors is a much more delicate challenge. What a person experiences when he or she is a project newcomer seems to have an important impact on the kind of contributor this person will become within a project.

There has been little research about what it takes to provide a greater newcomer experience to ensure that projects keep getting quality contributors. This is what I am trying to find out in my PhD thesis.

If you have joined Debian, GNOME, Gentoo, KDE, Mozilla, NetBSD, OpenSUSE, or Ubuntu since January 2010, and are willing to help out, I would like to invite you to take part in the project's [survey](#).

Find What Works

The research project will help to answer a variety of questions such as:

- Does formal mentorship really help in generating valued contributors? What about informal mentoring?
- How important is the support of a community towards its newcomers?
- How effective are formal training processes such as those online or in-person?

Figure 27: Survey invitation in Linux.com

thermore, responses were retained only from individuals who were direct contributors (technical or non-technical) to the software project and its related artifact. Responses from individuals who were only users of the software (including those who were involved in the project forums but who did not directly contribute to the software), were discarded. 34 such responses were then removed from the final dataset. In addition, 4 individuals mentioned not being currently active in their respective FLOSS projects. Their survey responses were discarded since this research targeted active contributors (the screening section clearly indicated that one had to have been active for the last 6 months to be able to participate to the survey). Finally, the total response time of all the remaining answers was examined to identify and remove potential ‘skimmers.’ Typically, such respondents completed the survey questionnaire in a very short period of time and rated most of the items with the same score. For instance, one respondent took 3 minutes and 32 seconds to complete the survey and answered 1 to nearly all items. For comparison, the median survey response time was around 15 minutes. This led to removing 2 additional respondents and gave a final dataset containing 327 valid responses.

The size of the obtained sample was judged appropriate for performing a satisfactory statistical analysis using Partial Least Squares. A common rule of thumb when assessing size adequacy of a sample using PLS, is to have at least a sample size of ten times the maximum number of structural paths leading into a construct (Chin, 1998a, 1998b). The next section presents the analysis of the data and the findings.

6.4.2 *Dataset release under a FLOSS license*

After consultation with the supervising staff, it was thought that releasing the dataset collected during the final survey under a FLOSS-like license would be perceived positively by FLOSS communities and their contributors. This strategy was aligned with Dillman’s idea of viewing survey administration from social exchange point of view. Indeed, providing free access to the survey dataset would allow participating FLOSS projects to further analyze the data to investigate certain newcomer-related issues of their own interest. Several FLOSS practitioners and community leaders were then consulted. All of them confirmed that such practice would considerably raise the interest from FLOSS communities in taking part in the study. Furthermore, the use of FLOSS-developed survey software in analysis of the survey data, together with the release of the dataset under a FLOSS license, would be perceived by potential survey respondents as practices aligned with the core FLOSS principles.

Various FLOSS dataset licenses were considered. FLOSS practitioners familiar with the use such licenses were also consulted in order to determine which license would be the most appropriate. The Open Data Commons initiative⁸ (supported by the Open Knowledge Foundation) provides an array of licenses that are more and more used when releasing datasets under FLOSS-like licenses. It was decided to use the ODbL 1.0 license⁹ (see Appendix E for details about the terms of the license) that allows using, copying, and distributing the dataset in addition to permitting users to adapt and modify the dataset. ODbL 1.0 is a share-alike license that requires people to distribute

⁸ <http://opendatacommons.org>

⁹ <http://opendatacommons.org/licenses/odbl/summary/>

the dataset or any modified versions under ODbL 1.0.

At that stage of the research, the Human Ethics Committees of the School of Information Management and Victoria University HEC were consulted as this was an uncommon practice in social science disciplines within the university. Some recommendations were provided by both committees regarding the content of the survey consent form and survey invitation. Moreover, it was requested that data items which might allow the identification of particular survey respondents, such as certain demographics, project roles, or the comment section, would have to be deleted before the release of the dataset.

6.4.3 SEM technique : Partial least squares

Once the conceptual models had been decided upon, the most appropriate statistical techniques to analyze the collected data and test the conceptual models could be determined.

In social sciences, structural equation modeling (or SEM) was first introduced in marketing literature [Bagozzi and Yi \(1988\)](#); [Fornell and Larcker \(1981\)](#) and has become widespread over the last decade in the IS field ([Marcoulides, Chin & Saunders, 2009](#); [Ringle, Sarstedt & Straub, 2012](#); [Wetzels, 2009](#)). SEM are second generation data analysis techniques ([Bagozzi & Fornell, 1982](#)) that are considered superior to first-generation methods such as linear regression models when analyzing path models that involve latent variables measured by multiple indicators ([Gefen & Rigdon, 2011](#)).

According to classical measurement theory ([Churchill, 1979](#)), latent variables are theoretical constructs that cannot be directly measured and that are indirectly assessed through attributes that characterize them (typically measured with questionnaire scales). As a result, the measurement of a given construct is always accompanied by a measurement error ([Gefen & Rigdon, 2011](#)). A common mistake is to not recognize the distinction between a construct being measured and its associated measures leading to invalid inferences ([Gefen & Rigdon, 2011](#)). As a result, editors and reviewers in IS have encouraged researchers to use statistical techniques that recognize such distinction ([Gefen & Rigdon, 2011](#)). SEM methods have been so far the most successful and effective candidate to address such measurement errors.

There are two main SEM techniques: covariance-based (usually using LISREL or EQS for instance) and component-based techniques using the Partial Least Squares approach ([Marcoulides et al., 2009](#); [Wetzels, 2009](#)). The main objective of Partial Least Squares (or PLS) analysis is to maximize the explained variance of the dependent latent constructs. Covariance-based techniques aim at reproducing the theoretical covariance matrix but without focusing on explained variance ([Hair, Ringle & Sarstedt, 2011](#)). PLS has come under some criticism, in that some researchers have viewed it as a “silver bullet” to address empirical research challenges ([Hair et al., 2011](#); [Marcoulides et al., 2009](#)).

PLS was chosen to analyze the survey data collected in this research for several reasons. First, the overall research goal plays an important role in determining the

choice of method. PLS is recommended in research that is exploratory (or an extension of an existing structural theory) and which involves theory development (Chin, 1998a; Hair et al., 2011). In addition, PLS techniques are usually more efficient for addressing complex structural models (many constructs and many indicators) (Hair et al., 2011). Furthermore, since a number of constructs in this study had been newly designed, there was no guarantee about the normality of the collected data; PLS is recommended in such situations (Chin, Marcolin & Newsted, 2003; Hair et al., 2011; Marcoulides et al., 2009).

6.5 CHAPTER SUMMARY

This chapter outlines the three-phase methodology that was implemented during this research. Section 6.1 explained the choice for opting for an online survey. Section 6.2 summarized the first phase of this research that consisted of the collection of qualitative data in order to develop two 'testable' conceptual models. Then, section 6.3 presented in details the development and validation of the survey instrument. Finally, section 6.4 described how the collection of the final data was conducted and justified the choice to use Partial Least Square techniques to run the statistical analyses.

RESULTS

A large-scale cross-community survey was conducted to test the two theoretical models. This chapter starts by presenting demographics about the survey respondents. Then, an in-depth analysis of the results is presented concluding with the evaluation of the two alternative models.

7.1 RESPONDENT DEMOGRAPHICS

Table 51 presents the respondents demographics. The sample included a large majority of males (88.9%): 281 men and 35 women. The proportion of women is slightly higher than past FLOSS works such as Wu et al. (2007), Ke and Zhang (2010), or Xu et al. (2009) respondents of, respectively, 97%; 96%; and 91%. Recent efforts have been made by large FLOSS communities to encourage women to contribute to FLOSS projects, including the ADA initiative¹ or the Outreach Program for Women² that are sponsored by a variety of FLOSS projects or by commercial entities such as Google, Mozilla, or Red Hat. This could explain why this survey gathered a proportion of women slightly higher than surveys that were performed some years back.

Fifty-five different countries were represented within the dataset. The United States were the most represented country with 54 individuals, followed by Germany (36 respondents), India (29 respondents), France (25 respondents), and the United Kingdom (19 respondents). The most represented FLOSS project was Debian with 66 respondents, followed by GNOME (42), KDE (39), and Fedora (39). A majority of respondents were less than 35 years old (about 80%), while 29% were between 21 and 25 years old. Forty-nine percent of the respondents had no previous FLOSS experience when they joined the FLOSS project about which they reported in their survey responses, while 20% had more than 3 years of experience (the rest of the respondents reported FLOSS experience between 0 and 3 years). The time it took for respondents to become established contributors was evenly spread out ranging from less than a month to more than a year (the highest frequency being between 1 to 3 months, with 88 individuals).

The survey participants were asked to indicate the different types of tasks and roles they performed in their respective FLOSS projects. Their responses could be categorized into two main types of tasks: direct contributions to the software artifact, and contributions to the overall project and its community. Among the direct contributions, three sub-groups could be distinguished: code-related (e.g. developers, maintainers...), non-code-related (e.g. documentation, translation...), and quality assurance (e.g. bug reporting, bug triage...) tasks. Project/community contributions involved tasks such as mentoring, community advocacy, or event organization. Table 52 presents an overview of the roles performed by the respondents.

¹ <http://adainitiative.org/>

² <https://live.gnome.org/OutreachProgramForWomen>

Table 51: Respondent demographics

		NUMBER OF RESPONDENTS / FLOSS PROJECT												Tot.
		1	2	3	4	5	6	7	8	9	10	11	12	
AGE	20 or younger	1	2	2		9	10	3	1	1		12	2	42
	21- 25	15	13	1	3	14	15	11	3	4	3	9	3	94
	26- 30	15	13	6	6	8	6		2	5	3	5	2	71
	31- 35	16	7	5		5	3	3	1	3	4	1	2	50
	36- 40	9	2	1		3	1	2	3		4	3	1	29
	41- 45	4	1	1	2	1	1		2	2	1	2	1	18
	46- 50	5	1	3										9
	51 or older				1	1	1		1	1				5
EDUC. BACKGROUND	PhD	8	2		1	1	2		2	1	1	2		20
	Master's degree	29	11	6	4	5	4	5	3	3	4	3	2	79
	Postgraduate certificate or diploma	3	2	4		5	3		2	3	4	5	1	32
	Undergraduate degree	12	13	4	3	15	10	7	1	4	6	8	4	87
	Postsecondary certificate or diploma	4	4	2	1	6	3	3	2			5		30
	Secondary or high school graduate	7	7	2	3	9	15	3	3	4		6	3	62
	None			1		1						2	1	5
	Other	2						1		1		1		5
PAST FLOSS EXPERIENCE	No experience	30	15	6	6	22	28	10	2	9	7	21	5	161
	Less than 6 months	4	5	2	3	9	2	1	1		1	5		33
	Between 6 months and 1 year	4	3			4	4	3	1	1		1	1	22
	Between 1 year and 2 years	1	3	5		4	1	2		2	3	3		24
	Between 2 years and 3 years	4	5		1	1	4	2	1	1		1		20
	More than 3 years	23	8	7	2	2		2	9	3	4	1	6	67
NEWCOMER EXPERIENCE LENGTH	Less than 1 month	5	11	3		10	4	4	6	2		3		48
	Between 1 month and 3 months	13	10	5	1	17	17	7	4	2	3	8	3	88
	Between 3 months and 6 months	20	10	5	6	7	3	5	2	7	5	5	6	81
	Between 6 months and 1 year	13	5	1	1	3	9	3	2	4	2	10	2	57
	More than 1 year	14	3	6	4	4	6	1		1	5	6	1	51
Total Respondents		66	39	20	12	42	39	20	14	16	15	32	12	327

PROJECT NUMBER: 1- Debian / 2- Fedora / 3- FreeBSD / 4- Gentoo / 5- GNOME 6- KDE
 7- Mozilla / 8- NetBSD / 9- OpenSUSE / 10- Python / 11- Ubuntu/ 12- Wikimedia

Table 52: Roles/Tasks performed by respondents

7.2 DATASET CHARACTERISTICS

While it is not essential to ensure the normality of the data distributions in PLS analysis (Chin, 1998b), the skewness and kurtosis ratings were nevertheless calculated for the dataset (Field, 2010). With the exception of ME₁, which had a skewness score within the desired $-2 / 2$ range but a Kurtosis rating slightly above 3, all the other socialization items satisfied the normality tests. The average skewness rating was 0.772 while the average Kurtosis score was -.072. The same conclusion was drawn about the *social identification* items for which all ratings were within the $-2 / 2$ range. In line with the preliminary results provided during the pilot study, the skewness ratings for *social integration* were all within the acknowledged range while 3 items evidenced some Kurtosis scores between 2 and 3, which has been argued by some researchers to still be within an acceptable range (Field, 2010; Tabachnick & Fidell, 2007).

The same pattern emerged for the *task performance* items, which had showed satisfactory skewness ratings but Kurtosis scores slightly higher than 2. All the CCB items had skewness ratings within the $-2 / 2$ range (with an average rating of 1.410). Six items had Kurtosis ratings within the $-2 / 2$ range, 4 items between 2 and 3, while 7 items had Kurtosis ratings higher than 3. This result led to the conclusion that the CCB items were overall not normally distributed (Field, 2010; Tabachnick & Fidell, 2007).

When analyzing a dataset, researchers in social sciences need to be particularly wary of common method bias, or CMB (Straub, 1989; Straub et al., 2004). CMB can 'inflate' the estimates of structural parameters estimates in a model, and can potentially lead to erroneous conclusions (Gefen & Rigdon, 2011). When it is not possible to collect data using different methods or sources (P. M. Podsakoff, MacKenzie, Lee & Podsakoff, 2003), Harman's one-factor (or single-factor) test is widely used to assess the existence of CMB in a dataset (P. M. Podsakoff et al., 2003). The test consists of loading all the measurement items in an exploratory factor analysis and examining the unrotated factor solution. The rationale behind the technique is that in the presence CMB, the EFA analysis would result in a single factor solution which will account for a large proportion of the covariance among the measurement items (P. M. Podsakoff et al., 2003).

In this study, the EFA with unrotated solution resulted in a 12-factor solution in which each factor corresponded to a construct in the model or a CCB dimension (accounting for 70% of the total variance). Meanwhile, the covariance explained by a forced one-factor solution was found to be 26.49%. These results argue strongly that common method bias is not a threat to the validity of the findings (Liang, Saraf, Hu & Xue, 2007).

Finally, the dataset was tested for non-response bias (Field, 2010) in order to ensure that the sample data adequately reflected the population under study (Lewis et al., 2005). Ideally, non-response bias in a dataset should be assessed by sampling a group of non-respondents and comparing their responses with those of the respondents (Malhotra, 2004). Because it is not usually feasible to do this in most research projects, non-response bias within a dataset is usually tested using the principle of extrapolation (Armstrong & Overton, 1977). This states that subjects who respond late are more like non-respondents than are the rest of the respondents (Armstrong & Overton, 1977). Typically, researchers compare 2 subsamples of 50 to 60 responses (one of the early respondents and the other of the late respondents) and determine whether significant differences exist between the two sets of responses (Churchill, 1979; Field, 2010).

In order to assess non-response bias, the researcher created two subsamples consisting of the first and last five responses for each project. The two subsamples of 60 responses (12 times 5 responses) each were then compared using a two-tailed t-test at 5% significance level (Field, 2010). Out of the 68 measurement items, only 3 (IN7, JS6, and ID2) presented some degree of statistical difference between the two groups. A close look at the two subsamples did not reveal any significant differences between the respondent profiles. The result of this test indicated that non-response bias is not a concern for the dataset.

7.3 VALIDATION AND MEASUREMENT OF CCB AS A SECOND-ORDER CONSTRUCT

The foundations of PLS analysis originate from Herman Wold (1966) who later extended his work (Wold, 1975) to cases involving more than one block of variables (Vinzi, Trinchera & Amato, 2010). Wold's original design did not consider higher-order constructs and each latent variable had to be directly related to a set of indicators (or observed variables). Later, Lohmöller (1989) designed a procedure entitled "Hierarchical Component Model Repeated Indicators Approach" which has become the most popular approach when dealing with higher-order constructs in PLS (Chin et al., 2003; Vinzi et al., 2010). The procedure relies on the principle that "a second-order factor is directly measured by observed variables for all the first-order factors" (Chin, 2010, p. 668). Practically, it means that the manifest indicators of the associated lower order constructs are repeated in the higher-order construct (Tenenhaus, Vinzi, Chatelin & Lauro, 2005; Wetzels, 2009).

For example, suppose a second-order latent variable that consists of three underlying first-order latent variables with each having two manifest variables. The second-order latent variable is then specified using all six manifest variables (three times two indicators). As a result, the manifest variables are used twice: once for the first-order latent variables (providing "primary" loadings) and a second time for the second-order latent variable ("secondary" loadings) (Wetzels, 2009). A prerequisite for this procedure is that all the indicators of the first and second-order constructs shall be reflective (Chin, 2010) which was the case in this research. Using this approach, the parameters of a model involving one or several second-order constructs can be estimated using the standard PLS algorithm (Bassellier & Benbasat, 2004; Chin et al., 2003).

An alternative "two-step" approach has been introduced by some researchers to estimate second-order constructs in PLS (e.g. Agarwal & Karahanna, 2000; Bock, Zmud, Kim & Lee, 2005; Croteau & Bergeron, 2001). The procedure consists of initially estimating the latent variable scores in a model without the second-order constructs. Then, the provided latent variable scores are subsequently used as indicators in a model that include the second-order constructs. The first-order latent variable scores can be computed by either averaging the manifest variable scores of each latent variable (Croteau & Bergeron, 2001) or by directly using the latent variable scores provided during the first step of the PLS analysis without the second-order constructs (Agarwal & Karahanna, 2000; Bock et al., 2005). This procedure has the advantage, contrary to the Repeated Indicators method, of being able to estimate higher-order models with formative indicators (Reinartz, Krafft & Hoyer, 2004). Figure 28 summarizes the two different approaches.

This research estimated the scores for CCB (conceptualized as a second-order construct) using the "Hierarchical Component Model Repeated Indicators Approach" initially introduced by Lohmöller (1989). However, CCB scores were also estimated using the Two-Step approach in a subsequent PLS analysis to ensure that the final results of this research were stable and remained consistent independently of the method used to assess CCB scores. The comparison of the results is presented in the subsection 7.6.1.

Figure 28: PLS Model building for second-order constructs- Repeated Indicators and Two-Step approaches

7.4 MEASUREMENT MODEL EVALUATION

In PLS, a structural equation model consists of 2 models (Hair et al., 2011; Wetzels, 2009). The outer model (or measurement model) specifies the relationships between the constructs and their associated indicators. The inner model (or structural model) connects the various constructs together. PLS models are then defined by two sets of linear equations, one for the inner and one for the outer model. In PLS, the evaluation of the outer model must be first performed before an inner model can be legitimately assessed (Chin, 1998a; Haenlein & Kaplan, 2004; Henseler, Ringle & Sinkovics, 2009).

7.4.1 Criteria for measurement model evaluation

The first step consisted of an exploratory factor analysis similar to the one performed during the pilot study. This was done for several reasons. First, it made sense to follow an evaluation strategy that was consistent with the one adopted during the pilot study. Second, the novelty of the socialization constructs as well as the CCB dimensions argued for performing a particularly thorough examination of both convergent and discriminant validity. The EFA was thus performed as a preliminary assessment of construct convergent and discriminant validity. Following the EFA analyses, internal consistency reliability, indicator reliability, convergent and discriminant validity were assessed for each model separately using SmartPLS 2.0. The implemented procedures followed best practices in terms of measurement model validation (Field, 2010; Hair et al., 2009; Straub, 1989; Straub et al., 2004) as well as specific guidelines associated with PLS analysis (Gefen & Straub, 2005; Hair et al., 2011; Henseler et al., 2009; Hulland, 1999; Ringle et al., 2012). All the measures that were implemented for evaluating the two measurement models are summarized in Table 53.

Internal consistency was assessed using Cronbach’s alpha coefficients and composite reliability. A threshold of 0.70 was used to validate the reliability of measurement scales (Chin, 1998b; Fornell & Larcker, 1981; Hair et al., 2011; Nunnally, 1978;

Table 53: Outer model evaluation strategy

Test	Technique	Criteria	Sources
<i>preliminary construct convergent / discriminant validity</i>	Exploratory Factor Analysis <i>Parameters</i> Factor extraction method: Principal Component Analysis (PCA) Number of factors to retain: Eigenvalue > 1 Rotation method: oblique (Oblimin) and orthogonal (Varimax) (IBM SPSS Statistics 20)	KMO > 0.5 significance of Bartlett's test of sphericity < 0.05 item loadings > 0.50 item cross-loadings < 0.40	Conway and Huffcutt (2003); Field (2010); Hair et al. (2009); Hayton, Allen and Scarpello (2004); Lewis et al. (2005); Nunnally (1978); Straub (1989)
<i>internal consistency reliability</i>	Cronbach's α composite reliability (SmartPLS 2.0)	$\alpha > 0.70$ CR > 0.70	Chin (1998b); Fornell and Larcker (1981); Hair et al. (2011); Nunnally (1978); Werts, Linn and Jöreskog (1974); Wetzels (2009)
<i>indicator reliability</i>	Confirmatory Factor Analysis <i>Parameters</i> Bootstrap: 1000 samples (SmartPLS 2.0)	if (indicator loadings > .707 and loading significance < 0.05) then KEEP item. if (0.50 < indicator loadings < .707 and loading significance < 0.05) then DISCARD item only if discarding item increases CR and content validity not affected. if (loading < 0.50) then DISCARD item.	Gefen and Straub (2005); Hair et al. (2011); Hulland (1999); Ringle et al. (2012); Straub et al. (2004)
<i>convergent validity</i>		AVE > 0.50	Chin (1998b); Fornell and Larcker (1981); Gefen and Straub (2005); Straub et al. (2004); Wetzels (2009)
<i>discriminant validity</i>		square root of every AVE > construct correlations loadings should be higher than all of its cross loadings.	Barclay, Thompson and Higgins (1995); Chin (1998b); Fornell and Larcker (1981); Gefen and Straub (2005); Hulland (1999); Wetzels (2009)

Werts et al., 1974; Wetzels, 2009).

In PLS, individual item reliability is assessed by examining the loadings of the measurement items with their respective construct (Hair et al., 2011; Hulland, 1999). A common rule of thumb is to retain the items which loadings are greater than 0.707 (Gefen & Straub, 2005; Straub et al., 2004). The reason behind such rationale is that because a squared factor loading corresponds to the percent of variance explained by the item, a loading of 0.707 implies that 50 percent of the variance in the observed variable is due to the item (Straub et al., 2004). Caution is recommended when dealing with items with low loadings since they add little explanatory power to the model, and more importantly, they can deflate the path coefficients (Hulland, 1999; Nunnally, 1978). As a result, items with loadings of less than 0.4 or 0.5 are usually dropped by researchers (Hair et al., 2011; Hulland, 1999; Ringle et al., 2012). This research adopted a 0.5 threshold for the item loadings (bearing in mind that Chin (1998a) states that loadings of 0.6 are still acceptable).

Furthermore, researchers have encouraged careful consideration of items displaying loadings between 0.5 and 0.7, which may or may not be discarded based on several criteria (Hair et al., 2011). The rules implemented in this research are the ones suggested by Henseler et al. (2009), Hair et al. (2011), and Ringle et al. (2012) which compile recent PLS analysis best practices. The authors recommend retaining an item whose loading is between 0.5 and 0.7 only when the composite reliability of the corresponding construct substantially decreases when the item is discarded as well as when the content validity of the construct is not affected by omitting the item.

Following the assessment of indicator reliability, construct convergent validity was evaluated for each construct by looking at Average Extracted Variance (or AVE), with a threshold value of 0.50 (Chin, 1998b; Fornell & Larcker, 1981; Gefen & Straub, 2005; Straub et al., 2004; Wetzels, 2009). An AVE of 50 percent or above signifies that the corresponding latent variable explains more than half of its indicators' variance (Hair et al., 2011).

Finally, construct discriminant validity was evaluated through two measures. First, the Fornell–Larcker criterion (Fornell & Larcker, 1981) postulates that a latent construct should share more variance with its indicators than with any of the other constructs of the structural model. In statistical terms, this means that the square root of the AVE of each latent construct should be greater than the highest correlation of the construct with any other constructs (Hair et al., 2011; Hulland, 1999). The second measure regards the item cross-loadings. A rule of thumb is that an indicator's loadings with its associated construct should be higher than all its loadings with the other constructs of the model (Hair et al., 2011; Hulland, 1999). This rule was implemented in this research.

7.4.2 *Exploratory Factor Analysis*

The rules and parameters that were used to run the EFA were identical to the ones implemented during the analysis of the pilot study data:

- Run a separate EFA for each causal stage of each model (Straub et al., 2004)

- Factor extraction method: Principal Component Analysis (PCA)
- Number of factors to retain: Eigenvalue > 1
- Rotation method: Oblique (Oblimin) and Orthogonal (Varimax)
- Factor loading threshold: 0.50
- Cross-loading Threshold: 0.40

Using the preceding rules, 3 separate EFAs were conducted.

7.4.2.1 *Socialization constructs*

As shown in Table 54, the EFA fully satisfied the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's tests (KMO = 0.907 and Bartlett's Test $p = 0.000$). The factor solution resulted in 6 distinct factors, each corresponding to a socialization construct. All items excepted IN8 loaded higher than 0.50 on their targeted factor and had cross-loadings smaller than 0.40. IN8 had a loading of .458 in the EFA simulation with oblique rotation and cross-loadings below 0.40. An EFA with orthogonal rotation was performed to further investigate the issue, leading to a stable 6-factor structure with item characteristics identical to the oblique rotation simulation. IN8 had in this case a loading of .507 and a cross-loading slightly above 0.4 (0.416) with the *mentoring* factor. Consequently, the item was dropped.

7.4.2.2 *Social identification and social integration*

The EFA with oblique rotation regarding the *social identification* and *social integration* constructs (see Table 55) were fully satisfactory with a KMO score of 0.855, a Bartlett's Test significance of 0.000, and a clean 2-factor solution. All the items were retained for the next steps of the analysis.

7.4.2.3 *Dimensionality of Community Citizenship Behaviours*

The most satisfactory and promising results emerged from the EFA of the CCB items. The analysis of the pilot study data had suggested a 3-dimension structure, which differed somewhat from the prior citizenship behaviour research tradition. Previous studies by Coleman and Borman (2000) and Williams and Anderson (1991) provided substantial empirical evidence supporting the existence of just two OCB behavioural dimensions: OCB-I, or citizenship behaviours directed towards the benefit of individuals, and OCB-O, or citizenship behaviours directed towards the benefit of the organization (Byrne & Hochwarter, 2008; Eatough, Chang, Miloslavic & Johnson, 2011; Lepine et al., 2002; N. P. Podsakoff et al., 2009, 2011).

The EFA was conducted by including all CCB items as well as the *task performance* ones (see Table 56). Both Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were fully satisfactory (KMO = 0.855 and Bartlett's Test $p = 0.000$). The EFA resulted into 4 factors each corresponding to either a CCB dimension or *task performance*. All items loaded higher than 0.50 on their targeted factor and less than 0.40 on the others. Only CCB-I8 had an item loading less than 0.50. An additional EFA simulation with varimax rotation confirmed the results by identifying CCB-I8 as being the only item not satisfying the loading criteria (item loading of .444 but no cross-loading higher than 0.40). The

Table 54: Results of EFA - Socialization constructs

CONSTRUCT/ITEM	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Task Segregation						
SE1			.892			
SE2			.789			
SE4			.899			
SE5			.855			
SE6			.805			
Task Purposefulness						
TP1				.843		
TP3				.703		
TP4				.815		
TP8				.830		
Interaction Intensity						
IN2						.684
IN4						.855
IN5						.726
IN7						.769
IN8						.458
Mentoring						
ME1	.896					
ME2	.891					
ME3	.774					
ME4	.834					
ME8	.654					
Joining Structuredness						
JS1		.777				
JS5		.640				
JS6		.877				
JS7		.858				
JS8		.839				
Supportiveness						
SU1					.788	
SU2					.813	
SU3					.799	
SU5					.859	
SU7					.547	

Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .907

Bartlett's Test of Sphericity Approx. Chi-Square = 6142.297 / df = 406 / Sig = .000

Table 55: Social identification/integration

CONSTRUCT/ITEM	Factor 1	Factor 2
Social Identification		
IDE1	.846	
IDE2	.616	
IDE3	.553	
IDE4	.701	
IDE5	.717	
IDE6	.813	
Social Integration		
INT1		.844
INT2		.852
INT3		.901
INT4		.837

item was therefore discarded.

Two subsequent EFA analyses were performed: one including the items of the constructs of Model 1 and the other one with the items of all Model 2 constructs. In both cases, the results were fully satisfactory, producing the expected factor structure and acceptable item loadings. When including the items of Model 1, a 10-factor solution was found with a KMO score of 0.895 and a Bartlett's Test significance of .000. All items loaded higher than 0.50 on their intended factor with no cross-loadings higher than 0.40. In the case of Model 2, similar results were generated with a 12-factor solution (KMO = 0.887, Bartlett's Test significance = .000). ID2, ID3, and INT2 had factor loadings slightly less than 0.50 (.458, .463, .490). All the other factors had appropriate factor loadings and cross-loadings.

The findings concerning the dimensionality of citizenship behaviours which emerged from the pilot study were thus confirmed. CCB in the particular context of FLOSS communities was found to consist of three distinct behavioural dimensions. CCB-I (or citizenship behaviours directed towards the benefit of individuals) consisted of helping behaviours towards other contributors. CCB-P (or citizenship behaviours directed towards the benefit of the project) related to notions around compliance or adherence to the behavioural rules, norms, and objectives of a FLOSS project. CCB-C (or citizenship behaviours directed towards the benefit of the overall community) comprised a set of behaviours which supported the interests of the community overall, such as the promotion of the community to the outside world and to potential new contributors.

Since this research developed two alternative conceptual models in an attempt to examine factors affecting the socialization of newcomers in FLOSS communities, the measurement model evaluation procedure (see Table 53) had to be implemented twice:

Table 56: Results of EFA - CCB Dimensions

DIMENSION / CONSTRUCT NAME	ITEM ID	Factor 1	Factor 2	Factor 3	Factor 4
<i>CCB-I</i>	CCB-I1			.845	
	CCB-I2			.740	
	CCB-I3			.882	
	CCB-I4			.782	
	CCB-I5			.829	
	CCB-I8			.401	
<i>CCB-P</i>	CCB-I7	.622			
	CCB-I9	.768			
	CCB-I10	.800			
	CCB-I11	.796			
	CCB-P2	.716			
	CCB-P4	.789			
<i>CCB-C</i>	CCB-P5	.805			
	CCB-P1		.587		
	CCB-P14		.865		
	CCB-P15		.823		
	CCB-P16		.862		
TASK PERFORMANCE	CCB-P17		.825		
	PER1				.782
	PER3				.861
	PER4				.858
	PER5				.632

once for each model. The two following subsections present the results of the outer model evaluation for Model 1, then for Model 2.

7.4.3 Measurement model 1: Results

The software SmartPLS 2.0 was used to perform a confirmatory factor analysis. The assessment of internal consistency reliability, indicator reliability, and convergent validity is summarized in Table 57 and Table 58. All Cronbach's alpha coefficients and composite reliability scores were above the 0.70 threshold. The average variance extracted for each of the socialization constructs, *task performance*, and the CCB dimensions were all above 0.50. The significance of all loadings was lower than 0.05 while only three items, JS6, PE3, and CCB-P14 had loadings below the 0.707 limit (respectively 0.562, 0.683, and 0.677).

In terms of discriminant validity, both tests were satisfactory (see Tables 59, 60, and 61). The square root of every construct's AVE was greater than the construct correlations with the other latent variables. Meanwhile, all cross-loadings were significantly lower than the item loadings.

A further investigation was carried out for the three items that presented some loading issues (JS6, PER5, and CCB-P14), using the decision rules previously discussed (see Table 62). After careful consideration, JS6 was discarded because the loading was relatively low, the CR was positively affected after removal of the item and the content validity of the construct was not threatened. PER5 was retained because the loading was very close to the threshold and because it was judged that the coverage of the construct domain would be affected. Finally, CCB-P14 was retained for similar reasons. The loading was close to the limit (0.68) and the content coverage of the CCB dimension would be impacted by its removal. Finally, all the retained items did not load higher with any of the other constructs (which discarded any discriminant validity concerns), strengthening confidence in the appropriateness of the items.

7.4.4 *Measurement model 2: Results*

The evaluation of the measurement model for Model 2 was performed following an identical sequence of tests. Tables 63 and Table 64 summarize the results of the assessment of internal consistency reliability and indicator reliability. All Cronbach's alpha coefficients and composite reliability scores were satisfactory, ensuring internal consistency reliability. The AVE of all constructs was above 0.50, demonstrating the convergent validity of the latent constructs. Regarding item loadings, seven items had loadings below the 0.7 threshold: JS5, JS6, IDE1, IDE2, IDE6, PER3, and CCB-P14. It is interesting to note that JS6, PER3, and CCB-P14 encountered a similar issue while evaluating indicator reliability in Model 1. This confirmed that regardless of the model, the items manifested identical loading issues and some decisions had to be taken.

The assessment of the discriminant validity of the constructs led to fully satisfactory results (see Tables 67, 65, and 66). The square root of each construct's AVE was found to be higher than the construct's cross-correlations while all cross-loadings were found to be smaller than the loadings of the items with their associated construct.

The following decisions were taken regarding the items whose loadings were below the 0.707 limit (see Table 68). JS6 was discarded since the loading was much lower than the 0.5 minimum (Hair et al., 2011; Ringle et al., 2012). JS5 was also discarded; the researcher judged that the domain coverage of the construct was unaffected when discarding the item. IDE1, IDE2, and IDE6 were all retained because the loading of each item was very close to the 0.7 threshold, but also because the composite reliability of the construct was negatively affected by removing any of the items. PER3 was kept since the loading was close to 0.7 (0.68) and the content coverage of the dependent variable was judged to have been adversely impacted by removing it. Finally, CCB-P16 was kept for the same reasons.

Table 57: MODEL 1- consistency reliability, indicator reliability, convergent validity of social-ization constructs

Task Segregation					Task Purposefulness				
$\alpha = 0.919$ / CR = 0.939 / AVE = 0.757					$\alpha = 0.850$ / CR = 0.899 / AVE = 0.691				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
SE1	0.917	0.012	76.2607	p<0.001	TP1	0.8708	0.0182	47.768	p<0.001
SE2	0.8587	0.0201	42.6637	p<0.001	TP3	0.823	0.0276	29.890	p<0.001
SE4	0.9048	0.0142	63.92	p<0.001	TP4	0.7785	0.0364	21.436	p<0.001
SE5	0.9015	0.0144	62.748	p<0.001	TP8	0.8436	0.031	27.3411	p<0.001
SE6	0.7546	0.0342	22.1033	p<0.001					
Interaction Intensity					Mentoring				
$\alpha = 0.821$ / CR = 0.882 / AVE = 0.651					$\alpha = 0.906$ / CR = 0.930 / AVE = 0.727				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
IN2	0.7664	0.0364	21.138	p<0.001	ME1	0.7823	0.028	27.832	p<0.001
IN4	0.8219	0.0257	32.139	p<0.001	ME2	0.8988	0.021	42.929	p<0.001
IN5	0.7915	0.0365	21.744	p<0.001	ME3	0.8662	0.0245	35.3814	p<0.001
IN7	0.8387	0.023	36.491	p<0.001	ME4	0.8848	0.015	59.038	p<0.001
					ME8	0.8207	0.0319	25.721	p<0.001
Joining Structuredness					Supportiveness				
$\alpha = 0.871$ / CR = 0.903 / AVE = 0.656					$\alpha = 0.916$ / CR = 0.938 / AVE = 0.752				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
JS1	0.8632	0.0324	26.795	p<0.001	SU1	0.7645	0.0356	21.422	p<0.001
JS5	0.7501	0.0665	11.437	p<0.001	SU2	0.8465	0.0264	32.141	p<0.001
JS6	0.5621	0.0887	6.3807	p<0.001	SU3	0.8813	0.0199	44.206	p<0.001
JS7	0.9078	0.0213	42.756	p<0.001	SU5	0.9279	0.0083	112.13	p<0.001
JS8	0.8869	0.0253	35.281	p<0.001	SU7	0.9053	0.0136	66.560	p<0.001

Table 58: MODEL 1- consistency reliability, indicator reliability, convergent validity of performance constructs

Task Performance					CCB-I				
$\alpha = 0.804$ / CR = 0.873 / AVE = 0.634					$\alpha = 0.882$ / CR = 0.914 / AVE = 0.681				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
PER1	0.7661	0.0365	21.025	p<0.001	CCB-I5	0.8456	0.0263	32.238	p<0.001
PER3	0.8611	0.0213	40.446	p<0.001	CCB-I2	0.8061	0.0254	31.730	p<0.001
PER4	0.8568	0.0244	35.202	p<0.001	CCB-I4	0.8614	0.0204	42.267	p<0.001
PER5	0.6831	0.0558	12.24	p<0.001	CCB-I1	0.7867	0.0336	23.405	p<0.001
					CCB-I3	0.8205	0.0301	27.278	p<0.001
CCB-P					CCB-C				
$\alpha = 0.884$ / CR = 0.91 / AVE = 0.591					$\alpha = 0.871$ / CR = 0.908 / AVE = 0.664				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
CCB-I7	0.7626	0.0457	16.717	p<0.001	CCB-P1	0.7991	0.0397	20.175	p<0.001
CCB-I9	0.8006	0.0302	26.578	p<0.001	CCB-P14	0.6766	0.0448	15.112	p<0.001
CCB-I10	0.7724	0.037	20.939	p<0.001	CCB-P15	0.8542	0.0236	36.22	p<0.001
CCB-I11	0.7097	0.0419	16.990	p<0.001	CCB-P16	0.8565	0.0188	45.631	p<0.001
CCB-P2	0.7796	0.032	24.402	p<0.001	CCB-P17	0.8674	0.0164	52.952	p<0.001
CCB-P4	0.8111	0.037	21.924	p<0.001					
CCB-P5	0.7381	0.0422	17.334	p<0.001					

7.4.5 About the dimensionality of CCB

CCB has been hypothesized as a second-order construct, based on previous research from the organizational citizenship behaviour literature (Coleman & Borman, 2000; N. P. Podsakoff et al., 2009; Williams & Anderson, 1991). The result of the pilot study suggested the existence of three dimensions that contradicts the traditional bidimensional nature of citizenship behaviours in organizational contexts. The analysis of the final dataset confirmed the structural nature of CCB. First, an exploratory factor analysis generated a three-factor solution identical to the one found during the pilot study. Also, the final loadings of the CCB dimensions, and their statistical significance, support the argument that CCB is indeed a second-order construct (Chin, 1998a; Wetzel, 2009). While evaluating the two measurement models, the loadings of each CCB dimension were all above 0.75 and were significant with a p-value smaller than 0.001 (see Table 69). As a consequence, the results strongly confirmed the dimensionality of CCB hypothesized at the conclusion of the pilot study.

7.5 STRUCTURAL MODEL EVALUATION

7.5.1 Criteria for structural model evaluation

Once the two measurement models were evaluated, the next step involved analyzing the structural portion of the two research models (Hulland, 1999; Vinzi et al., 2010). The next subsections start with a summary of the procedures that were per-

Table 59: MODEL 1 - Cross-loadings 1

ITEM / CONSTRUCT	SE	TP	IN	ME	JS	SU	PER	CCB-I	CCB-P	CCB-C
SE1	0.92	0.41	0.34	0.27	0.15	0.35	0.22	0.29	0.22	0.26
SE2	0.86	0.40	0.36	0.25	0.24	0.33	0.20	0.25	0.21	0.22
SE4	0.91	0.40	0.30	0.24	0.13	0.31	0.27	0.23	0.21	0.23
SE5	0.90	0.41	0.31	0.26	0.21	0.35	0.25	0.25	0.24	0.25
SE6	0.75	0.32	0.26	0.19	0.08	0.26	0.14	0.23	0.15	0.13
TP1	0.35	0.87	0.37	0.32	0.18	0.41	0.35	0.34	0.33	0.27
TP3	0.47	0.78	0.32	0.18	0.08	0.28	0.26	0.29	0.33	0.27
TP4	0.38	0.85	0.34	0.27	0.19	0.36	0.39	0.25	0.23	0.17
TP8	0.29	0.82	0.27	0.26	0.2	0.37	0.33	0.25	0.29	0.21
IN2	0.27	0.35	0.77	0.29	-0.12	0.32	0.17	0.34	0.22	0.26
IN4	0.32	0.25	0.82	0.34	-0.03	0.40	0.19	0.33	0.18	0.18
IN5	0.26	0.31	0.79	0.44	0.10	0.49	0.19	0.25	0.18	0.21
IN7	0.31	0.35	0.84	0.43	0.03	0.44	0.25	0.30	0.21	0.17
ME1	0.19	0.28	0.31	0.78	0.26	0.50	0.26	0.20	0.20	0.08
ME2	0.27	0.26	0.41	0.90	0.24	0.50	0.24	0.22	0.21	0.14
ME3	0.24	0.26	0.38	0.87	0.25	0.45	0.22	0.17	0.17	0.13
ME4	0.25	0.30	0.46	0.89	0.23	0.53	0.30	0.22	0.26	0.17
ME8	0.25	0.23	0.39	0.82	0.25	0.44	0.22	0.13	0.14	0.11
JS1	0.19	0.18	-0.01	0.23	0.87	0.13	0.20	-0.01	0.11	-0.03
JS5	0.06	0.06	-0.12	0.22	0.76	-0.02	0.13	-0.09	-0.02	0.00
JS6	0.09	0.02	0.01	0.18	0.57	-0.03	0.07	-0.07	0.00	0.02
JS7	0.18	0.19	0.02	0.26	0.91	0.22	0.24	0.03	0.11	0.06
JS8	0.19	0.22	0.02	0.28	0.89	0.22	0.17	0.04	0.12	0.04
SU1	0.28	0.33	0.41	0.51	0.17	0.76	0.31	0.31	0.31	0.16
SU2	0.26	0.35	0.43	0.49	0.13	0.85	0.31	0.29	0.34	0.11
SU3	0.36	0.39	0.47	0.47	0.20	0.88	0.36	0.28	0.34	0.15
SU5	0.37	0.41	0.47	0.54	0.13	0.93	0.39	0.40	0.36	0.18
SU7	0.32	0.37	0.42	0.48	0.12	0.91	0.37	0.32	0.33	0.12

Table 60: MODEL 1 - Cross-loadings 2

Item / Construct	SE	TP	IN	ME	JS	SU	PER	CCB-I	CCB-P	CCB-C
PER1	0.17	0.29	0.10	0.15	0.23	0.22	0.77	0.29	0.35	0.16
PER3	0.17	0.35	0.19	0.25	0.15	0.37	0.86	0.32	0.39	0.25
PER4	0.16	0.28	0.18	0.25	0.20	0.37	0.86	0.31	0.39	0.13
PER5	0.30	0.35	0.31	0.27	0.13	0.29	0.68	0.25	0.32	0.18
CCB-I5	0.19	0.29	0.30	0.13	-0.04	0.27	0.36	0.85	0.33	0.31
CCB-I2	0.20	0.25	0.28	0.14	-0.02	0.27	0.28	0.81	0.35	0.45
CCB-I4	0.23	0.31	0.33	0.22	-0.01	0.31	0.28	0.86	0.36	0.36
CCB-I1	0.28	0.23	0.29	0.27	0.10	0.43	0.32	0.79	0.37	0.27
CCB-I3	0.29	0.33	0.35	0.17	-0.03	0.26	0.30	0.82	0.33	0.36
CCB-I7	0.19	0.27	0.17	0.16	0.11	0.30	0.30	0.33	0.76	0.33
CCB-I9	0.15	0.27	0.19	0.16	0.03	0.27	0.32	0.37	0.80	0.37
CCB-I10	0.15	0.24	0.20	0.18	0.08	0.29	0.35	0.32	0.78	0.32
CCB-I11	0.20	0.28	0.13	0.17	0.08	0.23	0.37	0.31	0.71	0.36
CCB-P2	0.23	0.37	0.24	0.23	0.09	0.33	0.38	0.39	0.78	0.42
CCB-P4	0.20	0.27	0.20	0.18	0.07	0.32	0.38	0.29	0.81	0.39
CCB-P5	0.19	0.21	0.18	0.19	0.06	0.36	0.34	0.25	0.73	0.29
CCB-P1	0.20	0.19	0.24	0.11	0.01	0.19	0.20	0.35	0.34	0.80
CCB-P16	0.27	0.26	0.15	0.16	0.08	0.17	0.18	0.26	0.43	0.68
CCB-P14	0.16	0.21	0.22	0.07	0.02	0.10	0.25	0.36	0.38	0.85
CCB-P17	0.20	0.19	0.19	0.13	-0.04	0.08	0.15	0.36	0.35	0.86
CCB-P15	0.23	0.27	0.23	0.15	0.05	0.15	0.16	0.40	0.39	0.87

Table 61: MODEL 1 - Fornell-Larcker criterion

Construct	SE	TP	IN	ME	JS	SU	PER	CCB-I	CCB-P	CCB-C
SE	0.870									
TP	0.449	0.831								
IN	0.363	0.391	0.807							
ME	0.282	0.311	0.461	0.852						
JS	0.189	0.195	-0.01	0.289	0.810					
SU	0.369	0.428	0.508	0.572	0.172	0.867				
PER	0.255	0.400	0.247	0.296	0.220	0.402	0.797			
CCB-I	0.287	0.341	0.379	0.225	-0.001	0.373	0.370	0.825		
CCB-P	0.243	0.355	0.245	0.236	0.101	0.388	0.455	0.422	0.769	
CCB-C	0.258	0.276	0.255	0.152	0.027	0.167	0.229	0.427	0.464	0.815

Note: The square root of AVE is shown in the diagonal.

Table 62: MODEL 1 - Item loading issues

Item	Loading	CR with item	CR without item	Content validity	Final Decision
JS6	0.5621	0.903	0.92	not affected	DROP
PER5	0.6831	0.873	0.898	affected	KEEP
CCB-P14	0.6766	0.908	0.912	affected	KEEP

Table 63: MODEL 2- Consistency reliability, indicator reliability, and convergent validity of socialization constructs

Task Segregation					Task Purposefulness				
$\alpha = 0.9218$ / CR = 0.939 / AVE = 0.755					$\alpha = 0.849$ / CR = 0.899 / AVE = 0.690				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
SE1	0.9164	0.012	76.6443	p<0.001	TP1	0.8211	0.0278	29.5745	p<0.001
SE2	0.8565	0.0206	41.6647	p<0.001	TP3	0.7771	0.0374	20.806	p<0.001
SE4	0.9041	0.0139	65.3005	p<0.001	TP4	0.8453	0.0314	26.9431	p<0.001
SE5	0.9002	0.0142	63.5214	p<0.001	TP8	0.8711	0.0189	46.1024	p<0.001
SE6	0.7525	0.0353	21.3445	p<0.001					
Interaction Intensity					Mentoring				
$\alpha = 0.822$ / CR = 0.882 / AVE = 0.652					$\alpha = 0.905$ / CR = 0.930 / AVE = 0.726				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
IN2	0.7536	0.0338	22.3071	p<0.001	ME1	0.7574	0.0324	23.3918	p<0.001
IN4	0.8362	0.0218	38.3602	p<0.001	ME2	0.9048	0.016	56.6401	p<0.001
IN5	0.8183	0.0275	29.8121	p<0.001	ME3	0.8794	0.0197	44.685	p<0.001
IN7	0.818	0.0263	31.0761	p<0.001	ME4	0.8745	0.0176	49.6701	p<0.001
					ME8	0.8367	0.0267	31.4046	p<0.001
Joining Structuredness					Supportiveness				
$\alpha = 0.874$ / CR = 0.853 / AVE = 0.562					$\alpha = 0.917$ / CR = 0.938 / AVE = 0.752				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
JS1	0.7416	0.1817	4.5178	p<0.001	SU1	0.749	0.0368	20.4325	p<0.001
JS5	0.5505	0.1972	2.8841	p<0.01	SU2	0.8533	0.0249	34.3097	p<0.001
JS6	0.3568	0.1976	1.4185	n.s.	SU3	0.8827	0.0193	45.7528	p<0.001
JS7	0.8168	0.1884	4.7993	p<0.001	SU5	0.9247	0.0087	106.279	p<0.001
JS8	0.8249	0.1883	4.8351	p<0.001	SU7	0.9095	0.0123	74.0323	p<0.001

Table 64: MODEL 2- Consistency reliability, indicator reliability, and convergent validity of remaining constructs

Social Identification					Social Integration				
$\alpha = 0.825$ / CR = 0.8721 / AVE = 0.532					$\alpha = 0.897$ / CR = 0.928 / AVE = 0.764				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
IDE1	0.6749	0.0438	15.477	p<0.001	INT1	0.8815	0.0167	52.9541	p<0.001
IDE2	0.6358	0.0482	13.2293	p<0.001	INT2	0.8696	0.0196	44.4117	p<0.001
IDE3	0.7645	0.0292	26.225	p<0.001	INT3	0.8659	0.0234	36.9949	p<0.001
IDE4	0.817	0.0211	38.7411	p<0.001	INT4	0.8759	0.0168	52.0262	p<0.001
IDE5	0.8063	0.0276	29.221	p<0.001					
IDE6	0.6503	0.0451	14.4534	p<0.001					
Task Performance					CCB-I				
$\alpha = 0.805$ / CR = 0.874 / AVE = 0.636					$\alpha = 0.883$ / CR = 0.914 / AVE = 0.681				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
PER1	0.7729	0.0383	20.2134	p<0.001	CCB-I1	0.8447	0.026	32.562	p<0.001
PER3	0.8699	0.0187	46.6543	p<0.001	CCB-I2	0.8069	0.0243	33.2204	p<0.001
					CCB-I3	0.8611	0.0209	41.2052	p<0.001
PER4	0.8617	0.0229	37.6365	p<0.001	CCB-I4	0.7845	0.0342	22.9837	p<0.001
PER5	0.6676	0.0589	11.3086	p<0.001	CCB-I5	0.8205	0.0299	27.4805	p<0.001
CCB-P					CCB-C				
$\alpha = 0.884$ / CR = 0.910 / AVE = 0.590					$\alpha = 0.871$ / CR = 0.908 / AVE = 0.664				
Item	Loading	Std. Err.	t-stat	Sig.	Item	Loading	Std. Err.	t-stat	Sig.
CCB-I7	0.7628	0.0484	15.7801	p<0.001	CCB-P1	0.8028	0.0375	21.377	p<0.001
CCB-I9	0.7994	0.0312	25.6746	p<0.001	CCB-P14	0.6767	0.045	15.0468	p<0.001
CCB-I10	0.7719	0.0362	21.3825	p<0.001	CCB-P15	0.855	0.0236	36.1768	p<0.001
CCB-I11	0.7091	0.0411	17.2949	p<0.001	CCB-P16	0.8578	0.0189	45.3764	p<0.001
CCB-P2	0.7778	0.0318	24.5201	p<0.001	CCB-P17	0.868	0.0171	50.7137	p<0.001
CCB-P4	0.8129	0.0347	23.3984	p<0.001					
CCB-P5	0.7324	0.0409	17.889	p<0.001					

Table 65: MODEL 2 - Cross-loadings 1

ITEM / CONSTRUCT	SE	TP	IN	ME	JS	SU	IDE	INT	PER	CCB-I	CCB-P	CCB-C
SE1	0.92	0.40	0.34	0.27	0.17	0.35	0.26	0.21	0.22	0.29	0.22	0.26
SE2	0.86	0.39	0.36	0.25	0.24	0.32	0.21	0.17	0.19	0.25	0.21	0.22
SE4	0.90	0.40	0.30	0.24	0.14	0.31	0.24	0.21	0.27	0.24	0.21	0.23
SE5	0.90	0.41	0.31	0.26	0.23	0.34	0.24	0.23	0.25	0.25	0.24	0.25
SE6	0.75	0.32	0.26	0.18	0.07	0.25	0.18	0.19	0.14	0.23	0.15	0.13
TP1	0.35	0.87	0.37	0.31	0.22	0.40	0.32	0.32	0.35	0.34	0.33	0.27
TP3	0.29	0.82	0.26	0.26	0.24	0.37	0.30	0.28	0.33	0.26	0.29	0.21
TP4	0.47	0.78	0.32	0.17	0.11	0.28	0.32	0.24	0.25	0.29	0.32	0.27
TP8	0.38	0.85	0.33	0.26	0.23	0.36	0.28	0.28	0.39	0.25	0.22	0.17
IN2	0.27	0.35	0.75	0.28	-0.10	0.32	0.30	0.37	0.16	0.34	0.21	0.26
IN4	0.32	0.25	0.84	0.34	0.01	0.40	0.32	0.41	0.19	0.33	0.18	0.18
IN5	0.27	0.31	0.82	0.45	0.13	0.49	0.34	0.41	0.18	0.25	0.18	0.21
IN7	0.31	0.35	0.82	0.42	0.04	0.44	0.26	0.35	0.24	0.30	0.21	0.17
ME2	0.18	0.27	0.31	0.76	0.27	0.50	0.20	0.38	0.26	0.20	0.20	0.08
ME3	0.27	0.25	0.41	0.91	0.22	0.50	0.32	0.40	0.24	0.22	0.21	0.14
ME4	0.24	0.25	0.39	0.88	0.23	0.45	0.36	0.33	0.22	0.17	0.16	0.13
ME5	0.25	0.29	0.45	0.87	0.20	0.53	0.32	0.44	0.30	0.22	0.25	0.17
ME8	0.24	0.23	0.40	0.84	0.23	0.43	0.26	0.37	0.22	0.13	0.14	0.11
JS1	0.19	0.18	-0.01	0.23	0.82	0.13	0.02	0.11	0.20	-0.01	0.10	-0.03
JS5	0.07	0.06	-0.12	0.22	0.57	-0.02	-0.06	0.00	0.13	-0.09	-0.02	0.00
JS6	0.09	0.01	0.01	0.18	0.28	-0.03	-0.08	-0.05	0.07	-0.07	0.00	0.02
JS7	0.17	0.19	0.03	0.26	0.90	0.22	0.07	0.15	0.24	0.03	0.11	0.06
JS8	0.19	0.22	0.02	0.28	0.91	0.22	0.08	0.19	0.17	0.04	0.12	0.04
SU1	0.27	0.33	0.41	0.50	0.19	0.75	0.33	0.51	0.31	0.31	0.30	0.16
SU2	0.26	0.35	0.43	0.48	0.20	0.85	0.37	0.60	0.31	0.29	0.34	0.11
SU3	0.36	0.39	0.48	0.46	0.29	0.88	0.40	0.61	0.36	0.28	0.34	0.14
SU5	0.37	0.41	0.47	0.53	0.22	0.93	0.44	0.62	0.38	0.40	0.36	0.18
SU7	0.32	0.37	0.42	0.48	0.21	0.91	0.38	0.67	0.37	0.32	0.33	0.12

Table 66: MODEL 2 - Cross-loadings 2

ITEM / CONSTRUCT	SE	TP	IN	ME	JS	SU	IDE	INT	PER	CCB-I	CCB-P	CCB-C
IDE1	0.19	0.15	0.20	0.18	-0.05	0.23	0.68	0.18	0.13	0.27	0.18	0.30
IDE2	0.23	0.23	0.26	0.18	0.05	0.21	0.64	0.28	0.23	0.33	0.30	0.37
IDE3	0.20	0.34	0.39	0.31	0.08	0.42	0.76	0.49	0.29	0.42	0.32	0.35
IDE4	0.21	0.31	0.29	0.30	0.14	0.40	0.82	0.40	0.24	0.39	0.37	0.44
IDE5	0.18	0.34	0.28	0.27	0.12	0.37	0.81	0.42	0.24	0.40	0.36	0.41
IDE6	0.12	0.16	0.19	0.26	0.03	0.27	0.65	0.18	0.17	0.18	0.14	0.19
INT1	0.23	0.37	0.49	0.44	0.23	0.64	0.48	0.88	0.42	0.45	0.37	0.22
INT2	0.17	0.24	0.46	0.43	0.12	0.59	0.41	0.87	0.42	0.40	0.44	0.26
INT3	0.19	0.26	0.35	0.34	0.20	0.57	0.34	0.87	0.35	0.28	0.45	0.19
INT4	0.23	0.31	0.38	0.36	0.18	0.63	0.43	0.88	0.43	0.36	0.45	0.22
PER1	0.18	0.30	0.09	0.15	0.23	0.22	0.14	0.33	0.77	0.29	0.35	0.16
PER3	0.17	0.35	0.19	0.25	0.18	0.37	0.29	0.43	0.87	0.32	0.39	0.25
PER4	0.16	0.28	0.18	0.25	0.21	0.37	0.25	0.42	0.86	0.31	0.39	0.13
PER5	0.30	0.35	0.30	0.27	0.11	0.29	0.29	0.27	0.67	0.25	0.32	0.18
CCB-I1	0.19	0.29	0.30	0.12	0.02	0.27	0.39	0.36	0.35	0.85	0.33	0.31
CCB-I2	0.21	0.26	0.27	0.14	0.02	0.27	0.39	0.33	0.28	0.81	0.36	0.45
CCB-I3	0.23	0.31	0.33	0.22	0.04	0.31	0.42	0.38	0.28	0.86	0.36	0.36
CCB-I4	0.29	0.23	0.29	0.26	0.13	0.42	0.34	0.41	0.32	0.79	0.37	0.27
CCB-I5	0.30	0.33	0.35	0.17	0.03	0.26	0.41	0.31	0.30	0.82	0.33	0.36
CCB-I7	0.19	0.26	0.17	0.15	0.14	0.30	0.30	0.40	0.30	0.33	0.76	0.33
CCB-I9	0.14	0.27	0.19	0.15	0.06	0.27	0.28	0.36	0.33	0.38	0.80	0.37
CCB-I10	0.15	0.24	0.20	0.18	0.12	0.29	0.29	0.31	0.35	0.33	0.77	0.32
CCB-I11	0.20	0.28	0.13	0.16	0.10	0.23	0.23	0.26	0.37	0.31	0.71	0.36
CCB-P2	0.22	0.36	0.24	0.22	0.12	0.33	0.39	0.43	0.38	0.39	0.78	0.42
CCB-P4	0.19	0.27	0.20	0.18	0.10	0.31	0.36	0.45	0.38	0.29	0.81	0.39
CCB-P5	0.18	0.20	0.18	0.18	0.12	0.36	0.28	0.40	0.34	0.25	0.73	0.29
CCB-P1	0.20	0.20	0.24	0.11	0.03	0.19	0.40	0.22	0.19	0.35	0.34	0.80
CCB-P16	0.27	0.26	0.16	0.15	0.07	0.17	0.36	0.20	0.18	0.26	0.43	0.68
CCB-P14	0.16	0.21	0.22	0.07	0.03	0.10	0.35	0.20	0.25	0.36	0.38	0.85
CCB-P17	0.20	0.19	0.19	0.13	-0.04	0.07	0.41	0.18	0.15	0.36	0.35	0.86
CCB-P15	0.23	0.27	0.23	0.15	0.05	0.15	0.44	0.24	0.16	0.40	0.39	0.87

Table 67: MODEL 2 - Fornell-Larcker criterion

Construct	SE	TP	IN	ME	JS	SU	IDE	INT	PER	CCB-I	CCB-P	CCB-C
SE	0.869											
TP	0.444	0.830										
IN	0.362	0.386	0.807									
ME	0.279	0.305	0.464	0.853								
JS	0.202	0.243	0.029	0.267	0.738							
SU	0.366	0.426	0.511	0.564	0.258	0.867						
IDE	0.262	0.368	0.380	0.348	0.102	0.446	0.729					
INT	0.234	0.338	0.484	0.452	0.209	0.695	0.477	0.807				
PER	0.251	0.399	0.239	0.290	0.228	0.400	0.308	0.461	0.797			
CCB-I	0.290	0.344	0.376	0.223	0.054	0.371	0.475	0.430	0.370	0.825		
CCB-P	0.237	0.352	0.242	0.228	0.141	0.387	0.400	0.486	0.455	0.424	0.768	
CCB-C	0.258	0.276	0.255	0.151	0.035	0.165	0.484	0.256	0.228	0.427	0.464	0.815

Note: The square root of AVE is shown in the diagonal.

Table 68: MODEL 2 - Item loading issues

Item	Loading	CR with item	CR without item	Content validity	Final Decision
JS5	0.5505	0.842	0.9034	not affected	DROP
JS6	0.3568	0.842	0.849	not affected	DROP
IDE1	0.6749	0.871	0.859	affected	KEEP
IDE2	0.6358	0.871	0.871	affected	KEEP
IDE6	0.6503	0.871	0.862	affected	KEEP
PER5	0.6831	0.874	0.899	affected	KEEP
CCB-P14	0.6767	0.908	0.912	affected	KEEP

Table 69: Loadings of CCB dimensions

CCB DIMENSION	Model 1				Model 2			
	loading	std. error	t-stat	significance	loading	std. error	t-stat	significance
CCB-I	0.762	0.0284	26.83	p<0.001	0.764	0.0301	25.36	p<0.001
CCB-P	0.836	0.0225	36.95	p<0.001	0.832	0.0214	38.84	p<0.001
CCB-C	0.772	0.0363	21.287	p<0.001	0.775	0.0353	21.96	p<0.001

formed when evaluating the two inner models, followed by the results for Model 1, then Model 2. A final subsection assesses the overall quality of the models and summarizes the findings. The procedures and tests that were implemented to evaluate the two inner models are summarized Table 70.

In PLS, the primary objective when evaluating a structural model is the maximization of variance explained in all endogenous variables (Hulland, 1999). The degree to which the variance explained in a PLS model is maximized is determined through the examination of the R^2 measures (or coefficients of determination) associated with all the dependent (or endogenous) constructs (Chin, 1998b; Hair et al., 2011; Hulland, 1999). Chin (1998a) and Henseler et al. (2009) suggest comparing R-square values to the following benchmark levels: 0.67 (substantial), 0.33 (moderate), and 0.19 (weak). These thresholds were used in this research to assess the strength of each endogenous variable's coefficient of determination.

In addition, the path coefficients generated by PLS can be interpreted as standardized beta coefficients of ordinary least squares regressions (Hair et al., 2011). Consequently, the path coefficients and their associated significance (determined through a bootstrap procedure) are used to confirm or reject the hypotheses associated with the conceptual model (Chin, 1998b; Hair et al., 2011; Hulland, 1999). Standardized paths are evaluated by examining the coefficients' sign, magnitude, and significance (Chin, 1998a; Henseler et al., 2009). Path coefficients whose signs are consistent with the associated hypothesis, provide initial empirical evidence about the existence of the theoretically assumed relationship between two constructs (Henseler et al., 2009). Furthermore, Chin (1998a) argues that standardized paths should be at least 0.20 (and ideally above 0.30) so as to be considered meaningful. The author mentions that a 0.1

Table 70: Summary of the criteria for structural model evaluation

Criteria	Definition	Critical Value	Sources
R^2	Amount of explained variance of an endogenous construct by its associated exogenous constructs	0.67 = "substantial" 0.33 = "medium" 0.19 = "weak"	(Chin, 1998b; Henseler et al., 2009)
sign, magnitude, significance of path coefficients	Strength of the relationship between constructs	sign +: path coefficients ≥ 0.1 ++: path coefficients ≥ 0.2 +++: path coefficients ≥ 0.3 *: significant at $p < 0.05$ **: significant at $p < 0.01$ ***: significant at $p < 0.001$	(Chin, 1998b; Henseler et al., 2009)
f^2	Substantial influence of an exogeneous variable on an endogeneous variable	0.02 = "small" 0.15 = "medium" 0.35 = "large"	Chin (1998a); Cohen (1988); Hulland (1999); Ringle et al. (2012)
Q^2	Predictive relevance of a structural model	$Q^2 > 0$	Chin (1998a); Fornell and Cha (1994); Hair et al. (2011); Henseler et al. (2009); Ringle et al. (2012); Tenenhaus et al. (2005)
q^2	Relative influence of the relationships in a structural model	0.02 = "small" 0.15 = "medium" 0.35 = "large"	Chin (1998a); Henseler et al. (2009); Ringle et al. (2012); Tenenhaus et al. (2005)
GoF	Global fit measure	0.02 = "small" 0.15 = "medium" 0.35 = "large"	Henseler et al. (2009); Vinzi et al. (2010); Wetzels (2009).

path coefficient may demonstrate some empirical evidence but should be interpreted with caution. Finally, bootstrapping procedures allow the generation of a significance level for each path coefficient.

Recent reviews of PLS analyses in social sciences research (Hair et al., 2011; Henseler et al., 2009; Hulland, 1999; Ringle et al., 2012) have criticized the lack of consideration of a model's predictive capability and the absence of assessment of the paths' effect size (Chin, 1998a; Hair et al., 2011; Henseler et al., 2009; Ringle et al., 2012). For instance, Ringle et al. (2012) reviewed the MIS Quarterly articles published between 1992 and 2011 that used PLS analysis techniques. The authors identified 65 such articles containing 109 structural equation model estimations. Only 13 models were found to assess effect size while none of the studies evaluated the predictive ability of the models.

Cohen's (1988) pseudo F-test (f^2 effect size) allows researchers to evaluate the incremental explanation of an independent variable on a dependent variable (Hair et al., 2011; Henseler et al., 2009). This test has been recommended when assessing a structural model using PLS (Hair et al., 2011; Henseler et al., 2009; Ringle et al., 2012). Following Cohen (1988), the effect size of each path was calculated as:

$$f^2 = \frac{R_{included}^2 - R_{excluded}^2}{1 - R_{included}^2}$$

where $R_{included}^2$ represents the coefficient of determination of an exogenous variable including all exogenous variables, and where $R_{excluded}^2$ represents the coefficient of determination of the same endogenous variable excluding a given exogenous variable.

Cohen's (1988) approach was followed when assessing the strength of each effect size. Cohen states that f^2 of 0.02, 0.15, and 0.35 respectively correspond to small, medium, and large effects (f^2 values were computed only for the significant paths generated by the PLS analysis).

Recent reviews of PLS analysis in social sciences have also strongly criticized the lack of assessment of a model's predictive relevance (Hair et al., 2011; Henseler et al., 2009; Hulland, 1999; Ringle et al., 2012). In the review of MISQ articles from Ringle et al. (2012), the authors expressed serious concern that none of the reviewed studies had assessed the predictive relevance of their structural models. Recently, researchers using PLS have been encouraged to assess model predictive relevance using Stone's (1974) and Geisser's (1975) cross-validated redundancy measure Q^2 (Hair et al., 2011; Henseler et al., 2009; Hulland, 1999; Ringle et al., 2012). Stone-Geisser's Q^2 can be measured through a blindfolding procedure (Tenenhaus et al., 2005) that can be performed with SmartPLS. The Stone-Geisser criterion relies on the principle that a model should be able to provide a prediction of the indicators of the endogenous constructs (Henseler et al., 2009). Chin (1998b) earlier remarked that:

the prediction of observables or potential observables is of much greater relevance than the estimator of what are often artificial construct-parameters.
(p. 320)

When computing cross-validated redundancy measures (Q^2), a sample reuse technique is used (also called 'blindfolding') which skips every n^{th} data point and uses the computed estimates to predict the omitted data (Hair et al., 2011). The n values have been recommended to be between 5 and 10. Overall, positive Q^2 scores indicate that a model has predictive relevance whereas a negative Q^2 means a lack of predictive relevance (Chin, 2010; Fornell & Cha, 1994; Vinzi et al., 2010). The Stone-Geisser's Q^2 test was performed in this study for evaluating the predictive relevance of the structural models.

Q^2 procedures generate two scores: cross-validated redundancy and cross-validated communality (Hair et al., 2011). Cross-validated communality scores H_j^2 (j representing an endogenous construct) measure the capacity of a path model to predict the construct indicators directly from their associated latent variable by cross-validation (Chin, 1998). Cross-validated redundancy scores F_j^2 measure the quality of the structural model (for each structural path) by taking into account the measurement model (Tenenhaus et al., 2005). F_j^2 measures assess the extent to which the indicators of a given endogenous construct are indirectly predicted from a prediction of the associated latent variable using the related structural relation, through cross-validation (Chin, 1998b; Tenenhaus et al., 2005). The F_j^2 index is also commonly referred as Stone-Geisser Q^2 (Tenenhaus et al., 2005) and is the measure used to assess the predictive relevance of a model (Chin, 2010; Henseler et al., 2009; Tenenhaus et al., 2005).

Furthermore, changes in Q^2 allow to assess the relative impact of the structural model for predicting the observed measures of an endogenous construct by the q^2 effect size (Chin, 1998a; Henseler et al., 2009; Ringle et al., 2012). q^2 measures are evaluated in analogy to the way Cohen's (1988) f^2 effect size scores are assessed. q^2 values of 0.02, 0.15, and 0.35 are associated with a small, medium, or large predictive relevance of the respective latent variable (Henseler et al., 2009). Following the recommendations, q^2 effect size was assessed for each of the hypothesized model paths:

$$q^2 = \frac{Q_{included}^2 - Q_{excluded}^2}{1 - Q_{included}^2}$$

Finally, researchers have recently encouraged the use of a global goodness-of-fit (GoF) criterion to evaluate the quality of a model in PLS (Tenenhaus et al., 2005). GoF is the geometric mean of the average communality (equal to AVE in PLS path analysis) and the average R^2 of endogenous latent variables. GoF is normed between 0 and 1, where a higher value represents better path model estimations:

$$GOF = \sqrt{R^2 * Communality}$$

Relying on the fact that communality equals average variance extracted (AVE) in PLS, and using a cut-off value of 0.5 as suggested by Fornell and Larcker (1981), GoF benchmark levels have been established by researchers to help assess model fit (Henseler et al., 2009; Vinzi et al., 2010; Wetzels, 2009). In line with the f^2 effect sizes for R^2 (small: 0.02; medium: 0.13; large: 0.26) introduced by Cohen (1988), the GoF criteria for small, medium, and large effect sizes were derived for corresponding GoF scores of 0.1, 0.25, and 0.36. These criteria were used in this research to evaluate

the Goodness-of-Fit of both conceptual models.

The next subsections summarize the results of the outer model evaluation for Model 1, then for Model 2.

7.5.2 Structural Model 1: Results

Figure 29 and Table 71 present a summary of the results generated while evaluating Model 1. Tables 72 and 73 describe the f^2 and q^2 calculations. The six socialization constructs were found to explain about 24% of the variance of *task performance* and 26% of the variance of CCB. The cross-validated redundancy scores for *task performance* and CCB were found to be positive (0.13 and 0.11), thus demonstrating the predictive relevance of the model. Finally, the overall GoF was found to be large (0.36), providing strong evidence concerning the overall quality of the model.

Task purposefulness and *supportiveness* were both found to have a significant path between 0.2 and 0.3 with *task performance* and *Community Citizenship Behaviours* (respectively 0.25 / 0.23 for *task purposefulness* and 0.24 / 0.20 for *supportiveness*). The associated hypotheses were therefore all supported. Moreover, *task segregation* was found to have a 0.11 significant path with CCB ($p < 0.05$) while *joining structuredness* was found to have a significant impact on *task performance* (0.11, $p < 0.05$). However, no significant path was found regarding the two socialization constructs: *interaction intensity* and *mentoring*.

Table 71: MODEL 1 - Results summary

Hypothesis	Path coefficient	Standard Error	T-stat	Sig. (2-tailed)	Effect size f^2	Predictive relevance q^2
$SE \rightarrow CCB$	0.1083	0.0499	2.2101	p < 0.05	0.01	0.001
$SE \rightarrow PERF$	0.0473	0.0336	0.5499			
$TP \rightarrow CCB$	0.2367	0.0637	3.6785	p < 0.001	0.05	0.01
$TP \rightarrow PERF$	0.2538	0.071	3.5492	p < 0.001	0.06	0.03
$IN \rightarrow CCB$	0.1422	0.0751	1.7925			
$IN \rightarrow PERF$	0.0582	0.0435	0.0096			
$ME \rightarrow CCB$	-0.057	0.0425	0.2544			
$ME \rightarrow PERF$	0.0612	0.046	0.8878			
$JS \rightarrow CCB$	-0.0546	0.0412	0.9064			
$JS \rightarrow PERF$	0.1178	0.0525	2.1674	p < 0.05	0.01	0.01
$SU \rightarrow CCB$	0.2002	0.0736	2.784	p < 0.01	0.03	0.01
$SU \rightarrow PERF$	0.2436	0.0757	3.1988	p < 0.01	0.04	0.02

Table 72: MODEL 1 - Cohen's (1988) pseudo F-test (f^2 effect size)

Endogenous Constructs	Exogenous Constructs Excluded								
	Original R ²	SE		TP		JS		SU	
		R ²	f ²	R ²	f ²	R ²	f ²	R ²	f ²
Task Performance	0.241			0.198	0.06	0.231	0.01	0.212	0.04
CCB	0.260	0.251	0.01	0.222	0.05			0.236	0.03

Table 73: MODEL 1 - Predictive relevance (Q^2 and q^2)

Endogenous Constructs	Exogenous Constructs Excluded								
	Original Q ²	SE		TP		JS		SU	
		Q ²	q ²	Q ²	q ²	Q ²	q ²	Q ²	q ²
Task Performance	0.13			0.10	0.03	0.12	0.00	0.11	0.02
CCB	0.11	0.10	0.001	0.09	0.01			0.09	0.01

7.5.3 Structural Model 2: Results

Figure 30 and Table 74 summarize the results of Model 2's inner model evaluation. The outcome of the f^2 and q^2 calculations are presented in Tables 75 and 76. Overall, the exogenous variables (socialization variables, social identification, and social integration) explained 28% of the variance of task performance while they explained about 44%

of the variance of CCB. By comparing those results to the coefficients of determination provided for Model 1, it appears that Model 2 did a better job at explaining the two dependent variables. Additionally, the Q^2 scores of *task performance* and CCB were both positive (0.16 and 0.17), providing positive evidence for the predictive ability of the model. Finally, Model 2's overall goodness of fit was found to be higher than Model 1 with a GoF index of about 0.433 (corresponding to a large goodness-of-fit). This result overall suggests that Model 2 provides greater explanatory power than Model 1, and supports the importance of the two socialization proximal outcomes, *social identification* and *social integration*.

Social identification was found to have a significant impact on CCB (0.39, $p < 0.001$), which supports the associated hypothesis. As well, *social integration* was found to have a significant influence on both *task performance* (0.35, $p < 0.001$) and CCB (0.24, $p < 0.001$). The effect size and predictive relevance indices provided additional support for the associated hypotheses.

Consistent with the Model 1 results, *task segregation* was found to significantly influence *task performance* (0.1, $p < 0.001$) while *task purposefulness* positively impacted *task performance* (0.25, $p < 0.001$) and CCB (0.14, $p < 0.01$). As a result, the hypotheses $SE \rightarrow CCB$, $TP \rightarrow PERF$, and $TP \rightarrow CCB$ were supported, even though some of the paths should be interpreted with caution (since the path coefficients were below 0.2) (Chin, 1998a, 2010).

Interaction intensity, *mentoring*, *joining structuredness*, and *supportiveness* were found to explain 24% of the variance of *social identification* and about 51% of the variance of *social integration*. Both constructs were found to have positive Q^2 scores (0.12 and 0.34), providing further evidence concerning the explanatory power of the exogenous latent variables.

Social identification was found to be significantly affected by *supportiveness* (0.30 with $p < 0.001$) and *interaction intensity* (0.20 with $p < 0.01$). The effect size and predictive relevance scores provided additional evidence supporting the associated hypotheses. The results showed that the same constructs, *supportiveness* (0.58 with $p < 0.001$) and *interaction intensity* (0.18 with $p < 0.01$) were positively associated with *social integration*.

Neither *mentoring* nor *joining structuredness* was found to have any significant influence on *social identification* or *social integration*, meaning that there was insufficient statistical evidence to support the associated hypotheses.

A summary of the hypothesis tests is presented in Tables 78 and 79. The effect size and predictive relevance corresponding to each of the hypotheses are indicated in the tables in order to provide additional information about the amount of support gathered when an hypothesis was supported (Hair et al., 2011; Henseler et al., 2009; Ringle et al., 2012).

7.6 SUBSEQUENT ANALYSES

Some further analyses were run to investigate the appropriateness of estimating CCB using the Repeated Items approach (Chin et al., 2003; Lohmöller, 1989; Vinzi et al.,

Figure 30: Structural Model 2

Table 74: MODEL 2 - Results summary

Hypothesis	Path coefficient	Standard Error	T-stat	Sig. (2-tailed)	Effect size f^2	Predictive relevance q^2
$SE \rightarrow CCB$	0.1059	0.0402	2.6705	$p < 0.01$	0.02	0.01
$SE \rightarrow PER$	0.0614	0.0408	1.2704			
$TP \rightarrow CCB$	0.1374	0.0527	2.5784	$p < 0.01$	0.02	0.01
$TP \rightarrow PER$	0.2474	0.0689	3.582	$p < 0.001$	0.06	0.02
$IN \rightarrow IDE$	0.1993	0.073	2.6543	$p < 0.01$	0.03	0.01
$IN \rightarrow INT$	0.1776	0.0557	3.048	$p < 0.01$	0.04	0.02
$ME \rightarrow IDE$	0.0916	0.0617	1.3669			
$ME \rightarrow INT$	0.0517	0.038	0.9437			
$JS \rightarrow IDE$	-0.0462	0.0352	0.5956			
$JS \rightarrow INT$	0.0482	0.0339	0.9634			
$SU \rightarrow IDE$	0.3045	0.0719	4.2213	$p < 0.001$	0.07	0.04
$SU \rightarrow INT$	0.5775	0.0498	11.6693	$p < 0.001$	0.40	0.19
$IDE \rightarrow CCB$	0.3894	0.0554	7.0057	$p < 0.001$	0.19	0.05
$IDE \rightarrow PER$	0.0574	0.0408	0.7345			
$INT \rightarrow CCB$	0.2403	0.0545	4.4213	$p < 0.001$	0.11	0.02
$INT \rightarrow PER$	0.3534	0.0582	6.0209	$p < 0.001$	0.07	0.06

Table 75: MODEL 2 - Cohen's (1988) pseudo F-test (f^2 effect size)

Endogenous Constructs	Exogenous Constructs Excluded												
	Original R ²	SE		TP		IN		SU		IDE		INT	
		R ²	f ²	R ²	f ²	R ²	f ²	R ²	f ²	R ²	f ²	R ²	f ²
Social Identification	0.238					0.212	0.03	0.183	0.07				
Social Integration	0.508					0.488	0.04	0.310	0.40				
Task Performance	0.282			0.239	0.06							0.202	0.11
CCB	0.441	0.432	0.02	0.428	0.02					0.337	0.19	0.400	0.07

Table 76: MODEL 2 - Predictive relevance (Q² and q²)

Endogenous Constructs	Exogenous Constructs Excluded												
	Original Q ²	SE		TP		IN		SU		IDE		INT	
		Q ²	q ²	Q ²	q ²	Q ²	q ²	Q ²	q ²	Q ²	q ²	Q ²	q ²
Social Identification	0.12					0.11	0.01	0.08	0.04				
Social Integration	0.34					0.32	0.02	0.21	0.19				
Task Performance	0.16			0.14	0.02							0.11	0.06
CCB	0.17	0.16	0.01	0.16	0.01					0.13	0.05	0.15	0.02

Table 77: Overall model quality (GoF)

MODEL 1			MODEL 2		
Endogeneous variable	AVE	R ²	Endogeneous variable	AVE	R ²
PER	0.643	0.225	IDE	0.560	0.238
CCB	0.408	0.267	INT	0.763	0.508
			PER	0.636	0.282
			CCB	0.401	0.441
Average	0.525	0.246	Average	0.518	0.362
GoF	0.360		GoF	0.433	

Table 78: MODEL 1 - Hypothesis tests

Hypothesis	Path coefficient	Sig. (2-tailed)	Effect size f^2	Predictiverelevance q^2	Conclusion
H1a $SE \rightarrow PER$					Not supported
H1b $SE \rightarrow CCB$	+	*			Supported
H2a $TP \rightarrow PER$	++	***	small	small	Supported
H2b $TP \rightarrow CCB$	++	***	small		Supported
H3a $IN \rightarrow PER$					Not supported
H3b $IN \rightarrow CCB$					Not supported
H4a $ME \rightarrow PER$					Not supported
H4b $ME \rightarrow CCB$					Not supported
H5a $JS \rightarrow PER$	+	*			Supported
H5b $JS \rightarrow CCB$					Not supported
H6a $SU \rightarrow PER$	++	**	small	small	Supported
H6b $SU \rightarrow CCB$	++	**	small		Supported

Table 79: MODEL 2 - Hypothesis tests

Hypothesis	Path coefficient	Sig. (2-tailed)	Effect size f^2	Predictiverelevance q^2	Conclusion
H1a $SE \rightarrow PER$					Not supported
H1b $SE \rightarrow CCB$	+	**	small		Supported
H2a $TP \rightarrow PER$	++	***	small	small	Supported
H2b $TP \rightarrow CCB$	+	**	small		Supported
H7a $IDE \rightarrow PER$					Not supported
H7b $IDE \rightarrow CCB$	+++	***	medium	small	Supported
H8a $INT \rightarrow PER$	+++	***	small	small	Supported
H8b $INT \rightarrow CCB$	++	***	small	small	Supported
H9a $IN \rightarrow IDE$	++	**	small		Supported
H9b $IN \rightarrow INT$	+	**	small	small	Supported
H10a $ME \rightarrow IDE$					Not supported
H10b $ME \rightarrow INT$					Not supported
H11a $JS \rightarrow IDE$					Not supported
H11b $JS \rightarrow INT$					Not supported
H12a $SU \rightarrow IDE$	+++	***	small	small	Supported
H12b $SU \rightarrow INT$	+++	***	large	medium	Supported

2010). In addition, the findings suggested the existence of several mediating effects. As a consequence, some additional analyses were performed to assess the extent to which the highlighted relationships were full mediations or only partial ones.

7.6.1 CCB estimation: Repeated items vs. Two-Step

In this research, Lohmoller's (1989) "Hierarchical Component Model Repeated Indicators Approach" was used to compute the CCB scores (conceptualized as a second-order construct). The approach consists of repeating the manifest indicators of the associated lower order constructs (the three CCB dimensions) in the CCB second-order construct (Tenenhaus et al., 2005; Wetzels, 2009). For both models, the Two-Step approach was used in a subsequent analysis to assess the CCB scores and to verify the stability and consistency of the final results. In both cases, the generated R^2 coefficients for each of the endogenous variables were very close to the ones found when using a Repeated-Items approach. In addition, the path coefficients and significance levels were almost identical to the results provided when using the Repeated Indicators strategy.

For instance, the PLS path analysis in Model 2 using the Two-Step approach provided an R^2 of 0.434 for CCB ($R^2 = 0.436$ when using the Repeated Items procedure). *Social integration* was found to be significantly related to CCB with a path of 0.219 and a significance level of $p < 0.001$ (0.241 with $p < 0.001$ when using Repeated Items). A significant path of 0.405 was also found between *Social integration* and CCB with a significance level of $p < 0.001$ (compared to 0.388 with $p < 0.001$ with the Repeated Items method). It is important to note that the path coefficient (and significance level) differences between the Repeated Items and Two-Step approaches were all of similar size for all the hypothesized relationships.

7.6.2 Full mediation vs. partial mediation

This research introduced two conceptual models that predict *task performance* and *Community Citizenship Behaviours*. The first model relies on the direct impact of six socialization variables on the two dependent variables. Model 2 introduced two variables, *social identification* and *social integration*, that were hypothesized to mediate the relationship between four socialization constructs (*interaction intensity*, *mentoring*, *supportiveness*, and *joining structuredness*) and the two performance-related dependent variables. Supplemental analyses were then run to determine whether the data supported the posited full mediation of the effects of the four socialization constructs on *task performance* and CCB by *social identification* and *social integration* (Agarwal & Karahanna, 2000; Holmbeck, 1997; Wetzels, 2009). Consequently, a third model (referred as Model 3) that included all direct and indirect paths was tested.

The results confirmed the full mediating effect of the two proximal socialization variables (see Table 80). First, the coefficient of determination of *task performance* and CCB was not much affected: from 0.282 to 0.292 for *task performance* and from 0.441 to 0.448 for CCB. The GoF index was also close to the model quality score provided in Model 2 (0.469 for Model 3 and 0.433 for Model 2). More importantly, the PLS analysis did not result in any significant path coefficient for any of the direct paths

Table 8o: Model test with direct and indirect paths

MODEL 2			MODEL 3 (direct and indirect paths)		
Endogeneous variable	AVE	R ²	Endogeneous variable	AVE	R ²
IDE	0.560	0.238	IDE	0.560	0.237
INT	0.763	0.507	INT	0.763	0.507
PER	0.636	0.282	PER	0.636	0.297
CCB	0.401	0.441	CCB	0.401	0.448
Average	0.518	0.362	Average		
GoF	0.433		GoF	0.469	

associated with the four socialization variables. These results demonstrate that the relationships between *interaction intensity* and *task performance/CCB* are fully mediated by *social identification* and *social integration* (Agarwal & Karahanna, 2000). Similarly, the relationships between *supportiveness* and *task performance/CCB* were found to be fully mediated by *social identification* and *social integration*. No conclusion could be drawn for the two remaining variables, *mentoring* and *joining structuredness*, since no significant path was found in either Model 2 or Model 3.

When assessing Model 1, a significant relationship was found between *joining structuredness* and *task performance* (0.118, $p < 0.05$). This relationship did not hold when analyzing Model 3. However, the provided path coefficient was close to the 0.1 threshold (0.096) with a p-value of 5.3% which is close to the 5% limit. The results thus suggest that there is no mediating effect between *joining structuredness* and *task performance* through *social identification* and *social integration*. Meanwhile, the findings suggest the existence of a direct effect of moderate amplitude between *joining structuredness* and *task performance*.

7.7 CHAPTER SUMMARY

The chapter started by presenting some demographics of the survey respondents. Then, dataset characteristics were evaluated followed by explanations about how the dimensionality of the CCB construct was tested. The next section presented an overview of the validation of the two measurement models. Finally, the results from the validation of the two structural models are described leading to accepting some of the hypotheses derived in this research and rejecting others. The next chapter consists of a discussion of the findings, followed by the main conclusions of the study and the limitations of the research.

DISCUSSION AND INTERPRETATION OF THE FINDINGS

8.1 INTRODUCTION

This chapter discusses and interprets the findings generated during the various stages of this research. To this end, the chapter first justifies how this research has successfully answered the main research question and its subsidiary questions. Then, the nature of FLOSS socialization and its relation with contributor performance are revisited in light of the findings. Implications for Information Systems researchers, and practitioners, are then considered. Lastly, this chapter discusses the limitations of this research as well as directions for future research.

8.2 ANSWERING THE RESEARCH QUESTION

The purpose of this study was to develop and validate a model of the socialization factors that influence contributor behaviour in the particular context of large and complex FLOSS projects. Consequently, the initial research question that guided this study was:

What socialization factors influence contributor behaviour in FLOSS communities?

In order to answer the question, there was first a need to shed some light on the FLOSS socialization phenomenon to delineate the key aspects that characterize the FLOSS socialization experience. This led to derive the first subsidiary research question of this research project:

1- What are the key factors that characterize the socialization experience of newcomers in FLOSS communities?

The review of the literature presented in Chapter 2 allowed to gain a firm understanding of the accumulated socialization literature in both the FLOSS and organizational behaviour field. In addition, Van Maanen and Schein's (1979) theory appeared as a promising theoretical tool that could be used as a guiding lens to understand better FLOSS socialization. The qualitative component of this research involving the collection and analysis of interview data from 11 experienced FLOSS practitioners helped answer the first subsidiary research question by deriving a socialization framework that consists of six distinct factors: task segregation, task purposefulness, interaction intensity, mentoring, joining structuredness, and supportiveness.

The second step of the investigation consisted of revisiting the notion of performance in the FLOSS community context by adopting a wide perspective that encompasses all contributor actions that benefit the functioning and well-being of FLOSS communities, leading to introducing two complementary dependent variables: task performance and community citizenship behaviours. The novelty of the CCB concept

led to the second sub-question:

2- What are the different instances of citizenship behaviour in the FLOSS community context?

A stream of research, organizational citizenship behaviour, appeared particularly insightful in understanding and conceptualizing the idea of extra-role behaviour in FLOSS communities. This research drew insights from the OCB tradition and introduced the notion of Community Citizenship Behaviours (CCB). The qualitative phase of the project resulted in identifying two sets of citizenship behaviours characterizing a bi-dimensional view of citizenship behaviours. Five behaviours pertain to CCB-I (or community citizenship behaviours oriented towards the benefits of other individuals) while 7 behaviours characterize CCB-P (or community citizenship behaviours oriented towards the benefits of the project and its community). The results of the pilot study raised some suspicion about the dimensionality of the CCB construct suggesting a three-dimension structure instead. The main survey confirmed the preliminary findings by identifying a third dimension, labeled CCB-C, that characterized citizenship behaviours that directly benefit the overall community as a whole.

The next step of the research project involved understanding the causal relation between the two notions addressed in the two previous subsidiary research questions: FLOSS socialization and contributor performance. The review of the literature resulted in identifying two candidate approaches: a direct influence of FLOSS community socialization experience on contributor performance, and a mediation of proximal outcomes between socialization factors and contributor behaviour. The reflection about identifying key FLOSS-specific proximal outcomes led to the following subsidiary research question:

3- What are the important proximal socialization factors in the FLOSS community context?

The review of the FLOSS and organizational behaviour literatures allowed to identify two important proximal outcomes, social identification and social integration, as potential mediators of the relationship between socialization and individual performance.

Once the proximal factors identified, this research could be carried further towards answering the main research question by decomposing the investigated phenomenon into answering two complementary subsidiary research questions:

4- What is the influence of such proximal socialization factors on contributor behaviour?

5- What socialization factors influence the identified proximal socialization factors?

The analysis of the main survey data allowed to answer the fourth subquestion. It was found that task performance is positively influenced by social integration while CCB was found to be influenced by both social identification and social integration. In addition, interaction intensity and supportiveness were both found to influence the

two proximal outcomes, thus answering the final subsidiary question. Finally, task performance was found to be directly impacted by task purposefulness, joining structuredness, and supportiveness. Meanwhile, CCB was found to be directly influenced by task segregation, task purposefulness, and supportiveness.

Consequently, it can be confidently concluded that this research has fulfilled its purpose by having successfully developed and validated two conceptual models of the socialization factors that influence contributor behaviour in FLOSS communities.

The following sections provide a more detailed discussion and interpretation of the findings of this research project.

8.3 FLOSS SOCIALIZATION UNDER THE SOCIAL SCIENTIST'S THEORETICAL MICROSCOPE

Past FLOSS research has demonstrated the importance of studying FLOSS socialization but the literature review revealed that a large majority of research projects were exploratory in nature, qualitative, and often relied on a small number of individuals from a single community (or a few of them). Additionally, past research was found to have strongly relied on the sole use of archival data drawn from software artifacts (e.g. code repositories, mailing lists...), excluding the social side of the phenomenon. Meanwhile, large and complex FLOSS projects have for long seized the importance of facilitating the inception of their newcomers. They have designed a wealth of socialization initiatives that could not benefit from insights provided by academics since past FLOSS socialization studies were overall characterized by a lack of generalizability and applicability. There was thus a need to gain a broader and more generalizable comprehension of FLOSS socialization, allowing the generation of practical recommendations to FLOSS communities to facilitate the effective integration of their newcomers.

This research developed a socialization framework by conducting a qualitative research process in which Van Maanen and Schein's theory of socialization was used as a theoretical lens (see Figure 31). From the analysis of the data, a framework consisting of six distinct factors was generated (see Figure 31). In this framework, two factors are related to the tasks performed by community newcomers (task segregation and task purposefulness), two factors pertain to interactions with other community members (interaction intensity and mentoring), and the last two (supportiveness and joining structuredness) characterize the interactions between a newcomer and the community itself.

Through the use of a card sorting procedure, an expert panel, a pilot study, and a large-scale survey, measurements for each of the socialization variables were developed and validated. All six variables demonstrated acceptable psychometric properties with satisfactory scale reliability, convergent and discriminant validity (Cronbach, 1971; Field, 2010; Fornell & Larcker, 1981).

Figure 31: FLOSS socialization framework

8.3.1 Task segregation

Task segregation was defined as *the degree to which a newcomer has performed tasks that are specifically tailored for newcomers*. The definition was derived from the results of the qualitative phase of the research that concluded that it has become a common practice for FLOSS projects to encourage their newcomers to perform special tasks that have been identified or even particularly designed to help newcomers learn. The extent to which newcomers take part to such tasks was then hypothesized to have some influence on the socialization of this individual. The final validated scale consisted of five indicators and demonstrated solid psychometric properties ($\alpha = 0.922$ / CR = 0.939 / AVE = 0.755) and successfully passed through several rounds of exploratory and confirmatory factor analysis.

8.3.2 Task purposefulness

The results generated from the interview data emphasized the discrepancy between individuals performing a sequence of project tasks with the overall objective to learn about a project and other individuals selecting tasks in a more random manner. The notion of *task purposefulness* was defined as *the degree to which the sequence of tasks performed by a newcomer has been purposefully selected by the newcomer in order to learn how to*

become a contributor. The final scale consisting of four items remained unchanged after the pretest and pilot study, having demonstrated satisfactory reliability, convergent and discriminant validity levels. The scale also proved satisfactory results during the full-scale survey ($\alpha = 0.849$ / $CR = 0.899$ / $AVE = 0.690$) leading to a scale with strong psychometric properties.

8.3.3 *Mentoring*

This research defined *mentoring* as *the degree to which a newcomer has been taken under the wing of one or more experienced members while learning how to become a project contributor*. Best exemplified by the ever growing success of the Google Summer of Code programme, formal and informal mentoring is a growing practice in the FLOSS world. In spite of the importance of the phenomenon and common acknowledgement of the effectiveness of mentoring practices, the literature review concluded that no study had so far quantitatively verified the impact of mentoring on FLOSS newcomer socialization. The final five-item scale provided satisfactory reliability, convergent and discriminant validity tests ($\alpha = 0.905$ / $CR = 0.930$ / $AVE = 0.726$) thus leading to the full validation of the scale.

8.3.4 *Interaction intensity*

This construct was defined as *the degree to which a newcomer is actively engaged with other newcomers and community members while learning how to become a project contributor*. The final measure consisted of five items that manifested acceptable levels of reliability, convergent and discriminant validity during both the pilot and final studies ($\alpha = 0.821$ / $CR = 0.882$ / $AVE = 0.651$ in the final study).

8.3.5 *Joining structuredness*

The analysis of the qualitative data emphasized the role played by the nature of the joining process and allowed to identify *joining structuredness* that pertains to *the degree to which a newcomer has to adhere to a step-by-stepwise joining process in order to become a community contributor*, as an important socialization factor. Refinements of the scale led to a final three-item measure resulted in satisfactory psychometric characteristics ($\alpha = 0.923$ / $CR = 0.923$ / $AVE = 0.814$).

8.3.6 *Supportiveness*

Finally, the amount of community support perceived by a newcomer defined as *the degree to which a newcomer has perceived a community to be supportive while learning how to become a project contributor*, emerged from the analysis of the qualitative data. The final scale included 5 items and was fully validated with coefficients of $\alpha = 0.917$, $CR = 0.958$, and $AVE = 0.752$ as determined from analysis of the final survey data.

8.4 IT'S NOT ONLY ABOUT WRITING CODE.

This research adopts the view point that performance in large and complex FLOSS projects extends beyond writing lines of code. Indeed, it was shown that a software-

centric view is too restrictive as it omits important behaviours that support the effective functioning of communities. Drawing insights from the seminal works from Daniel Katz, this research introduced two performance constructs, one specifically concentrating on task-related performance while the other notion, entitled *community citizenship behaviours*, encompasses all types of extra-role behaviours that are required for the well-being of FLOSS communities.

Task performance was defined as *the extent to which an individual carries out assigned tasks by meeting some minimal level of performance defined by the FLOSS community to which he/she belongs*. Seven indicators were initially derived inspired by Williams and Anderson (1991)'s measures and the results of the qualitative component of the project. The validation process resulted into a final four-item measure that demonstrated solid psychometric properties (see Figure 32).

The definition of community citizenship behaviours (or CCB) was inspired from Organ (1988): *behaviours that are discretionary, not directly or explicitly recognized by the formal reward system of the community and that in the aggregate promote the effective functioning of the community*. The review of the FLOSS literature as well as results from the interview phase confirmed the appropriateness of the CCB construct and also the relevance of two CCB dimensions: CCB-I (or community citizenship behaviours directed towards project members) and CCB-P (or community citizenship behaviours directed towards the project and its community as a whole). For each CCB dimension, the generation of the measurement items was performed using a two-phase procedure. First, items from existing scales for each dimension were carefully analyzed in line with the findings of the qualitative phase of the project. Second, new items were generated from the findings of the qualitative component of the research. The initial pool of items resulted in 11 CCB-I and 17 CCB-P items.

The dimensionality of the CCB construct was first explored during the pilot study, which suggested a three-dimension structure that contradicted the literature. Several rounds of EFA resulted in retaining 18 items, of which six belonged to CCB-I, seven to CCB-P (in relation to behaviours reflecting the compliance to compliance to a project's behavioural rules, norms, and objectives), and five to a new dimension labelled CCB-C capturing behaviours that directly benefit the overall community as a whole. Since the results of the pilot study were derived from a sample of modest size consisting of contributors from only two large FLOSS communities, the 18 items from the three identified dimensions were retained for the final study. One CCB-I item was then dropped during the EFA simulation since its factor loading was below the set threshold.

The analysis of the data collected during the final survey confirmed the 3-dimension structure of the CCB construct. Both EFA and CFA analyses indicated that CCB in the FLOSS community context comprised three distinct dimensions: CCB-I, CCB-P, CCB-C (see Figure 32). The results have a range of theoretical and practical implications that are discussed below.

Figure 32: Performance-related measures

8.5 MODELLING THE SOCIALIZATION PHENOMENON

The development of two FLOSS socialization conceptual models provides a deeper comprehension of the socialization phenomenon and its relation to performance. The first model hypothesizes a direct relation between socialization factors and the two performance measures. It explained about 24% of the variance in *task performance* and 26% of the variance in *community citizenship behaviours*. Following recent efforts to standardize PLS analysis and encouragements to use a global criterion to assess model quality (Henseler et al., 2009; Tenenhaus et al., 2005), the goodness-of-fit (GoF) score was computed for both models. Model 1 was found to have a GoF of 0.36 that corresponds to « large » effect sizes of R^2 .

The second model examines the impact of socialization factors on performance as mediated by proximal socialization outcomes. In this research, two proximal outcomes were introduced in light of findings from the review of the FLOSS-related literature. *Social identification* was defined as *the extent to which one identifies with a FLOSS community, leading to viewing the community's successes and failures as one's own* (inspired by Mael and Ashforth, 1992). Mael and Ashforth's measure was tailored to the context of this study and used to capture the construct. The measure demonstrated satisfactory psychometric properties ($\alpha = 0.825$ / $CR = 0.8721$ / $AVE = 0.532$), confirming the applicability of the measure to the FLOSS context.

Social integration was defined as *the extent to which an individual perceives himself/herself to be trusted and accepted by the other FLOSS community contributors*. The measure was adapted from Wanberg and Kammeyer-Mueller (2000) and provided satisfactory results during both the pilot and the final studies ($\alpha = 0.897$ / $CR = 0.928$ / $AVE = 0.764$). Again, this confirms the relevance and usability of the measure in the FLOSS

community context.

In Model 2, about 29% of the variance in *task performance* was explained by the hypothesized paths, while the model explained 44% of the variance in CCB. The GoF score was found to characterize large effect sizes, with an R^2 value of 0.433, indicating good predictive power of the model. As a consequence, the results demonstrated that Model 2 was a slightly better model in terms of predicting *task performance*, and a much stronger model in predicting CCB. This confirmed the superiority of Model 2 over Model 1. Figure 33 presents a summary of the final results.

8.6 YES, FLOSS SOCIALIZATION MATTERS.

Through the test of the two models using a web-based survey involving 12 large and complex FLOSS projects, this research has provided empirical evidence about the causal link that exists between newcomer socialization and performance. The nature of the link is investigated through the parallel analysis of Model 1 (direct effect) versus Model 2 (indirect effect) results. The findings overall confirms the relevance of studying FLOSS socialization as solid empirical evidence has demonstrated its impact on the performance of FLOSS contributors.

8.6.1 Predictors of task performance

In Model 1, *task performance* was found to be positively impacted by *task purposefulness* (path coefficient of 0.279 with $p < 0.001$), *supportiveness* (0.239 with $p < 0.01$), and *joining structuredness* (0.114 with $p < 0.05$), thus supporting the three corresponding hypotheses. In Model 2, the same result was found with *task purposefulness* (path coefficient of 0.244 with $p < 0.001$) while *social integration* was also found to positively influence *task performance* (0.348, $p < 0.001$).

In Model 2, *social integration* was thus found to be the strongest predictor of *task performance*, confirming earlier socialization studies such as Bauer et al. (2007), J. Chen and Eldridge (2011), and Kammeyer-Mueller and Wanberg (2003) who all demonstrated the positive impact of *social integration* on distal socialization outcomes such as performance. The result then allows to assert that the effect of *social integration* on *task performance* holds in the FLOSS community context. The integration of newcomers within FLOSS communities is then proven to be of high importance when trying to have project newcomers become better achievers in terms of direct project contributions.

Task purposefulness was found to be the second best predictor. The result suggests that even though individuals in FLOSS communities are generally free to perform any task they want, the structuring of tasks in a way that optimizes learning is expected to increase newcomer performance. FLOSS communities should then encourage their newcomers to perform project tasks following a logic that supports their learning about the project. In addition, dedicated resources shall provide guidance to newcomers about the order and types of tasks that shall be performed.

Figure 33: Summary of the results

*. $p < 0.05$ / **. $p < 0.01$ / ***. $p < 0.001$

Finally, the analysis of the main survey data revealed the existence of a positive relationship between *joining structuredness* and *task performance*. The finding confirms that the ritualization of newcomer socialization has some impact on the performance of individuals (Trice & Beyer, 1984; Van Maanen, 1973). In the same way that ritual processes have an effect on the effectiveness of a new recruit's socialization within organizations (Islam & Zyphur, 2009), the use of a formal and structured joining process within a FLOSS community was found to directly impact the task-related performance of contributors.

8.6.2 Predictors of Community Citizenship Behaviours

This research identified five predictors of CCB. The results generated by the testing of Model 1 emphasized the positive influence of *task purposefulness* (path coefficient of 0.245 with $p < 0.001$), *supportiveness* (0.200 with $p < 0.01$), and *task segregation* (0.116 with $p < 0.01$). The analysis of the collected data in light of the structural characteristics of Model 2 confirmed the effect of *task purposefulness* (path coefficient of 0.147 with $p < 0.01$) and *task segregation* (0.112 with $p < 0.01$) while *social identification* and *social integration* were both found to have a positive influence (with path coefficients of respectively 0.362/0.275 with $p < 0.001$ for both paths).

The two task-related factors, *task purposefulness* and *task segregation* were found to positively influence CCB. These results strongly suggest that the nature of the tasks performed as well as the structure governing the links among them play an important role in shaping contributors in a way that benefits the well-being of a FLOSS community. This argues that practices such as the GnomeLove initiative¹ that help newcomers identify tasks that are most suitable should be encouraged. Meanwhile, structuring the sequence of tasks performed by newcomers in an attempt to optimize learning was found to not only directly influence project-related performance but also the additional behaviours that help FLOSS communities to function.

The results of this research are consistent with past research that has demonstrated the impact of organizational support on work outcomes in organizational settings (Rhoades, Eisenberger & Armeli, 2001) as well as in FLOSS-related production (Carillo & Okoli, 2011). This research has identified a positive causal relation between the community support perceived by newcomers and community citizenship behaviours. This finding opens several research avenues about the importance of supportiveness in organizational settings, and in particular in FLOSS and other online communities.

Both socialization proximal outcomes were shown to be strong predictors of CCB. *Social integration*, or the feeling of acceptance of a newcomer within a given FLOSS community, was found to strongly predict the extent to which newcomers will become good community citizens. Combined with the effect of *social integration* on *task performance*, this finding strongly echoes considerations about the crucial importance of social integration in retaining FLOSS contributors (von Hippel & Von Krogh, 2003) but also the integration of newcomers in digitally enabled communities (Kraut, Riedl & Burke, 2011). This is the first time, to our knowledge, that such a finding has emerged from empirical data, and encourages future research to further explore the

¹ <https://live.gnome.org/GnomeLove>

key role played by social integration when assimilating new recruits into a social group.

Past FLOSS research has demonstrated the influence of social identification on various measures of performance in FLOSS projects such as participation (Bagozzi & Dholakia, 2006), contributor task performance (Ke & Zhang, 2009), or team effectiveness (Stewart & Gosain, 2006). This research project adds to this body of knowledge by making explicit the influence of the construct on CCB. This research also extends results from past socialization research that found social identification to be one of the influential proximal outcomes in organizational socialization (e.g. Saks & Ashforth, 1997a).

8.6.3 Predictors of social identification

The socialization tradition has demonstrated the important role played by social identification within an organization in influencing distal outcomes such as job performance or OCB (Ashforth & Mael, 1989; Saks & Ashforth, 1997b). In addition, past FLOSS research has also highlighted the importance of engendering a sense of social identify among FLOSS community members as it positively influences participation and contributions in FLOSS projects (Bagozzi et al., 2007; Dholakia et al., 2004; Liu & Iyer, 2007).

The statistical analysis performed in the assessment of Model 2 demonstrated the importance of developing a sense of identification with a FLOSS community in order to generate citizen-like behaviours from its new contributors. The findings suggest that FLOSS communities should attempt to influence each new member's socialization experience in an attempt to nurture such feelings of identification. The six socialization factors introduced in this research explained about 24% of the variance of the *social identification* construct while *interaction intensity* (path coefficient of 0.186 with $p < 0.01$) and *supportiveness* (0.300 with $p < 0.001$) were found to have a significant impact, thus confirming the associated hypotheses.

Due to the size of the effect of *supportiveness* on *social identification*, the finding strongly encourages FLOSS communities to be especially supportive during the socialization of new contributors. The result is consistent with past socialization research that emphasized the role played by investiture tactics on the related concept of organizational commitment (N. J. Allen & Meyer, 1990a; Cooper-Thomas et al., 2004; Simosi, 2010). Meanwhile, the amount of project learning-related social interaction was also found to be associated with higher levels of identification towards a community. The finding is aligned with the seeking tradition of socialization research (Ashford & Black, 1996). Robert Kraut and colleagues' hypothesized that using collective socialization tactics would engender a higher sense of commitment in virtual communities (Kraut, Burke & Riedl, 2011). This study's findings (albeit limited to the FLOSS community context) provide the first empirical evidence supporting those authors' hypothesis.

In the FLOSS context, the two findings together emphasize the importance of the social side of socialization experience, during which the interaction with community members as well as the support provided by a community to a newcomer help gen-

erate a higher sense of identification towards the community. This result extends past socialization research that has seen social identification as one of the most important socialization proximal outcomes (Bauer et al., 2007; Chan, 2006; Chan & Schmitt, 2000).

8.6.4 *Predictors of social integration*

One of the critical socialization challenges emphasized in the socialization literature is for newcomers to become integrated into their work group (Feldman, 1976; C. D. Fisher, 1986; Louis, 1980). The importance of social integration in the FLOSS context was identified by von Hippel and Von Krogh (2003) who hypothesized that it helps sustain participation within projects.

The six socialization factors were found to explain about 51% of the variance of the *social integration* construct while *interaction intensity* (path coefficient of 0.170 with $p < 0.01$) and *supportiveness* (0.581 with $p < 0.001$) were found to be significant predictors. First, the findings are consistent with past socialization literature that demonstrated the influence of the extent to which a newcomer interacts with individuals on the person's perception towards his or her integration within a group or an organization (D. G. Allen, 2006; Morrison, 1993b; Wanberg & Kammeyer-Mueller, 2000). Second, the results are aligned with past socialization studies that demonstrated the positive impact of investiture tactics on on-the-job embeddedness (D. G. Allen, 2006) and social acceptance (Bauer et al., 2007) thus suggesting that the positive support provided by an organization helps increase the integration of newcomers. This study extends such past findings to the FLOSS community context and raises new FLOSS research avenues.

8.7 CONTRIBUTIONS OF THE STUDY

This study has investigated phenomena - newcomer socialization and individual performance - that are important to the survival and sustainability of in large and complex Free/Libre/Open Source Software projects but which have been generally lacking in solid theoretical foundations when studied in the IS discipline. This section presents a summary of the contributions of this research project to academics as well as to practitioners.

8.7.1 *Theoretical contributions*

THEORIZATION OF LARGE AND COMPLEX FLOSS PROJECTS The first theoretical contribution of this research resides in the theorization of the social entity under investigation: large and complex FLOSS projects. The term 'complex' characterizes the produced software artifact whereas the term 'large' addresses the size of the population of contributors taking part in the project. Moving away from the outdated bazaar-view, this research conceptualizes large and complex FLOSS projects as organization-community hybrids which function is the commons-based peer production of highly modular software artifacts. In such projects, contributing involves a variety of tasks that go well beyond coding.

This research is unusual, in that it gathered data from a relatively large number of different projects (fourteen); past FLOSS research works have tended to focus on single projects. The approach adopted in this research improves the generalisability and validity of the findings within the theoretical boundaries set in this research: large and complex FLOSS projects.

UNDERSTANDING FLOSS SOCIALIZATION Despite a substantial body of knowledge on FLOSS community practices, FLOSS socialization research has lacked well-grounded theoretical investigations. The organizational socialization literature contains a wealth of results, a number of which were applied in the derivation of the two FLOSS-specific socialization models featured in this study. Previous FLOSS socialization research had suffered from: a lack of strong theoretical underpinnings encompassing the complexity of large and complex FLOSS projects, the use of archival data drawn from project repositories, as well as a tendency to rely on single projects, thus compromising the generalizability of the findings.

The development of a FLOSS socialization framework and the associated measurement of the six socialization variables provide a deeper understanding of the experience through which a FLOSS community newcomer goes. This project also demonstrated the relevance of using a well-grounded socialization model, [Van Maanen and Schein \(1979\)](#), as a theoretical lens in order to explore the socialization phenomenon in a non-organizational context. The sub-categorization of socialization factors into task, individuals, and community-related factors also provides helps in gaining insights about the various dimensions that comprise characterize FLOSS socialization.

The development and validation of this study's two models extends the body of knowledge of socialization theories in virtual environments. The previous chapter demonstrated the superiority of the second model (with higher R^2 coefficients for the two dependent variables and a higher GoF score), which relies on the assumption that the relationship between socialization experience factors and individual performance is mediated by both social identification and social integration. The model comprises a theoretical tool that can be used by researchers to study the FLOSS socialization phenomenon in a given FLOSS community or a set of communities.

THEORETICAL FINDINGS The theorization of large and complex FLOSS projects as organization-community hybrids was reflected on several aspects in this research. First, the analysis of the gathered qualitative data revealed that socialization in large and complex FLOSS projects included a community dimension within which supportiveness and joining structuredness were shown to be key aspects. Furthermore, both factors were found to play an important role by directly or indirectly influencing contributor performance.

The theoretical particularities of the social entities under investigation aslo led to re-investigate the notion of individual performance in the context of large and complex FLOSS projects in light of the organizational behaviour literature. Performance was then conceptualized into two distinct but complementary constructs: task performance and community citizenship behaviour. This view contributes to the FLOSS

research body of knowledge by introducing a community-level view of performance. Furthermore, this research contributes to the OCB body of knowledge by extending the notion of citizenship behaviours to non-organizational and virtual social contexts.

Some theoretical developments were also made in the conceptualization of citizenship behaviours in large and complex FLOSS projects. This research highlights the particularities of FLOSS organization-community hybrids by showing that unlike the two-dimensional OCB structure which has been shown to apply to traditional organizations, CCB is characterized by a 3-dimension structure that consists of: CCB-I (citizenship behaviours directed towards the benefit of individuals), CCB-P (citizenship behaviours directed towards the benefit of the software project), and CCB-C (citizenship behaviours directed towards the community).

The first CCB dimension is overall common with traditional organizational settings highlighting that large and complex FLOSS projects are characterized by some organizational features. The second CCB dimension addresses aspects that characterize the compliance to the behavioural rules and norms of the software project. This aspect is unique to the context under investigation and emphasizes the theoretical specificities of large and complex FLOSS projects.

The third dimension of CCB taps into the community-side of large and complex FLOSS projects. A possible reason for such dimensional discrepancy may lie in the nature itself of FLOSS communities. An essential characteristic that significantly differentiates a FLOSS community from a traditional organization is that a FLOSS community is somehow 'viral' in the way it attracts new software users and potential project contributors. The survival of a community then depends on its capacity to increase the user base of its associated software project combined with its ability to constantly renew its contributor population.

The review of the literature revealed that social identification and social integration are important factors influencing performance in both organizational settings as well as digitally-enabled environments. This research validated measures for both constructs adapted to the FLOSS context. Besides, the final results demonstrated their key role in predicting both task performance and CCB confirming the results from the literature and their extension to the context of large and complex FLOSS projects.

LINK BETWEEN SOCIALIZATION AND CITIZENSHIP BEHAVIOURS This project has helped to address a significant literature gap in both FLOSS as well as broader organizational behaviour research concerning the relationship between socialization and citizenship behaviours.

USING MIXED-METHODS IN IS Finally, this research contributes to the current discussion on the use of multi-methods in Information Systems research. Eminent IS scholars have urged the need for IS researchers to conduct research that rely on the use of mixed methods (Venkatesh et al., 2013). This research contributes to the overall IS body of knowledge by answering this call.

8.7.2 *Theoretical conjectures*

It can be argued that the theoretical models developed in this research could be extended to contexts which scope is close to the theoretical boundaries set in this research. However, the specificities of large and complex FLOSS projects render certain aspects of the models simply irrelevant or only partially applicable. This research may provide theoretical insights into studying socialization in the following contexts: other types of FLOSS projects and also non-software commons-based peer production. This subsection discusses some of the potential differences among the various types of on-line groups. However, the various statements that follow are highly speculative due to the lack of empirical evidence.

APPLICABILITY TO OTHER TYPES OF FLOSS PROJECTS When presenting the theoretical boundaries of this research, two dimensions were taken into consideration: the complexity of the produced software artifact and the size of the population of contributors. By focusing on *large* and *complex* FLOSS projects, this research assumes the existence of three other types of FLOSS projects: complex projects with a small population of contributors, and projects of lesser complexity with a large but also a small number of contributors.

In the case of projects of modest complexity, it may be argued that the software artifact is overall 'simpler' in terms of modularity but also variety of skills necessary to contribute. Task segregation and task purposefulness seem to be particularly relevant in complex projects with high barriers to entry which is less the case in FLOSS projects that are less complex. Task segregation relates to performing smaller tasks to learn better whereas task purposefulness pertains to determining a certain sequence of tasks that help in getting the 'big picture' of a project. As a result, both task-related factors might not be pertinent in the case of FLOSS projects that are less complex.

In FLOSS projects consisting of a limited number of individuals, the community perspective adopted in this research becomes invalid; a group perspective appearing much more appropriate. As a consequence, the notion of joining structuredness may be irrelevant since small projects (in terms of number of contributors) may not be as concerned with 'filtering' potential new contributors; attracting the attention of individuals already being a challenge. However, the importance of 'group supportiveness' may still play a role in successfully integrating newcomers. For similar reasons, social identification is pertinent in large social entities such as organizations or communities but may not be for groups of individuals. Nonetheless, the notion of integration and acceptance being important in group settings (C. O'Reilly et al., 1989), one can argue that the importance of social integration may hold for small FLOSS projects. However, the associated construct's measure shall be revamped to be adapted it to the group context.

Whatever the size and complexity of a given FLOSS project, it might be assumed that they all expect their newcomers to provides quality contributions making the notion of task performance a key socialization outcome. However, the scope of citizenship behaviours in large and complex FLOSS projects seems much broader than in FLOSS projects with a small number of contributors. Whereas actions related to interpersonal helping would appear equally important in all types of projects, certain

aspects of CCB-P may not apply in small projects such as behaving in a way that is aligned with the shared values of the project or else respecting the behavioural rules. Both ideas might be more related to larger populations of contributors. The community-focused dimension of CCB, CCB-C, may be hypothesized to be less relevant in small projects.

As a result, future research may start the investigation of socialization in other types of FLOSS projects by considering the conjectured socialization models presented in Figures 34, 35, and 36. Each FLOSS project being unique, both project complexity and project size (in terms of number of contributors) vary along a continuum. The dichotomous view *low/high* introduced above is an obvious oversimplification of the reality but its purpose is solely to speculate on the degree of applicability of the theoretical models. Besides, FLOSS projects are no static entities; project complexity and size thus fluctuate over time.

Figure 34: *Low number of contributors / Low complexity* - Conjectured socialization models

Dashed constructs: required conceptual and operational modifications

Figure 35: *Low number of contributors / High complexity* - Conjectured socialization models

Dashed constructs: required conceptual and operational modifications

Figure 36: *High number of contributors / Low complexity* - Conjectured socialization models

Dashed constructs: required conceptual and operational modifications

APPLICABILITY TO NON-SOFTWARE COMMONS-BASED PEER PRODUCTION

Commons-based peer production has now reached beyond software to include the production of commons-based information, knowledge, and culture goods through digitally networked environments. The online encyclopedia Wikipedia, for instance, is among the most successful non-software projects that is based on commons-based peer production. Other examples include Slashdot², a collective commentary website, OpenStreetMap³ a project generating a free editable map of the world, or Distributed ProofReaders⁴, a project based on the collective proofreading of public domain texts for e-books.

It seems that overall, the barriers to entry of non-software goods that are generated through commons-based peer production initiatives might be overall lower than those of FLOSS projects in general and also lower than those of complex FLOSS projects. As a consequence, the two task-related notions of task segregation and task purposefulness might not be essential for a number of commons-based peer production projects. Nonetheless, such non-software projects tend to rely on large communities suggesting that both factors characterizing interactions between a newcomer and other contributors as well as those with the community itself may be important in determining newcomer performance. Furthermore, the mediated effect of social identification and social integration between socialization and performance may hold (or at least deserves to be investigated) since both notions are particularly relevant in large commu-

² <http://slashdot.org>
³ <http://www.openstreetmap.org>
⁴ <http://www.pgdp.net>

Figure 37: Non-software commons-based peer production projects - Conjectured socialization models

Dashed constructs: required conceptual and operational modifications

nities.

Finally, task performance may still be a key socialization outcome in the context of commons-based peer production because of the overall productive objective of such entities. When reflecting on the notion of CCB in non-software commons-based peer production, the importance of CCB-I and CCB-C may hold. The conceptual definition and measurement CCB-P may only partially apply because certain aspects of the developed definition and measure pertain to the complexity of the FLOSS environment. For example, avoiding creating problems when contributing or else aligning contributions with the shared goals of a project tend to imply a high degree of complexity in the project. The conjectured models are presented in Figure 37.

8.7.3 Practical contributions

This research project addresses two important issues faced by large FLOSS projects. On the one hand, it will help FLOSS communities in understanding the socialization factors that contribute to the successful socialization of new members. The results of this research thus help communities to tailor effective socialization initiatives that are likely to generate contributor behaviours which support the values and sustainability of the communities. For instance, this study has demonstrated the direct and indirect influence of *task purposefulness* on individual performance. This finding en-

courages FLOSS community managers to determine sequences of tasks that optimize project learning and strongly encourage newcomers to perform them. This research has demonstrated that such guidance would enhance significantly the contributions provided by a “socialized” newcomer. Besides, a number of FLOSS communities have started realizing that being positive and supportive with potential project newcomers is important. This research confirms the benefit of such practice and goes beyond by demonstrating the overall positive effect on the community well-being of being consistently supportive with all project newcomers.

On the other hand, in spite of the importance of citizenship behaviours to ensure the survival of a community, this study’s investigation of the individual factors that lead to citizenship behaviours is new in FLOSS research. The finer understanding of participation and performance provided in this research will also benefit communities in helping them to understand more thoroughly how individuals contribute. This research clearly demonstrates the importance of performance-related behaviours to go well beyond the common idea that contributing to a FLOSS project is only about writing code. Such behaviours, labelled *community citizenship behaviours*, are categorized and identified along each category. The results provided in this research help FLOSS communities to realize that three overall types of behaviour benefit a community: behaviours aiming at helping individuals, those directed toward the direct benefit of the software project, and also those targeted towards the benefit of the overall community. The identification of such behaviours will help FLOSS communities to evaluate a contributor’s level of performance more objectively and using a community-wide perspective. Besides, it can also help identify and support the individuals that contribute heavily to the functioning and sustainability of a given community even though their code contribution is limited or even non-existent.

This research will also benefit FLOSS participants themselves by providing them indications about the practices into which they should engage when joining a new FLOSS community. For instance, this research strongly encourages newcomers to engage with as many project contributors as possible when learning about a FLOSS project. Such behaviour helps become ‘better’ contributors by developing both the feeling of being accepted by the community and the sense of identification to this community. Besides, newcomers should also take part in tasks that are specifically designed for them as it affects their overall contributor performance.

Finally, this research indirectly contributes to our understanding of virtual communities more generally. Both, socialization and contribution behaviour have some overlapping aspects with FLOSS communities. The findings may be to some extent extrapolated to the domain of other virtual communities.

8.8 LIMITATIONS

The limitations of this study include those common to quantitative survey research and others particular to this research project.

The conceptualization of socialization adopted in this research, has obvious organizational connotations. As a result, this research particularly applies to FLOSS communities that have some organizational aspects that are usually present within com-

munities that consist of more than a handful of programmers collaborating on some FLOSS-licensed software project. In other words, this study applies to large FLOSS communities by assuming the presence of a minimum of structural and hierarchical features. The hypothetical application of the findings to smaller FLOSS projects shall then be considered very carefully as it breaches the validity boundaries of this research.

The data used to test the hypotheses are cross-sectional, which means that causality is only inferred, not proven. Future research studies may extend this research by adopting a complimentary approach such as longitudinal studies that inherently lead to stronger causal conclusions.

Survey research is characterized by potential challenges such as measurement error, sampling error, internal validity error, and statistical conclusion error (Straub, 1989). This research attempted to mitigate these errors through the use of commonly accepted methods such as judgment rounds, card sorting procedures, and a pilot test to develop the survey instrument. Nevertheless, while mitigated, these sources of error cannot be entirely eliminated.

Limitations exist about the implemented measurement strategy. For instance, participants were asked to assess various aspects of their socialization within a given community based on the recollection of past perceptions (up to three years back) but also at one particular point of time and without regard to differences in time. This view is overall simplistic since socialization is a collection of actions and interactions that occur at different times, and thus perceptions are bound to vary through time. A similar point can be made about using a 'snapshot measure' of contributor behaviour. A contributor can be a good community citizen for a certain period of time and then be a "bad citizen" for the rest of the time. This research has also done its best to mitigate the inherent social desirability risk of performance-based measures but it cannot be said that a social desirability bias was totally absent from the responses provided by the participants.

Finally, this research relied on the use of a self-selecting sample consisting of individuals whose socialization experience resulted into a 'positive' outcome since they eventually became project contributors. In other words, the sample did not include individuals who left a project while being in the socialization phase. This is a limitation inherent to the methodology implemented in this research.

8.9 DIRECTIONS FOR FUTURE RESEARCH

This research has opened new research avenues and suggested areas for further research along three main areas of development: extending the current models, re-examining and/or validating the constructs and their associated measurement scales; and finally assessing the applicability of the conceptual models in different but related contexts.

The socialization framework developed in this research unveils new possibilities for studying socialization in FLOSS communities. First, the framework and its associated measurement can be used to assess newcomer socialization in other FLOSS commu-

nities (other than the 12+2 communities which took part in this study). It can also be used in the case of smaller FLOSS projects so as to assess its applicability to other FLOSS contexts. Second, this research has left room for other important socialization factors that could be discovered through the collection of additional qualitative data. The sub-categorization into task, individuals, and community factors introduced in this research acknowledges the possible existence of other factors and leaves room for further refinements of the framework.

The results which emerged from analysis of the survey data concerning the non-relevance of mentoring are somewhat surprising, since the qualitative phase of this research concluded that mentoring was important. Further research should investigate further the potential role of mentoring when inducing FLOSS newcomers; the measurement of the mentoring construct in particular should be re-examined in order to confirm or infirm the results found in this study.

Some further research could be conducted in order to identify other possible predictors of both social identification and social integration in the FLOSS socialization context, since the socialization variables accounted for respectively 24% and 51% of the variance of the two proximal socialization variables. Individual factors (such as intrinsic/extrinsic motivation) or project-related factors (such as license type, license restrictiveness, or project size) that have been shown by past FLOSS research to influence FLOSS project participation could be investigated. Another possibility could be in importing constructs from well-acknowledged attitude-theoretic models in IS models such as the theory of planned behaviour (Ajzen, 1991) or the theory of reasoned action (Ajzen & Fishbein, 1980). Initial studies along these lines have been undertaken in past FLOSS research. For example Bagozzi and Dholakia (2006) found that attitudes and perceived behavioural control had both a significant positive impact on participation intentions in the case of Linux user groups.

Furthermore, this research suggests the possibility of uncovering proximal socialization outcomes, since the model explained respectively 29% and 44% of the variance of task performance and community citizenship behaviours. Consequently, the socialization literature could be consulted in order to find other potential factors such as perceived fit (Saks et al., 2007), self-efficacy (Bauer et al., 2007), task mastery (Kammeyer-Mueller & Wanberg, 2003), or role conflict (Ashforth et al., 1997). The socialization tradition also provides an array of distal outcomes that could be worth investigating in relation to the socialization framework developed in this research. Investigation of socialization distal outcomes such as intentions to remain, turnover, satisfaction, or commitment (Bauer et al., 2007; Gruman et al., 2006; Saks & Ashforth, 1997a; Saks et al., 2007) could provide additional insights into the FLOSS socialization phenomenon.

Finally, the applicability of the FLOSS-specific conceptual models developed in this research could be assessed in related contexts such as FLOSS projects of lesser size, online communities, and virtual or geographically-distributed organizations.

8.10 SUMMARY

This chapter concludes this thesis by summarizing and discussing the main findings of the research. It then identifies the contributions to theory and to practice made by

the results, and presents some of the limitations of the research. The chapter ends by suggesting research areas which would deserve further investigation based on the research findings of this thesis. We hope this research project will trigger such new research efforts.

APPENDICES

HUMAN ETHICS COMMITTEE APPROVALS

HEC APPROVAL FOR QUALITATIVE COMPONENT - INTERVIEWS

SIM HUMAN ETHICS COMMITTEE Comments on Application for Human Ethics Approval

Date: **14 Dec. 11**

Principal Researcher: **Kevin Carillo**

Research Project: **A socialization perspective towards understanding contribution behaviours within Free/Libre/Open Source Software communities**

Supervisor: **Sid Huff; Brenda Chawner**

Reference No: **#19075**

Your HEC application has been reviewed and the Committee's decision is the following:

Category 2 - Application approved subject to the following minor change(s).

Human Ethics Approval valid until: (Date: *as in application or no more than 3 years*)

The committee requires the change(s) noted below, if these are acceptable you may begin immediately but you will need to send an updated application within one month. If the changes are not acceptable to you, please email me detailing your objections. Note that the updated hard copy of your application needs to be signed by both you and your supervisor and should be submitted to me at:

Wendy Chen
School of Information Management
Victoria University of Wellington
Room 518, Rutherford House
23 Lambton Quay
Wellington

Required Changes (dealing with ethical issues only)

The email invitations for the focus group and interview note, "I will also send you a summary of the results after my thesis is completed". As this could be a long wait, participants should be given the opportunity to receive a report with interim results from the study.

Other Suggested Changes:

The tick boxes on the consent forms are unnecessary to indicate agreement with points if people don't tick all the boxes problems will arise. A signature at the bottom should be sufficient.

SIM Human Ethics Committee

**SCHOOL OF INFORMATION MANAGEMENT
FACULTY OF COMMERCE AND ADMINISTRATION**

HEC APPROVAL FOR CARD SORTING PROCEDURES

SIM HUMAN ETHICS COMMITTEE
Comments on Application for Human Ethics Approval

Date: **29 May 2012**

Principal Researcher: **Kevin Carillo**

Research Project: **A socialization perspective towards understanding contribution behaviours within Free/Libre/Open Source Software communities**

Supervisor: **Sid Huff; Brenda Chawner**

Reference No: **#19337**

Your HEC application has been reviewed and the Committee's decision is the following:

Category 1 - Application accepted.

Human Ethics Approval valid until: (Date: *as in application or no more than 3 years*)

You may begin your data collection immediately. Please note that a hard copy of your application signed by both you and your supervisor is required within one month to ensure that we have a complete record of the approval of your application. This should be submitted to me at:

Wendy Chen
School of Information Management
Victoria University of Wellington
Room 518, Rutherford House
23 Lambton Quay
Wellington

SIM Human Ethics Committee

SCHOOL OF INFORMATION MANAGEMENT
FACULTY OF COMMERCE AND ADMINISTRATION

HEC APPROVAL FOR PILOT STUDY

Feedback on Kevin Carillo's HEC proposal

This research can proceed subject to the following minor changes (Category 2):

1. 5 (j) (ii) should say No. (So it does not contradict 5(j)(i))
2. Please clarify how the feedback will be given to respondents given that this is an anonymous survey. Respondents should be able to request a copy of the results at the time of participation. It is recommended that a separate optional final section of the online survey is created that allows people to add their details and request a copy of the final report. This should be designed so the information goes into a separate database and respondents should be made aware (in the letter and the survey) about how to do this and the fact that the data cannot be linked with their responses. Qualtrix has this functionality.

Best wishes for your research

HEC APPROVAL FOR MAIN STUDY

SIM HUMAN ETHICS COMMITTEE
Comments on Application for Human Ethics Approval

Date: 1 Nov 2012

Principal Researcher: **Kevin Carillo**

Research Project: **A socialization perspective towards understanding contribution behaviours within Free/Libre/Open Source Software communities**

Supervisor: **Sid Huff; Brenda Chawner**

Reference No: **#19569**

	Accept
x	Accept with minor changes. (Stated below)
	Accepted with required changes. (Stated below)
	Do not accept in present form. (Changes required as below)

Required Changes (dealing with ethical issues only)

Please revise section 5(k)(ii) to read N

In section 5(l) please make clear the distinction between procedures for storage of and access to the data and any notes during the study and the procedures after the study. (At present it appears to be contradictory.) I suggest that the written section spells out clearly how data will be stored, managed and made available after it has been released, and that a note is added to explain that the multi-choice responses relate to the period pre-release.

Please amend the consent form (at the start of the survey) and the information sheet so that people are given clear information about how and when they will receive and how they could to request a copy of the results.

Additional Comments (dealing with non-ethical issues only)

Best wishes for your study.

QUALITATIVE DATA ANALYSIS

OPEN CODING - DESCRIPTIVE CODES

PARTICIPANT_BACKGROUND
PARTICIPANT_DEMOG
PARTICIPANT_SOCIALIZATION_EXP
PARTICIPANT_CONTRIBUTION
PARTICIPANT_HISTORY

OPEN CODING - SOCIALIZATION / FINAL CODES

	CODES	PROPERTIES	CODING EXAMPLE
CONTEXT	COLLECTIVE initial	groups of newcomers going through formal common experience	<i>Things like Ubuntu, for example, have very active forums which means that there is actually a good point and a bad point. You have the collective learning experience of newcomers because that's typically the first place that a lot of people who aren't familiar with the project will go.</i>
	FORMAL initial	formal segregation of newcomers going through newcomer-tailored tasks	<i>Yes, there are things like the Summer of Code, there is things like these "Build It" events that we ran last year with ###. Yes ... formal programmes are quite rare</i>
	JOINING new	newcomers taking part in formal joining process	<i>So, how it works from our recruiting people is that we file a developer bug. It says "this person is going to become a developer and here is what he is going to work on. He is working on the quizzes right now". 30 days from then, as soon as they are eligible to be processed from one of our recruiters, "go to the quizzes in details". Then following that they can become a developer with a certain amount of access.</i>
CONTENT	SEQUENTIAL initial	newcomers being provided a given sequence of steps leading to target role	<i>There is not really a sequence of tasks. It depends on what the person is interested in. We have experience, so this is what we think you should be doing to not have all the issues that we have had in the past.</i>
	FIXED initial	newcomers perform steps involving an acknowledged timetable	<i>Those projects have no sort of timetable for new contributors especially because the sort of high watermark for a project to make somebody go to the next stage of power in your project, say commit access, or maybe read access to the primary mailing list, or write access to the bug tracker is very subjective.</i>
SOCIAL	SERIAL initial	experienced members grooming newcomers	<i>There is other people that are a bit more shy and they will wait for long time before they ask a question, they will talk to people privately, but there are like kind of leaders within online forums that tend to emerge and often you see newcomers asking these people directly or repeating what they have said to other people. I think there are kind of role models.</i>
	INVESTITURE initial	confirming/disconfirming past newcomer identity	<i>Most communities basically disregard them. Because in a sense, nobody cares about your qualifications, they care about the way you interact with people, that proves that you have knowledge. When people submit a patch you require a high quality C-style.</i>
	MENTORING new	newcomers being informally/formally mentored by experienced contributors	<i>Explicit mentoring works very well. I had my own of informal mentoring network and learnt what was going on and who is who. I think it does happen a lot.</i>
	COMMUNITY SUPPORT new	community being supportive towards newcomers during socialization experience	<i>Some communities are supportive and other communities are not. Other than that, some people are supportive in communities and some people somehow don't receive the support from supportive communities.</i>
	INTERACTION new	newcomers interacting with other project contributors in order to learn about project	<i>There is lots of spaces where people interact but nothing just for newcomers. They are welcome everywhere. They get integrated.</i>

OPEN CODING - CITIZENSHIP / FINAL CODES - 1

CODES	PROPERTIES	CODING EXAMPLE
ALTRUISM initial	altruistic and voluntary actions intended to help individuals for project-related problems	<i>So being a good citizen is just doing things like helping other contributors</i>
CONSCIENTIOUSNESS initial	behaviour indicating acceptance and adherence to rules and regulations	<i>I don't think that complying with rules is an end goal, the rules are usually a codification of community practices - and respecting your community and the way it works is very important.</i>
CIVIC VIRTUE initial	behaviours indicating interest in taking part in project life	<i>When you have a multifaceted project like GNOME, you don't necessarily need to know every detail happening in every team. But it is important to know kind of the big picture, the big announcements, how things are going, how things are changing.</i>
SPORTSMANSHIP initial	tolerating difficult circumstances without complaining	<i>I don't think it is important for a good community citizen to tell other people when their workload is very heavy. But if the workload is so heavy that it is considered heavy, then that person might have been managing himself poorly and then will probably burn out.</i>
COURTESY initial	behaviours that help others avoid problems	<i>You do have to be self-aware of what you are going to do. Because everybody can change everything, you need to make sure that what you are changing, you have to think about it. You cannot just change some major stuff before thinking before. Yes, I think it is important to be aware of what they are doing.</i>
FRIENDLINESS new	being friendly with others	<i>Behaviour would be someone who is friendly, someone who is not too verbose. They tell others when they need help, they are very friendly with other people, they can work with other people.</i>
BEING NICE new	being nice with other contributors	<i>The idea or contributor or community citizen would be someone who is a nice person as well.</i>

OPEN CODING - CITIZENSHIP / FINAL CODES - 2

CODES	PROPERTIES	CODING EXAMPLE
ADVOCACY new	promoting the project to others	<i>Somebody advocating the project outside, that could be perhaps an aspect of citizenship.</i>
PROJECT COMPLIANCE new	contributing in accordance with project goals and direction	<i>...preferably drives the direction of the project in some meaningful way. What makes him good ... Good as if they are trying ... if they know what the goals of the overall project are, and then pushing in the same direction. ...</i>
CODE OF CONDUCT COMPLIANCE new	complying to the behavioural expectations of project	<i>We have a code of conduct that actually lays down some of our community rules. It is very important especially for long-term community members who are the role models that you comply to our community standards.</i>
CONSIDERATENESS new	being considerate towards other project contributors	<i>I would say "considerate" rather than "conscientious" - I think it's important to do what you say, or say as soon as you think you won't be able to do it as promised.</i>
TIME RESPECT new	behaving in a way that respects others' time	<i>I guess someone would say maybe respectful of other people's time. I can't think of a word that would describe that. It's like being respectful in a social setting but also you do realize that people are spending free time on your problem.</i>
MAINTENANCE new	performing maintenance tasks	<i>A lot of the really good contributors are people that clean up other people's messes. For example merging bug reports that are duplicates, cleaning up code ...</i>
SOLVE CONFLICT new	facilitating conflicts among contributors	<i>If you just turn up with something which they could be persuaded whether they wanted but if you say: "take it or leave it", it is quite confrontational, it is quite aggressive. [...] So, an extra layer of person management again is ... "What do you mean 'take it or leave it'? Oh, what I meant was I don't mind if you don't take the patch. Oh, great. That's positive.</i>
INFORM PROJECT new	keeping the community informed about one's work and progress	<i>If you are saying you'll do something by January 31 and then you don't, this is important but it is even more important to say that you can't and let people know about that.</i>
DECISION INVOLVEMENT new	responsible involvement in project decisions	<i>I do think that there is some kind of civic responsibility to if you disagree with something that is being decided at the project level, you must say that you disagree and why before beforehand rather than wait until the decision is made and then go off complaining. That's something which I think is not a good community behaviour.</i>

AXIAL CODING - FINAL CODES (OPEN/AXIAL CODES MATCHING)

FINAL CODES (AXIAL CODES)			PROPERTIES	INITIAL CODES (OPEN CODING)
SOCIALIZATION	TASK	SEGREGATION	performing tasks specifically tailored for newcomers	COLLECTIVE FORMAL
		STRUCTURE	performing a sequence of tasks following a pattern intended to learn about a project	SEQUENTIAL
	INDIVIDUALS	INTERACTION	actively interacting with newcomers and community members while learning how to become a project contributor.	INTERACTION
		MENTORING	being taken under the wing of an experienced contributor	FORMAL SERIAL
	COMMUNITY	JOINING STRUCTURE	having to follow a structured joining process while joining a project	JOINING FIXED
		SUPPORT	perceiving community support while learning how to become a contributor	INVESTITURE COMMUNITY SUPPORT
CITIZENSHIP	INDIVIDUALS	INTERPERSONAL HELP	helping and assisiting other community members	ALTRUISM
		PREVENT PROBLEMS	helping to prevent project problems to occur for other contributors	COURTESY CONSIDERATENESS
		RESPECT	treating others with respect	COURTESY TIME RESPECT
		POSITIVE ATTITUDE	maintaining a positive attitude	FRIENDLINESS
		CONFLICT FACILITATION	facilitating conflicts among contributors	SOLVE CONFLICT
	PROJECT	PROJECT DIRECTION	contributing with the best interest for the community	PROJECT COMPLIANCE
		CULTURE COMPLIANCE	complying to the project's behavioural expectations	CONSCIENTIOUSNESS CODE OF CONDUCT COMPLIANCE BEING NICE
		INFORM PROJECT	keeping the community informed about one's work and progress	INFORM PROJECT
		MAINTENANCE	keeping project resources to a good standard	MAINTENANCE
		CIVIC RESPONSABILITY	behaviours indicating interest in taking part in project life	CIVIC VIRTUE
		RESPONSIBLE INVOLVEMENT	responsible involvement in project decisions and discussions	DECISION INVOLVEMENT
		ADVOCACY	promoting the project to potential contributors and to the outside world	ADVOCACY

EXAMPLES OF DATA DISPLAY (NVIVO 9 REPORTS)

Nodes

citizen_friendly [X]

<Internals\IntervieweeA> - § 4 references coded [3.33% Coverage]

Reference 1 - 0.19% Coverage

Other kinds of good contributors for a community have been people who just hang out on the IRC channels and are friendly

Reference 2 - 0.17% Coverage

Even then you still end up with things like maintaining the culture of the project, being nice to people.

Reference 3 - 1.36% Coverage

You said a good citizen is somebody who is nice to people. Can you elaborate a little bit? Why is it important?

IntervieweeA (17.49):
Yes. I think that it varies with the culture of your project, first of all. I think that would be interesting to interview contributors to the Linux kernel. It is a product where that frequently feels like it is very tough in its communication. But it does not prevail to building a product. When I say nice here, I mostly mean treats people in accordance with the community expectations for how people are to be treated. In OpenHatch in particular and in the Debian-mentor mailing list, there is a big focus on attracting new contributors. Being nice when people are making mistakes means that they are more likely to correct those mistakes and stick around. That's one of the reasons why it is important to be nice.

Reference 4 - 1.61% Coverage

Kevin (19.04):
So, would you say that being nice is for you behaving according to the community code of conduct?

IntervieweeA (19.12):
Well code of conduct usually refers to some written document. Most products have a written document but it is about how people are expected to behave. If you are walking to a party, you can infer how people expect each other to be treated. The same happens on a mailing list, or IRC, or in a commit box of a project. I mean I would say that being nice is not complying with the expectations of a project ... but also having those expectations being somehow positive and rewarding. I have only participated in projects where the expectations are to be friendly with people. There are surely some open source projects somewhere that are based on mutual abuse and some other things where people are not treated nicely. The community members would find it very draining if somebody showed up and tries to be friendly and excessively verbose, et cetera. So it's kind of slippery

<Internals\IntervieweeC> - § 2 references coded [3.68% Coverage]

Reference 1 - 1.20% Coverage

IntervieweeC (38.06):
I think it is important that to be a good citizen that you get on well with the community, that you get on well with the other people. So, think of all the things that you can or cannot do in a local community to get on with your neighbours or not to get on well with them. If your neighbour is doing something that annoys you and you don't tell him, is that being a bad neighbour or a good one? If you don't tell him, it

Kevin (38-35):
Here, we are thinking about overall contributors so we are not speaking about newcomers. It's just a regular community member, full-time member, a full contributor, if he is submitting a patch, is it important that he thinks a little bit about how it is going to impact the other modules of the project, or the other people working on the project?

IntervieweeA (39-10):
It is only important that this person acts in a way that is maximally respectful of the time other people have already put into the project and all the time that people are putting into changes that have not been published yet. I think that there are ways to be that respectful that don't come from being considerate that instead come from habits and defaults. Of course, somebody ten years ago once told you: "Submit every change in a separate patch", so you do that, it happens to be an extremely considerate thing to do because it means that packages in the middle that don't depend on each other can get merged now, some can get merged later and get merged later because that's ... you have been respectful by default. This is just as true if it is a long time contributor than if you are a non long-time one.

Kevin (40-05):
Ok, let's move to something slightly different. Is it important for you to have contributors taking voluntary actions such as helping other members?

IntervieweeA (40-15):
Almost all kinds of contributions are voluntary. Submitting a patch is a voluntary action. I can't think of any non-voluntary compulsory action.

Kevin (40-25):
Here, it is more in terms of altruistic behaviours like helping.

IntervieweeA (40-30):
Yes, but there is a bunch of perspectives on free software and personal philosophy. One of the things about free software is that you can justify things on different personal philosophies. The same action can be justified in different ways. Some people believe that helping a new contributor is not an altruistic thing but instead it is a desperate act of self-promotion in order to gain a new contributor. So, it is radically not altruistic. If I can pay this person in kindness, then I buy a new contributor, and my project will have doubled in contributors because I am the only person right now.

Kevin (41-17):
Now, let's take a different angle. From the community perspective and again about a contributor about being or not a good citizen. Would you like to have this person to have real altruistic behaviours which is helping other members or helpina newcomers?

JONES' (1986) MEASUREMENT SCALES

Collective versus Individual	
CI ₁	In the last six months, I have been extensively involved with other new recruits in common, job related training activities.
CI ₂	Other newcomers have been instrumental in helping me to understand my job requirements.
CI ₃	This organization puts all newcomers through the same set of learning experiences.
CI ₄	Most of my training has been carried out apart from other newcomers. (R)
CI ₅	There is a sense of being in the same boat amongst newcomers in this organization.
Formal versus informal	
FII	I have been through a set of training experiences which are specifically designed to give newcomers a thorough knowledge of job related skills.
FI ₂	During my training for this job I was normally physically apart from regular organizational members.
F ₁₃	I did not perform any of my normal job responsibilities until I was thoroughly familiar with departmental procedures and work methods.
FI ₄	Much of my job knowledge has been acquired informally on a trial and error basis. (R)
FI ₅	I have been very aware that I am seen as learning the ropes in this organization.
Investiture versus divestiture	
ID ₁	I have been made to feel that my skills and abilities are very important in this organization.
ID ₂	Almost all of my colleagues have been supportive of me personally.
ID ₃	I have had to change my attitudes and values to be accepted in this organization. (R)
ID ₄	My colleagues have gone out of their way to help me adjust to this organization.
ID ₅	I feel that experienced organizational members have held me at a distance until I confirm to their expectations.

Sequential versus random	
SR ₁	There is a clear pattern in the way one role leads to another or one job assignment leads to another in this organization.
SR ₂	Each stage of the training process has, and will, expand and build upon the job knowledge gained during the preceeding stages of the process.
SR ₃	The movement from role to role and function to function to build up experience and a track record is very apparent in this organization.
SR ₄	This organization does not put newcomers through an identifiable sequence of learning experiences. (R)
SR ₅	The steps in the career ladder are clearly specified in this organization.
Serial versus disjunctive	
SD ₁	Experienced organizational members see advising or training newcomers as one of their main job responsibilities in this organization.
SD ₂	I am gaining a clear understanding of my role in this organization from observing my senior colleagues.
SD ₃	I have received little guidance from experienced organizational members as to how I should perform my job. (R)
SD ₄	I have little or no access to people who have previously performed my role in this organization. (R)
SD ₅	I have been generally left alone to discover what my role should be in this organization. (R)
Fixed versus variable	
FV ₁	I can predict my future career path in this organization by observing other people's experiences.
FV ₂	I have a good knowledge of the time it will take me to go through of the training process in this organization.
FV ₃	The way in which my progress through is organization will follow of events has been clearly communicated to me.
FV ₄	I have little idea when to expect a new job assignment or training exercise in this organization. (R)
FV ₅	Most of my knowledge of what may happen to me in the future comes informally, through the grapevine, rather than through regular organizational channels. (R)

CARD SORTING PROCEDURE

ITEMS FOR THE PRACTICE EXERCISE

PERCEIVED DRIVING ABILITY

- I consider myself a good driver.
- I can drive well. I am not good at driving.
- I am confident in my driving ability.
- I am able to control my car better than most people.

PERCEIVED RISKS (IN RELATION TO DRIVING)

- It is usually safe to drive.
- Driving a car is perilous.
- I consider that driving is dangerous.
- I find that it is risky to use a car.

AMBIGUOUS/NOT FITTING

- I usually avoid taking my car.
- I like cars that are very comfortable.

CARD SORTING PROTOCOL - CLOSED ROUND

Perspectives on encouraging newcomers to contribute to FLOSS communities**Introduction**

Thank you for your interest in participating in this research project which investigates the factors that lead a Free/Libre/Open Source Software (FLOSS) community newcomer to become a 'good' contributor from the community perspective.

In this research, a FLOSS community is defined as a group of individuals that collectively contribute over the Internet to a software project in which the license allows access and modification of the source code as well as free redistribution.

Your participation in this session will help me construct a survey instrument that will be administered to contributors from various FLOSS communities.

Today's session will start with an explanation of the procedure. You will then be given time to ask questions in case certain aspects of the procedure are not clear.

There will then be two exercises during which you will have to assign statements to categories characterizing FLOSS newcomer experience in the first exercise and FLOSS contributor behaviour in the second one.

Once you have completed the first exercise, you will be asked to start the second exercise.

In total, both exercises require between 30-45 minutes to complete.

While performing Exercise 1 or Exercise 2, you are free to read these instructions any time you want.

Thank you for your contribution to my PhD research project. Please do not discuss about this exercise with colleagues and students as they may also be invited to take part in this phase of the research project.

Exercise 1 – Newcomer experience

Instructions

Two stacks of randomly ordered index cards will be provided to you. The yellow pile contains the names and definitions of six categories about aspects of FLOSS community newcomer experience whereas.

Each white card contains a statement intended to reflect actions that individuals performed when they were newcomers in a given FLOSS community.

1. Sorting cards into categories

- Please start by reading all the category names and definitions (yellow cards) and placing them in front of you.
- Read each statement of the white pile and assign the corresponding card to the category which you think corresponds the best.
- Make sure that within each group of cards, the statements are more similar to one another than they are to the statements on the cards of the other categories.
- The number of cards within each category can vary. You are free to change your mind as many times as you want when assigning a card to a category.
- Once you have assigned each card to a category, please review each pile one more time to ensure that all cards are in the category that you think is the most appropriate.

2. Ambiguous, repeated and/or indeterminate items

- Indicate, by writing on the cards, the items that you have found to be either ambiguous, indeterminate (fitting in no particular category), or those that seem to be repeated.
- Finally, point out the items that you think could be better worded by writing your remarks on the identified index cards.

Exercise 2 – Contributor behaviour

Instructions

The procedure is the same as in Exercise 1. This time, the yellow pile contains the names and definitions of three categories about FLOSS community contributor behaviours.

Each white card contains a statement intended to reflect behaviours that members perform when contributing to a FLOSS project.

Important remark:

***Community citizenship behaviour** is defined as the degree to which a community member performs discretionary behaviours that promote the effective functioning of the community.*

TERMS OF THE ODC OPEN DATABASE LICENSE (ODBL)

ODC Open Database License (ODbL) Summary

This is a human-readable summary of the [ODbL 1.0 license](#). Please see the disclaimer below.

You are free:

To Share: To copy, distribute and use the database.

To Create: To produce works from the database.

To Adapt: To modify, transform and build upon the database.

As long as you:

Attribute: You must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database.

Share-Alike: If you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL.

Keep open: If you redistribute the database, or an adapted version of it, then you may use technological measures that restrict the work (such as DRM) as long as you also redistribute a version without such measures.

Disclaimer

This is not a license. It is simply a handy reference for understanding the ODbL 1.0 — it is a human-readable expression of some of its key terms. This document has no legal value, and its contents do not appear in the actual license. Read the full [ODbL 1.0 license text](#) for the exact terms that apply.

Source: <http://opendatacommons.org/licenses/odbl/summary/> (for the full terms of the license, see <http://opendatacommons.org/licenses/odbl/1.0>)

SURVEY INSTRUMENT USED FOR THE MAIN SURVEY

PAGE 1

Newcomer experience and contributor behavior in FOSS communities - Survey

Thank you for accessing the questionnaire!

The goal of this research is to understand how a person's experience as a newcomer to a Free/Open Source Software (FOSS) community influences that person's behavior and contributions within that community. I am interested in hearing from people who are either technical or non-technical contributors, and who have had either positive or negative newcomer experiences.

About the survey

- This survey is **anonymous**, and no information that would identify you is being collected.
- Your participation is voluntary, and you are implying consent to participate by completing and submitting this online survey.
- You will not be able to withdraw data once you submitted the survey, but it will be possible to withdraw anytime while completing the survey by simply closing the web browser.
- It should take you around **20 minutes** to complete the survey.

Data and results

- The collected data will be released under a '**share-alike' Open Data Commons Open Database License** (ODbL) ensuring that the dataset can be shared (copied, used, and distributed to others), used to produce derived works, or adapted. Any data that may help identify participants will be removed before the dataset is released.
- The software used for this survey is issued under the **GNU General Public License** (GPL version 3).
- The results of this research may be deposited in the library's institutional repository or presented at conferences or published as articles in professional or academic journals. Only aggregate data will be used in any presentations or publications that result from this research.

Do not hesitate to let other people know about the study. I will post news about my progress with this research, and the results on my blog: <http://kevincarillo.org>.

If you have any questions, please contact me at Kevin.Carillo@vuw.ac.nz or my supervisors Sid Huff at Sid.Huff@vuw.ac.nz and Brenda Chawner at Brenda.Chawner@vuw.ac.nz.

PAGE 2

Newcomer experience and contributor behavior in FOSS communities - Survey

0% 100%

CONDITIONS TO PARTICIPATE

Please complete this survey only if you satisfy the following conditions:

- You started contributing to the community after January 2010 (within approximately the last 3 years).**
- You have been contributing to the project on a regular basis: on average once a month or more during the last 6 months.**

Here the term 'contribution' encompasses any sort of technical/non technical contributions to the project (e.g. submitting a patch, editing a project documentation page, participating in a mailing list or an IRC channel, representing the project to outsiders, providing feedback to someone, redirecting a person with a problem to another person in the project...).

This survey will ask you questions about your experience when you were in the process of joining the project as well as about your current contributions.

If you meet the conditions above, please select the FOSS community in which you are involved and click the 'Next' button to continue to the survey.

1 Please select the FOSS community in which you are a contributor. In case you are contributing to more than one, please select the one to which you relate the most.

Please choose... ▼

2 In which sub-project(s) of the FOSS community selected above have you contributed?

<< Previous
Next >>

Resume later

PAGE 3

Survey

Newcomer experience and contributor behavior in FOSS communities - Survey

0% 100%

free motives (most important motive on top)

PART I - Project background

3 When did you become a contributor in Debian?

☐ Less than 6 months ago

☐ Between 6 months and 1 year ago

☐ Between 1 year and 1.5 years ago

☐ Between 1.5 years and 2 years ago

☐ Between 2 years and 2.5 years ago

☐ Between 2.5 years and 3 years ago

4 What were your main motives for deciding to join Debian community? You may select up to the top).

Click on an item in the list on the left, starting with your highest ranking item, moving through to your lowest ranking item.

Your choices:	Your ranking:
I wanted to have some fun.	1: <input style="width: 80%;" type="text"/>
I wanted to be intellectually challenged.	2: <input style="width: 80%;" type="text"/>
It would increase my salary.	3: <input style="width: 80%;" type="text"/>
I wanted to learn some new skills.	
I wanted to feel part of the community.	
I wanted to increase job opportunities or advance my career.	
I had a particular software problem to solve.	
I wanted to help others.	
I wanted to gain the recognition of others in this project.	
I adhered to the values and ideology of this project.	
I wanted to give back to the project after having been helped.	

Click on the scissors next to each item on the right to remove the last entry in your ranked list

4 What were your main motives for deciding to join Debian community? You may select up to three motives (most important motive on top).

Click on an item in the list on the left, starting with your highest ranking item, moving through to your lowest ranking item.

Your choices:	Your ranking:
I wanted to have some fun.	1: <input type="text"/>
I wanted to be intellectually challenged.	2: <input type="text"/>
It would increase my salary.	3: <input type="text"/>
I wanted to learn some new skills.	
I wanted to feel part of the community.	
I wanted to increase job opportunities or advance my career.	
I had a particular software problem to solve.	
I wanted to help others.	
I wanted to gain the recognition of others in this project.	
I adhered to the values and ideology of this project.	
I wanted to give back to the project after having been helped.	

Click on the scissors next to each item on the right to remove the last entry in your ranked list

5 How much time did it approximately take you to become an established contributor since the time you decided to join Debian?

☐ Less than 1 month
☐ Between 1 month and 3 months
☐ Between 3 months and 6 months
☐ Between 6 months and 1 year
☐ More than 1 year

📌 An *established contributor* is a person who has completed a series of tasks that establishes him or her as a contributor and that gives the person some confidence when interacting within the Debian community.

6 Did you take part in any official mentoring programme in the Debian community (as a mentee)?

☐ Yes ☐ No

7 Did you take part in any community event before you became an established contributor?

☐ Yes ☐ No

8 When you decided to join the Debian community, how many years of experience did you have as a FOSS project contributor in other projects?

☐ No experience
☐ Less than six months
☐ Between 6 months and a year
☐ Between 1 year and 2 years
☐ Between 2 years and 3 years
☐ More than 3 years

9 When you decided to join the Debian community, how knowledgeable were you about Free/Open Source Software?

☐ Very knowledgeable
☐ Somehow knowledgeable
☐ Neutral
☐ Not very knowledgeable
☐ Not knowledgeable at all

10 How would you describe your current role in the Debian community?

📌 Examples of roles include developer, maintainer, project leader, release manager, document writer, web artists, translators, community advocate... If you have more than one role, please choose the one that takes up most of your time.

10

How would you describe your current role in the Debian community?

Examples of roles include developer, maintainer, project leader, release manager, document writer, web artists, translators, community advocate... If you have more than one role, please choose the one that takes up most of your time.

11

What other roles have you held in the Debian community, if any?

12

What proportion of your time working on Debian has been part of your paid employment?

☐ None

☐ Less than 20%

☐ Between 20% and 50%

☐ Between 50% and 80%

☐ Between 80% and 100%

<< Previous

Next >>

Resume later

PAGE 4

Newcomer experience and contributor behavior in FOSS communities - Survey

0% 100%

PART II - Your experience as a project newcomer

The remaining questions are all about your experience within the Debian community.

This section will help us understand your experience when you were in the phase of becoming a contributor for the Debian community.

Think back about the time from when you decided to join Debian until you became an established contributor. Answer the items below by reflecting on what your experience was during that period.

14 Please indicate to what extent the following statements characterize the period when you were joining the Debian community.

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
The way I have learned to become a contributor in this project has been highly collaborative.	☐	☐	☒	☐	☐	☐	☐
I have been active in the project communication channels (such as forums, mailing lists, or IRC channels) to help me learn about the project.	☐	☐	☒	☐	☐	☐	☐
I have been engaged with many other community members to gain project-related knowledge.	☐	☐	☒	☐	☐	☐	☐
Interacting with a lot of other project members has helped me understand how to become a project contributor.	☐	☐	☒	☐	☐	☐	☐
Working with a lot of other community members as a newcomer has contributed to my learning.	☐	☐	☒	☐	☐	☐	☐

18 Please indicate to what extent the following statements characterize the period when you were joining the Debian community.

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
This community is very supportive towards its newcomers.	☐	☐	☐	☐	☐	☐	☐
This community cared a lot about me when I was a newcomer.	☐	☐	☐	☐	☐	☐	☐
I have been made to feel that my contributions are valuable in this project.	☐	☐	☐	☐	☐	☐	☐
Most community members that I have been in contact with have gone out of their way to help me become a contributor.	☐	☐	☐	☐	☐	☐	☐
I have felt valued by the community even though I was a newcomer.	☐	☐	☐	☐	☐	☐	☐

19 Please indicate to what extent the following statements characterize the period when you were joining the Debian community.

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
The pattern of the sequence of tasks that I have chosen to undertake helped me gain an understanding of the project.	☐	☐	☐	☐	☐	☐	☐
Almost every task that I have chosen to perform has expanded and built upon the knowledge gained from the preceding tasks.	☐	☐	☐	☐	☐	☐	☐
To learn how to become a contributor, I have selected jobs with gradually increasing complexity.	☐	☐	☐	☐	☐	☐	☐
The logical connection among the tasks I have opted to perform helped me gain an understanding of the project.	☐	☐	☐	☐	☐	☐	☐

<< Previous Next >>

Resume later

PAGE 5

Newcomer experience and contributor behavior in FOSS communities - Survey

0% 100%

PART III - Your current perceptions of, and contributions to the Debian community.

The remaining questions are not about your experience as a newcomer. They are all about your current perceptions and contributions to the Debian community.

IMPORTANT: Some of the following questions are about your behavior within the community. Your answers will in no way be used to judge whether you are a 'good' or a 'bad' contributor. Besides, this survey is anonymous.

To be able to effectively study how a person's newcomer experience influences this person's behaviour within a FOSS community, it is crucial that you answer the following questions by being as honest as possible.

20 Please indicate to what extent do you agree with the following statements:

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
I am very interested in what outsiders think about the community.	☐	☐	☐	☐	☐	☐	☐
When I talk about this community, I usually say 'we' rather than 'they'.	☐	☐	☐	☐	☐	☐	☐
When someone criticizes the community, it feels like a personal insult.	☐	☐	☐	☐	☐	☐	☐
When someone praises this community, it feels like a personal compliment.	☐	☐	☐	☐	☐	☐	☐
If a story in the media criticized the community, I would feel embarrassed.	☐	☐	☐	☐	☐	☐	☐
This community's successes are my successes.	☐	☐	☐	☐	☐	☐	☐

21 Think about the people within the Debian community who are important sources of knowledge for you, including all individuals whom you might approach if you have a project-related problem.

How many people have you identified?

27 Please indicate to what extent you agree with the following statements:

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
I redirect people who have a problem for which I do not know the answer to members who are in a position to help.	☐	☐	☐	☐	☐	☐	☐
I assist newcomers in becoming familiar with the community.	☐	☐	☐	☐	☐	☐	☐
I consult people who might be affected by my actions.	☐	☐	☐	☐	☐	☐	☐
I lend a hand to members who express a need for help in project areas for which I am knowledgeable.	☐	☐	☐	☐	☐	☐	☐
I go out of my way to help other community members with project-related problems.	☐	☐	☐	☐	☐	☐	☐
I help community members who seek support from other members.	☐	☐	☐	☐	☐	☐	☐

28 Please indicate to what extent you agree with the following statements:

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
I try to avoid creating problems for other members when contributing to the project.	☐	☐	☐	☐	☐	☐	☐
I try to align my contributions with the shared goals of the project.	☐	☐	☐	☐	☐	☐	☐
I pay attention to how my actions and contributions will affect other members.	☐	☐	☐	☐	☐	☐	☐
I respect the behavioral rules of the community.	☐	☐	☐	☐	☐	☐	☐
I behave in a way that is aligned with the shared values of this community.	☐	☐	☐	☐	☐	☐	☐
I treat other community members with respect.	☐	☐	☐	☐	☐	☐	☐
I maintain a positive attitude when interacting with other members.	☐	☐	☐	☐	☐	☐	☐

28 Please indicate to what extent you agree with the following statements:

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
I try to avoid creating problems for other members when contributing to the project.	☐	☐	☐	☐	☐	☐	☐
I try to align my contributions with the shared goals of the project.	☐	☐	☐	☐	☐	☐	☐
I pay attention to how my actions and contributions will affect other members.	☐	☐	☐	☐	☐	☐	☐
I respect the behavioral rules of the community.	☐	☐	☐	☐	☐	☐	☐
I behave in a way that is aligned with the shared values of this community.	☐	☐	☐	☐	☐	☐	☐
I treat other community members with respect.	☐	☐	☐	☐	☐	☐	☐
I maintain a positive attitude when interacting with other members.	☐	☐	☐	☐	☐	☐	☐

29 Please indicate to what extent you agree with the following statements:

	Strongly agree	Agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Disagree	Strongly disagree
I promote the project to potential users.	☐	☐	☐	☐	☐	☐	☐
I spend time talking about the project to project outsiders.	☐	☐	☐	☐	☐	☐	☐
I actively promote the achievements of the project's community.	☐	☐	☐	☐	☐	☐	☐
I encourage potential users to use the software produced by the community.	☐	☐	☐	☐	☐	☐	☐
I think about the whole community when contributing to the project.	☐	☐	☐	☐	☐	☐	☐

<< Previous Next >>

Resume later

Newcomer experience and contributor behavior in FOSS communities - Survey0% 100%**FINAL PART - Background and education****30 How old are you?**

- ☐ 20 or younger
- ☐ 21- 25
- ☐ 26- 30
- ☐ 31- 35
- ☐ 36- 40
- ☐ 41- 45
- ☐ 46- 50
- ☐ 51 or older

31 What is your gender? (This question is optional. If you choose to leave the question blank, your answer will be counted as 'not specified').

- ☐ Female
- ☐ Male

BIBLIOGRAPHY

- Aberdour, M. (2007). Achieving quality in open source software. *IEEE Software*, 24(1), 58–65. doi: 10.1007/s10729-010-9141-8
- Adkins, C. L. (1995). Previous work experience and organizational socialization: A longitudinal examination. *The Academy of Management Journal*, 38(3), 839–862. doi: 10.2307/256748
- Agarwal, R. & Karahanna, E. (2000). Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24(4), 665–694. doi: 10.2307/3250951
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. doi: 10.1016/0749-5978(91)90020-T
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Al Marzouq, M., Zheng, L., Rong, G. & Grover, V. (2005). Open source: Concepts, benefits, and challenges. *Communications of the AIS*, 16(1), 37. Retrieved from <http://aisel.aisnet.org/cais/voll6/iss1/37/>
- Alge, B. J., Ballinger, G. A., Subrahmaniam, T. & Oakley, J. L. (2006). Information privacy in organizations: Empowering creative and extrarole performance. *Journal of Applied Psychology*, 91(1), 221–32. doi: 10.1037/0021-9010.91.1.221
- Allen, D. G. (2006, April). Do organizational socialization tactics influence newcomer embeddedness and turnover? *Journal of Management*, 32(2), 237–256. doi: 10.1177/0149206305280103
- Allen, N. J. & Meyer, J. P. (1990a). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63(1), 1–18. doi: 10.1111/j.2044-8325.1990.tb00506.x
- Allen, N. J. & Meyer, J. P. (1990b). Organizational socialization tactics: A longitudinal analysis of links to newcomers' commitment and role orientation. *Academy of Management Journal*, 33(4), 847–858. doi: 10.2307/256294
- Allen, N. J. & Meyer, J. P. (1996). Affective, continuance, and normative commitment to the organization: An examination of construct validity. *Journal of Vocational Behavior*, 49(3), 252–276. doi: 10.1006/jvbe.1996.0043
- Allen, T. D., Eby, L. T. & Lentz, E. (2006). Mentorship behaviors and mentorship quality associated with formal mentoring programs: closing the gap between research and practice. *Journal of Applied Psychology*, 91(3), 567–78. doi: 10.1037/0021-9010.91.3.567
- Allen, T. D., Eby, L. T., Poteet, M. L., Lentz, E. & Lima, L. (2004). Career benefits associated with mentoring for protégé: a meta-analysis. *Journal of Applied Psychology*, 89(1), 127–136. doi: 10.1037/0021-9010.89.1.127
- Allen, T. D. & Rush, M. C. (1998, April). The effects of organizational citizenship behavior on performance judgments: a field study and a laboratory experiment. *The Journal of applied psychology*, 83(2), 247–260. doi: 10.1037//0021-9010.83.2.247
- Alreck, P. A. & Settle, R. B. (1995). *The survey research handbook* (2nd ed.). Chicago: McGraw-Hill.
- Amrit, C. & van Hillegersberg, J. (2010). Exploring the impact of socio-technical core-periphery structures in open source software development. *Journal of Information*

- Technology*, 25(2), 216–229. doi: 10.1057/jit.2010.7
- Angeles, P. A. (1981). *Dictionary of philosophy*. New York, USA: Harper Perennial.
- Apache Software Foundation. *Meritocracy. Glossary of Apache- related terms*. (2012).
- Armstrong, J. S. & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(3), 396. doi: 10.2307/3150783
- Ashford, S. J. (1986). Feedback-seeking in individual adaptation: A resource perspective. *The Academy of Management Journal*, 29(3), 465–487. doi: 10.2307/256219
- Ashford, S. J. & Black, J. S. (1996). Proactivity During Organizational Entry: The Role of Desire for Control. *Journal of Applied Psychology*, 81(2), 199–214. doi: 10.1037//0021-9010.81.2.199
- Ashford, S. J. & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32(3), 370–398. doi: 10.1016/0030-5073(83)90156-3
- Ashforth, B. E. (1998). Socialization and newcomer adjustment: The role of organizational context. *Human Relations*, 51(7), 897–926. doi: 10.1177/001872679805100703
- Ashforth, B. E. & Mael, F. (1989). Social identity theory and the organization. *The Academy of Management Review*, 14(1), 20–39. doi: 10.2307/258189
- Ashforth, B. E. & Saks, A. M. (1996). Socialization tactics: Longitudinal effects on newcomer adjustment. *The Academy of Management Journal*, 39(1), 149–178. doi: 10.2307/256634
- Ashforth, B. E., Saks, A. M. & Lee, R. T. (1997). On the Dimensionality of Jones ' (1986) Measures of Organizational Socialization Tactics. *International Journal of Selection and Assessment*, 5(4), 200–214. doi: 10.1111/1468-2389.00061
- Ashforth, B. E., Sluss, D. & Saks, A. M. (2007). Socialization tactics, proactive behavior, and newcomer learning: Integrating socialization models. *Journal of Vocational Behavior*, 70(3), 447–462. doi: 10.1016/j.jvb.2007.02.001
- Bagozzi, R. P. & Dholakia, U. M. (2006). Open source software user communities: A study of participation in Linux user groups. *Management Science*, 52(7), 1099–1115. doi: 10.1287/mnsc.1060.0545
- Bagozzi, R. P., Dholakia, U. M. & Pearo, L. R. K. (2007). Antecedents and consequences of online social interactions. *Media Psychology*, 9(1), 77–114. doi: 10.1080/15213260709336804
- Bagozzi, R. P. & Fornell, C. (1982). Theoretical Concepts, Measurement, and Meaning. In C. Fornell (Ed.), *A second generation of multivariate analysis* (2nd ed., pp. 5–23). Westport, CT: Praeger.
- Bagozzi, R. P. & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94. doi: 10.1007/BF02723327
- Barclay, D., Thompson, R. & Higgins, C. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(2), 285–309.
- Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.
- Bassellier, G. & Benbasat, I. (2004). Business competence of information technology professionals: Conceptual development and influence on IT-business partnerships. *MIS Quarterly*, 28(4), 673–694.
- Bateman, T. S. & Organ, D. W. (1983). Job satisfaction and the good soldier: The relationship between affect and employee citizenship. *The Academy of Management Journal*, 26(4), 587–595. doi: 10.2307/255908

- Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M. & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology*, 92(3), 707–721. doi: 10.1037/0021-9010.92.3.707
- Bauer, T. N. & Green, S. G. (1994, April). Effect of newcomer involvement in work-related activities: A longitudinal study of socialization. *Journal of Applied Psychology*, 79(2), 211–23. doi: 10.1037//0021-9010.79.2.211
- Bauer, T. N. & Green, S. G. (1998). Testing the combined effects of newcomer information seeking and manager behavior on socialization. *Journal of Applied Psychology*, 83(1), 72–83. doi: 10.1037//0021-9010.83.1.72
- Bauer, T. N., Morrison, E. W. & Callister, R. R. (1998). Organizational socialization: A review and directions for future research. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management (volume 16)* (pp. 149–214). Greenwich, CT: JAI Press.
- Benbasat, I. & Weber, R. (1996). Research commentary: Rethinking diversity in information systems research. *Information Systems Research*, 7(4), 389. doi: 10.1287/isre.7.4.389
- Benkler, Y. (2002). Coase's penguin, or, Linux and "The nature of the firm". *The Yale Law Journal*, 112(3), 369. doi: 10.2307/1562247
- Benkler, Y. (2006). *The wealth of networks. How social production transforms markets and freedom*. Yale University Press.
- Benkler, Y. & Nissenbaum, H. (2006). Commons-based peer production and virtue. *Journal of Political Philosophy*, 14(4), 394–419. doi: 10.1111/j.1467-9760.2006.00235.x
- Bergami, M. & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *The British Journal of Social Psychology / The British Psychological Society*, 39(4), 555–577. doi: 10.1348/014466600164633
- Bergquist, M. & Ljungberg, J. (2001). The power of gifts: Organizing social relationships in open source communities. *Information Systems Journal*, 11(4), 305–320. doi: 10.1111/j.1365-2575.2001.00111.x
- Bernard, H. R. (1988). *Research methods in cultural anthropology*. Beverly Hills, CA: Sage Publications.
- Berry, D. M. & Moss, G. (2008). *Libre culture: Meditations on free culture*. Pygmalion Books.
- Bock, G. W., Zmud, R. W., Kim, Y. G. & Lee, J. N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 29(1), 87–111.
- Bogdan, R. & Biklen, S. K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston: Allyn & Bacon.
- Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10(2), 99–109. doi: 10.1207/s15327043hup1002_3
- Boswell, W. R., Shipp, A. J., Payne, S. C. & Culbertson, S. S. (2009). Changes in newcomer job satisfaction over time: Examining the pattern of honeymoons and hangovers. *Journal of Applied Psychology*, 94(4), 844–858. doi: 10.1037/a0014975
- Boudreau, M. C., Gefen, D. & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly*, 25(1), 1–16. doi: 10.2307/3250956

- Brodbeck, F. C. & Greitemeyer, T. (2000). Effects of individual versus mixed individual and group experience in rule induction on group member learning and group performance. *Journal of Experimental Social Psychology*, 36(6), 621–648. doi: 10.1006/jesp.2000.1423
- Buchanan II, B. (1974). Building organizational commitment: The socialization of managers in work organizations. *Administrative Science Quarterly*, 19(4), 533–546. doi: 10.2307/2391809
- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, 5(1), 3–52. doi: 10.1177/004912417600500101
- Byrne, Z. S. & Hochwarter, W. A. (2008). Perceived organizational support and performance: Relationships across levels of organizational cynicism. *Journal of Managerial Psychology*, 23(1), 54–72. doi: 10.1108/02683940810849666
- Cable, D. M. & Parsons, C. K. (2001). Socialization tactics and person-organization fit. *Personnel Psychology*, 54(1), 1–23. doi: 10.1111/j.1744-6570.2001.tb00083.x
- Carillo, K. D. & Okoli, C. (2011). Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, 48(6), 208–219. doi: 10.1016/j.im.2011.04.004
- Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology*, 91(2), 475–81. doi: 10.1037/0021-9010.91.2.475
- Chan, D. & Schmitt, N. (2000). Interindividual differences in intraindividual changes in proactivity during organizational entry: A latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology*, 85(2), 190–210. doi: 10.1037//0021-9010.85.2.190
- Chang, J., Chang, W. & Jacobs, R. (2009). Relationship between participation in communities of practice and organizational socialization in the early careers of South Korean IT employees. *Human Resource Development International*, 12(4), 407–427. doi: 10.1080/13678860903135805
- Chao, G. T., O'Leary-Kelly, A. M., Wolf, S., Klein, H. J. & Gardner, P. D. (1994). Organizational socialization: Its content and consequences. *Journal of Applied Psychology*, 79(5), 730–743. doi: 10.1037//0021-9010.79.5.730
- Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36(3), 459. doi: 10.2307/2393204
- Chen, G. (2005). Newcomer adaptation in teams: Multilevel antecedents and outcomes. *Academy of Management Journal*, 48(1), 101–116. doi: 10.5465/AMJ.2005.15993147
- Chen, G. & Klimoski, R. J. (2003). The Impact of expectations on newcomer performance in teams as mediated by work characteristics, social exchanges, and empowerment. *Academy of Management Journal*, 46(5), 591–607. doi: 10.2307/30040651
- Chen, J. & Eldridge, D. (2011). The missing link in newcomer adjustment: The role of perceived organizational support and leader-member exchange. *International Journal of Organizational Analysis*, 19(1), 71–88. doi: 10.1108/19348831111121312
- Chen, X.-P., Hui, C. & Sego, D. J. (1998). The role of organizational citizenship behavior in turnover: Conceptualization and preliminary tests of key hypotheses. *Journal of Applied Psychology*, 83(6), 922–931. doi: 10.1037//0021-9010.83.6.922
- Chengalur-Smith, I. & Nevo, S. (2010). An empirical analysis of the business value of open source infrastructure technologies. *Journal of the Association for Information*

- Systems*, 11(11), 708–729. Retrieved from <http://aisel.aisnet.org/jais/vol11/iss11/3/>
- Chengalur-Smith, I., Sidorova, A. & Daniel, S. L. (2010). Sustainability of free/libre open source projects: A longitudinal study. *Journal of the Association for Information Systems*, 11(11), 657–683. Retrieved from <http://aisel.aisnet.org/jais/vol11/iss11/5/>
- Chin, W. W. (1998a). Issues and opinion on structural equation modeling. *MIS Quarterly*, 22(1), vii–xvi.
- Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chin, W. W. (2010). How to write up and report PLS analyses. In *Handbook of partial least squares* (pp. 655–690). Springer Handbooks of Computational Statistics. doi: 10.1007/978-3-540-32827-8_29
- Chin, W. W., Marcolin, B. L. & Newsted, P. R. (2003). Partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion. *Information Systems Research*, 14(2), 189–217. doi: 10.1287/isre.14.2.189.16018
- Chiu, C., Hsu, M. & Wang, E. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3), 1872–1888. doi: 10.1016/j.dss.2006.04.001
- Chou, S. & He, M. Y. (2010). The factors that affect the performance of open source software development: The perspective of social capital and expertise integration. *Information Systems Journal*, 21(2), 195–219. doi: 10.1111/j.1365-2575.2009.00347.x
- Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64–73. doi: 10.2307/3150876
- Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16), 386–405. doi: 10.1111/j.1468-0335.1937.tb00002.x
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. doi: 10.1177/001316446002000104
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Associates, Lawrence Erlbaum.
- Colazo, J. & Fang, Y. (2010). Following the sun: Temporal dispersion and performance in open source software project teams. *Journal of the Association for Information Systems*, 11(11), 684–707. Retrieved from <http://aisel.aisnet.org/jais/vol11/iss11/4/>
- Coleman, V. I. & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resource Management Review*, 10(1), 25–44. doi: 10.1016/S1053-4822(99)00037-6
- Conway, J. M. & Huffcutt, J. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147–168. doi: 10.1177/1094428103251541
- Cooper-Thomas, H. D. & Anderson, N. (2002). Newcomer adjustment: The relationship between organizational socialization tactics, information acquisition and attitudes. *Journal of Occupational and Organizational Psychology*, 75(4), 423–437. doi: 10.1348/096317902321119583
- Cooper-Thomas, H. D. & Anderson, N. (2005). Organizational socialization: A field study into socialization success and rate. *International Journal of Selection and Assessment*, 13(2), 116–128. doi: 10.1111/j.0965-075X.2005.00306.x

- Cooper-Thomas, H. D. & Anderson, N. (2006). Organizational socialization: A new theoretical model and recommendations for future research and HRM practices in organizations. *Journal of Managerial Psychology*, 21(5), 492–516. doi: 10.1108/02683940610673997
- Cooper-Thomas, H. D., van Vianen, A. & Anderson, N. (2004). Changes in person-organization fit: The impact of socialization tactics on perceived and actual P-O fit. *European Journal of Work and Organizational Psychology*, 13(1), 52–78. doi: 10.1080/13594320344000246
- Cornford, T., Shaikh, M. & Ciborra, C. (2011). Hierarchy, laboratory and collective: unveiling Linux as innovation, machination and constitution. *Journal of the Association for Information Systems*, 11(12), 809–837.
- Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 7(10), 1–9.
- Coyle-Shapiro, J. A. (2004). Exploring organizationally directed citizenship behaviour: Reciprocity or it is my job? *Journal of Management Studies*, 41(1), 85–106. doi: 10.1111/j.1467-6486.2004.00422.x
- Creswell, J. W. & Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W. & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130. doi: 10.1207/s15430421tip3903_2
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.; J. DeVito, Ed.). New York: Harper and Row.
- Cronbach, L. J. (1971). Test Validation. In R. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, D.C.: American Council on Education.
- Croteau, A.-M. & Bergeron, F. (2001). An information technology trilogy: business strategy, technological deployment and organizational performance. *Journal of Strategic Information Systems*, 10(2), 77–99. doi: 10.1016/S0963-8687(01)00044-0
- Crowston, K., Annabi, H. & Howison, J. (2003). Defining open source software project success. In *Proceedings of the 24th international conference on information systems* (pp. 327–340). 14–17 December, Seattle, WA.
- Crowston, K., Heckman, R., Annabi, H. & Masango, C. (2005). A structural perspective on leadership in Free/Libre Open Source Software teams. In *Oss2005: Open source systems*. 11–15 July, Genova, Italy.
- Crowston, K. & Howison, J. (2005). The social structure of free and open source software development. *First Monday*, 10(2), 1–100.
- Crowston, K. & Howison, J. (2006). Assessing the health of a FLOSS community. *IEEE Computer*, 39(5), 89–91. doi: 10.1002/etc.412
- Crowston, K., Howison, J. & Annabi, H. (2006). Information systems success in free and open source software development: Theory and measures. *Software Process: Improvement and Practice*, 11(2), 123–148. doi: 10.1002/spip.259
- Crowston, K. & Scozzi, B. (2004). Coordination practices within FLOSS development teams: The bug fixing process. In *1st international workshop on computer supported activity coordination, 6th international conference on enterprise information systems* (Vol. 4, pp. 21–30). Porto, Portugal.
- Crowston, K., Wei, K., Howison, J. & Wiggins, A. (2011). Free/libre open source software development: What we know and what we do not know. *ACM Computing Surveys*(forthcoming), 1–37. doi: 10.1145/2089125.2089127

- Crowston, K., Wei, K., Li, Q., Eseryel, U. Y. & Howison, J. (2005). Coordination of free/libre open source software development. In *Proceedings of the international conference on information systems*. 11-14 December, Las Vegas, NV.
- Dahlander, L. & O'Mahony, S. (2011). Progressing to the center: Coordinating project work. *Organization Science*, 22(4), 1-19. doi: 10.1287/orsc.1100.0571
- Daniel, S., Agarwal, R. & Stewart, K. (2006). An absorptive capacity perspective of open source software development group performance. *ICIS 2006 Proceedings*, 59.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. doi: 10.2307/249008
- Denzin, N. K. & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.
- Dholakia, U., Bagozzi, R. P. & Pearo, L. (2004). A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing*, 21(3), 241-263. doi: 10.1016/j.ijresmar.2003.12.004
- Dillman, D. A. (2009). *Mail and Internet surveys: The Tailored Design Method* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Dillman, D. A., Smyth, J. D. & Christian, L. M. (2008). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- Dinh-Trong, T. T. & Bieman, J. M. (2005). The FreeBSD project: A replication case study of open source development. *IEEE Transactions on Software Engineering*, 31(6), 481-494. doi: 10.1109/TSE.2005.73
- Ducheneaut, N. (2005). Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work*, 14(4), 323-368. doi: 10.1007/s10606-005-9000-1
- Eatough, E. M., Chang, C.-H., Miloslavic, S. A. & Johnson, R. E. (2011). Relationships of role stressors with organizational citizenship behavior: A meta-analysis. *Journal of Applied Psychology*, 96(3), 619-32. doi: 10.1037/a0021887
- Ellemers, N., Kortekaas, P. & Ouwerkerk, J. W. (1999). Self categorisation, commitment to the group and group self esteem as related but distinct aspects of social identity. *European Journal of Social Psychology*, 29(2-3), 371-389. doi: 10.1002/(SICI)1099-0992(199903/05)29:2/3<371::AID-EJSP932>3.3.CO;2-L
- Endres, M. L., Endres, S. P., Chowdhury, S. K. & Alam, I. (2007). Tacit knowledge sharing, self-efficacy theory, and application to the Open Source community. *Journal of Knowledge Management*, 11(3), 92-103. doi: 10.1108/13673270710752135
- English, R. & Schweik, C. M. (2007). Identifying success and tragedy of FLOSS commons: A preliminary classification of Sourceforge. net projects. In *First international workshop on emerging trends in floss research and development*. 20-26 May 2007, Minneapolis, MN. doi: 10.1109/FLOSS.2007.9
- Erdogan, B., Kraimer, M. L. & Liden, R. C. (2004). Congruence and intrinsic career success: the compensatory roles of leader-member exchange and perceived organizational support. *Personnel Psychology*, 57, 305-332. doi: 10.1111/j.1744-6570.2004.tb02493.x
- Evans, J. R. & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195-219. doi: 10.1108/10662240510590360
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological*

- Methods*, 4, 272–299. doi: 10.1037//1082-989X.4.3.272
- Fang, Y. & Neufeld, D. (2009). Understanding sustained participation in open source software projects. *Journal of Management Information Systems*, 25(4), 9–50. doi: 10.2753/MIS0742-1222250401
- Farh, J. L., Earley, P. C. & Lin, S. C. (1997). Impetus for action: A cultural analysis of justice and organizational citizenship behavior in Chinese society. *Administrative Science Quarterly*, 42(3), 421–444. doi: 10.2307/2393733
- Feeny, D., Berkes, F., McCay, B. B. & Acheson, J. M. (1990). The tragedy of the Commons: Twenty-two years later. *Human Ecology*, 18(1), 1–19. doi: 10.1007/BF00889070
- Feldman, D. (1976). A practical program for employee socialization. *Organizational Dynamics*, 5(2), 64–80. doi: 10.1016/0090-2616(76)90055-3
- Feldman, D. (1981). The multiple socialization of organization members. *Academy of Management Review*, 6(2), 309–318. doi: 10.2307/257888
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Field, A. (2010). *Discovering Statistics Using SPSS* (3rd ed.). Sage Publications Ltd.
- Fielding, R. T. (1999). Shared leadership in the Apache project. *Communications of the ACM*, 42(4), 42–43. doi: 10.1145/299157.299167
- Fisher, C. D. (1985). Social support and adjustment to work: A longitudinal study. *Journal of Management*, 11(3), 39–53. doi: 10.1177/014920638501100304
- Fisher, C. D. (1986). Organizational socialization: An integrative review. *Research in Personnel and Human Resources Management*, 4, 101–145.
- Fisher, R. J. & Katz, J. E. (2000). Social-desirability bias and validity of self-reported values. *Psychology and Marketing*, 17, 105–120.
- Fitzgerald, B. (2006). The Transformation of Open Source Software. *MIS Quarterly*, 30(3), 587.
- Fogel, K. (1999). *Open Source Development with CVS*. Scottsdale, AZ: Coriolis Open Press.
- Follett, M. P. (2003). The Early Sociology of Management and Organizations, Book 3. In H. C. Metcalf & L. Urwick (Eds.), *Dynamic administration: The collected papers of Mary Parker Follett*. Routledge, N.Y.
- Ford, J. K., MacCallum, R. C. & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(1), 291–314. doi: 10.1111/j.1744-6570.1986.tb00583.x
- Fornell, C. & Cha, J. (1994). Partial least squares. In R. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 52–78). Cambridge, MA: Basil Blackwell.
- Fornell, C. & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18, 39–50. doi: 10.2307/3150980
- Gacek, C. & Arief, B. (2004). The many meanings of open source. *IEEE Software*, 21(1), 34–40. doi: 10.1109/MS.2004.1259206
- Ge, J., Su, X. & Zhou, Y. (2010). Organizational socialization, organizational identification and organizational citizenship behavior: An empirical research of Chinese high-tech manufacturing enterprises. *Nankai Business Review International*, 1(2), 166–179. doi: 10.1108/20408741011052573
- Gefen, D. & Rigdon, E. E. (2011). Editor's comment: An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly*, 35(2).
- Gefen, D. & Straub, D. W. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the AIS*, 16, 91–109.

- Retrieved from <http://aisel.aisnet.org/cais/vol16/iss1/5/>
- Gefen, D., Straub, D. W. & Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4(1), 7. Retrieved from <http://aisel.aisnet.org/cais/vol4/iss1/7/>
- Geisser, S. (1975). A predictive approach to the random effect model. *Biometrika*, 61(1), 101–107. doi: 10.2307/2334290
- Geller, D. & Bamberger, P. (2012). The impact of help seeking on individual task performance: the moderating effect of help seekers' logics of action. *Journal of Applied Psychology*, 97(2), 487–97. doi: 10.1037/a0026014
- George, J. M. & Brief, A. P. (1992). Feeling good-doing good: A conceptual analysis of the mood at work-organizational spontaneity relationship. *Psychological Bulletin*, 112(2), 310–29. doi: 10.1037//0033-2909.112.2.310
- George, J. M. & Jones, G. R. (1997). Organizational spontaneity in context. *Human Performance*, 10(2), 153–170. doi: 10.1207/s15327043hup1002_6
- Ghosh, R. A., Glott, R., Krieger, B. & Robles, G. (2002). *Free/libre and open source software: Survey and study: Survey of developers* (No. June). Retrieved from <http://www.infonomics.nl/FL0SS/report/Final4.htm>
- Gorden, W. I., Anderson, C. M. & Bruning, S. D. (1992, March). Employee perceptions of corporate partnership: An affective-moral quid pro quo. *Employee Responsibilities and Rights Journal*, 5(1), 75–85. doi: 10.1007/BF01407810
- Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68, 532–560. doi: 10.1207/s15327752jpa6803_5
- Graham, J. W. (1986). Organizational citizenship informed by political theory. In *Paper presented at academy of management meetings*. Chicago, IL.
- Graham, J. W. (1991, December). An essay on organizational citizenship behavior. *Employee Responsibilities and Rights Journal*, 4(4), 249–270. doi: 10.1007/BF01385031
- Grant, E. S. & Bush, A. J. (1996). Salesforce socialization tactics: Building organizational value congruence. *Journal of Personal Selling and Sales Management*, 16(3), 17–32.
- Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 30(3), 611.
- Greguras, G. J. & Ford, J. M. (2006). An examination of the multidimensionality of supervisor and subordinate perceptions of leader-member exchange. *Journal of Occupational and Organizational Psychology*, 79(3), 433–465. doi: 10.1348/096317905X53859
- Grewal, R., Lilien, G. L. & Mallapragada, G. (2006). Location, location, location: How network embeddedness affects project success in open source systems. *Management Science*, 52(7), 1043–1056. doi: 10.1287/mnsc.1060.0550
- Gruman, J., Saks, A. & Zweig, D. (2006). Organizational socialization tactics and newcomer proactive behaviors: An integrative study. *Journal of Vocational Behavior*, 69(1), 90–104. doi: 10.1016/j.jvb.2006.03.001
- Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development*, 29(2), 75–91.
- Guba, E. & Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (Vol. 2, pp. 163–194). London: Sage.
- Gwebu, K. L. & Wang, J. (2011). Adoption of open source software : The role of social identification. *Decision Support Systems*, 51(1), 220–229. doi: 10.1016/j.dss.2010

.12.010

- Haenlein, M. & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics*, 3(4), 283–297. doi: 10.1207/s15328031us0304_4
- Hahn, J., Moon, J. Y. & Zhang, C. (2008). Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Information Systems Research*, 19(3), 369–391. doi: 10.1287/isre.1080.0192
- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Prentice Hall.
- Hair, J. F., Ringle, J. F. & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *The Journal of Marketing Theory and Practice*, 19(2), 139–152. doi: 10.2753/MTP1069-6679190202
- Hann, I., Roberts, J. & Slaughter, S. (2004). Why developers participate in open source software projects: An empirical investigation. In *Proceedings of the 25th international conference on information systems*. 9–12 December, Charlottesville, Virginia.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243–1248.
- Hars, A. & Ou, S. (2002). Working for free? Motivations for participating in open-source projects. *International Journal of Electronic Commerce*, 6(3), 25–39.
- Hayton, J. C., Allen, D. G. & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205. doi: 10.1177/1094428104263675
- Henseler, J., Ringle, C. M. & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*(20), 277–320.
- Herraiz, I., Robles, G., Amor, J. J., Romera, T. & González Barahona, J. M. (2006). The processes of joining in global distributed software projects. In *Proceedings of the 2006 international workshop on global software development for the practitioner* - (pp. 27–33). New York, New York, USA: ACM Press. doi: 10.1145/1138506.1138513
- Higgins, M. & Kram, K. E. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *The Academy of Management Review*, 26(2), 264–288. doi: 10.2307/259122
- Himanen, P. (2002). *The hacker ethic*. New York, New York, USA: Random House Trade Paperbacks.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104–121. doi: 10.1177/109442819800100106
- Hirschheim, R. (1985). Information systems epistemology: An historical perspective. In *Research methods in information systems* (Vol. 25, pp. 13–35). New York, USA: North-Holland, Amsterdam.
- Hirschheim, R., Klein, H. & Lyytinen, K. (1995). *Information systems development and data modeling: conceptual and philosophical foundations*. Cambridge, UK: Cambridge university press.
- Holland, R. G., Smith, A., Stephen, F., Hasselback, J. R. & Payne, B. (2010). Survey responses: Mail versus email solicitations. *Journal of Business & Economics Research*, 8(4), 95–98.
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Counseling and Clinical Psychology*, 65(4), 599–610. doi: 10.1037//0022-006X.65.4.599

- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, 20(2), 195–204. doi: 10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.3.CO;2-Z
- Iivari, J., Hirschheim, R. & Klein, H. K. (1998). Information systems development approaches and methodologies. *Information Systems Research*, 9(2), 164–193. doi: 10.1287/isre.9.2.164
- Islam, G. & Zyphur, M. J. (2009). Rituals in organizations: A review and expansion of current theory. *Group & Organization Management*, 34(1), 114–139. doi: 10.1177/1059601108329717
- Jarvenpaa, S. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science*, 35(3), 285–303. doi: 10.1287/mnsc.35.3.285
- Jensen, C. & Scacchi, W. (2005). Modeling recruitment and role migration processes in OSSD projects. In *Proceedings of 6th international workshop on software process simulation and modeling*. 14–15 May, Institute for Software Research, University of California, St. Louis, CA.
- Jensen, C. & Scacchi, W. (2007). Role migration and advancement processes in OSSD projects: A comparative case study. In *29th international conference on software engineering (icse'07)* (pp. 364–374). 20–26 May, Minneapolis, MN: IEEE. doi: 10.1109/ICSE.2007.74
- Johns, G. & Saks, A. M. (2001). Social influence, socialization, and culture. In *Organizational behaviour: Understanding and managing life at work* (5th ed.). Reading, MA: Addison-Wesley.
- Jones, G. R. (1986, June). Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *The Academy of Management Journal*, 29(2), 262–279. doi: 10.2307/256188
- Kammeyer-Mueller, J. D. & Judge, T. (2008). A quantitative review of mentoring research: Test of a model. *Journal of Vocational Behavior*, 72(3), 269–283. doi: 10.1016/j.jvb.2007.09.006
- Kammeyer-Mueller, J. D. & Wanberg, C. R. (2003). Unwrapping the organizational entry process: disentangling multiple antecedents and their pathways to adjustment. *The Journal of Applied Psychology*, 88(5), 779–94. doi: 10.1037/0021-9010.88.5.779
- Kanter, R. M. (1968). Commitment and social organization: A study of commitment mechanisms in utopian communities. *American Sociological Review*, 33(4), 499–517. doi: 10.1684/san.2010.0206
- Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral science*, 9(2), 131–46. doi: 10.1002/bs.3830090206
- Katz, D. & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York: Wiley.
- Katz, D. & Khan, R. L. (1966). *The social psychology of organizations*. New York: Wiley.
- Ke, W. & Zhang, P. (2009). Motivations in open source software communities: The mediating role of effort intensity and goal commitment. *International Journal of Electronic Commerce*, 13(4), 39–66. doi: 10.2753/JEC1086-4415130403
- Ke, W. & Zhang, P. (2010). The effects of extrinsic motivations and satisfaction in open source software development. *Journal of the Association for Information Systems*, 11(12), 784–808. Retrieved from <http://aisel.aisnet.org/jais/vol11/iss12/5/>

- Kim, T.-Y., Cable, D. M. & Kim, S.-P. (2005). Socialization tactics, employee proactivity, and person-organization fit. *Journal of Applied Psychology*, 90(2), 232–241. doi: 10.1037/0021-9010.90.2.232
- Klein, H. J. & Weaver, N. A. (2000). The effectiveness of an organizational-level orientation training program in the socialization of new hires. *Personnel Psychology*, 53(1), 47–66. doi: 10.1111/j.1744-6570.2000.tb00193.x
- Knippenberg, D. & Schie, E. C. M. (2000). Foci and correlates of organizational identification. *Journal of Occupational and Organizational Psychology*, 73(2), 137–147. doi: 10.1348/096317900166949
- Konovsky, M. A. & Organ, D. W. (1996). Dispositional and contextual determinants of organizational citizenship behavior. *Journal of Organizational Behavior*, 17(3), 253–266. doi: 10.1002/(SICI)1099-1379(199605)17:3<253::AID-JOB747>3.3.CO;2-H
- Korte, R. (2010). First, get to know them: A relational view of organizational socialization. *Human Resource Development International*, 13(1), 27–43. doi: 10.1080/13678861003588984
- Kram, K. E. (1983). Phases of the mentor relationship. *Academy of Management Journal*, 26(4), 608–625. doi: 10.2307/255910
- Kram, K. E. & Isabella, L. (1985). Mentoring alternatives: The role of peer relationships in career development. *Academy of Management Journal*, 28(1), 110–132. doi: 10.2307/256064
- Kraut, R. E., Burke, M. & Riedl, J. (2011). Dealing with newcomers. In *Evidence-based social design: Mining the social sciences to build online communities* (pp. 1–45). Cambridge, MA: MIT Press.
- Kraut, R. E. & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, 28(3), 377–408. doi: 10.1177/0170840607076007
- Kraut, R. E. & Resnick, P. (2011). *Evidence-based social design: Mining the social sciences to build online communities*. Cambridge, MA: MIT Press. In preparation.
- Kraut, R. E., Riedl, J. & Burke, M. (2011). Dealing with newcomers. In R. Kraut & P. Resnick (Eds.), *Designing from theory: Using the social sciences as the basis for building online communities*. MIT Press.
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy*, 45(3), 214–222. doi: 10.5014/ajot.45.3.214
- Kreiner, G. E. & Ashforth, B. E. (2004). Evidence toward an expanded model of organizational identification. *Journal of Organizational Behavior*, 25(1), 1–27. doi: 10.1002/job.234
- Kuk, G. (2006). Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Science*, 52(7), 1031–1042. doi: 10.1287/mnsc.1060.0551
- Kvale, D. (1996). *InterViews: An Introduction to Qualitative Research Interviewing*. London: Sage Publications.
- Lakhani, K. & Wolf, R. G. (2007). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In J. Feller, B. Fitzgerald, S. A. Hissam & K. R. Lakhani (Eds.), *Perspectives on free and open source software*. Boston, MA: MIT Press.
- Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi: 10.2307/2529310
- Landry, M. & Banville, C. (1992). A disciplined methodological pluralism for MIS research. *Accounting, Management and Information Systems*, 2(2), 77–97. doi: 10

- .1016/0959-8022(92)90002-A
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society* (M. Keynes, Ed.). Open University Press.
- Laughlin, P. & Sweeney, J. (1977). Individual to group and group to individual transfer in problem solving. *Journal of Experimental Psychology: Human Learning and Memory*, 3(2), 246–254. doi: 10.1037//0278-7393.3.2.246
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge university press.
- Lederer, A. L. & Sethi, V. (1992). Root causes of strategic information systems planning implementation problems. *Journal of Management Information Systems*, 9(1), 25–45.
- Lee, G. L. & Cole, R. E. (2003). From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, 14(6), 633–649. doi: 10.1287/orsc.14.6.633.24866
- Lee, K. & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology*, 87(1), 131–142. doi: 10.1037//0021-9010.87.1.131
- Lee, S.-Y. T., Kim, H.-W. & Gupta, S. (2009). Measuring open source software success. *Omega*, 37(2), 426–438. doi: 10.1016/j.omega.2007.05.005
- Lee, T., Mitchell, T. & Sablinski, C. (2004). The effects of job embeddedness on organizational citizenship, job performance, volitional absences, and voluntary turnover. *The Academy of Management Journal*, 47(5), 711–722. doi: 10.2307/20159613
- Lepine, J. A., Erez, A. & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, 87(1), 52. doi: 10.1037//0021-9010.87.1.52
- Lepine, J. A. & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology*, 83(6), 853–868. doi: 10.1037//0021-9010.83.6.853
- Lerner, J. & Tirole, J. (2001). The open source movement: Key research questions. *European Economic Review*, 45(4-6), 819–826. doi: 10.1016/S0014-2921(01)00124-6
- Levinson, D. J., Darrow, C. M., Klein, E. G., Levinson, M. H. & McKee, B. (1978). *Seasons of a man's life*. New York: Knopf.
- Levy, S. H. (1984). *Heroes of the Computer Revolution* Anchor Press. Doubleday Garden City, NY.
- Lewis, B. R., Snyder, C. A. & Rainer, R. (1995). An empirical assessment of the information resource management construct. *Journal of Management Information Systems*, 12(1), 199–223.
- Lewis, B. R., Templeton, G. F. & Byrd, T. A. (2005, December). A methodology for construct development in MIS research. *European Journal of Information Systems*, 14(4), 388–400. doi: 10.1057/palgrave.ejis.3000552
- Liang, H., Saraf, N., Hu, Q. & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, 31(1), 59–87.
- Lin, N., Ensel, W. M. & Vaughn, J. C. (1981). Social resources and occupational status attainment. *Social Forces*, 59, 1163–1181. doi: 10.2307/2577987
- Lincoln, Y. S. & Guba, E. A. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Liu, X. & Iyer, B. (2007). Design architecture, developer networks and performance of open source software projects. In *Proceedings of 28th international conference on information systems*. 9-12 Decembre, Montreal, Canada.

- Lohmöller, J. (1989). *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica-Verlag.
- Lopez, T. & McMillan-Capehart, A. (2009). Elements of salesperson control: an organization theory perspective. *Journal of Business & Industrial Marketing*, 24(2), 98–107. doi: 10.1108/08858620910931712
- Louis, M. R. (1980). Surprise and sense making: What newcomers experience in entering unfamiliar organizational settings. *Administrative Science Quarterly*, 25(2), 226–251. doi: 10.1300/J111Vo4No1_08
- Louis, M. R., Posner, B. Z. & Powell, G. N. (1983, December). The availability and helpfulness of socialization practices. *Personnel Psychology*, 36(4), 857–866. doi: 10.1111/j.1744-6570.1983.tb00515.x
- Luong, A. & Rogelberg, S. G. (1998). How to increase your survey response rate. *The Industrial Organizational Psychologist*, 36, 61–65.
- Lynch, P. D., Eisenberger, R. & Armeli, S. (1999). Perceived organizational support: Inferior versus superior performance by wary employees. *Journal of Applied Psychology*, 84(4), 467–483. doi: 10.1037//0021-9010.84.4.467
- Ma, M. & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, 18(1), 42–67. doi: 10.1287/isre.1070.0113
- MacKenzie, S., Podsakoff, P. & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons' performance. *Organizational Behavior and Human Decision Processes*, 50(1), 123–150. doi: 10.1016/0749-5978(91)90037-T
- MacKenzie, S., Podsakoff, P. & Fetter, R. (1993). The impact of organizational citizenship behavior on evaluations of salesperson performance. *The Journal of Marketing*, 57(1), 70–80. doi: 10.2307/1252058
- MacKenzie, S., Podsakoff, P. & Paine, J. (1999). Do citizenship behaviors matter more for managers than for salespeople? *Journal of the Academy of Marketing Science*, 27(4), 396–410. doi: 10.1177/0092070399274001
- Mael, F. & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, 13(2), 103–123. doi: 10.1002/job.4030130202
- Major, D. a. & Kozlowski, S. W. (1997). Newcomer information seeking: Individual and contextual influences. *International Journal of Selection and Assessment*, 5(1), 16–28. doi: 10.1111/1468-2389.00042
- Malhotra, N. K. (2004). Marketing Research - An Applied Orientation. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 176–202). Thousand Oaks, California, USA: Sage Publications.
- Marcoulides, G. A., Chin, W. W. & Saunders, C. (2009). A critical look at partial least squares modeling. *MIS Quarterly*, 33(1), 171–175.
- Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.
- Méndez-Durón, R. & García, C. E. (2008). Returns from social capital in open source software networks. *Journal of Evolutionary Economics*, 19(2), 277–295. doi: 10.1007/s00191-008-0125-5
- Mertens, D. M. (2009). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (3rd ed.). Sage.
- Miles, M. & Huberman, A. (1994). *An expanded sourcebook qualitative data analysis*. Thousand Oaks, California, USA: Sage Publications.

- Miller, V. & Jablin, F. (1991). Information seeking during organizational entry: Influences, tactics, and a model of the process. *The Academy of Management Review*, 16(1), 92–120. doi: 10.2307/258608
- Mockus, A., Fielding, R. & Herbsleb, J. (2000). A case study of open source software development: the Apache server. In *Proceedings of the international conference on software engineering* (pp. 263–272). June, Limerick, Ireland: ACM Press. doi: 10.1109/ICSE.2000.870417
- Mockus, A., Fielding, R. & Herbsleb, J. D. (2002). Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, 11(3), 309–346. doi: 10.1145/567793.567795
- Moerke, K. A. (2000). Free speech to a machine. *Minnesota Law Review*, 84(4), 1007–1008.
- Moody, G. (2002). *Rebel code: Inside Linux and the open source revolution*. New York, USA: Basic Book.
- Moore, G. C. & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192–222. doi: 10.1287/isre.2.3.192
- Moorman, R. & Blakely, G. (1995). Individualism-collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior*, 16(2), 127–142. doi: 10.1002/job.4030160204
- Morrison, E. W. (1993a). Longitudinal study of the effects of information seeking on newcomer socialization. *Journal of Applied Psychology*, 78(2), 173–183. doi: 10.1037//0021-9010.78.2.173
- Morrison, E. W. (1993b). Newcomer information seeking: Exploring types, modes, sources, and outcomes. *The Academy of Management Journal*, 36(3), 557–589. doi: 10.2307/256592
- Morrison, E. W. (1994). Role definitions and organizational citizenship behavior: The importance of the employee's perspective. *The Academy of Management Journal*, 37(6), 1543–1567. doi: 10.2307/256798
- Morrison, E. W. (2002a). Information seeking within organizations. *Human Communication Research*, 28(2), 229–242. doi: 10.1093/hcr/28.2.229
- Morrison, E. W. (2002b). Newcomers' relationships: The role of social network ties during socialization. *The Academy of Management Journal*, 45(6), 1149–1160. doi: 10.2307/3069430
- Moser, C. A. (1958). *Survey methods in social investigation* (Heinemann, Ed.). London, UK: The English Language Book Society.
- Motowidlo, S. J. & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79(4), 475–480. doi: 10.1037//0021-9010.79.4.475
- Nahapiet, J. & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *The Academy of Management Review*, 23(2), 242–266. doi: 10.2307/259373
- Nakakoji, K., Yamamoto, Y., Nishinaka, Y., Kishida, K. & Ye, Y. (2002). Evolution patterns of open-source software systems and communities. In *Proceedings of the international workshop on principles of software evolution* (pp. 76–85). January, Orlando, Florida, USA: ACM Press. doi: 10.1145/512054.512055
- Nelson, M., Sen, R. & Subramaniam, C. (2006). Understanding open source software: A research classification framework. *Communications of the AIS*, 17(1), 2–37. Retrieved from <http://aisel.aisnet.org/cais/vol17/iss1/12/>

- Netemeyer, R. G., Boles, J. S., McKee, D. O., McMurrian, R. & McKee, D. (1997). An investigation of organizational antecedents behaviors in a personal citizenship context selling. *The Journal of Marketing*, 61(3), 85–98. doi: 10.2307/1251791
- Niehoff, B. P. & Moorman, R. H. (1993). Justice as a Mediator of the Relationship between Methods of Monitoring and Organizational Citizenship Behavior. *Academy of Management Journal*, 36(3), 527–556. doi: 10.2307/256591
- Noe, R. A. (1988). An investigation of the determinants of successful assigned mentoring relationships. *Personnel Psychology*, 41(3), 457–479. doi: 10.1111/j.1744-6570.1988.tb00638.x
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw Hil.
- Oh, W. & Jeon, S. (2007). Membership herding and network stability in the open source community: The Ising perspective. *Management Science*, 53(7), 1086–1101. doi: 10.1287/mnsc.1060.0623
- Okoli, C. & Oh, W. (2007). Investigating recognition-based performance in an open content community: A social capital perspective. *Information & Management*, 44(3), 240–252.
- O’Leary, Z. (2004). *The Essential Guide to Doing Research* (1st ed.). Sage Publications.
- Olivera, F. & Straus, S. G. (2004). Group-to-individual transfer of learning: Cognitive and social factors. *Small Group Research*, 35(4), 440–465. doi: 10.1177/1046496404263765
- O’Reilly, C., Caldwell, D. & Barnett, W. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly*, 34, 21–37.
- O’Reilly, C. & Chatman, J. (1986). Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. *Journal of Applied Psychology*, 71(3), 492–499. doi: 10.1037//0021-9010.71.3.492
- O’Reilly, T. (1999). Lessons from open-source software development. *Communications of the ACM*, 42(4), 33–37. doi: 10.1145/299157.299164
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Organ, D. W. (1990a). The motivational basis of organizational citizenship behavior. *Research in Organizational Behavior*, 12(1), 43–72.
- Organ, D. W. (1990b). The motivational basis of organizational citizenship behavior. *Research in organizational behavior*, 43–72.
- Organ, D. W. (1997). Organizational citizenship behavior: It’s construct clean-up time. *Human performance*, 10(2), 85–97. doi: 10.1207/s15327043hup1002_2
- Organ, D. W. & Konovsky, M. (1989). Cognitive versus affective determinants of organizational citizenship behavior. *Journal of Applied Psychology*, 74(1), 157–164. doi: 10.1037//0021-9010.74.1.157
- Organ, D. W. & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behaviour. *Personnel Psychology*, 48(4), 775–802. doi: 10.1111/j.1744-6570.1995.tb01781.x
- Orlikowski, W. J. & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, 2(1), 1–28. doi: 10.1287/isre.2.1.1
- Orlikowski, W. J. & Iacono, C. S. (2001). Research Commentary: Desperately Seeking the IT in IT Research - A Call to Theorizing the IT Artifact. *Information Systems Research*, 12(2), 121–134. doi: 10.1287/isre.12.2.121.9700

- Ostroff, C. & Kozlowski, S. (1992). Organizational socialization as a learning process: The role of information acquisition. *Personnel Psychology*, 45(4), 849–874. doi: 10.1111/j.1744-6570.1992.tb00971.x
- Parasuraman, A. & Zinkhan, G. M. (2002). Marketing to and serving customers through the Internet: An overview and research agenda. *Journal of the Academy of Marketing Science*, 30(4), 286–295. doi: 10.1177/009207002236906
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Peter, J. P. (1981). Construct validity: a review of basic issues and marketing practices. *Journal of Marketing Research*, 18(2), 133–145.
- Pinsonneault, A. & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. *Journal of Management Information Systems*, 10(2), 75–105.
- Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M. & Blume, B. D. (2009). Individual and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 94(1), 122–41. doi: 10.1037/a0013079
- Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M. & Mishra, P. (2011). Effects of organizational citizenship behaviors on selection decisions in employment interviews. *Journal of Applied Psychology*, 96(2), 310–26. doi: 10.1037/a0020948
- Podsakoff, P., MacKenzie, S. B. & Ahearne, M. (1997). Moderating effects of goal acceptance on the relationship between group cohesiveness and productivity. *Journal of Applied Psychology*, 82(6), 374–983. doi: 10.1037//0021-9010.82.6.374
- Podsakoff, P. M., Ahearne, M. & MacKenzie, S. B. (1997). Organizational citizenship behavior and the quantity and quality of work group performance. *Journal of Applied Psychology*, 82(2), 262–270. doi: 10.1037//0021-9010.82.2.262
- Podsakoff, P. M. & MacKenzie, S. (1994). Organizational citizenship behaviors and sales unit effectiveness. *Journal of Marketing Research*, 31(3), 351–363.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. doi: 10.1037/0021-9010.88.5.879
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B. & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and future research. *Journal of Management*, 26(3), 513–563. doi: 10.1016/S0149-2063(00)00047-7
- Podsakoff, P. M., Scott, B. M., Moorman, R. H. & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, 1(2), 107–142. doi: 10.1016/1048-9843(90)90009-7
- Porter, L., Lawler, E. & Hackman, J. (1975). *Behavior in organizations*. New York. McGraw-Hill.
- Qureshi, I. & Fang, Y. (2010). Socialization in open source software projects: A growth mixture modeling approach. *Organizational Research Methods*, 14(1), 208–238. doi: 10.1177/1094428110375002
- Raghunathan, S., Prasad, A., Mishra, B. & Chang, H. (2005). Open source versus closed source: Software quality in monopoly and competitive markets. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 35(6), 903–918. doi: 10.1109/TSMCA.2005.853493
- Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. Sebastopol, CA: O' Reilly. doi: 10.1007/s12130-999

-1026-0

- Reinartz, W., Krafft, M. & Hoyer, W. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research*, 41(3), 293–305. doi: 10.1509/jmkr.41.3.293.35991
- Rhoades, L. & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, 87(4), 698–714. doi: 10.1037//0021-9010.87.4.698
- Rhoades, L., Eisenberger, R. & Armeli, S. (2001). Affective commitment to the organization: The contribution of perceived organizational support. *Journal of Applied Psychology*, 86(5), 825–836. doi: 10.1037//0021-9010.86.5.825
- Riehle, D., Ellenberger, J., Menahem, T., Mikhailovski, B., Natchetoi, Y., Naveh, B. & Odenwald, T. (2009). Open collaboration within corporations using software forges. *IEEE Software*, 52. doi: 10.1109/MS.2009.44
- Ringle, C., Sarstedt, M. & Straub, D. W. (2012). Editor's comments: a critical look at the use of PLS-SEM in MIS quarterly. *MIS Quarterly*, 36(1), 1–18.
- Roberts, J. A., Hann, I.-H. & Slaughter, S. A. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science*, 52(7), 984–999. doi: 10.1287/mnsc.1060.0554
- Robey, D. (1996). Research commentary: diversity in information systems research: threat, promise, and responsibility. *Information Systems Research*, 7(4), 400–408. doi: 10.1287/isre.7.4.400
- Robinson, S. & Morrison, E. W. (1995). Psychological contracts and OCB: The effect of unfulfilled obligations on civic virtue behavior. *Journal of Organizational Behavior*, 16(3), 289–299. doi: 10.1002/job.4030160309
- Robles, G., Gonzalez-Barahona, J. & Michlmayr, M. (2005). Evolution of volunteer participation in libre software projects: Evidence from Debian. In *Proceedings of the first international conference on open source systems* (pp. 100–107). July, Genova, Italy.
- Rosemann, M. & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: the role of applicability checks. *MIS Quarterly*, 32(1), 1–22.
- Saks, A. M. (1995). Longitudinal field investigation of the moderating and mediating effects of self-efficacy on the relationship between training and newcomer adjustment. *Journal of Applied Psychology*, 80(2), 211–225. doi: 10.1037//0021-9010.80.2.211
- Saks, A. M. & Ashforth, B. (1997a). Organizational socialization: Making sense of the past and present as a prologue for the future. *Journal of Vocational Behavior*, 51(2), 234–279. doi: 10.1006/jvbe.1997.1614
- Saks, A. M. & Ashforth, B. E. (1997b). Socialization tactics and newcomer information acquisition. *International Journal of Selection and Assessment*, 5(1), 48–61. doi: 10.1111/1468-2389.00044
- Saks, A. M. & Gruman, J. a. (2011). Getting newcomers engaged: the role of socialization tactics. *Journal of Managerial Psychology*, 26(5), 383–402. doi: 10.1108/026839411111139001
- Saks, A. M., Uggerslev, K. & Fassina, N. (2007). Socialization tactics and newcomer adjustment: A meta-analytic review and test of a model. *Journal of Vocational Behavior*, 70(3), 413–446. doi: 10.1016/j.jvb.2006.12.004

- Scacchi, W. (2007). Free/open source software development: Recent research results and methods. *Advances in Computers*, 69, 243–295.
- Sen, R. (2007). A strategic analysis of competition between open source and proprietary software. *Journal of Management Information Systems*, 24(1), 233–257. doi: 10.2753/MIS0742-1222240107
- Sethi, V. & King, W. R. (1991). Construct measurement in information systems research: an illustration in strategic systems. *Decision Sciences*, 22(4), 455–472. doi: 10.1111/j.1540-5915.1991.tb01274.x
- Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, 52(7), 1000–1014. doi: 10.1287/mnsc.1060.0553
- Shaikh, M. & Cornford, T. (2003). Version management tools: CVS to BK in the Linux kernel. In *3rd workshop on open source software engineering* (pp. 127–131). 3–10 May, Portland, Oregon, USA: Citeseer.
- Shibuya, B. & Tamai, T. (2009). Understanding the process of participating in open source communities. In *Workshop on emerging trends in free/libre/open source software research and development* (pp. 1–6). 18 May, Vancouver, BC, Canada. doi: 10.1109/FLOSS.2009.5071352
- Sidorova, A. & Evangelopoulos, N. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, 32(3), 1–20.
- Siegel, S. & Castellan, N. J. (1988). *Nonparametric statistics for the social sciences*. New York: McGraw-Hill.
- Simon, E. (1996). Innovation and intellectual property protection: The software industry perspective. *The Columbia Journal of World Business*, 31(1), 30–37.
- Simosi, M. (2010). The role of social socialization tactics in the relationship between socialization content and newcomers' affective commitment. *Journal of Managerial Psychology*, 25(3), 301–327. doi: 10.1108/02683941011023758
- Simsek, Z. & Veiga, J. F. (2000). The electronic survey technique: An integration and assessment. *Organizational Research Methods*, 3(1), 93–115. doi: 10.1177/109442810031004
- Simsek, Z. & Veiga, J. F. (2001). A Primer on internet organizational surveys. *Organizational Research Methods*, 4(3), 218–235. doi: 10.1177/109442810143003
- Singh, P., Tan, Y. & Mookerjee, V. (2011). Network effects: The influence of structural social capital on open source project success. *MIS Quarterly*, 35(4), 813–829.
- Smith, C. A., Organ, D. W. & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, 68(4), 653–663. doi: 10.1037//0021-9010.68.4.653
- Sowe, S., 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), 431–446. doi: 10.1016/j.jss.2007.03.086
- Sparrowe, R. & Liden, R. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal*, 44(14), 216. doi: 10.2307/3069458
- Stallman, R. (2002). *Free software, free society: Selected essays of Richard M. Stallman*. Cambridge, MA: GNU Press.
- Stamper, C. L. & Masterson, S. S. (2002). Insider or outsider? How employee perceptions of insider status affect their work behavior. *Journal of Organizational Behavior*, 23(8), 875–894. doi: 10.1002/job.175
- Stanton, J. M. & Rogelberg, S. G. (2001). Using internet/intranet web pages to collect organizational research data. *Organizational Research Methods*, 4(3), 200–217. doi:

- 10.1177/109442810143002
- Stewart, K. J., Ammeter, A. P. & Maruping, L. M. (2006). Impacts of license choice and organizational sponsorship on user interest and development activity in open source software projects. *Information Systems Research*, 17(2), 126–144. doi: 10.1287/isre.1060.0082
- Stewart, K. J., Darcy, D. P. & Daniel, S. L. (2006). Opportunities and Challenges Applying Functional Data Analysis to the Study of Open Source Software Evolution. *Statistical Science*, 21(2), 167–178. doi: 10.1214/088342306000000141
- Stewart, K. J. & Gosain, S. (2006, June). The impact of ideology on effectiveness in open source software development teams. *MIS Quarterly*, 30(2), 291–314.
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36, 111–147.
- Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, 13(2), 147. doi: 10.2307/248922
- Straub, D. W., Boudreau, M.-C. & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the AIS*, 13(24), 380.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications.
- Subramaniam, C., Sen, R. & Nelson, M. L. (2009). Determinants of open source software project success: A longitudinal study. *Decision Support Systems*, 46(2), 576–585. doi: 10.1016/j.dss.2008.10.005
- Subramanyam, R. & Xia, M. (2008). Free/Libre open source software development in developing and developed countries: an exploratory study. *Decision Support Systems*, 46(1), 173–186. doi: 10.1016/j.dss.2008.06.006
- Sudman, S. & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago: Aldine.
- Tabachnick, G. B. & Fidell, S. R. (2007). *Using Multivariate Statistics* (5th ed.). Pearson Education.
- Tajfel, H. & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations*, 2nd ed. (pp. 7–24). Nelson-Hall, Chicago, IL.
- Tajfel, H. & Turner, J. C. (1985). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 7–24). Chicago: Nelson-Hall.
- Tan, Y., Mookerjee, V. & Singh, P. (2007). Social capital, structural holes, and team composition: Collaborative networks of the open source software community. In *Proceedings of the 28th international conference on information systems*. 14–17 December, Paris, France.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M. & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48(1), 159–205. doi: 10.1016/j.csda.2004.03.005
- Trice, H., Belasco, J. & Alutto, J. (1969). Role of Ceremonials in Organizational Behavior. *Industrial and Labor Relations Review*, 23(1), 40–51. doi: 10.2307/2521970
- Trice, H. & Beyer, J. (1984). Studying organizational cultures through rites and ceremonials. *Academy of Management Review*, 9(4), 653–669. doi: 10.2307/258488
- Turner, J., Hogg, M., Oakes, P., Reicher, S. & Wetherell, M. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.
- Turnley, W., Bolino, M., Lester, S. W. & Bloodgood, J. M. (2003). The impact of psychological contract fulfillment on the performance of in-role and organizational citizenship behaviors. *Journal of Management*, 29(2), 187–206. doi:

- 10.1016/S0149-2063(02)00214-3
- Van Dyne, L., Graham, J. W. & Dienesch, R. M. (1994). Organizational Citizenship Behavior: Construct redefinition, measurement, and validation. *The Academy of Management Journal*, 37(4), 765. doi: 10.2307/256600
- Van Maanen, J. (1973). Observations on the making of policemen. *Human Organization*, 32, 407–417.
- Van Maanen, J. & Schein, E. H. (1979). Towards a theory of organizational socialization. In B. Staw (Ed.), *Research in organizational behavior* (Vol. 1, pp. 209–264). Greenwich, CT: Elsevier Science.
- Van Scotter, J. & Motowidlo, S. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81(5), 525–531. doi: 10.1037//0021-9010.81.5.525
- van de Mortel, T. (2008). Faking it: Social desirability response bias in self report research. *Australian Journal of Advanced Nursing*, 25(4), 40–48.
- Vanyperen, N. W., Berg, A. E. & Willering, M. C. (1999). Towards a better understanding of the link between participation in decision making and organizational citizenship behaviour: A multilevel analysis. *Journal of Occupational and Organizational Psychology*, 72(3), 377–392. doi: 10.1348/096317999166734
- Venkatesh, V., Brown, S. & Bala, H. (2013). Bridging the qualitative quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, 37(1), 1–34.
- Vinzi, V. E., Trinchera, L. & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of partial least squares* (pp. 47–83). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-32827-8
- Von Krogh, G. (2003). Open-source software development. *MIT Sloan Management Review*, 44(3), 14.
- von Hippel, E. & Von Krogh, G. (2003). Open source software and the private-collective innovation model: Issues for organization science. *Organization Science*, 14(2), 209–223. doi: 10.1287/orsc.14.2.209.14992
- von Krogh, G., Spaeth, S. & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7), 1217–1241.
- Wanberg, C. & Kammeyer-Mueller, J. (2000). Predictors and outcomes of proactivity in the socialization process. *Journal of Applied Psychology*, 85(3), 373–38. doi: 10.1037//0021-9010.85.3.373
- Wanberg, C., Kammeyermueller, J. & Marchese, M. (2006). Mentor and protégé predictors and outcomes of mentoring in a formal mentoring program. *Journal of Vocational Behavior*, 69(3), 410–423. doi: 10.1016/j.jvb.2006.05.010
- Wand, Y. & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, 3(4), 217–237. doi: 10.1111/j.1365-2575.1993.tb00127.x
- Wasko, M. & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1), 35–57.
- Welbourne, T., Johnson, D. & Erez, A. (1998). The role-based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal*, 41(5), 540–555. doi: 10.2307/256941

- Werts, C. E., Linn, R. L. & Jöreskog, K. G. (1974). Intraclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement*, 34, 25–33. doi: 10.1177/001316447403400104
- Wetzels, M. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, 33(1), 177–195.
- Wiesenfeld, B. M., Raghuram, S. & Garud, R. (2001, April). Organizational identification among virtual workers: the role of need for affiliation and perceived work-based social support. *Journal of Management*, 27(2), 213–229. Retrieved from <http://jom.sagepub.com/cgi/doi/10.1177/014920630102700205> doi: 10.1177/014920630102700205
- Williams, L. & Anderson, S. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17(3), 601–617. doi: 10.1177/014920639101700305
- Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics*, 22(2), 233–261. doi: 10.1086/466942
- Wold, H. (1966). Estimation of principal component and related models by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 391–420). New York: Academic Press.
- Wold, H. (1975). Path Models with Latent Variables: The NIPALS Approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon & V. Cappecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307–357). New York: Academic Press.
- Worchel, S., Rothgerber, H., Day, E. a., Hart, D. & Butemeyer, J. (1998). Social and identity and individual productivity within groups. *The British Journal of Social Psychology*, 37(4), 389–413. doi: 10.1111/j.2044-8309.1998.tb01181.x
- Wu, C. G., Gerlach, J. H. & Young, C. E. (2007). An empirical analysis of open source software developers motivations and continuance intentions. *Information & Management*, 44(3), 253–262. doi: 10.1016/j.im.2006.12.006
- Xu, B., Jones, D. R. & Shao, B. (2009). Volunteers' involvement in online community based software development. *Information & Management*, 46(3), 151–158. doi: 10.1016/j.im.2008.12.005
- Yang, J., Mossholder, K. W. & Peng, T. K. (2007). Procedural justice climate and group power distance: an examination of cross-level interaction effects. *Journal of Applied Psychology*, 92(3), 681–92. doi: 10.1037/0021-9010.92.3.681
- Yang, Z. & Jiang, M. (2007). A tool-integration platform for software development. *IEEE Software*, 24(2), 87–90. doi: 10.1109/MS.2007.58
- Ye, T. & Kishida, K. (2003). Toward an understanding of the motivation of open source software developers. In *Proceedings of the 25th international conference on software engineering* (pp. 419–429). 3–10 May, Portland, Oregon, USA. doi: 10.1109/ICSE.2003.1201220
- Yi, X. & Uen, J. F. (2006). Relationship between organizational socialization and organization identification of professionals: Moderating effects of personal work experience and growth need strength. *Journal of American Academy of Business, Cambridge*, 10(1), 362–371.