2020

Interaction, collaboration and content creation in informal online learning environments: multidimensional analyses of longitudinal data from the scratch coding community

Seung B. Lee
seungboklee@gmail.com

Pepperdine University

Graduate School of Education and Psychology

INTERACTION, COLLABORATION AND CONTENT CREATION IN INFORMAL ONLINE LEARNING ENVIRONMENTS: MULTIDIMENSIONAL ANALYSES OF LONGITUDINAL DATA FROM THE SCRATCH CODING COMMUNITY

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy in Global Leadership and Change

by

Seung B. Lee

October, 2020

Eric Hamilton, Ph.D. – Dissertation Chairperson

This dissertation, written by

Seung B. Lee

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Doctoral Committee:

Eric Hamilton, Ph.D., Chairperson

Andrew Ruis, Ph.D.

David Strong, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of many individuals.

I would like to express my deep gratitude to my dissertation chair, Dr. Eric Hamilton. If I had to pick the most important aspect of my doctoral studies, it would no doubt be my experience in the IC4 research team. It changed the trajectory of my time at Pepperdine, and I am so grateful to have had this opportunity to learn and grow. Thank you, Dr. Hamilton for also guiding me through the dissertation process. There were many moments when I was lost in the weeds, and you helped me to see the bigger picture and continue forward.

I am extremely grateful to the members of my dissertation committee, Dr. Andrew Ruis and Dr. David Strong, for sharing with me your time and expertise for improving my research. Your insightful feedback has helped me to broaden my perspective and push my thinking beyond my boundaries.

I want to thank Dr. Danielle Espino for all your encouragement, support and advice over the past three years. You have been the voice of positivity and reason throughout–I am so thankful to have you as a friend and colleague.

I am grateful to my family for their unconditional love and support. To my mother, thank you for always believing in me and encouraging me to follow my heart. This degree is for you mom. To my two babies, Lauren and Jay, thank you for your love and smiles. You make me happy. To the love of my life, Anna—thank you for accepting and loving me in all my faults and weaknesses. Thank you for being there for me always.

VITA

EDUCATION

Pepperdine University, Graduate School of Education and Psychology — Los Angeles, CA
Ph.D., Global Leadership and Change — Expected Fall 2020

Columbia University, School of International and Public Affairs — New York, NY
Master of International Affairs, Economic and Political Development — May 2008

University of California, Berkeley, College of Engineering — Berkeley, CA
B.S., Mechanical Engineering — May 2000

RESEARCH EXPERIENCE

Pepperdine University
Graduate Research Assistant — 2017–Present

- Research on an international network for STEM media making and student-led participatory teaching: International Community for Collaborative Content Creation (IC4) [NSF Grant #1612824]

Northern Illinois University
Research Consultant — 2019

- Inclusive Design for Engaging All Learners (IDEAL): Designing technology for cultural brokering [NSF Grant #1839194]

TEACHING EXPERIENCE

Pepperdine University, Graduate School of Education and Psychology
Adjunct Faculty Member — 2020

- Inferential Statistics (EDD 734) - Fall 2020
- Disseminating Knowledge and Publishing (PGLC 801B) - Fall 2020

FELLOWSHIPS, GRANTS & AWARDS

- Doctoral Consortium Selection, First International Conference on Quantitative Ethnography, 2019
- Provost Grant for Ph.D. Research, Pepperdine University, 2019
- Randy Clark Endowed Scholarship, Graduate School of Education and Psychology, Pepperdine University, 2017-2018
- Chiang Kai-shek Scholarship, Graduate School of Education and Psychology, Pepperdine University, 2017-2018
- Leous-Parry Award for Progressive Sustainability, School of International and Public Affairs, Columbia University, 2008
- Teaching Assistantship (Course: Economic Development for International Affairs), School of International and Public Affairs, Columbia University, 2008

PUBLICATIONS

JOURNAL ARTICLES

Espino, D. P., **Lee, S. B.**, Van Tress, L., Baker, T. T., & Hamilton, E. R. (2020). Analysis of US, Kenyan, and Finnish discourse patterns in a cross-cultural digital makerspace learning community through the IBE-UNESCO global competences framework. *Research in Social Sciences and Technology, 5*(1), 86-100.

Calhoun, D. & **Lee, S. B.** (2019). Computer usage and cognitive capability of older adults: Analysis of data from the Health and Retirement Study. *Educational Gerontology*, *45*(1), 22–33.

CONFERENCE PROCEEDINGS

Lee, S. B., Gui, X., & Hamilton, E. R. (2020). Application of AutoML in the automated coding of educational discourse data. In M. Gresalfi, & I. S. Horn (Eds.), *Proceedings of the 14th International Conference of the Learning Sciences (ICLS)* (pp. 2597-2600). Nashville, TN.

Lee, S. B., Gui, X., Manquen, M., & Hamilton, E. R. (2019). Use of training, validation, and test sets for developing automated classifiers in quantitative ethnography. In B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Eds.), *Advances in Quantitative Ethnography: Proceedings of the First International Conference on Quantitative Ethnography* (pp. 117-127). Madison, WI.

Espino, D. P., **Lee, S. B.**, Van Tress, L., & Hamilton, E. R. (2019). Examining the dynamic of participation level on group contribution in a global, STEM-focused digital makerspace community. In B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Eds.), *Advances in Quantitative Ethnography: Proceedings of the First International Conference on Quantitative Ethnography* (pp. 55-65). Madison, WI.

Espino, D. P., **Lee, S. B.**, Eagan, B., & Hamilton, E. R. (2019). An initial look at the developing culture of online global meet-ups in establishing a collaborative, STEM media-making community. In K. Lund, G. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *Proceedings of the 13th International Conference on Computer Supported Collaborative Learning (CSCL)* (pp. 608-611). Lyon, France.

Espino, D. P., **Lee, S. B.**, Van Tress, L., & Hamilton, E. R. (2019). Application of the IBE-UNESCO global competences framework in assessing STEM-focused, global collaborative learning within a digital makerspace environment. In K. Lund, G. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *Proceedings of the 13th International Conference on Computer Supported Collaborative Learning (CSCL)* (pp. 612-615). Lyon, France.

PRESENTATIONS, POSTERS & WORKSHOPS

CONFERENCE PAPERS & PRESENTATIONS

Lee, S. B., Espino, D. P., Van Tress, L. & Hamilton, E. R. (2020, April). Examining connections between positive affect and discourse in a global, STEM-focused digital makerspace community. Paper accepted for presentation at the 2020 Annual Meeting of the American Educational Research Association (AERA). San Francisco, CA.

Lee, S. B., Espino, D. P., & Hamilton, E. R. (2020, April). Initial examination of roles in online meet-ups in a global collaborative STEM-focused, digital makerspace community. Paper accepted for presentation at the 2020 Annual Meeting of the American Educational Research Association (AERA). San Francisco, CA.

Lee, S. B. (2019, July). Examining collaborative creativity in informal global online learning environments through epistemic network analysis. Presentation at the 2nd Southern Oregon University (SOU) Creativity Conference, Ashland, OR.

Lee, S. B., Espino, D. P., & Hamilton, E. R. (2019, April). Exploratory research application of epistemic network analysis for examining international virtual collaborative STEM learning. Paper presented at the 2019 Annual Meeting of the American Educational Research Association (AERA). Toronto, Canada.

CAMPUS PRESENTATIONS

Calhoun, D. & **Lee, S. B.** (2019, May). Envisioning the future role of technology for older Americans: An empirical analysis of data from the Health and Retirement Study. Presentation at the Pepperdine GSEP Research and Project Symposium, Malibu, CA.

Lee, S. B. (2018, June). Quantitative ethnography and use of epistemic network analysis (ENA) to study international collaboration in STEM learning. Presentation at the Pepperdine GSEP Research and Project Symposium, Malibu, CA.

POSTER PRESENTATIONS

Eagan, B., **Lee, S. B.**, Lux, K., & Hamilton, E. R. (2019, April). Measuring connections between affect and motivation in informal STEM learning. Poster presented at the 2019 Annual Meeting of the American Educational Research Association (AERA), Toronto, Canada.

WORKSHOP FACILITATION

Siebert-Evenstone, A. & **Lee, S. B.** (2019, October). Introduction to Automated Coding and nCoder. Workshop at the First International Conference on Quantitative Ethnography, Madison, WI.

Siebert-Evenstone, A., Eagan, B., Swiecki, Z., **Lee, S. B.** & Hamilton, E. R. (2019, June). Creating, refining, and validating automated discourse codes: An introduction to nCoder and rho. Workshop at the 13th International Conference on Computer Supported Collaborative Learning, Lyon, France.

PROFESSIONAL AFFILIATIONS / SERVICE

- Co-Chair, Program Committee for the Second International Conference on Quantitative Ethnography, January 2021
- Member, American Educational Research Association (AERA), 2019 ~
- Member, International Society of the Learning Sciences (ISLS), 2018 ~

OTHER PROFESSIONAL EXPERIENCE

2017–2018	Program Advisor, Daum Foundation, Seoul, Korea
2012–2016	Director of Education Initiatives, Asan Nanum Foundation, Seoul, Korea
2011–2012	Researcher / Head of Public and Media Relations, IOM Migration Research and Training Centre, Goyang, Korea
2008–2011	Programme Officer, Social Policy and Economic Analysis Unit, Division of Policy and Practice, UNICEF, New York, NY
2000–2002	Math Resource Teacher, U.S. Peace Corps, Nepal

ABSTRACT

Despite rising levels of participation by children and adolescents in large, informal online learning communities, there has been limited research examining the role that social dynamics play on the online behavior of young users. In this context, this mixed-methods longitudinal study aimed to investigate the relationship between interaction, collaboration and content creation through the analysis of user-generated comments and log-data from the Scratch platform. The research focused on more than 45,000 comments associated with the online activity of 200 randomly selected participants over a period of three months in early 2012. A combination of methodological techniques was applied in the analysis of the data. Epistemic network analysis was used to identify patterns in the discourse of the comments shared by users in the online community. In addition, generalized linear mixed models (GLMM) and vector autoregression (VAR) models were developed to assess the temporal associations that may exist between interaction and content creation. Analysis of the comment data revealed that young users participate in the Scratch community in creative and diverse ways that involve actively interacting and collaborating with others on the platform. It was also found that the discourse of participants who engaged in higher levels of content creation were more likely to emphasize information seeking and collaboration while those exhibiting lower levels of content creation tended to focus on socially-oriented exchanges, including those seeking to build of relationships with other members of the community.

Chapter 1: Introduction

Background

Instagram, YouTube, Snapchat, and TikTok—only to name a few—fill the days of our youth, as they spend countless hours consuming, sharing, creating, and discussing digital media content. According to the Pew Research Center (2018), 95% of teens aged 13 to 17 in the U.S. either own or have access to a smartphone and 45% mentioned that they are almost always online. With so much time spent online, social media has become the main mode of communication for many individuals. Digital content is everywhere, created by virtually everyone. And it is not just that young people are consuming and creating content. They are sharing them with their friends and the wider public as well as posting their thoughts and exchanging opinions through the comments. In this manner, online interactions have come to constitute a defining part of the digital lives that they lead today.

How young people create digital media and how they interact and collaborate with one another in these virtual environments have major implications for their learning and development. Digital media offer new ways to obtain knowledge and information, access entertainment, and communicate and network with others (Erstad, 2012). While the potential exists for such technologies to support learning and promote empowerment for young people, they can also be associated with risks and vulnerabilities (Craft, 2012; Livingstone, 2010). For this reason, it has become ever more critical to have a better understanding of the activities and interactions that the youth engages in while they are online.

In the educational context, a variety of websites and mobiles applications allow learners to create and share their own digital content, such as wikis, videos and games (Roque, Rusk, et al., 2016; Wheeler et al., 2008). Many online platforms have also developed functionalities that allow users to comment on each other's activities as well as to exchange relevant information, ideas and resources. Through participation in these online communities, learners are able to interact with a wide range of fellow members, including those who possess different types of skills and experiences. Some of the interactions may lead to a process of collaborative learning, whereby the participants engage in the co-construction of shared knowledge (Järvelä, Järvenoja, et al., 2016). Such meaningful experiences can play an integral role in a user's creation process, by offering not only content-related input but also social and cultural cues that can influence future creative behavior (Sawyer & DeZutter, 2009).

What might such experiences entail? Consider the case of Nadia and Katie, two teenage users from different countries participating in an informal online community that focuses on the creation and sharing of interactive media projects:

> [In the online community] Nadia found hundreds of these static images [that were created by Katie], with accompanying notes describing the stories behind the images, focusing on the adventures of a superheroine named Jodie. Nadia imagined the possibilities of bringing these stories to life through animation and proposed a collaboration to Katie by leaving a comment on one of Katie's projects: "Can I try to make moving sprites of your characters? We could work together to make this animated if you want. But only if you want to. Thanks. (I like these drawings you do.)" Katie responded positively to Nadia's suggestion, and for more than a year and a half, the girls have collaborated on animating Jodie the Superheroine, producing ten episodes in a series about Jodie. (Brennan et al., 2010, pp. 79-80)

While it may be illustrative of the best-case scenario, the narrative above nevertheless points to the potential that online interactions can have in stimulating content creation. The encouraging comment and collaborative proposition from one member led to a cooperative relationship that was productive for both participants. On the other hand, it is also possible that certain types of interactions can have negative effects. Comments that express aggression, disagreement or disapproval may lead not only to the stifling of creative efforts but also to the participant's disengagement from future content creation activities.

Problem Statement

The use of online educational technology has increased significantly in recent years. While some are mainly focused on providing instruction and educational content to its users through various media, others allow users to engage in content creation and provide the ability for users to interact and collaborate with one another. Analyses of the online communication among participants can contribute to a better understanding of how interaction, collaboration and content creation are related in the learning processes as they take place. While numerous prior studies have examined issues around participation, interaction and collaboration in online communities, many have focused on descriptive approaches (Malinen, 2015). Based on a meta-analysis of 83 studies from 2002 to 2014, Malinen (2015) noted a gap in the literature on the social influence of user interactions in online communities. Furthermore, there has been limited research on large-scale online communities, particularly in assessing the varying modes of interactions that occur in these settings (Jeong et al., 2017).

One example of a large, informal online community is the Scratch platform, launched by the Media Lab at the Massachusetts Institute of Technology (MIT) in 2007 to develop computational thinking and programming skills among young people (Resnick et al., 2009). Scratch is a programming environment designed for young learners to create and share interactive digital projects, such as animations and video games (Hill & Monroy-Hernández, 2017). Scratch applies a visual approach to programming, which is also known as block-based coding. Since its launch, the Scratch online community has had more than 50 million registered users, 48 million shared projects and 233 million comments as of January 2020.

Several modes of social interaction and collaboration are available in the Scratch online community, which was designed to foster participation and engagement among users (Brennan et al., 2011). All shared projects are open for others to view and download. Users are not only able to create new projects, but also make changes to downloaded projects to create and share their own "remixed" versions of projects. In this way, the entire community is driven by an open-source philosophy, whereby all users have access to the source codes and other digital resources contained in shared projects.

In addition, users are able to react on others' projects by clicking on the "love-it" or "favorite" buttons to express their interest or appreciation for a particular project. A way to engage directly with other users is to post comments about projects to communicate their opinions, exchange ideas, and provide feedback. Lastly, users are able to send friend requests to other members whose projects they are interested in following. In this manner, the interaction and collaboration among users in this environment constitute key elements not only in the creation of the digital artifacts but also in the construction of collective knowledge that is reflective of the culture of the larger community.

Earlier research on the Scratch online community has focused mainly on identifying different types of users and analyzing patterns in their participation behavior (Fields et al., 2013; Fields et al., 2016) as well as assessing development in programming skills (Dasgupta et al., 2016; Fields et al., 2014). Although several studies have investigated the role of online interactions on media creation, they have relied on narrative case study methods that focused on a handful of user experiences (Brennan et al., 2010; Brennan et al., 2011).

Analyses of comments from the Scratch online community have provided insights on the information types (Aragon et al., 2009), functional focus (Fields et al., 2015) and linguistic characteristics (Velasquez et al., 2014) of the user-generated text. However, the studies were conducted using de-contextualized data, which meant that it was not possible to effectively model the interactions that occurred among the participants. Also prevented was the analysis of any potential relationship between creative output and the interactive discourse that was reflected in the user comments.

Purpose Statement

This mixed-methods study aimed to examine the extent to which social interaction and collaboration are related to content creation by users in the Scratch online community. Focusing on the online activities of 200 Scratch users over a 3-month period, the research consisted of two main components. First, epistemic network analysis (ENA) was used to model the connections between the key functional areas present in the interactive comments. The resulting ENA models, reflective of the interactive and collaborative behavior of the users in the sample, was used in comparative analyses to investigate differences among groups and changes over time. Second, two linear models were utilized to investigate the temporal relationship between interaction, collaboration and creativity. Negative binomial generalized linear mixed models (GLMMs) were used to examine the contemporaneous effects, while vector autoregression (VAR) models were applied to explore the time-lagged effects.

Research Questions

This exploratory study was designed to address the following research questions:

- **RQ 1:** What patterns of discourse are observed in the comments of Scratch users? Are there differences in the discourse patterns among users with varying modes of participation?
- **RQ 2:** What relationships, if any, exist between the different elements of interaction and content creation?

Significance of the Study

This study aimed to contribute to the literature on the role that social interaction plays in creative online learning activities. Despite the rising level of participation by children and adolescents in large, informal online learning communities, there has been limited research examining both 'what' is discussed among users in these settings and 'how' it might affect their behavior. As such, it is expected that the findings of this research will contribute to an enhanced understanding of the role that social dynamics play in determining how young people engage in online activities, particularly those involving creative tasks.

From a methodological perspective, the research design encompassed several novel approaches to studying interactive discourse among participants in online learning communities. First was the analysis of contextualized discourse data, where each user-generated comment was situated within the context of its interactions. Accordingly, threaded discussions consisting of the initial comment and subsequent responses were grouped together into a single conversational unit. Second, the study examined the connections between constructs identified in the data through ENA, which allowed for a more nuanced analysis of the temporal dimensions of discourse (e.g. Csanadi et al., 2018). The technique also enabled the exploration of the data at varying levels of granularity, from investigating individual conversational units at the microgenetic level of analysis to the assessment of broad trends at the aggregate level for different groups and/or time periods. Third was the examination of the longitudinal association between key aspects of social interaction and content creation on the Scratch platform. This required the analysis of irregular time series data, which was reflective of the informal, user-driven nature of the online community. Quantitative methods, including GLMMs, VARs and the Granger causality test (Granger, 1969; Hu et al., 2012), were used to explore temporal relationships that may be present in the data.

Definition of Terms

- *Interaction*: Interaction is defined as the situation in which actions of individuals are mutually dependent on each other (Baker, 1999). Interaction is closely related to participant agency and is therefore fundamentally interpersonal in nature. Interactions often consist of reciprocal exchanges that involve at least two participants and two actions (Järvelä, Järvenoja, et al., 2016; Wagner, 1994). Such exchanges indicate that the participants have effectively established a level of shared understanding needed for the purposes of communication (Greeno & Van De Sande, 2007).
- *Collaboration*: Collaboration represents interactions that contribute to the co-construction of meaning that can be interpreted by the group members or preserved in artifacts (Stahl, 2004). This notion encompasses engagement in conversational interaction that facilitates the incremental development of a common understanding among participants about a particular topic (Roschelle, 1992). This is in alignment with one of the functional themes that was applied in the qualitative coding of the data: *Personalized Tutoring*. Project

comments that provide suggestions, tips, explanations or resources were coded for this theme (Fields et al., 2015).

- *Content creation*: Content creation refers to the generation of digital artifacts that are made publicly available, involve creative effort and are produced outside the context of professional settings or practices (OECD, 2007). For the purposes of this study, content creation was operationalized as the act of producing and sharing of projects in the Scratch online community. This includes new projects as well as projects that have been remixed from others. However, remixed projects that are identical to the original project were excluded.
- *Online community*: Online communities refer to virtual social spaces that bring together individuals to share information and resources, learn and engage in social activities (Preece, 2001). Online communities are driven by a shared purpose among its members and supported by technological tools, including communication software and web-interfaces (Preece et al., 2004). The Scratch online community, which is the focus of this study, gathers young participants who are interested in coding and digital media creation from all parts of the globe.
- *Epistemic frame theory*: Epistemic frame theory argues that a culture should be understood not only by the ways of thinking, acting and being that characterize a community of practice, but also by the configuration of the relationships that exist between those elements (Shaffer, 2017; Shaffer & Ruis, 2017). Conceptualizing learning as a process of enculturation (Brown et al., 1989), epistemic frame theory posits that “learning can be characterized by the structure of connections that students make among elements of authentic practice” (Shaffer & Ruis, 2017, p. 182).

- *Computer-supported collaborative learning*: Computer-supported collaborative learning (CSCL) refers to joint activity and interaction among peers for purposes of learning that is facilitated by information and communication technologies and tools (Koschmann, 2002; Suthers, 2012). The term also describes the field of research that focuses on such activities, including interactive processes that occur in online learning environments.

Conceptual Framework

The study applied a constructivist approach to examine how social interaction and collaboration among members affect their engagement in content creation in online learning communities. Based on the notion that all knowledge is constructed from an individual's interaction with their surroundings (Fosnot & Perry, 2005), constructivism provides a useful conceptual framework for understanding how interactive discourse between individuals might influence cognitive, social and motivational factors that are associated with the learning process. Interactions provide opportunities to not only experience and integrate new perspectives but also confront one's own ideas and perceptions, paving way for sociocognitive conflict that stimulates higher order knowledge and thinking (Bell et al., 1985; Piaget, 1977). Similarly, interacting with others who possess greater expertise can enable engagement in the zone of proximal development, where learning and enculturation occur through participation in higher level activities (Rogoff, 1998; Vygotsky, 1978).

Moreover, theoretical consideration of creativity has evolved in the last several decades to place greater significance on the social and cultural dimensions. This perspective focuses on how creativity is shaped by contextual factors and the experiences that emerge from them, moving beyond the earlier emphasis of research on individual traits and abilities. Several conceptual frameworks have been put forth to capture the various aspects of the interrelationship between creativity and the sociocultural context, including from social psychological (Amabile, 1983) and systems-based (Csikszentmihalyi, 1988) perspectives. Furthermore, other concepts such as collaborative creativity (Mamykina et al., 2002), group creativity (Sawyer, 2003) and distributed creativity (Sawyer & DeZutter, 2009) have been posited to describe the collective efforts of individuals in creative activities.

Two additional theoretical perspectives were relevant for this study's exploration of content creation in the Scratch online community. First was the theory of constructionism, which stresses that learning processes are enhanced when carried out through the creation of public artifacts (Papert, 1993). It takes a 'learning by making' approach, whereby the construction of shared objects represents a way for individuals to express their ideas in concrete form (Harel & Papert, 1991). Learners are seen to be able to develop their own tools and symbols needed to interpret and engage with the world around them (Ackermann, 2001). In this process, the surrounding culture becomes the material with which knowledge is constructed by the learner (Papert, 1993). In the context of this study, the Scratch projects created by the young users serve not only as the medium through which their thoughts and ideas are manifested in tangible form but also as the object of shared interest and discussion that connects them to other participants on the platform.

Second was the notion of communities of practice, in which learners become full members through a process of legitimate peripheral participation (Lave & Wenger, 1991). Learning is seen not as an internalization of knowledge, but instead as a situated activity occurring within the continually shifting social, cultural and historical circumstances of the lived-in world (Brown & Duguid, 1991). From this situated perspective, learning can also be regarded as a process of enculturation, through which participants adopt the values and behaviors of a particular community (Brown et al., 1989). Skills and experiences gained through the learner's engagement become reflected in their identity as they move from being a novice toward full participation in the sociocultural practices of the community (Lave & Wenger, 1991). An in-depth analysis of the activities of Scratch users in the online community, including the interactive discourse captured in the project-related comments, offered valuable insights about the evolving dynamics of participation within this community of practice.

Analytical Approach

Studying the interaction and collaboration in the Scratch online community requires the consideration of several facets in the operationalization of the theoretical perspectives discussed above. As such, the analytical approach taken for this study brought together several different methods and tools to carry out the investigation. This was also reflected in the dataset analyzed in the study, which comprised of both quantitative and qualitative data. Log data of online activities consisted of quantitative data while the text data from the user-generated comments constituted the qualitative portion of the dataset. Analyses of both types of data was carried out in a complementary manner, with quantitative findings informing qualitative analyses in later stages and vice versa.

The analytical approach was grounded in three interrelated theoretical premises. First was that discourse—representing language in use—is an essential element in understanding the social practice of individuals, including their actions and interactions (Gee, 2005). Building on this notion, the second premise asserted that connections between elements of social practice constitute a critical part of understanding the culture of a community (Shaffer, 2017). The third and final premise was that temporal dimensions of social practices and the connections among them are important for understanding the impact of these sociocultural dynamics on individual behavior (Damşa, 2014). The discussion below elaborates on each of these premises.

- *Premise 1:* Language enables social activities and identities, while also allowing for human relations to take place within cultures, communities and institutions (Gee, 2005). Language functions not only to communicate content but also express social contexts and relationships as well as personal values and attitudes (Brown & Yule, 1983). In this process, discourse serves as signs that point to the network of meanings that are embedded within shared activities and artifacts (Stahl, 2004). Analysis of discourse, therefore, can shed light not only on how language was used in specific contexts of social interaction (Gee, 2018) but also on the function that the language served in any given situation (Brown & Yule, 1983). In this study, discourse analysis, through qualitative coding of the data, was applied to the project comments shared by Scratch users to identify the functional themes present within the text. The results of this analysis provided insights on the prevalence of the different types of social interactions that were undertaken by Scratch users included in the sample.
- *Premise 2:* The culture of a community is defined not only by the elements of social practices that are enacted by its members, but also the connections that are made among those elements (Shaffer, 2017). Moreover, Gee and Green (1998) observed that while cultural models constitute the "stories" that allow groups to interpret and apply situated meanings of discourse, the models themselves are continuously modified and expanded by group members in their social practices. As such, gaining a comprehensive understanding of culture also requires the recognition of its evolving dynamic. In this study, these two fundamental aspects of culture (connections among its constituent

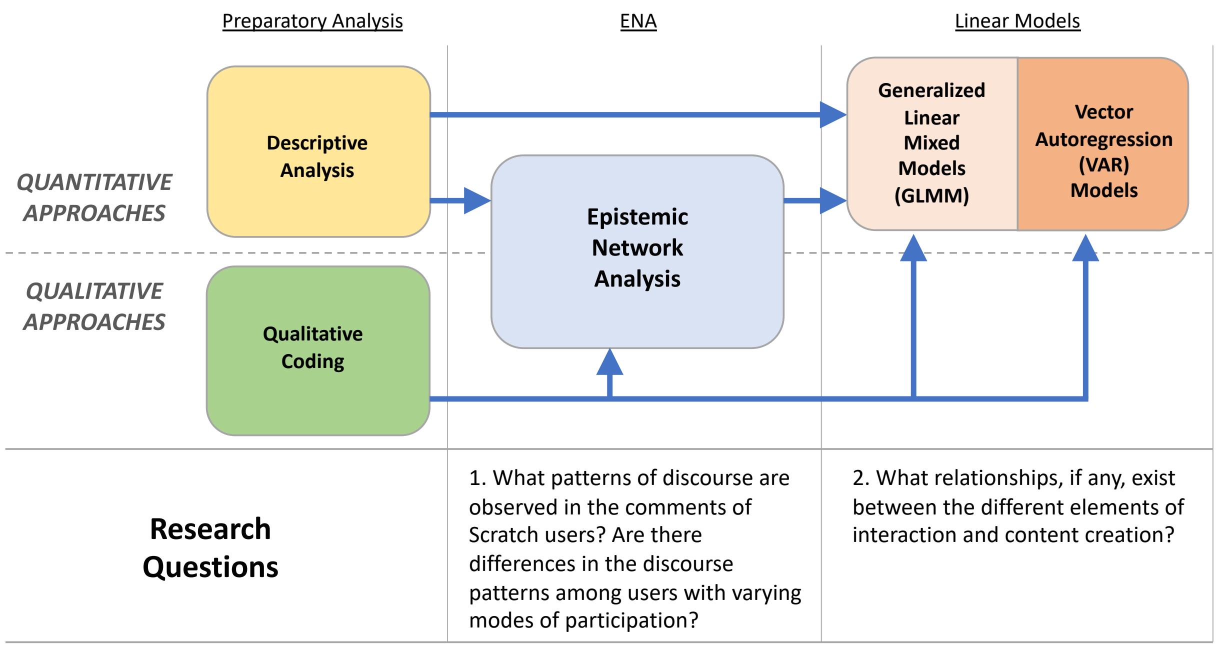
elements, change over time) were operationalized through the application of ENA. ENA combines the qualitative methods of discourse analysis with statistical and visualization techniques to model the co-occurrences between salient elements of discourse and other actions of social practice (Shaffer, 2017). In this process, ENA uses a moving window to identify linkages among constructs occurring within a recent temporal context (Ruis et al., 2019; Siebert-Evenstone et al., 2017). As a result, ENA models are able to capture the structure of connections that are generated for any given unit of analysis. It is possible to make statistical comparisons of these “snapshots” to examine differences among groups as well as to assess changes in the networks over time.

- *Premise 3:* Temporal and sequential dimensions need to be taken into account when examining how social interactions unfold over time (Damşa, 2014). Reimann (2009) noted that temporality has been often overlooked in CSCL research, despite the availability of pertinent datasets in many cases. Time was an important element of the current study, particularly in assessing the impact that different types of interaction has on the creation of content by individual users. This emphasis was operationalized through the use of generalized linear mixed models (GLMM) and vector autoregression (VAR) to explore the contemporaneous and time-lagged effects of interaction on content creation, respectively. Analyses were conducted at three levels of social practice related to the user-generated comments: comment activity, functional focus, and network location. Comment activity refers simply to the posting or receiving of a comment, while functional focus and network location refer to the frequency of coded values and the x- and y-coordinates of the plotted points obtained from ENA, respectively. Different GLMM and VAR models created in this process were compared to identify the elements of social practice that may be exerting the greatest influence on an individual's creation of media content.

An overview of the study's analytical design is presented in Figure 1. Preparatory analysis consisted of descriptive analysis of the log data as well as qualitative coding of the interactive comments. This was followed by epistemic network analysis and the application of GLMM and VAR models. The arrows indicate how the results of earlier phases contributed to the subsequent stages of analysis.

Figure 1

Stages of Analysis with Associated Research Questions



Limitations

This study utilized secondary data in the form of a publicly released dataset from the MIT Media Lab (Hill & Monroy-Hernández, 2017). For this reason, the scope of the analysis was contingent on the breadth of information provided in the dataset. The discussion below pertains to several constraints posed by the dataset that were considered most salient for the study.

The first limitation was the potential for missing information in the Scratch longitudinal dataset. Due to privacy issues, the dataset contains only the information that was publicly available on the Scratch platform at the time of the data collection in May 2013. Variables such as self-reported age and gender were omitted from the dataset along with any data that was removed by the user. In addition, user activities that were not made public, including views and downloads of other members' projects, were not included in the dataset. The availability of such information may have allowed for further analyses on the influence that passive participation behavior, e.g. viewing the projects of others, may have had on outcomes related to content creation.

One of the key variables in the study relates to the creation of projects. As an outcome for the linear regression models utilized in this study, this construct was operationalized as the number of projects created and shared in the Scratch online community. Using project count as the outcome variable offers efficiencies in the quantification of the construct for analytical purposes. At the same time, this simplified approach raises the question of whether it is the most appropriate given its narrow scope of interpretation. Projects created by users on the Scratch platform vary widely, not only in terms of its type and the kinds of digital resources that may be involved, but also in its complexity and creativity. For this reason, other outcomes variables for content creation may be considered that take into account the quality of the projects produced. One such variable could examine the type, frequency and application technique of coding blocks used in the project programming script. Another approach might consider measures related to the reception of the project by the online community, through variables such as the number of views, love-it's or downloads. While the integration of such dimensions into the outcome variable may provide a more comprehensive measure of content creation, it may also introduce additional complexities into the study. As such, it was decided to maintain the discrete count of projects as the outcome variable, despite its inherent limitations.

Another aspect was the temporal dimension of project creation, which also played an important role in this study. In particular, the dataset provides a timestamp for when projects were created as well as when certain types of comments were exchanged between the creator and other users. For each project, a single timepoint is given, indicating when it was shared in the Scratch online community. This moment of transfer from the private to the public domain may represent a significant point of distinction in terms of the effect of social interaction; however, the single data point nevertheless does not adequately capture the evolving nature of projects, as many creators continue to revise and tinker with their projects after sharing it publicly.

Another limitation relates to the user-generated nature of the project comments that were analyzed in this study. For researchers, user-generated text can present a double-edged sword. On the one hand, such text can be extremely insightful, as they are direct expressions of the author's thoughts and reactions. On the other hand, typing errors and other irregularities in the text can present challenges for the researcher in accurately interpreting the text. During the qualitative coding of the comments, numerous instances were encountered where the user had intentionally misspelled words in their message. In one comment, for example, a user wrote, "ZO WAT. NOW I PLAI LEGND OV ZLDA. ITZ AWSOM ON DA 3DS" instead of "So what. Now I play Legend of Zelda. It's awesome on the 3DS."

In addition, two other difficulties were confronted in the Scratch online community. First was related to the global reach of the platform, with the participation of users from around the world. For this reason, language use—both in terms of *which* languages are used and *how* they are used in the comments—was a key factor that needed to be taken into account in the analysis.

Second was that the Scratch platform was designed to be an informal community of young users. As such, it is likely that the comments were reflective of the sociocultural norms and practices of its members. Lack of familiarity with all of such elements on the part of the researcher was a limitation that could hinder the valid interpretation of the comments data.

Delimitations

Capturing the online activities of more than 1 million Scratch users over a period of 5 years, the full longitudinal dataset contains a massive amount of data, including over 2 million projects and 10 million comments. However, in order to focus the research on the linkages between social interaction, collaboration and content creation, the sample dataset was delimited by applying several parameters. Time was a key delimiting factor, with specific criteria applied to ensure that users included in the sample dataset have been actively and continuously engaged in the Scratch online community during the period under study. Active and continuous engagement was defined as creating at least one project in each of the 3 months from January to March 2012. Two hundred users from this subset was randomly selected for inclusion in the study. Subsequently, only the comments within conversations that contained at least one comment sent or received by a sample user between January and March 2012 were included in the study dataset.

Assumptions

This study viewed the Scratch platform as an online learning community that enables members to not only enhance their mathematical and programming abilities, but also develop social and communication skills through their interaction with other users (Resnick et al., 2009). In many ways, the Scratch platform provides a learning environment that is dissimilar to other online educational sites. No explicit efforts are made to provide instruction or disseminate educational content. However, as young users create and share their own digital media artifacts on the Scratch platform, they are engaged in active forms of knowledge building and meaning making, in alignment with the constructionist notions of learning (Papert, 1993).

The Scratch website allows for open registration for anyone with an email address that has not been previously registered. While it is possible for any individual to create multiple accounts using different email addresses, this study assumed that each registered user was unique. As such, all activities associated with any given account were assumed to be attributable to the registered user.

Similarly, it was assumed that the primary users of the Scratch online community are young people for whom the platform was designed. Based on self-reported data collected during the registration process, previous research noted that the generally accepted range for the majority of Scratch users is between 11 and 18 (Fields et al., 2016). Given that no age-related variable was provided in the dataset used for this research, it was assumed that the users, whose activities are included in this analysis, were representative of this group.

Another assumption was that the user-generated comments were representative of the entirety of the social interactions between Scratch users pertaining to a particular project. As the Scratch platform is also often utilized in face-to-face instruction in both formal and informal educational settings (e.g. Gutierrez et al., 2018), it is possible that the some of the comments in the dataset may have been composed by co-located users. Research has shown that young users often choose to use text-based communication even when in close proximity to one another (Brown & Larson, 2009). In such cases, the online interactions may capture only a small portion of the actual level of interaction that is taking place between the users. However, for the purposes of this study, it was assumed that only online interactions had occurred among users in the sample.

Summary

The Scratch platform was established in 2007 to support the development of computational thinking and coding skills among youth. The programming environment allows young users to create and share interactive digital media projects. In order to foster participation and engagement in the online community, the platform provides several functionalities for social interaction and collaboration, including the ability to exchange comments with other members. Through an analysis of user comments and activity data from the Scratch online community, this study aimed to explore the influence of social interaction and collaboration on the creation of content by participants. The analysis focused on a sample of 200 users who were actively and continuously engaged in the Scratch online community during a 3-month period from January to March 2012.

A mixed methods approach was utilized in this research, applying a diverse set of analytical techniques to examine qualitative as well as quantitative data. Epistemic network analysis (ENA) was used to examine the discourse patterns produced by the interactive comments. In addition, generalized linear mixed models (GLMM) and vector autoregression (VAR) were utilized to explore the temporal dynamics present between content creation and different elements of social interaction. The study is expected to contribute to an improved understanding of the role that sociocultural processes can play in fostering creative behavior among children and adolescents, particularly in the context of large, informal online learning communities.

Organization of the Study

The subsequent sections of this study are organized in the following manner. Chapter 2 presents an in-depth review of theoretical frameworks and research literature relevant to the topics of interaction, collaboration and content creation in online learning environments. Pertinent insights from prior research related to the Scratch platform are also summarized. Chapter 3 provides a detailed discussion on research design and methodology. Issues related to the sampling strategies and selection criteria are also addressed. Chapter 4 reports the results generated by each stage of the analysis. A discussion of the key findings along with recommendations for further study are presented in Chapter 5.

Chapter 2: Literature Review

Overview

This chapter provides a review of literature relevant for examining social interaction, collaboration, and content creation in online learning environments. The review begins with a theoretical background that serves as the overarching conceptual foundation for the study. Emphasis is given to cognitive and social constructivism as well as constructionism, situated learning and communities of practice. Following a review of literature on the impact of social interaction and collaboration on learning in online contexts, subsequent sections focus on content creation and online learning communities. The final section presents a summary of previous findings related to interaction, collaboration and content creation in the Scratch online community.

Theoretical Background

Constructivism

Constructivism characterizes learning as a process whereby individuals actively make meaning through their interactions with the physical and social environment (Fosnot & Perry, 2005). From this perspective, learning is not viewed as a linear process but rather as a complex, iterative process of knowledge construction and reconstruction. In addition to Piaget's seminal work on elaborating the progressive stages of intellectual development in children and adolescents (Piaget & Inhelder, 1969), his later research focused on learning mechanisms, particularly on the processes that facilitated the generation of new perspectives (Fosnot & Perry, 2005).

Piaget posited that cognitive change was brought forth through equilibration, a dynamic, self-regulated process through which the two contrasting behaviors of assimilation and accommodation are balanced (Fosnot & Perry, 2005; Piaget, 1977). With assimilation, individuals assertively act on new experiences using their existing cognitive constructions, whereas accommodation refers to the reflective and integrative behavior that leads to structural changes and adaptations (Piaget, 1977). He perceived that equilibration was a nonlinear mechanism in constant flux, shifting among instances of self-organization, adaptation, growth and change (Fosnot & Perry, 2005). Piaget's notion of equilibration is the basis for the concept of sociocognitive conflict, which proposes that the contradictions between the individual's prior understanding and new experiences of the individual give rise to higher forms of thinking and learning (Bell et al., 1985; Palincsar, 1998). As Piaget (1985) noted, "disequilibria alone force the subject to go beyond his current state and strike out in new directions" (p. 10). Bell et al. (1985) argued that the social environment played a significant factor behind such conflicts responsible for cognitive development. Through interactions with others, children are exposed to different ways of thinking—which in turn promotes the development and reconstruction of their cognitive structures (Bell et al., 1985). In this way, social interaction can lead to enhanced levels of cognition in individuals (Palincsar, 1998). In addition, the interaction between peers was believed to provide children not only with the opportunity to assess one's ideas against that of another, but also to gain knowledge and experience about relevant social situations (Bell et al., 1985).

On the other hand, Vygotsky emphasized the social, cultural and historical dimensions of knowledge construction (Fosnot & Perry, 2005; John-Steiner & Mahn, 1996). Vygotsky highlighted the role of social interaction in the learning and development of individuals, particularly through the use of tools and signs (Vygotsky, 1978, 1994). Vygotsky (1978) believed that: "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level…All the higher functions originate as actual relations between human individuals" (p. 57). In this process, Vygotsky argued, language is often used as a means of communication between the child and others in their environment as well as a way for organizing their thinking in the form of internal speech (Vygotsky, 1978).

Based on this logic, Vygotsky (1978) emphasized that learning precedes and provides the basis for the growth of higher cognitive functions: "Learning awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers" (p. 90). While the child is initially dependent upon the guidance and support of others with greater experience and expertise, such higher-level functions are eventually matured and internalized (Fosnot & Perry, 2005; John-Steiner & Mahn, 1996). Vygotsky (1978) referred to this formative, transitional state as the "zone of proximal development" (p. 86). Interactions that take place in the zone of proximal development enable the child to participate in activities beyond their competence while learning to use the cultural tools that are specific to the activity (Rogoff, 1998). From this perspective, learning and development are interrelated processes occurring within social and cultural contexts that are constantly changing (Palincsar, 1998).

Constructionism

Constructionism builds on the concepts of constructivism, but places greater emphasis on the role that the creation of shared artifacts plays in learning (Papert, 1993). While constructionism shares the constructivist notion that knowledge structures are developed and transformed through the individual's interaction with their environment, it argues that such processes are further enhanced in contexts where the learner is actively engaged in the construction of a public objects (Harel & Papert, 1991). It is grounded on the idea that knowledge is situated, continually shaping and being shaped by the surroundings (Ackermann, 2001). The expression of ideas through different media serves as a means to build knowledge in particular contexts. As such, constructionism views individuals as builders, makers and creators, who construct for themselves the tools and symbols needed to understand and interact with the world around them (Ackermann, 2001; Papert, 1993).

In this regard, Papert was especially interested in the application of computational models and tools in supporting the development of higher-order thinking in children, particularly those related to mathematical knowledge and concepts. Papert (1993) posited that by allowing students to build their own models through the use of computers, learning could be transformed into a process that is more active and self-directed. As students are able to apply their knowledge in the construction of concrete ideas and artifacts, learning becomes an empowering experience which can further motivate learning (Harel & Papert, 1991; Kafai, 1996; Papert, 1993). From a social interaction perspective, Papert (1993) observed that the computer can function as a tool that mediates the relationship between students, and that such connections made among learners may have positive influence on their level of engagement and collaboration: "Although the work at the computer is usually private it increases the children's desire for interaction. The children want to get together with others engaged in similar activities because they have a lot to talk about" (p. 179). Technology, in this role, can be considered as a tool to facilitate not only communication and interaction among students but also something that motivates and supports their learning, collaboration and creativity.

Situated Learning and Communities of Practice

Closely related to constructivism is the concept of situated learning, which argues that knowledge is developed and applied in a continuous process of acting in social situations (Brown et al., 1989; Lave & Wenger, 1991). From this point of view, learning is seen as being inseparable from the activity that takes place within a historical and culturally defined context (Brown & Duguid, 1991). Brown et al. (1989) posited that knowledge should be conceived as a set of tools, which “can only be fully understood through use, and using them entails both changing the user's view of the world and adopting the belief system of the culture in which they are used” (p. 33). Rather than conceiving learning as a process of transmission or internalization of knowledge, the situated perspective takes learning as an integral component of social practice (Lave & Wenger, 1991). Drawing on the interpretation of Geertz (1973) that people are entangled in a socially constructed web of culture, Brown et al. (1989) observed that culture and the use of knowledge and tools continually interact to create an increasingly complex and intricate understandings of the world. In this manner, learning can be viewed as a process of enculturation, whereby individuals adopt the behaviors and norms of a particular community (Brown et al., 1989).

Based on these notions, Lave and Wenger (1991) elaborated the concept of legitimate peripheral participation, a process through which learners come to function as fully participating members in a community of practice. Hence, learning is perceived as increasing participation in the social world, which is constituted in social practices that are constantly reproduced, reshaped and transformed (Lave & Wenger, 1991). In this conceptualization of learning, the individual’s development of knowledge and skills is incorporated within their changing identity as a more central participant in the community (Lave, 1991). As such, the community of practice remains in a state of continuous transformation and innovation, reflective of the changing relations among its members, activity and the world (Brown & Duguid, 1991). Nevertheless, the community of practice—grounded in the shared understanding of its purpose and significance—

provides the necessary conditions for interpretation of knowledge and meaning-making within its sociocultural context (Lave & Wenger, 1991).

Social Interaction & Collaboration in Learning

Social Interaction in Learning

Interaction can be defined as a situation in which the actions of individuals involved are mutually dependent on each other (Baker, 1999). Observing that action is distinct from behavior mainly in its premise of intentionality, Baker (1999) further argued that interaction is closely related to participant agency and is therefore fundamentally interpersonal in nature. Researchers have distinguished between cognitive and socioemotional dimensions of interpersonal interaction (Järvelä, Järvenoja, et al., 2016; Kreijns et al., 2013). The former can include task-related discussion among individuals in a collective activities, while the latter may encompass expressions of emotion or motivation that takes place at the group level (Järvelä, Järvenoja, et al., 2016).

Learning and social interaction are closely intertwined. Interaction among individuals not only provide information and content needed for knowledge building but can also produce important social and motivational effects. At the same time, interactions allow participants to attain a shared understanding and co-construct knowledge (Kent et al., 2016). Related to the constructivist concept of socio-cognitive conflict, King (1990) posited that the mental dissonance stemming from social interactions can have more cognitive benefits than when it is caused by contradicting ideas within an individual.

A relevant theory from the field of sociology is symbolic interactionism, which emphasizes the importance of social interactions in how humans construct and interpret the meanings of the world around them (Blumer, 1986; Mead, 1934). The concept is grounded in three main premises: (a) human actions upon the world are based on meanings that the world presents to them; (b) the meanings of the world are created through social interactions that humans have with one another; and (c) the meanings of the world are managed and revised through a process of interpretation (Blumer, 1986). From this perspective, social interaction is seen to shape human behavior, and is not simply a manifestation of it.

Numerous studies have found positive linkages between interaction and learning outcomes in various contexts. For example, results from a study examining the effect of interaction in building Lego block structures among preschool learners showed that working with a partner can result in greater learning than independent work, especially when paired with someone who has greater expertise (Azmitia, 1988). An analysis of dyadic discussions between adolescent and young adult learners by Kuhn et al. (1997) found that a series of interactive dialogues about a specific topic resulted in an improvement in the quality of reasoning as well as in the metacognitive awareness of the existence of multiple perspectives. Based on a review of studies focused on argumentative reasoning skills, Anderson and Soden (2001) concluded that interaction between students provides benefits for enhancing their critical thinking and problem-solving capacity.

Of special interest among educational researchers has been the effects of peer interaction on learning. Research has shown that peer interaction can result in enhanced cognitive activity, which is reflective of learning processes (King, 2002; Webb, 1987). King (2002) distinguished between the different types of peer interaction that can influence different learning processes and outcomes. Peer-assisted rehearsal is adequate for acquiring factual knowledge through the sharing of information; however, more complex levels of learning require interactions that are of higher order of cognition that involve the sharing of multiple perspectives, elaborations, inferences, predictions, justifications. Studies have also shown that peer interactions can be structured, e.g. through approaches using guided peer questioning strategies, to facilitate higher level mental functions (King, 2002). Furthermore, Piaget posited that social interactions between children is more likely to contribute to cognitive development than the exchanges between children and adults, based on the belief that the former presented a context where the peers can exert a shared control of the interaction (Palincsar, 1998).

Collaboration in Learning

Social interaction serves as a process through which collaboration and learning can take place (Dillenbourg et al., 1996). As individuals communicate and interact with one another, they are able to co-construct new meaning, which in turn becomes materialized into shared knowledge objects (Damşa, 2014; Kreijns et al., 2013). Kreijns et al. (2003) noted that social interaction and collaboration are necessary conditions for one another, arguing that without social interaction, there can be no genuine collaboration, and vice versa. Roschelle (1992) used the notion of convergent conceptual change to describe how individuals engage in conversational interaction to coordinate their efforts in building a common understanding of a particular topic. In this context, Roschelle suggested that the collaborative construction of knowledge is carried out incrementally, as participants progressively obtain a better grasp of complex concepts through the formulation and refinement of partial meanings. From the perspective of agency, Damşa (2014) posited that active participation in interaction enables individuals to engage in collective knowledge building. Agency in collaborative situations are thus based on social participation involving acts of negotiation and the development of intersubjectivity, which facilitate the pursuit of shared efforts at the group level.

Regarding the social influences on motivation for engaging in learning tasks, Järvelä et al. (2010) underscored that there have been two main perspectives. One views motivation as being socially influenced, whereby the social environment impacts the motivation of individuals to participate in learning activities. The other is the social construction perspective, which is grounded on the notion that motivation is socially constructed in a continually evolving process of interpersonal interaction and shared regulation. In this respect, Järvelä, Kirschner, et al. (2016) suggested that there are three types of regulation are needed to successfully carry out collaborative tasks: self-regulated learning, co-regulated learning and socially shared regulation of learning. Whereas self-regulated learning relies on the individual's ability to understand and control the learning process, co-regulated learning is facilitated by group dynamics, technology and contextual factors. In socially shared regulation, members collaborate to organize and coordinate their collective thinking, behavior and affect to achieve their goal.

In collaborative learning, knowledge is constructed as a result of a dynamically evolving process of social interaction that takes place among individuals and communities. Stahl (2004) emphasized that while knowledge building is always situated within a context that provides meaning to the activities, tools and symbols, the social situation in itself is continuously transformed through the interactions in a process of social reproduction. Stahl (2004) noted that:

> Out of the social interaction among people, the following elements get produced, re-produced and habituated: the group itself as an interactive unit, the individuals as roles and mental subjects, the situation as network of artefacts and space/time as dimensions of reality. (p. 76)

Similarly, Gunawardena et al. (1997) observed that shared learning comprises of two related and concurrent processes of knowledge creation that take place at the social and individual levels. At the social level, shared knowledge is co-constructed by the group. At the same time, the individual's own meaning-making is continually informed and updated as he or she interacts with the group's shared knowledge. Thus, Gunawardena et al. (1997) emphasized the importance of understanding the interrelationship between the cognitive and social processes associated with knowledge building. Likewise, Järvelä, Järvenoja, et al. (2016) argued that cognitive and socioemotional interaction are complementary dimensions needed for effective collaborative engagement.

In addition, it is important to recognize that not all social interactions—and ensuing collaborations—will be related to learning. Different types of interactions will have varying levels of significance on learning processes and outcomes, and will also depend on the particular objectives and tasks that are associated with the activity (Cohen, 1994). In this regard, prior research has examined the concept of 'constructive interaction,' referring to interaction between peers that leads to an improved understanding of relevant knowledge and concepts (Miyake, 1986). For constructive interaction to occur, a shared objective as well as interaction among participants with diverse ideas, knowledge and viewpoints is required (Miyake & Kirschner, 2014). Baker (1999) suggested two situations under which interactions can be considered to be constructive. First is when the interaction directly results in the co-construction of meaning or knowledge. The second case involves when the interaction contributes to a collective goal or collaborative activity. A similar notion is that of 'meaningful interaction,' which describes interaction that directly affects learning through the stimulation of intellectual exchanges and engagement in educational activities (Vrasidas & McIsaac, 1999; Woo & Reeves, 2007).

Computer-Supported Learning Contexts

Interaction in Computer-Supported Learning Contexts

With advances in technology in the past several decades, interactions among learners in online environments have not only increased significantly but there also have been changes in how the interactions take place. The field of distance education, for example, has evolved from correspondence courses in the past and now encompasses online learning programs including massive open online courses (MOOCs). Online learning offers greater accessibility to content and resources, allowing learners to engage in the process at the time and location of their choosing (Garrison & Cleveland-Innes, 2005). At the same time, interactions in such contexts can also become more difficult when compared with face-to-face environments due to limitations imposed by the technology (Angeli et al., 2003; Woo & Reeves, 2007). Moreover, Kreijns et al. (2003) stressed that group interactions in computer-supported learning environments are often ineffective due to problems related to the social dynamics among members. For these reasons, it is ever more important that the study of interaction in online learning settings take into account how well technologies can support educational communication. Valkenburg et al. (2016) argued that the evolution and diversification of online communication tools and methods necessitate changes in the theoretical approaches used in research.

Interactions in online settings were initially examined within the framework of computer-mediated communication (CMC). One area of focus was to consider how personal CMC was relative to face-to-face communication. In this regard, Walther (1996) distinguished three types of CMC: impersonal, interpersonal, and hyperpersonal. Impersonal communication, referring to cases in which the communication was less personal than the face-to-face ones, mainly involved task-oriented nature and reduced social dimensions such as interpersonal affect or group cohesion (Kiesler et al., 1984; Walther, 1996). Impersonal interactions can have favorable influence for increasing effectiveness in certain task-oriented group work, as the medium was able to minimize the socioemotional elements that can distract or hinder the group processes; however, it can also have negative effects such as reducing self-regulation and worsen aggressive behavior (Kiesler et al., 1984). Interpersonal communication, referring to instances that are as personal as face-to-face communication, occur when the participants develop social relationships over time. In this process, Walther (1996) posited that the rate—not the amount—of social information exchange plays a more significant role. In hyperpersonal communication, where the social and emotional interaction goes beyond that of face-to-face interactions, takes place when the CMC allows for the participant to engage in “selective self-presentation, idealization and reciprocation” (p. 28). On the part of the sender, they may choose to selectively send communications that are more socially desirable; on the part of the receiver, such messages may lead to idealized construction of the sender’s image and their relationship, leading to a validation through their response (Walther, 1996).

In this regard, social presence can be seen as a key factor that determines how individuals interact in electronically mediated communication (Gunawardena, 1995; Lowenthal, 2010; Tu, 2000). Social presence was defined by Short et al. (1976) as the “degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships” (p. 65). The level of social presence can be influenced by the type of media in which the interaction takes place, which determines the availability of transmissions channels available (Walther, 1992). As fewer channels are available in a medium, the less focus is given by the individual on the presence of the other (Short et al., 1976). From this viewpoint, Walther (1992) noted that decrease in social presence will result in a more impersonal communication. However, Gunawardena (1995) argued that the participant’s perception of the social qualities of the online communication will depend mainly on the degree of social presence created by the other participants, including the instructor or facilitator. In this regard, Garrison and Cleveland-Innes (2005) found that peer interaction had a stronger effect in establishing social presence in online learning environments.

Furthermore, the impact of interaction on learning is also affected by the type of interaction in which the student is engaged. Research indicates that the exchange of feedback can be useful for learners in online settings. Feedback not only provides a means to exchange task-related information but also affects the student's continued engagement in the online discussions. Timely feedback was considered to be important element for encouraging further participation (Vrasidas & McIsaac, 1999). Wagner (1994) distinguished between the behavioral and cognitive functions of feedback. The former emphasizes reinforcement and shaping of behavior through action-eliciting cues, while the latter focuses on the role of feedback in providing learners with information about the accuracy of a response (Wagner, 1994). One way to foster such interaction is through authentic tasks, which require engagement and communication between students to share their ideas and opinions, coordination and cooperation, agreeing on a shared meaning, and accommodating differing viewpoints (Woo & Reeves, 2007).

Collaboration in Computer-Supported Learning Contexts

Interactive processes are central to collaborative learning (Stahl, 2000). Examining collaborative engagement, therefore, requires paying attention not only to the interactions among learners but also how such interactions influence the shared process of engagement (Järvelä, Järvenoja, et al., 2016). Research focusing on interaction and collaboration in online learning contexts, particularly in the CSCL field, have sought to address the various dimensions of this complex dynamic. In this regard, Jeong and Hmelo-Silver (2016) outlined seven key affordances that are provided by technology for collaborative learning. They posited that technologies give learners the opportunity to (a) undertake joint tasks, (b) communicate and (c) share resources with one another, (d) engage in productive learning activities, (e) participate in co-construction, (f) monitor and regulate their learning processes, and (g) identify and create groups and communities. Noting that technology only plays a partial role in supporting each of these affordances, Jeong and Hmelo-Silver stressed the importance of integrating technology with appropriate pedagogical approaches to strengthen collaborative learning outcomes.

Similarly, researchers have also flagged the need for greater emphasis on the social dimensions of collaborative learning in online contexts. For example, Miyake and Kirschner (2014) have pointed out that the social aspects of learning in these environments have not garnered the same focus as the technological and pedagogical dimensions. Kreijns et al. (2003) attributed this relative inattention to two main reasons. First, it is assumed that social interaction will occur naturally due to technological affordances inherent to the setting. Second, social interaction tends to focus on cognitive aspects while the socioemotional processes are often overlooked.

According to Kreijns et al. (2013), the interplay among the dimensions of sociability, social space and social presence determine how interactions take place in CSCL environments. Sociability refers to the capacity for the environment to support and promote socioemotional interaction. In turn, interactions facilitate the establishment of a social space where interpersonal relationships, trust, and a sense of community can develop among members of the group (Kreijns et al., 2004; McMillan & Chavis, 1986; Rovai, 2002). Kreijns et al. (2013) further observed that this is a mutually reinforcing process, whereby the creation of robust social spaces can lead to greater social interactions among those involved. Social presence also plays a key role in this interactive process. While social presence is determined in part by the technological environment, it is also strongly influenced by the manner in which participants engage one another.

Engagement, as a mediator between learning and motivation, can be seen as a central element in a group process that defines the sociocultural environment (Sinha et al., 2015). Sinha and colleagues conceptualized collaborative group engagement as consisting of behavioral, social and cognitive engagement. They defined behavioral engagement as referring to continuous on-task behavior on learning activities (including the exhibition of focus), social engagement as the quality of social and emotional interaction in the group effort, and cognitive engagement to be associated with the planning, regulation and assessment of learning by the student. In addition, Sinha and colleagues proposed the notion of conceptual-to-consequential engagement, which reflects the progression toward the completion of a task through use of disciplinary content and strategies as conceptual tools. Considering the issue of small groups in completely online learning environments, Goggins et al. (2011) highlighted a number of challenges in initiating learning activities, including the unavailability of technology that enables the creation of experiences akin to a common physical work environment and the difficulties in simultaneously and meaningfully engaging the technology, the task and other participants.

From this perspective, Järvelä, Kirschner, et al. (2016) pointed out that collaborative learning should not be understood as group learning but rather interaction among individual participants gradually move toward acquiring shared meaning. Gunawardena et al. (1997) outlined five phases of the social construction of knowledge in CSCL environments. The first phase involves the exchange and comparing of information by the members of the group. This is followed by the identification and exploration of any contradictions or discrepancies that exist within the concepts and ideas presented. The third phase revolves around the negotiation of meaning and the co-creation of shared knowledge. In the next phase, the co-constructed knowledge is tested and revised, which is then summarized and applied in the fifth and final phase.

Another area of emphasis has been the roles of participants in CSCL contexts, particularly in understanding how learners interact and behave as members of a group or community (Hoadley, 2010). Strijbos and Weinberger (2010) observed that roles can foster the coordination and distribution of activities that support the group in reaching their shared objective. How the roles are assumed has been of interest to CSCL researchers, and special focus has been given to emergent and scripted roles. The former describes roles that are assumed spontaneously by participants to support collaborative learning and the latter refers to roles that have been prescribed to facilitate collaborative processes (Strijbos & Weinberger, 2010). Emerging roles are especially relevant to self-regulated learning, particularly in understanding how learners engage in the collaborative process.

Related to this work, Strijbos and De Laat (2010) proposed a conceptual framework that considers roles as participative stances, which takes into account the differing styles of participation based on the student's attitude toward the collaborative learning task. They defined the stances along three dimensions of group size (small or large), goal orientation (individual or group) and effort (high or low) to identify eight roles. For example, the 'lurker' role refers to a participant in a large group with an individual orientation exhibiting a low level of effort in the task. Based on a qualitative analysis of data examining unscripted groups participating in asynchronous CSCL tasks, Strijbos and De Laat (2010) asserted that the eight stances consistently reflected student behavior; nevertheless, it was also stressed that learner roles can affect one another and evolve over time.

The increasing availability of mobile computing devices in recent years has provided new opportunities for the application of technology in collaborative learning settings, particularly in the use of multimedia tools to deliver content and facilitate communication (Falloon & Khoo, 2014; Roschelle et al., 2007). In this regard, various curriculums have been developed to take advantage of the interactive functionalities of mobile devices along with the improved wireless capabilities of many instructional environments. At the same time, the capacity of students to comprehend and utilize emerging technologies has risen over time as the result of having greater exposure to electronic devices and software in their daily lives. Examining the role of mobile tablets in supporting collaborative learning at the secondary school level, Reychav and Wu (2015) focused on the effects of different content delivery modes (text versus video) and response submission types (individual versus group). Results showed that students who submitted individual responses demonstrated higher performance and satisfaction when learning content was provided through video, whereas text delivery was found to be more effective for students who had to negotiate a group response with team members. Based on these findings, Reychav and Wu concluded that content delivery methods need to be taken into consideration when designing collaborative learning interventions.

Assessment of Collaborative Learning in Online Contexts

Assessing collaborative learning requires taking into account not only the factors that affect the cognitive and social dimensions of the learning process but also the interplay between them. At the same time, collaboration in itself represents both a process and an outcome of learning. Therefore, it is essential that assessments properly capture both dynamics. Furthermore, collaborative learning in online contexts involve interactions that occur—synchronously and asynchronously—among students as well as between students and digital media and tools. For this reason, there is a need to go beyond simply measuring students' mastery of specific knowledge or skills, and place greater emphasis on understanding how learning unfolds within the complex conditions of practice (Shaffer et al., 2009).

Online environments present many opportunities for collecting different types of data that capture the learning process. These include interaction data generated during collaborative activities, frequently in the form of text data derived from written messages or transcribed discourse. User log records as well as learner-generated digital artifacts also provide a rich source of data which can be analyzed through a variety of methodological approaches. In a systematic review of 400 CSCL research publications from 2005 to 2009, Jeong et al. (2014) found that descriptive studies were most prevalent among a broad range of methodological approaches. In addition, quantitative analyses were most frequently utilized, either alone or in combination with qualitative methods, such as conversation and discourse analysis. Jeong and colleagues observed that numerous studies have applied different techniques in complementary fashion due to the large quantity and diverse nature of the datasets collected in CSCL settings.

Studies examining collaborative learning in online contexts have also incorporated network approaches to model and visualize interaction and engagement among participants. For example, social network analysis (SNA) techniques have been used to understand how actors in collaborative learning processes are related to one another. SNA can be used to identify patterns in the relationships that exist between participants as well as to assess the overall cohesion of a network through measures such as density and centrality (De Laat et al., 2007). Density provides information on the degree of connectedness among participants in a given network, while centrality is indicative of an individual's level of interaction with other members. Graphically, SNA results can be presented as a sociogram, in which the nodes represent the participants and the edges depict the intensity of interaction between them.

More recently, epistemic network analysis (ENA) has been utilized to model the co-occurrences of key concepts present in group discourse (Csanadi et al., 2018; Shaffer, 2017; Shaffer et al., 2016). ENA conceptualizes the collaborative thinking as cognitive connections made by individuals within a recent temporal context, which is operationalized through the use of a moving window that takes into account a prior segment of the discussion (Ruis et al., 2019; Siebert-Evenstone et al., 2017). Nodes of ENA network models represent the constructs, or codes, included in the discourse analysis, while the weighted edges indicate the strength of connection between each pair of constructs. Applying ENA to assess the relationship between cognitive and social presences in asynchronous online discussions, Rolim et al. (2019) found that social presence had stronger connections overall to the exploration and integration stages of cognitive presence. Additionally, analyses of the changes in the ENA network models over time provided insights into the evolving patterns of participant behavior exhibited during each week of a month-long online course.

Others have explored ways to integrate the study of cognitive and social dimensions of collaborative learning by combining SNA and ENA techniques. Gašević et al. (2019) utilized a method called the social epistemic network signature (SENS) to analyze more than 6,000 posts collected from a MOOC discussion forum. In this process, variables derived from SNA and ENA findings were used to build multiple linear regression models to predict the academic performance of students. Results showed that utilizing both SNA and ENA predictors produced the most effective model, implying the complementary nature of SNA and ENA variables in explaining the variance in student performance.

Another line of inquiry for CSCL researchers has been in the temporal analysis of longitudinal data. While the temporal dimension of activities and events constitutes a key element of collaborative learning, it has often been overlooked in the literature, despite the increasing availability of relevant data from CSCL environments (Reimann, 2009). Emphasizing the importance of the timing and sequence of experiences in learning processes and outcomes, Reimann (2009) argued for an event-centered approach—using techniques such as Markov chain modeling—to temporal analysis in CSCL research. Chiu and Khoo (2005) observed that temporal analysis of discourse data from collaborative learning environments often faces several challenges, including coding difficulties and the violation of statistical assumptions such as independence of observations and homogeneity of variance. A methodology proposed to address these issues is statistical discourse analysis (SDA), which models relationships among explanatory variables across different levels of time (Chiu & Khoo, 2005; Lu et al., 2011). SDA combines discourse analysis with different statistical techniques, such as vector auto-regression (VAR) and multi-level M-tests (Molenaar & Chiu, 2014). Applying SDA to the analysis of knowledge construction in online discussions, Wise and Chiu (2011) were able to identify "pivotal posts" that sparked new lines of discussion toward higher phases of knowledge construction.

Content Creation

Content Creation in Online Environments

Digital communities, including those involving adolescents and youth, have increased in recent years as a result of advances in technology that has facilitated the ability of users to create and share content (Harlan et al., 2012). As a result, the concept of content creation in the digital domain has garnered greater attention. However, there has been a lack of a widely accepted definition on content creation, which has presented challenges for those conducting research on the topic (Brake, 2014; Hoffmann et al., 2015). A survey carried out by the Pew Research Center on ‘online content creation’ in 2004 applied a broad interpretation of the concept, encompassing activities such as creating websites, posting materials to work-related and personal websites and blogs, and sharing information on online chat rooms and discussion groups (Lenhart et al., 2004). On the other hand, the definition of ‘user-created content’ adopted by the Organisation for Economic Cooperation and Development (OECD) was narrower in scope, focusing on digital artifacts characterized by the following three dimensions: public availability; creative effort; and production outside the context of professional settings or practices (OECD, 2007).

From a learning perspective, content creation can be viewed as a core component of media and information literacy (MIL), which according to UNESCO (2013) is the capacity to acquire, understand, assess, utilize, produce and share different types of information and media content (Drotner, 2020). MIL expands upon the concept of digital literacy, which refers to one’s capacity to comprehend and utilize information presented through computers in diverse formats (Gilster, 1997). Reynolds (2016) further argued that the skills-focused understanding of digital literacy is inadequate in capturing the evolving nature of the media environment. As such, Reynolds suggested that content creation should be understood within a concept of *social constructivist digital literacy*, in which people are viewed as autonomous agents who utilize technology for meaningful or productive intentions. Reynolds (2016) outlined six domains of practice within the framework of social constructivist digital literacy: create; manage; publish; socialize; research; and surf/play. Reynolds observed that the first three elements the constitute the primary activities of digital literacy, which are supported by the latter three components. Further applying social constructivist notions to digital content creation, Drotner (2020) underscored that content creation should be viewed as an activity situated within social practice, in which creators actively construct knowledge and meaning in their interactions with the social environment.

Content Creation as Creative Digital Participation

Content creation is also closely related to online participation, with many areas of overlap (Hoffmann et al., 2015). However, given that content creation may sometimes only involve a one-way dissemination of content, some have posited that online participation may represent a broader concept. For example, Lutz et al. (2014) defined online participation to not only encompass producing and sharing content but also having a particular audience and a social purpose. Based on ethnographic research on the new forms of digital media, Ito et al. (2009) identified two main modes of participation among youth: friendship-driven and interest-driven participation. The former refers to the everyday practices of young participants that focus on interactions and negotiations with their peers and friends, while the latter places emphasis on engagements the revolve around specific activities or topics of interest. Ito and colleagues further noted that transitions between these two modes can be understood through three additional genres of participation, which they termed hanging out, messing around, and geeking out. Hanging out and playing video games with friends can initially be friendship-driven but can result in interest-driven forms of participation. Likewise, engagements that are driven by interest can lead to meaningful interpersonal interactions and friendships that can go beyond any particular area or topic of focus. Ito and colleagues observed that the shifts between these genres of participation can be reflective of the continuously evolving nature of the participant's modes of learning and social engagement.

As such, content creation presents opportunities for "self-expression, sociability, community engagement, creativity and new literacies" (Livingstone, 2008, p. 394). Terras et al. (2015) noted that user-generated content is naturally reflective of the attitudes, behavior and identity of its creator. This process can support the identity development of young creators, as they are able to create and modify virtual characters and environments that are often representative of themselves (Beals, 2010). Craft (2012) further suggested that creativity is inherently present in the digital lives of young people, particularly in the adoption of different perspectives, construction of novel content, and pursuit of new possibilities. In this manner, content creation can be seen as a creative process resulting in the generation and sharing of digital artifacts in online environments. Applying the perspective of distributed creativity to content creation, Literat and Glaveanu (2018) outlined a framework for understanding online creative participation along three interconnected dimensions: social, material and temporal. The social dimension refers to the interactive and collaborative nature of online creativity, while the material dimension focuses on the technological affordances and commercial factors. The temporal dimension examines the different levels in which creative participation unfolds, from stages of creativity at the micro level to the evolving understanding of creativity at the macro level.

Conceptualizing Creativity. Creativity is a complex process involving numerous factors impacting its various phases, from idea generation to the of an innovative product or solution (Mumford, 2003). Over the past century, there have been numerous conceptualizations of creativity. Influential theorists include Graham Wallas, J. Paul Guilford and E. Paul Torrance. Wallas (1926) described creative thinking as a process comprised of four stages. A conscious, comprehensive yet unsuccessful analysis of the problem in the *preparation* stage is followed by the *incubation* phase during which no conscious thinking is done on the matter. The emergence of the new idea takes place in the *illumination* stage, after which it is tested and validated in the *verification* stage (Wallas, 1926). Guilford (1967) and Torrance (1965) proposed that creative potential could be assessed through divergent thinking, which refers to the ability to provide multiple responses to a given problem. Tools for measuring divergent thinking were developed, focusing on the components of fluency, originality, flexibility and elaboration (Runco, 2010).

Today, there is general consensus among creativity researchers that both originality and effectiveness are necessary conditions for creativity (e.g. Amabile et al., 2005; Robinson, 2011), where effectiveness can take the form of usefulness, suitability or appropriateness (Runco & Jaeger, 2012). In addition, there have been efforts to distinguish among different levels of creative impact, from what has been called the *big-c creativity* of extraordinary nature that are often attributed to eminent individuals to the *little-c creativity* referring to ideas that contribute to the enhancement of everyday life (Beghetto, 2010; Craft, 2003).

More recent research has recognized the significance of affect in creativity processes. Creativity and emotional experiences are often understood to mutually affect each other. Examining the relationship between affect and creativity using time-lagged analysis of longitudinal data, Amabile et al. (2005) found that positive affect preceded creative thought, with up to two days of incubation. On the other hand, qualitative analyses have indicated that positive affect occurred both concurrently and following creative thought events. Isen (1999) identified several ways through which positive affect was found to benefit cognitive functioning and creative problem solving. One was that positive affect enhances the capacity for organizing and classifying ideas. Another is that positive affect fosters flexibility in cognitive processing, allowing individuals to perceive different associations and dimensions. Moreover, positive affect was seen to contribute to innovative solutions through the ability to consolidate and combine ideas in novel and useful ways. At the same time, it may also be possible that negative affect can increase creativity. Citing the mood-as-input model, which suggests that current emotional state is used as an informational cue for determining the status of a situation (Martin et al., 1993), Amabile et al. (2005) observed that negative mood, as an indication of an unsatisfactory condition, may serve as a motivator for task-related behavior.

Social Dimensions of Creativity. While much of the earlier research on creativity focused on the creative outputs and processes of individuals (Amabile & Pillemer, 2012; Aragon & Williams, 2011), the conceptualization of creativity has evolved since the 1980s to emphasize the role of social and cultural contexts in influencing creative behavior. Creativity research has increasingly recognized the importance of interpersonal interaction and collaboration in creative work, embracing a more complex yet comprehensive notion of the concept (Kutaka-Kennedy, 2015; Mamykina et al., 2002). Applying a social-psychological lens to creativity, Amabile (1983) articulated a model of creativity which recognized the significance of the social context in creative activities. Referred to as the componential theory of creativity, this framework posits that social and environmental factors—along with the individual dimensions of personality characteristics, domain expertise and task motivation—formed the basic elements of creativity (Amabile, 1983). The model considers intrinsic motivation as being favorable to creativity, while extrinsic motivation is generally regarded as disadvantageous (Amabile & Pillemer, 2012; Hennessey & Amabile, 2010; Simonton, 2000).

A relevant framework is the systems model of creativity, which posits that creative behavior must be situated within the social, cultural and historical context in which they take place (Csikszentmihalyi, 1988). In this model, creativity is seen as the product of three interacting components: domain, field and individual. The domain is comprised of the body of knowledge and set of rules and values that define the area. The field is made up of persons engaged in the particular discipline, who play the role of gatekeepers (Csikszentmihalyi, 1997). Lastly, the individual is considered to have performed a creative act only when their new idea or product can effect change within a domain with consent from the field. From this perspective, all creativity can be understood as being socially-derived, in the sense that peer acceptance is needed to determine whether a novel idea or artifact is indeed worth incorporating into the particular domain (Csikszentmihalyi, 1988, 1997).

More recently, Sawyer and DeZutter (2009) proposed the notion of distributed creativity, which applies the concept of distributed cognition to describe the collaboration of individuals in collectively producing a shared creative output. They also define collaborative emergence as group processes that lead to unexpected creative outcomes. Such processes are characterized by unpredictable outcomes, influence of prior actions on the following ones, impact of subsequent actions on prior interactional effects, and the equal contribution of participants. Sawyer and DeZutter (2009) asserted that collaborative emergence results from the interaction among group members.

Learning and Creativity. Many consider creativity and learning to be closely connected. Guilford (1950) noted that a creative act can be seen as a case of learning, as it reflects a behavioral change resulting from a stimulus or response. Guilford, therefore, argued that creativity should to be taken into account in learning theories. Mumford (2003) observed that creativity involves the cognitive capacities for generating ideas and combining concepts that enable the creation of innovative outputs. Novel interpretations that emerge from the combination of knowledge and concepts lead to subsequent formulation of new ideas. Turvey

(2006) considered creativity as the integration of two seemingly contradictory actions, open exploration and focused cognition. As such, creativity consists of both divergent and convergent thinking processes (Coursey et al., 2018; Tan et al., 2014). The former allows for the search and discovery of new ideas and knowledge, while the latter supports their application in appropriate contexts (Turvey, 2006).

Considering the connections that exist between imagination—a creative activity—and reality, Vygotsky (2004) articulated four different types. First is that imagination is derived from experiences gained in real life. In order to enhance the creative potential of individuals, Vygotsky noted, it is important to expose them to a broad set of experiences that they will be able to integrate into their creative activities. Second is the linkage between imagination and the experiences of others. Based on descriptions and stories of what others have experienced—referred to as historical or social experiences—it becomes possible for individuals to use it in their own reconceptualization and formulation of imaginary situations and structures. The third type of connection is between imagination and affect, whereby imagination is closely related to the emergence of feelings that a person experiences in reality. Lastly, the fourth connection involves the embodiment of novel imaginary ideas in material form, which serves to complete the cycle.

Many have promoted the idea that enhancing creative potential should be a core educational goal (e.g. Beghetto, 2010; Robinson, 2011). Vygotsky (2004) argued that cultivating creativity in children should be the primary objective of school instruction. In the last several decades, there have been increasing policy attention to include creativity in the school curriculum (Beghetto, 2010; Craft, 2003). In addition to subjects such as visual and performing arts, opportunities for fostering creativity among learners are available throughout the curriculum

(Craft, 2003; Fasko, 2001). Innovative approaches to teaching, such as the use of teacher-led group improvisations, can foster greater creativity and collaborative learning in the classroom (Sawyer, 2006).

In a systematic analysis of 210 studies on creativity and learning in schools, Davies et al. (2013) found several crucial factors for enhancing the development of creativity in schools including: flexible utilization of time, space and other resources; integration of play- and game-based approaches that foster student autonomy; supportive teacher-student relationships; and opportunities for collaboration among learners. Emphasizing that group work with peers was effective for student creativity, Davies and colleagues also observed that engagement in creative activities in itself could encourage further collaborative behavior.

Meanwhile, others have highlighted a number of limitations for fostering creativity in classroom contexts. Craft (2003) outlined several main challenges, including limitations due to the central control of pedagogy as well as contradictions in policy and practice that relate to how creativity is mainstreamed into the curriculum. Beghetto (2010) observed that while schools do not necessarily stifle creativity, more needs to be done to support educators in fostering the creative potential of their students.

Outside of schools, efforts have been made in informal learning contexts to enhance creativity. One example is the Maker Movement, which refers to the rising interest in the general public for the creative production of physical and digital artifacts (Halverson & Sheridan, 2014). Grounded in the constructionist views of learning (Papert, 1993), the movement's experiential and problem-based approaches have been adopted in diverse learning settings, including museum and libraries.

Creative Digital Participation through Collaborative Interaction. Fostering creative digital participation through collaboration in online learning environments introduces new dimensions that need to be taken into consideration. In this regard, Fischer (2004) emphasized that social creativity in CSCL contexts can be distributed spatially (across physical distance), temporally (across time), conceptually (across communities), and technologically (between individuals and artifacts). While these dimensions can act as barriers, Fischer (2004) argued that building bridges across boundaries can enhance creativity by providing opportunities to integrate different ideas and generate new solutions. Online technologies, therefore, can serve as communication tools that support collaborative creativity by facilitating the distribution and diversification of creative activities and processes (Fischer, 2007).

Through an analysis of collaborative interactions among virtual math teams consisting of middle school students, Sarmiento and Stahl (2007) concluded that the creative engagement of learners in online environments are based on three key processes encompassing both synchronic and diachronic interactions. Synchronic interactions take place simultaneously and in parallel with one another, while diachronic exchanges occur over longer time periods and involve mediating objects. The first process consists of the establishment of a network of interrelated references that point to a shared meaningful object. The second relates to collective remembering, in which students participate dynamically to reconstruct past work and enable its continuation in the present. The third revolves around the bridging of discontinuities across the boundaries of time, perspectives, activities or teams through the use of shared artifacts, spaces and other resources. These interactive mechanisms, Sarmiento and Stahl argued, enable teams of participants to engage in collective creative work by facilitating group cognition and maintaining continuity throughout the process.

Furthermore, Aragon and Williams (2011) posited that creativity in distributed groups is supported by technological interfaces that facilitate social and affective communication. Highlighting the importance of distributed affect in group creative processes, Aragon and Williams proposed a new model for collaborative creativity in online settings that consists of four phases: focus, frame, create and complete. The group identifies the problem and determines the collective goal during the *focus* phase, then comes together as a cohesive team through a process of enculturation and affective integration in the subsequent *frame* phase. The *create* phase involves the generation, sharing and modification of ideas in an iterative manner. Interactions—between group members, people and information as well as among ideas—play a key role during this stage. In the *complete* phase, the final ideas are consolidated and evaluated, leading to an alignment with the collective goal.

In computer-supported learning contexts, Tan et al. (2014) conceptualized collective creativity as being comprised of three skill dimensions: (a) metacognitive: ability to jointly examine and regulate the group's shared goals and strategies; (b) cognitive: capacity to collectively generate and assess new ideas and solutions; and (c) socio-communicative: ability to participate in productive, prosocial interactions that facilitate the first two dimensions. Analyzing dialogic patterns among student pairs collaborating on creative problem-solving tasks, Tan and colleagues found that successful dyads engaged more frequently in discourse that established mutual understanding and involved reciprocal questioning and responding. In addition, the results also indicated that successful pairs engaged more often in on-task cohesive talk while unsuccessful dyads participated more often in playful off-task dialogue.

Online Learning Communities

Interaction and Collaboration in Online Learning Communities

Social interaction in communities are a fundamental part of understanding how individual learning shared meaning making (Jeong et al., 2017). Interaction and communication allow community members to understand one another, share information, build shared knowledge and co-create artifacts. Preece (2001) defined an online community as “virtual social space where people come together to get and give information or support, to learn or to find company” (p. 348). This definition was broadly conceived to encompass small or large communities that bring together members from local, national or international levels. Driven by a shared purpose and facilitated through technology, the characteristics and outcomes of online communities can vary depending on several factors (Preece et al., 2004). These factors include: the primary goal of the community; the technological tools, such as software and web-interfaces, used to support its activities; the size and history of the community; and the extent to which virtual presence is utilized.

A related concept is the community of inquiry (COI) framework, which is defined as “a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding” (Garrison, 2011, p. 15). The framework posits that online learning is the result of the interaction of cognitive, social and teaching presence (Garrison et al., 1999). Cognitive presence refers to how participants in the community make meaning through continuous communication and interaction. Social presence, according to Garrison et al. (1999), focuses on how well the participants can reveal their personal qualities to present themselves as a ‘real’ person. Finally, teaching presence involves the two functions of design and facilitation, which may be carried out by any member of the community. The design function includes selecting and organizing the content and activities related to the learning experience. Garrison et al. (1999) noted that the main role of the teaching presence is to support cognitive and social presence to enhance the educational outcomes of the computer-mediated community.

Related to the notion of learner interaction, Turvey (2006) put forth a conceptual framework for understanding the learner's activities and interactions within an online community, which can facilitate the shared construction of knowledge among its members. In this process, the online community allows participants to explore their personal identities, engage in collaboration through online media and tools, and develop collective goals through dialogue and negotiation. In this manner, students gain the skills needed to engage in learning that goes beyond the artificial boundaries imposed by school environment or curriculum. Online communities can be viewed as mechanisms that support the removal of such boundaries, empowering students to become lifelong learners. At the same time, Turvey stressed that although participants in online communities interact and communicate in the virtual context, they are nevertheless situated within the cultural norms and values of the real world. As individuals participating in online communities are also part of real communities, it is crucial to take into account the influence exerted by the real world on the virtual environment.

Based on a review of 83 empirical studies published between 2002 and 2014 on the topic of participation in online communities, Malinen (2015) concluded that the literature has been dominated by descriptive quantitative research, which may have led to less attention being given to theory development in the field. Also noted was a lack of consistency in defining participation, which was largely operationalized in the studies in terms of the amount of activity. Furthermore, the meta-analysis found a gap in research examining the social influence that participants exert on one another in the online communities. In this regard, Malinen recommended a greater focus on the quality of participant activity, particularly as it relates to the overall performance of the community.

Moreover, Malinen (2015) noted that participation is integral to the success of online communities, highlighting the important role that active engagement in social media platforms—through the creation and sharing of media—have played in their proliferation in recent years. One of the key elements to this trend has been the blurring of the distinction between producers and consumers of digital content (Malinen, 2015). As increasing numbers of users engage in creative activities in such settings, there is a need to reconsider how creativity is understood and assessed (Peppler & Solomou, 2011). Unlike traditional communities where expertise is concentrated around its core members, the virtual environment enables the distribution of relevant knowledge and skills across a large number of participants, who contribute to determining whether an idea or product is new and useful. Examining the creative behavior among students in a multi-user virtual environment called Quest Atlantis, Peppler and Solomou (2011) found that the community was able to foster cultures that supported the creative engagement of participants. Dialogue among participants was a critical component of this process, as it allowed for new ideas to spread throughout the community and for the members to collectively establish their cultural values. Based on these observations, Peppler and Solomou concluded that creativity is a social and cultural endeavor that is informed and sustained by the norms and behaviors of the community.

Large-scale Online Learning Communities

Mass collaboration refers to situations where a large number of individuals learn and work together (Fischer, 2016; Jeong et al., 2017). Fischer (2016) posited that mass collaboration takes place in socio-technical environments that are supported by the technical affordances of the

Internet and enable the collaborative action of numerous participants in information sharing, problem-solving, and knowledge building. From a social perspective, massive collaboration involves a network of a large group of participants who are distributed not only geographically but also conceptually and temporally. This differs from smaller communities where close relationships shape and reinforce communal goals and beliefs. Technically, massive collaboration takes advantage of the advances in the infrastructure of digital networks that can connect users across the world. As such, Fischer observed that mass collaboration enables learners to become active contributors and co-creators in an expansive, decentralized community where participation is valued and encouraged, innovation is shared openly as work-in-progress, and learning—rather than teaching—becomes the focus of education.

Focusing on large-scale online knowledge communities, Jeong et al. (2017) proposed a conceptual model for understanding joint interaction among members. Referred to as the A3C framework, the model identifies four types of interaction, namely attendance, coordination, cooperation and collaboration. According to Jeong and colleagues, the different types of interaction can be further distinguished along the dimensions of goals, processes and outcomes. *Attendance* refers to activities that are carried out without a clear intent to work together with other participants in the community. Individuals who only exhibit attendance behavior are often called lurkers (Sun et al., 2014). While there is a tendency to consider these passive members in a negative light, Jeong and colleagues pointed out that new participants may require time to become more familiar with community activities and processes. *Coordination* describes interactions in which members align their activities with others, but do not engage in cooperative or collaborative actions. While coordination behavior is motivated by individualistic goals, but its outcomes are interdependent. With *cooperation*, the goal become shared among those involved; however, as the work is carried out through a division of labor, the outcomes of cooperative interactions are the sum of the individual contributions. Finally, *collaboration* refers to situations in which participants undertake collective effort to achieve shared goals. Contributions of members are integrated into the outcome, with individual inputs becoming indistinguishable. Jeong and colleagues further noted that the different forms of interaction are not mutually exclusive; rather, each level of interaction expands upon the previous level (e.g. cooperation involves attendance and coordination). Likewise, learners are able to transition between levels depending on the community and specific tasks involved.

Scratch Online Community

The Scratch online community is a learning platform developed to enhance computational thinking and programming skills among young people (Brennan & Resnick, 2013; Resnick et al., 2009). Launched in 2007, it has evolved into a large-scale online community with more than 50 million registered users around the globe. In the community, users are able to create and share digital artifacts while interacting with one another through comments and other social practices (Hill & Monroy-Hernández, 2017).

Previous studies have examined various aspects of the Scratch online community, ranging from the promotion of youth agency (Kafai et al., 2012), leadership (Roque et al., 2013), and civic engagement (Roque, Dasgupta, et al., 2016) to the development of programming skills (Dasgupta et al., 2016; Fields et al., 2014; Gutierrez et al., 2018) and game design expertise (Kafai & Peppler, 2012). Other studies have explored the issue of diversity among users of digital media (Richard & Kafai, 2016) and the role of the audience in learning processes (Brennan, 2016). However, this review will focus on the select pieces of research that specifically address issues of interaction, collaboration and content creation in the Scratch online community.

Participant Roles and Behavior

One line of research on the Scratch online community has focused on identifying the various types of participants and assessing their pattern of behavior. Fields et al. (2013) examined the online activity of Scratch users to investigate the different patterns in their participation practices. Applying latent class analysis to the 3-month log file data of a random sample of 5,004 users, the study explored the interrelations among six activities: downloading, remixing, commenting, favorites, love-its, and friend requests. Downloading enables the user to view the programming code of another user's project. Remixing refers to the downloading, editing and sharing of projects initially posted by another user (Monroy-Hernández, 2009). Commenting can be done for individual projects or a gallery of curated projects. Favorites and love-it's both allow users to express interest and appreciation for another user's project. The difference lies in how they are presented. While favorites are displayed on a list belonging to the user making the selection, love-it's are shown as a heart icon on the project itself. Finally, friend requests can be made to follow another user. The results of the analysis by Fields and colleagues (2013) suggested that there are several classes of participants and that active users tend to concentrate on creating and downloading projects. Another key finding was that producing at least one project per month served as an entry point to other participatory behavior.

Building on this research, Fields et al. (2016) focused on the social practices of users to analyze shifts in participation patterns over time. Latent class and transition analyses were used to examine a subset of 2,225 users who had created at least one project during a 3-month period. For this study, user activities were grouped into three types of social practices: DIY participatory activities (sharing, remixing and downloading projects); socially supportive action (loving and favoriting projects); and socially engaging interaction (commenting and friending). Results identified five distinct latent classes present in the sample of users. It was found that the majority of low networkers, who only create and share projects, remained in the same class for the duration of the study. A similar pattern was also identified for most high networkers, who were consistently engaged in all activities. Finally, no differences in gender and length of membership were found among the classes of users.

Role of Online Interactions on Content Creation.

Another area addressed in the research has been on the role of online interactions on the creation of digital media artifacts in the Scratch platform. Using case study methodology, Brennan et al. (2010) explored how different modes of engagement can influence content creation in the Scratch online community. In the study, participation is conceptualized as falling within a spectrum between socializing and creating behaviors. Socializers are driven primarily by social interaction and group dynamics, whereas the creators are mainly focused on producing projects. Three narratives showcasing exemplary user experiences are presented to demonstrate how various configurations of collaborative participation in the online community—including working in pairs as well as in small and large groups—facilitated the construction of interactive digital media. In this manner, the processes of socializing and creating can be viewed as playing complementary roles to support and catalyze project creation among Scratch users.

Similarly, Brennan et al. (2011) examined the social dimensions of participation on the Scratch platform, focusing on the question of how users develop as creators within the online community. Taking a qualitative approach, the researchers initially conducted ethnographic observations of the community, then selected and interviewed six individual users about their experiences. This resulted in the elaboration of six case studies, each representing a different mode of participation: newcomer, remixer, collaborator, teacher, moderator, and contributor. The case studies placed emphasis on the role that the social context has played in supporting their growth as creators of digital media and members of the community. Overall, the findings indicated a progression in the roles that learners assume over time. However, the changing participation patterns were in no way linear; rather, learners exhibited a trajectory of development that was based on their relationship to other participants in the community. The first three types of participation (newcomer, remixer, collaborator) were seen to be driven primarily by *action*, including creating projects and interacting and working together with others. The last three types (teacher, moderator, contributor) were oriented towards *reflection-on-action*, which involves thinking about the meaning of being a creator and ways to support that process. Based on this analysis, Brennan and colleagues concluded that the social contexts afforded by the Scratch online community can contribute significantly to creative behavior.

Analyses of User Comments to Assess Interaction and Collaboration

Other studies have analyzed user-generated comments from the Scratch online community to assess interaction and collaboration. Aragon et al. (2009) analyzed user comments to explore which socio-technical conditions are necessary to foster collaboration on creative work. One of two communities examined in the study, the Scratch data was collected from a user-initiated online group working together to create projects. As part of the analysis, a total of 1,470 comments shared among group members over a period of 3 months were categorized into one of three categories: contextual, task, and socio-emotional. The 'contextual' category includes issues connected to work but not directly task-related, such as scheduling and the organization of the group. The 'task' category refers to discussions focusing on the job that needed to be completed. The ‘socio-emotional’ category involves conversations that are personal or intended for socializing, such as greetings, opinions, and jokes. The results found that 19% of the comments were task-related, 49% were socio-emotional and 32% were contextual. The high frequency of socio-emotional comments can be understood from the viewpoint that most of the participants are children who are engaged in Scratch as a hobby. An interesting phenomenon noted by the researchers was that the participants at times used the comments feature, which is asynchronous in design, in a manner resembling a real-time chat dialogue. This was done by constantly refreshing the web browser, and this may have enabled the participants to develop a greater sense of presence in their online communications. The contextual comments make up about a third of all conversations. This could be attributed to the need for establishing common ground among the participants, given that the collaboration is entirely virtual. The study also included a survey component, which provided insights on the importance of motivation in fostering collaboration among users in the group. Responses from group members indicated that a strong motivating factor was the potential for gaining social status in the Scratch online community. Another key driver was the sense of responsibility and accountability toward the group.

Building on this research, Velasquez et al. (2014) utilized automated coding techniques to analyze user-generated comments on the Scratch platform. In particular, message feature mining was carried out using the General Architecture for Text Engineering (GATE) tool to extract the linguistic cues associated with each comment. The final analytic sample included 4,536 comments, half of which were identified as being project-related as the result of a manual coding of the data. A total of 39 linguistic cues were determined using GATE. Based on a comparison of the means, statistically significant differences between the project-related and other comments were found for 14 of the linguistic indicators. Project-related comments were higher in 11 of the dimensions: quantity (words, verbs, modifiers), expressivity (emotiveness), diversity (content, redundancy), specificity (spatial indicators, imagery), and affect (affect words, pleasantness, activation). Comments not related to projects were higher in 3 dimensions: informality (misspelling) and pronoun (group, other). These results suggest that comments related to projects were richer in a number of ways, including length, verbs, adjectives, imagery, and pleasantness. The researchers observed that this could be indicative of a more thoughtful engagement by the participants. On the other hand, non-project comments were found to have higher usage of group pronouns, suggesting that they may be more closely associated with relationships and collaborative interactions between participants.

Utilizing the same dataset of 2,273 project-related comments, Fields et al. (2015) carried out in-depth analyses examining the quality of comments, focusing on the constructive, emotional and functional dimensions. Three different coding schemes were used to analyze the data. First, two categories, specific and simple, were used to identify whether a comment was constructive. This analysis was grounded on the understanding that while simple comments tend to close communication, detailed constructive comments can lead to further interaction. Results showed that more than half (58%) of the comments were constructive. Second, the emotional tone of each comment was annotated as being positive, negative or neutral. Findings showed that a large portion of the comments were positive (72%) while the negative and neutral comments each accounted for only 14% of the total. Third, the functional focus of the comments was categorized into one of six themes, which were identified through grounded analysis. The frequency for each theme was as follows: motivational feedback (58%); personalized tutoring (9%); agency in learning (5%); building a following (23%); cultural competence (2%); and conversational partners (3%). Overall, the findings of this study suggested that the majority of project-related comments shared on the Scratch platform were positive and encouraging in nature. The substantial focus of participants on developing a following could be indicative of the fact that while participants seek to achieve a certain level of social status in the community, difficulties exist for gaining recognition and visibility in the large-scale online environment. Nevertheless, Fields and colleagues concluded that commenting is an important facet of the participation and engagement of learners in the Scratch online community.

Summary

This chapter has presented an overview of the extant literature related to interaction, collaboration and content creation in online learning communities. From a constructivist perspective, learning can be perceived as a fundamentally interactive process that occurs as individuals make meaning from their social and physical surroundings. Interactions not only provide information and experiences that allow a person to build knowledge for themselves, but also enable collaboration among groups of individuals that facilitates the co-construction and negotiation of shared understanding. Learning, therefore, can be seen as being situated within cultural and historical contexts, which shape how individuals participate as members in communities of practice that are defined by a shared sense of significance and purpose.

While online learning environments provide greater access to resources and opportunities for interaction and collaboration, they can also present communication challenges due to limitations posed by the technology. Research indicates that factors such as the level of social presence and the roles of participants strongly influence learning in online contexts. Participation in online communities further allows students to engage with a large number of fellow learners who share similar goals and interests. Often informal in nature, these communities provide an interactive learning environment that transcends physical, temporal and conceptual boundaries that exist in traditional educational settings.

Content creation not only refers to the production and sharing of digital artifacts but also involves creative effort that takes place outside of the context of professional practices. Viewed from a social constructivist lens, it is an activity situated within social practice, in which the creators actively construct knowledge and meaning in their interactions with the social context. As such, content creation can be understood from the perspective of creative digital participation. Creativity has been increasingly understood as encompassing social and cultural dimensions. A growing body of research has suggested that interaction and collaboration among individuals can have substantial impact on creative behavior. According to this view, content creation can be seen as a social process. Fostering creative behavior is increasing seen as an educational goal, with efforts to incorporate into formal learning processes. At the same time, several interest-driven initiatives have also taken root, especially in informal settings, based on the constructionist notions of learning as a process of active and self-directed creation of concrete ideas and artifacts.

As an online community of media creators, the Scratch platform has attracted millions of young users worldwide. Prior research has examined the patterns of behavior of different roles and activities of participants in the Scratch community as well as changes in their patterns over time. Through analyses of user-generated comments, studies have focused on how online interaction and collaboration affect creative performance. Findings based on decontextualized data indicate commenting plays an important part of how learners participate and engage with one another in the community. However, there is a need for further studies examining the interactive discourse that are contextualized within conversations.

Chapter 3: Research Design and Methodology

Introduction

The purpose of this study was to explore how social interaction and collaboration are related to the creation of content by users in the Scratch online community. In particular, the research examined the online activity of 200 Scratch users over a 3-month period in early 2012. Analysis was focused on the interaction of participants through the exchange of comments as well as the creation and sharing of projects on the platform. The study was carried out in several stages, involving qualitative, quantitative and network approaches to provide an in-depth perspective at the relationship between interaction, collaboration and content creation in informal online learning environments. Specifically, the study sought to address the following research questions:

- **RQ 1:** What patterns of discourse are observed in the comments of Scratch users? Are there differences in the discourse patterns among users with varying modes of participation?
- **RQ 2:** What relationships, if any, exist between the different elements of interaction and content creation?

Research Design

Within a constructivist learning context, interaction can be viewed as the process that enables the negotiation of meaning and the co-construction of knowledge among participants (Damşa, 2014; Gunawardena et al., 1998). Language plays a critical role in this process, not only in communicating information and content but also in conveying social contexts and personal attitudes (Brown & Yule, 1983). In this study, the written comments of users in the Scratch online community were analyzed as the medium through which interpersonal interactions are enacted and embodied. The research design, therefore, focused on examining the interactional discourse contained in the user-generated comments and exploring its association with the user's content creation behavior. In addition, the contextualized nature of the interactive comments was taken into account in the study. In analyzing discourse, Gee (2018) stressed the importance of examining how language is used and understood in specific circumstances and situations involving social and cultural interactions. As such, the discourse corpus was constructed in a manner that preserves the situated meaning and context of each comment. Comments within threaded discussions were grouped together into a single conversational unit for analysis.

The research design also sought to apply and integrate multiple analytical approaches to examine the influence of social interaction and collaboration on the creation of content in the Scratch online community. The main modes of analysis involved epistemic network analysis and the linear modeling of time series data. Each analytic approach was used to address a corresponding research question, with the ENA results being utilized in the subsequent time-series analysis.

Setting

The data used for this study was collected on the Scratch online community, which was created in 2007 to support the development of computational thinking and programming skills among youth (Resnick et al., 2009). Since its launch, Scratch platform has been continuously upgraded. The latest version, Scratch 3.0, was released in January 2019. It should be noted, however, that the current study focuses on data from January to March 2012, during which Scratch 1.4 was in usage. As such, all description and information provided in this study refer to this earlier version of the Scratch platform.

The Scratch platform consists of the two interconnected components, the authoring environment and the online community (Hill & Monroy-Hernández, 2017). Users are able to produce interactive digital projects, including animations and video games, in the authoring environment. In creating the projects, users have the option to either construct a new project or to download an existing project (with all associated graphics and coding scripts) and make adjustments to it, in a process known as remixing (Monroy-Hernández, 2009).

The projects can then be shared with others on the Scratch online community, where the users can provide comments and communicate with one another about the posted projects. The online community includes a number of social networking features designed to facilitate participation and interaction among users. In this regard, Fields et al. (2016) identified three main categories of social practices available in the Scratch online community: (a) do-it-yourself participatory activities; (b) socially supportive actions; and (c) socially engaging interactions. Activities in the first category include sharing, downloading and remixing projects, which do not necessarily involve interactions with other users. The second category consists of socially oriented actions that are primarily unidirectional, such as clicking on the love-it (to express appreciation for the project) or favorite (to add it to one's own list of favorite projects) buttons. The final category contains actions that are directly interactive, including posting of comments and sending a friend request to another user.

Human Subjects Considerations

The Scratch dataset used for this study was published by Hill and Monroy-Hernandez (2017) and is accessible online for approved researchers. In order to ensure the privacy of users, the dataset only contains data made public by users on the Scratch website. The Committee on the Use of Humans as Experimental Subjects at MIT approved the protocol for the publication of the dataset. The protocol requires researchers working with the dataset to provide their name, email address, affiliation, title and a description of the project that will utilize the data. The researcher also needs to sign to the Scratch Research Data Sharing Agreement (SRDSA), which outlines that:

> researchers use the data only for scholarly and research purposes; prohibits sharing, redistribution, or republishing of the data; prohibits attempts to identify or contact individual Scratch users; and requires that researchers attempt to maintain the anonymity of individual users. (Hill & Monroy-Hernández, 2017, p. 12)

Furthermore, researchers are requested in the agreement to not quote directly from certain parts of the data in order to ensure the anonymity of the authors of the data. To this end, researchers are not allowed to quote text from the "project_text" and "galleries_text" tables. However, Hill & Monroy-Hernandez (2017) notes that "the text of comments (i.e., text stored in the pcomments_text and gcomments_text) tables, which are not included in search results, can be quoted" (p. 13). The agreement also asks the researchers to cite publications describing the Scratch project and dataset. This study was conducted in accordance with the requirements and guidelines set forth in the SRDSA. Furthermore, the analysis of the dataset for this study was carried out after receiving the approval of the Institutional Review Board (IRB) of Pepperdine University (see Appendix).

Data Management

The dataset obtained from the MIT Media Lab for this study was kept in a digital format on a password-protected computer managed by the researcher. The dataset will be deleted from the researcher's computer after 3 years of the completion of this study.

Data

Data used for the study was sampled from the longitudinal dataset of the public activity on the Scratch online community from 2007 to 2012 (Hill & Monroy-Hernández, 2017). The full dataset includes records related to 1,056,951 registered users. During the first 5 years of the operation of the Scratch website, 304,793 users (about 29% of registered users) created a total of 1,928,699 projects, which averages to approximately 6.33 projects for every creator. In the online community, 165,214 users (about 16% of registered users) posted a total of 7,788,414 comments, leading to an average of around 47.14 messages per commenter during the 5-year period.

Variables

The dataset contains numerous variables that provide information on user activities and projects shared on the Scratch website. Unique identifiers are available for each user, project and comment along with a timestamp indicating when the particular data object was created. Project-level variables include log data that reflect user actions involving a given project, such as the number of total views, likes, and downloads, as well as information related to the project's type and source code. Two variables from the latter category were utilized in the study to first determine whether a project was new or remixed and then to identify any remixed projects that were identical to the original. The comment-level data contain the full text of the user-generated comments and the identifier for the prior message to which the comment was composed in response. The "reply.to.pcomment.id" variable was used to organize the comments into conversations consisting of initial messages and their linked responses. Table 1 presents a summary of the main variables used in the study.

Table 1

List of the Main Variables Used in the Study

Category	Variable	Description
User	user.id	Unique identifier assigned to each user account created
	date.created	Date and time that the user account was created
Project	project.id	Unique identifier assigned to each project shared on the website
	user.id	Identifier for the user who created the project
	date.created	Date and time that the project was first shared on the website
	viewers.website	Number of times the project has been viewed
	lovers.website	Number of unique users who have clicked the "love-it" button
	downloaders.website	Number of unique users who have downloaded the project
	is.remix	A Boolean identifier (T/F) indicating whether the project is a remix
	is.remix.identical	A Boolean identifier (T/F) indicating whether the remixed project is identical to the original
Comment	pcomment.id	Unique identifier assigned to each project comment
	project.id	Identifier for the project that is the target of the comment
	user.id	Identifier for the user who posted the comment
	date.created	Date and time that the comment was posted
	text	Full text of the comment posted on the website
	reply.to.pcomment.id	Identifier for the prior comment to which the current comment is a response

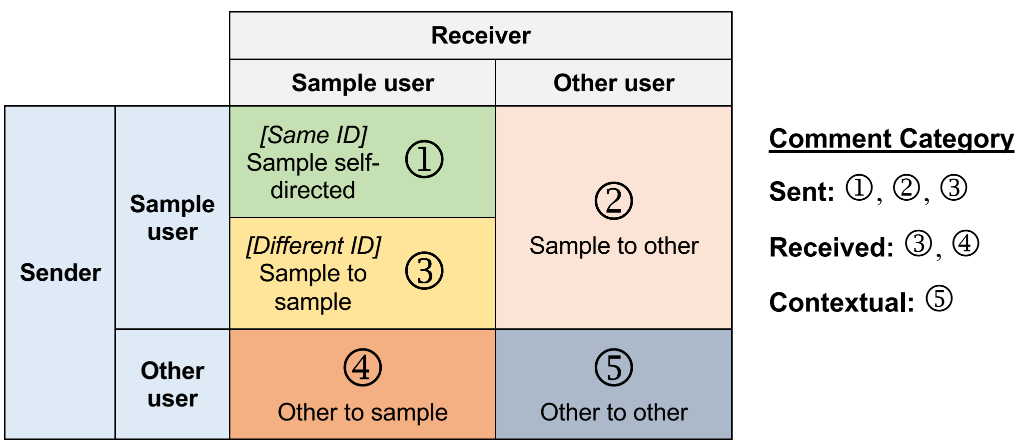
Data Structure

As noted earlier, the threaded nature of the comments presented some challenges for grouping related comments into conversational units. The comments can be added at one of three levels: (a) as a first level comment on the project; (b) as a reply to a first-level comment; and (c) as a reply to a second-level comment. Comments were displayed in the Scratch project sites using a mix of a hierarchical arrangement for the first level and a linear arrangement for the second and third levels. This meant that while a nested structure was visible between the first and second/third levels, all messages in the second and third levels were presented in chronological order. Given this data structure, the following approach was taken to organizing the comments into conversational units. First, all comments linked to a common first level comment were placed into a single conversational unit. Second, all second and third level comments falling within one first level comment were organized chronologically.

Given this organization of the comment data, each comment was classified into one of three typologies: comment sent, comment received, and contextual comment (see Figure 2). The first type (comment sent) refers to the comments that were composed by a sample user. The second type (comment received) denotes comments that have been explicitly directed at a sample user, including first level comments on the user's project page and second/third level responses to a comment posted by the user. The third type (contextual comment) was reserved for all other comments in the conversational unit that provide situational information for the comments sent and received by the user. All three types of comments that are part of a conversational unit included in the sample dataset were analyzed in the study.

Figure 2

Types and Categories of Comments



Sampling

Data Selection Criteria

Data for this study was confined to comments and projects related to user activity between January 1 and March 31, 2012, the most recent 3-month time frame for which data is available in the full dataset.

A criterion was applied at the user level to ensure that the sample would be drawn from users who had maintained an active and sustained engagement in the Scratch online community through the creation of projects throughout the period under research. For the purposes of this study, active and sustained engagement was operationalized as the creation at least one project in each calendar month between January 1 and March 31, 2012. This criterion was based on the finding by Fields et al. (2013) that creating at least one project per month served as an indicator for broader participation in the Scratch community.

Sample Dataset

Sample Users and Projects. From full dataset, a subset was generated consisting of users that created one or more projects during the period from January 1 to March 31, 2012. This subset of 38,282 users was further narrowed down by selecting those who had created at least one project in each calendar month, resulting in 2,561 users. From this group, a random sample of 200 users was drawn. As for the projects, both new and remixed projects created during the period were retained in the data. However, remixed projects that were identical to their original versions were excluded, as the sharing of such projects does not involve the contribution of additional content.

Sample Comments. The sample comments dataset was constructed to include all comments in conversations involving one or more messages that were either posted or received by sample users between January 1 and March 31, 2012. This process was complicated by the fact that relevant comments could be found under projects that were produced by both sample users and other creators as well as created during or prior to the research period.

The first step in the creation of the sample comments dataset involved identifying all projects in which sample users posted a comment from January 1 to March 31, 2012. The projects thus categorized contained not only the comments posted by sample users but also any comments received as responses to the sample users' messages. The resulting list consisted of 8,132 projects, of which 1,784 were created by sample users. The remaining 6,348 projects were produced by other users.

The second step was to generate a list of all projects containing comments that were produced by sample users at any time from 2007 to March 2012. This step ensured that all first-level comments composed by other users for projects created by sample users—categorized as "comments received" in this study—would be captured in the sample comments dataset. It was found that a total of 14,106 projects with comments were created by sample users, including 12,322 projects for which they had not posted any comments during the research period.

A summary of the number of projects identified in the first two steps are shown in Table 2. Combining the results produced a list of 20,454 projects that were linked to (a) the comments made by sample users during the research period, and (b) all projects created by sample users between 2007 and March 2012. Projects created by other users in which the sample users did not post comments in the 3-month period were not considered in this computation, as they would not contain any comments relevant to the study.

Table 2

Number of Projects Linked to the Comments and Projects of Sample Users

Type	Projects to which… Sample users posted comments in Jan-Mar 2012	Sample users did NOT post comments in Jan-Mar 2012	Total
Projects created by sample users	1,784	12,322	14,106
Projects created by other users	6,348	XXXX	6,348
Total	**8,132**	**12,322**	**20,454**

The third step began with the creation of a subset of all comments that were posted between January 1 and March 31, 2012. This subset consisted of a total of 611,611 comments. Of these comments, only those associated with the 20,454 projects identified earlier were preserved, resulting in a list of 9,028 projects in which relevant comments were present.

In the fourth step, all comments associated with the 9,028 projects were identified. This set of 420,356 comments was then organized into conversational units consisting of a first-level message and related second- and third-level replies, if any. As noted earlier, the second and third-level comments were arranged chronologically.

The fifth and final step involved the selection of the conversations containing at least one comment posted or received by sample users during period of study. This led to the final dataset, which was comprised of 45,617 comments within 18,354 conversations in 9,028 projects.

Data Analysis

Descriptive Analysis

Descriptive analyses were carried out to provide initial insights into the constitutive components of the sample as well as to inform subsequent analyses. Summary statistics were calculated at the level of the user, project, and comment. At the user level, the analysis focused on dimensions such as the number of projects created, the number of comments posted and received, and their length of membership. At the project level, measures related to originality of projects (new vs. remix) and community reception (views, downloads, like-it's, etc.) were examined. The comment-level analysis looked into elements such as types of comments (sent, received, contextual) and frequency of commenting behavior.

Qualitative Coding

Qualitative coding was conducted on the project comments to identify the salient topics and themes present within them. The study applied an adapted version of the coding scheme developed by Fields et al. (2015), which was used to annotate the functional focus of the project comments shared on the Scratch online community. Developed through a grounded analysis of 2,273 comments, the scheme consists of 21 functional categories that have been grouped into the following six themes: Motivational Feedback; Building a Following; Personalized Tutoring; Agency in Learning; Cultural Competence; and Conversational Partners. A key limitation in the data analyzed by Fields and colleagues was the decontextualized nature of the comments, which resulted in the inclusion of the Conversational Partners theme for comments that appear to be responses and disagreements to earlier messages.

Table 3 presents the codebook utilized for this study, which contains four key adaptations on the scheme advanced by Fields and colleagues. First was the removal of the Conversational Partners theme. Given that the comments analyzed in this study are contextualized within conversations, this theme was considered to be no longer necessary. Second was the creation of a new thematic group labeled "Conflict" which brought together two related subcategories of Criticism (from the Motivational Feedback theme) and Disagreement (from the Conversational Partners theme). The addition of this theme enabled the coding scheme to capture instances when negative responses were provided through the comments. The third adaptation was the renaming of the Motivational Feedback theme as “Supportive Feedback,” as it consisted only of the Encouragement and Sharing Personal Experience subcategories following the reassignment of the Criticism category into the new Conflict theme.

The fourth adaption was the addition of a theme called “Role playing.” This phenomenon was observed during the initial stage of the interrater reliability process, when both raters recognized that comments in certain projects included conversations that resemble a play script with characters, dialogue and stage directions. Below is a snippet from a conversation involving three Scratch users:

> User A: Nobu: I found a whole thing of jems. *Gives darkness a whole bag of jems full of power*
> User B: Darkness: thxs*takes bag*
> User A: Nobu: No problem. Anything to stop him will help.
> User B: Darkness: Hmhp...*walks awayand looks thourgh bag*
> User A: Nobu: Hopefully everything in that bag can be used....
> User C: dark kyuililbis: makes forse field and does chaos blast aims at prokyrate.
> User C: kyuil: darkness chaos shift hyper kyuil 3 blast Prokyrate.

Due to the interactive nature of the activity, the role-playing sequences were captured in extended conversations between the participants. In order to take this observed phenomenon into account, the two raters agreed during the social moderation process to add the Role Playing construct into the coding scheme.

The resulting codebook comprised seven themes, or codes, which were used to annotate each comment in the study dataset. As noted above, Supportive Feedback referred to comments that offer positive input through encouragements and the sharing of personal experiences and opinions about Scratch projects. The theme of Building a Following was applied to comments that seek to establish one’s own community of followers by expressing acknowledgement, gratitude, awareness and regret as well as making announcements and seeking support from other users. Comments coded for Personalized Tutoring were aimed toward facilitating the development of the capacity of community members through the exchange of explanations,

Table 3

Codebook for Qualitative Coding

Theme	Subcategory	Description	Example
Supportive Feedback	Encouragement	Sharing compliments and expressions of enthusiasm, motivation, etc.	*"Cool poem! Great animation this time too!"*
	Sharing personal experience	Communicating one's own experiences and opinions about Scratch projects	*"By holding space, I was able to see how the cone is drawn: with triangles. With the orbiting technology from my E Ject Cannon, I could make one of these"*
Building a Following	Acknowledgement	Expressing recognition and appreciation about another's comment	*"Wow, so many tips and so much to read! But anyway thanks I will try to do what you said, I think it will help"*
	Announcement	Giving information about a project by providing a link	*"This battle test has been epically updated!! http://scratch.mit.edu/projects/epninja/2325626!!"*
	Apology	Expressing regret for one's inability to fix errors, play a game, etc.	*"Sorry about that. I fixed it so it is possible to beat."*
	Disclaimer	Disclosing one's awareness about mistakes or poor quality of a project	*"Thanks :D and im having a little problems with Titan,its super slow offline and when i test it online its 10x times faster so its hard to program."*
	Expressing gratitude	Sharing thank you messages	*"Thanks you for playing it and like it!"*
	Seeking support	Making a request for others to visit a project and give reactions/feedback	*"Thx, by the way, check out the realtime square thing and please love it... Tell some other people about it too..".*
Personalized Tutoring	Explanation	Describing the project, creator, or project development process	*"This is to calculate the resistor value that is supposed to be used in a LED-Resistor circuit."*
	Gaming tip	Sharing tips on how to play a game on Scratch	*"You can't make the jump: you need to be smart! turn the "on" button of the light THEN try jumping up!!!"*
	Resource sharing	Providing tools and resources on for making projects	*"were do you get your music from i usually get mine from xtreamfilez.com"*
	Suggestion	Feedback on how to improve a project	*"lol i just spawn camped might i suggest that you make zombies come from all directions"*
	Scratching tip	Providing general tips about community practices	*"A curator is a person who chooses projects to be on the front page. :) You can become one by applying on the forums"*
Agency in Learning	Seeking collaboration	Asking or suggesting to work together on a project or task	*"Hey...this looks like a cool game...wanna collab? :D I dont really have anything to work on in scratch right now, and ive never done a collab…"*
	Seeking explanations	Posing questions to obtain clarification or more information	*"How do you do the multicolor variables??"*
	Seeking help	Requesting assistance, advice or resources	*"could you do some voice overs for me?"*
Cultural Competence	Ethics	Calling attention to ethical issues, such as cheating, copyright, etc.	*"hey! that project shouldent be famous! you copied it from google! you should at least give credit to whoever made this animation"*
	Seeking permission	Requesting permission to use some part of a project	*"Can i use the music for my new rpg game im working on?"*
Conflict	Criticism	Communicating disapproval or pointing out areas needing improvement without suggestions or encouragement	*"This is really glitchy. You can move before the race starts, and just go back and forth to win. And the CPUs don't turn the right way when they move"*
	Disagreement	Expressing disagreement with another user's comment	*"i didnt download your bobby game , why would i want to know how your game works? my games are more successful than yours"*
Role Playing	—	Participating in a collaborative story-telling activity	*"Star: there is some power left in me that i can break Valerie's spell that is on her. Kels.... get Charlie if he is feeling better... understand? Kelsey: understood! *kelsey starts to run*"*

suggestions, resources and tips. Agency in Learning was used to code for comments that seek out help and collaboration from others for the purpose of creating projects. Cultural Competence included comments that reflect the user's understanding of the informal rules of participation in the Scratch community, particularly in navigating ethical issues and demonstrating respect for the intellectual property of others. The theme of Conflict pertained to comments representing criticisms and other negative responses. Lastly, the theme of Role Playing was used for comments posted by users participating in a collaborative story-telling activity.

The coding process involved the qualitative coding of each comment in the sample dataset for the presence of the six themes. Binary coding was applied, meaning that the presence of a code within a comment was assigned a "1" and its absence was labeled with a "0." In this process, the codes were not considered to be mutually exclusive. Hence, if two or more themes were determined qualitatively to be present in a single comment, all corresponding codes received a positive annotation.

The interrater reliability (IRR) of the coding process was calculated using two statistical measures: Cohen's kappa and Shaffer's rho. The kappa statistic was used to determine the level of agreement between two raters, taking into account chance agreement (Cohen, 1960). Shaffer's rho, which estimates the likelihood of Type I error for a given IRR measure, was used to statically test for its generalizability (Eagan et al., 2017). Thresholds of $\kappa > 0.65$ and $\rho < 0.05$ were applied to the kappa and rho values, respectively.

About 5% of the comments in the sample dataset was initially coded independently by two raters. Once the thresholds have been met for the kappa and rho measures for each code, the remainder of the comments was annotated by one rater.

Epistemic Network Analysis

Epistemic network analysis (ENA) was used to model and compare discourse patterns exhibited by the interactive comments across groups and over time. A technique in quantitative ethnography, ENA utilizes visualization and statistical methods to identify meaningful patterns in discourse. ENA is a methodology grounded in epistemic frames theory, which posits that "learning can be characterized by the structure of connections that students make among elements of authentic practice" (Shaffer & Ruis, 2017, p. 182). ENA operationalizes this theoretical approach by generating networks of weighted connections among salient constructs in a recent temporal context (Shaffer, 2017; Siebert-Evenstone et al., 2017). In this process, ENA utilizes a moving window to model the co-occurrence of codes between a given line of data and a specified number of previous lines within the same conversation (Ruis et al., 2019). The ENA algorithm considers each code pair as a dimension that describes the connection between the two codes within a moving window. For example, having 4 codes (A, B, C, D) in the analysis will result in 6 dimensions, or code pairs (AB, AC, AD, BC, BD, CD).

For each line of data, the ENA algorithm creates a vector of n dimensions, with a binary code indicating whether one or more connections are present in each code pair within the defined temporal context. These high dimensional vectors, also referred to as adjacency vectors, are accumulated for each unit of analysis and are then subjected to spherical normalization to account for any differences in the frequency of data points making up each unit of analysis. The ENA algorithm applies dimensionality reduction to the normalize vector through singular value decomposition (SVD) to identify the two orthogonal axes that are able to explain the greatest amount of variance in the data.

The resulting ENA models are visualized through two interrelated graphs (Shaffer et al., 2016). First is in the form of plotted points that represent the location of each unit of analysis in the two-dimensional projected space. Second is the weighted network graph, in which the nodes represent the codes included in the analysis and the edges depict the relative frequency of co-occurrence between the codes. The fixed positions of the nodes in the network graphs are determined through an optimization algorithm that minimizes the distance between the plotted points and the network centroids. This co-registration of the two graphs enables the interpretation of the positions of the network nodes in the ENA space. Furthermore, statistical comparisons can be carried out on the locations of the group means of plotted points, thereby providing a means to quantitatively test whether significant differences exist between selected networks.

In this study, each comment was defined as the basic unit of analysis. The comments were aggregated into groups that will allow comparisons of the discourse patterns associated with them. In order to identify the appropriate moving window size for the ENA models, approaches developed by Ruis et al. (2019) were applied, which included both qualitative and quantitative techniques. Subsequently, comparative analyses of the resulting ENA models were carried out to investigate the differences among various groups of users and to assess changes in the networks over time.

Generalized Linear Mixed Models

Analyses involving two different linear regression models were utilized to examine the question of whether and to what extent certain aspects of social interaction and collaboration are temporally related to the creation of content by users on the Scratch platform. First, a generalized linear mixed model (GLMM) was used to assess the contemporaneous effects of the predictor variables (related to user interaction) on the outcome variables (related to content creation). For this analysis, time series data was generated for each individual by aggregating the variables at one-day intervals, resulting in 91 data points for each of the 200 individuals. The GLMM was used to accommodate this hierarchical grouping of the variables in exploring the relationships that may exist between various aspects of commenting and content creation taking place on the same day.

Given that the outcome variable represented a count of the projects created, Poisson and negative binomial GLMMs were considered. However, the negative binomial model was selected due to overdispersion (i.e. when the variance is greater than the mean) in the data. For the analysis, the outcome variable was defined as content creation while the predictor variables represented the various elements of social practice based on indicators generated during the earlier parts of the analysis. A total of 9 predictor variables were used to develop the regression models for the three categories of interactions: comment activity (1), functional focus (6), and network location (2). Comment activity referred to the number of comments posted or received, while functional focus was the number of code occurrences resulting from the qualitative coding. Network location was represented by the x- and y-coordinates of the project point obtained from the ENA models. Table 4 provides a list of the predictor variables.

Table 4

Predictor Variables Used in the GLMM and VAR Analysis

Category	Predictor Variables	Unit	Origin
Comment Activity	Comment frequency	Number of comments	Descriptive analysis
Functional Focus	Supportive Feedback (A), Building a Following (B), Personalized Tutoring (C), Agency in Learning (D), Cultural Competence (E), Conflict (F)	Number of code occurrences	Qualitative coding
Network Location	X and Y Coordinates of the projected point in the ENA space	SVD1 and SVD2 values for each unit	Epistemic network analysis

Vector Autoregression Models

The time-lagged effects of the multivariate time series data were analyzed through vector autoregression (VAR) models (Zivot & Wang, 2006). This analysis sought to address the gap in research in understanding the social influences among individuals in online communities, which generally requires longitudinal approaches (Malinen, 2015). Upon developing the VAR models, Granger causality tests were carried out to assess the influence of the time lagged measure of the predictor variable on the current measure of the outcome variable (Granger, 1969; Hu et al., 2012). However, it should be noted that Granger causality does not impute causality in the traditional sense, as the process does not account for the possible presence of latent confounding factors.

A general bivariate VAR model of lag order 1 takes the form:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} c_y \\ c_x \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix} \qquad (1)$$

where *c* is the intercept term, *A* is the regression coefficient, e is the error term, and *x* and *y* are time series variables. Value at the current time point is indicated by the *t* in the subscript, whereas *t-1* represents the value at the previous time point. The model in scalar notation breaks down to:

$$y_t = c_y + A_{11}y_{t-1} + A_{12}x_{t-1} + e_{yt} \qquad (2)$$
$$x_t = c_x + A_{21}y_{t-1} + A_{22}x_{t-1} + e_{xt} \qquad (3)$$

As can be seen in the equations above, the VAR takes into account the lagged values of the predictor variable as well as the autoregressive effect of the outcome variable. VAR also models for the potential of the occurrence of bidirectional causality between the two variables. The appropriate lag length to be applied in each VAR model was determined through the use of the Akaike Information Criterion (AIC; Bose et al., 2017).

Summary

This chapter provided a detailed discussion on the research design and methodology adopted for this study. The sample dataset consisted of comments and projects generated by 200 randomly selected users during a 3-month period from January to March 2012. The study utilized quantitative, qualitative and network approaches in an effort to present a comprehensive examination of the relationship between interaction, collaboration and content creation in informal online learning environments. Epistemic network analysis was used to examine the discourse patterns produced by the interactive comments, while negative binomial generalized linear mixed models and vector autoregression were applied to explore the temporal dynamics present between content creation and different elements of social interaction.

Chapter 4: Research Findings

Introduction

The purpose of this study was to explore how social interaction and collaboration are related to the creation of content by users in the Scratch online community. In particular, the research examined the online activity of 200 Scratch users over a 3-month period in 2012. Analysis was focused on the interaction of participants through the exchange of comments as well as on the creation and sharing of new projects on the platform. The study was carried out in several stages, involving qualitative, quantitative and network approaches to provide an in-depth perspective at the relationship between interaction, collaboration and content creation in informal online learning environments. Specifically, the study aimed to address the following research questions:

- **RQ 1:** What patterns of discourse are observed in the comments of Scratch users? Are there differences in the discourse patterns among users with varying modes of participation?
- **RQ 2:** What relationships, if any, exist between the different elements of interaction and content creation?

Descriptive Analysis

Descriptive analyses were conducted at the user, project and comment levels to examine the underlying characteristics of the sample dataset.

User Level

Overall, a total of 37,904 Scratch users created at least one project during the study period from January 1 to March 31, 2012. As shown in Table 5, the total number of projects created by these users was 162,320, resulting in an average of 4.28 projects per user (*SD* =

11.88). Nearly two thirds of the users had created their Scratch accounts during the first three months of 2012, indicating that the majority of project creators were new account holders. The proportion of user accounts created within 3 months prior to 2012 was about 14%, while those established earlier was approximately 21%.

The study focused on users who had created at least one project in each month from January to March 2012 in order to examine the online behavior of users who had sustained their content creation activities during the period of study. Applying this criterion resulted in 2,529 users, who created a total of 52,712 projects during the three-month period. The much higher average of 20.84 projects per user (*SD* = 33.13) indicates that this subgroup of Scratch users was more actively engaged in content creation. The length of membership is also reflective of a sustained participation in the Scratch platform, with about half of users in this group having established their account more than 3 months prior to 2012. Users having created their account within 3 months of January 1, 2012 consisted of about 26% of the group. The remaining 24% consisted of users who had joined Scratch in 2012, meaning that their account was opened in January (in order to meet the criterion of creating at least one project in each month).

For this study, a random sample of 200 users was drawn from the subgroup of 2,529 users who had created at least one project in each month from January to March 2012. It can be seen that the sample users exhibit characteristics that are similar to the population from which it was drawn, particularly with respect to the length of membership. A total of 3,619 projects were created by the 200 sample users, resulting in an average of 18.09 (*SD* = 22.57) projects per sample user. While slightly lower than the mean of 20.84 for the subpopulation of 2,529 users, no statistically significant difference was found between two mean values.

Table 5

User-level Statistics

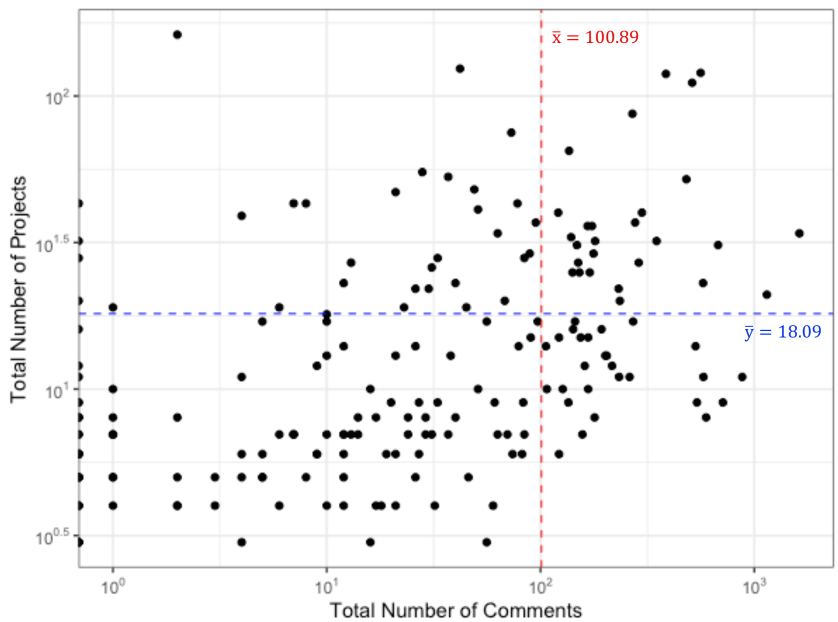
User Subset	Number of Users	Projects Created: Total	Projects Created: Per user	Length of Membership*: More than 3 months	Length of Membership*: Less than 3 months	Length of Membership*: New accounts
At least 1 project created in Jan-Mar 2012	37,904	162,320	4.28	8,099 *(21.37%)*	5,223 *(13.78%)*	24,582 *(64.85%)*
At least 1 project created in each month in Jan-Mar 2012	2,529	52,712	20.84	1,265 *(50.02%)*	650 *(25.70%)*	614 *(24.28%)*
— Sample Users	200	3,619	18.09	104 *(52.00%)*	47 *(23.50%)*	49 *(24.50%)*

*As of January 1, 2012. New accounts refer to those created between January and March 2012.

Figure 3 presents a logarithmic scale plot of the total number of comments posted (x-axis) and projects created (y-axis) by sample users during the 3-month period. The average number of comments posted per user is 100.89 (*SD*=198.68), which is indicated by the dashed line in red. The total number of comments for each sample user varies widely, ranging from 0 to 1,625 comments. The mean of 18.09 projects per sample user (*SD* = 22.57) is indicated by the dashed line in blue. Similar to the values for the number of comments, a broad range can be seen in the number of projects created, extending from a low of 3 to a high of 162 projects. user. Also visible on the left side of the graph are the points representing 34 sample users who did not post any comments during the period of study.

Figure 3

Log Scale Plot of Comments Posted and Project Created by Sample Users During Jan-Mar 2012



Project Level

Projects on the Scratch platform are classified as new or remixed, depending on whether or not the it was based on a previously shared project. Of the 3,619 projects created by sample users during the study period, 2,306 projects (63.72%) were new while 1,313 projects (36.28%) were remixed from other projects. Approximately one third of the remixed projects were self-remixes, meaning that they were remixed based on one of the creator's previous projects. The monthly totals indicate that the proportion of new to remixed projects created by the sample users remained at similar levels throughout the study period (see Table 6).

The number of views, downloads and love-it's can provide an indication of how well a project was received by the Scratch community. Table 7 presents the descriptive statistics for each measure, as displayed publicly on the project page. Overall, each project created by a sample user during the study period had about 28.18 views, 1.27 downloads, 0.73 love-it's, and 4.31 comments. While the number of views, downloads and love-it's show slight differences between new and remixed projects, none were found to be statistically significant at $\alpha=0.05$ level using a Welch's two-sample t-test. The number of comments, however, was statistically significantly higher for new projects, which had a mean of 5.06 comments per project compared to a mean of 2.99 comments for remixed projects.

Table 6

Monthly Totals of Projects Created During the Study Period by Type (New/Remixed)

	Projects		
	New	**Remixed**	**Total**
All	2,306 *(63.72%)*	1,313 *(36.28%)*	3,619
— Jan 2012	807 *(68.33%)*	374 *(31.67%)*	1,181
— Feb 2012	793 *(60.72%)*	513 *(39.28%)*	1,306
— Mar 2012	706 *(62.37%)*	426 *(37.63%)*	1,132

Table 7

Community Reception of Projects Created by Sample Users

							Two-sample t-test (Welch)		
		Mean	**Std. Dev.**	**Median**	**Min**	**Max**	**t**	**df**	**p-value**
Views	All	28.18	110.78	11	0	3683			
	— New	30.85	108.15	12	0	3268			
	— Remix	23.48	115.15	11	0	3683	1.8912	2589.2	0.0587
Downloads	All	1.27	10.90	0	0	475			
	— New	1.25	8.98	0	0	369			
	— Remix	1.31	13.63	0	0	475	-0.1396	1972.2	0.889
Love-it's	All	0.73	4.12	0	0	130			
	— New	0.81	4.12	0	0	123			
	— Remix	0.60	4.12	0	0	130	1.4366	2725.5	0.151
Comments	All	4.31	17.00	1	0	356			
	— New	5.06	18.67	1	0	356			
	— Remix	2.99	12.49	0	0	335	3.8451	3419.4	**<0.001*** **

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Comment Level

In order to examine all interactions involving the sample users from January to March 2012, this study analyzed all comments in conversations containing at least one comment that was sent or received by a sample user during this period. This meant that the full dataset included comments associated with not only projects created by sample users during the study period but also projects created by other users and by sample users before 2012. The sampling process described in Chapter 3 resulted in a dataset comprised of 45,617 comments contained within 18,354 conversations in 9,028 projects.

A cross-tabulation of the 45,617 comments by the project creator type and project creation period is presented in Table 8. Comments associated with projects created by sample users and other users account for approximately 46% and 54% of the dataset, respectively. The majority of comments in the dataset involve projects created during the study period, consisting of over 94% of the comments. Only about 6% of the comments are related to projects created before 2012. From this tabulation, it can be seen that most of the comments by the sample users revolved around current projects, regardless of whether a project was their own or created by another user.

Table 8

Cross-tabulation of Comments by Project Creator Type and Project Creation Period

	Number of comments related to…		
Project creation period	**Projects created by sample users**	**Projects created by other users**	**Total**
Before 2012	1,170 *(2.56%)*	1,490 *(3.27%)*	2,660 *(5.83%)*
Jan-Mar 2012	19,758 *(43.31%)*	23,199 *(50.86%)*	42,957 *(94.17%)*
Total	20,928 *(45.87%)*	24,689 *(54.13%)*	45,617 *(100%)*

The comments were also identified according to the three categories defined in Chapter 3: sent, received and contextual comments. The categorizations were based on the five types of comments determined by the user who posted the comment ("sender") and to whom the comment was directed ("receiver"). Examples of the five comment types include sample self-directed and other to sample. The frequency of occurrence for each comment type in the dataset is presented in Table 9. Comments exchanged between sample users and other users make up nearly 79% of all comments, with sample user to other users accounting for about 40% and comments in the opposite direction comprising approximately 39%. Comments exchanged among other users represented about 17% of the dataset, while comments posted by sample users to other sample users and self-directed comments by sample users only accounted for about 2% each.

Aggregating the relevant comment types for each category resulted the placement of 20,178 comments (44.23%) into the "sent comments" category and 18,670 comments (40.93%) into the "received comments" category. The remaining 7,752 comments (16.99%) were identified as contextual comments, representative of the "other to other" comment type. It should be noted that the percentages associated with the categories represent the proportion of the comments in each category in relation to the full dataset.

Table 9

Frequency of Comments by Type and Category

Comment Type	Frequency	Percentage
1. Sample self-directed	932	2.04 %
2. Sample to other	18,263	40.04 %
3. Sample to sample	983	2.15 %
4. Other to sample	17,687	38.77 %
5. Other to other	7,752	16.99 %
Total	**45,617**	**100 %**

Comment Categories

Sent comments: 20,178 (44.23%)

Received comments: 18,670 (40.93%)

Contextual comments: 7,752 (16.99%)

Qualitative Coding

In order to identify the functional themes present in the user-generated comments, the 45,617 comments in the dataset were annotated using a coding scheme containing seven themes, which was adapted from Fields et al. (2015). A total of 2,471 comments, comprising about 5.4% of all comments, was initially coded by two raters to ensure construct validity and interrater reliability. After reaching acceptable levels of agreement between the two raters for all codes, the remaining comments were annotated by one rater.

Interrater Reliability

The interrater reliability (IRR) process consisted of two stages of coding carried out by two raters with backgrounds in science, technology, engineering and math (STEM) education research. The first stage involved the coding of 1,298 comments within conversations which were related to 114 randomly selected projects. The comments were coded individually by both raters, followed by a process of social moderation undertaken for each comment to reach agreement on the final coding of the data (Frederiksen et al., 1998; Herrenkohl & Cornelius, 2013). No IRR measures were calculated, as this initial step was intended to allow the raters to become familiar with the data and to refine their understanding of the codes. It also provided an opportunity to identify issues concerning the application of the coding scheme. One such issue was related to the phenomenon of role-playing taking place within the comments, which led to the inclusion of the “Role Playing” code.

The second stage consisted of four rounds of double-coding, each followed by the calculation of IRR statistics and a process of social moderation to discuss and resolve any discrepancies in the coding. Rounds 1 and 2 involved the annotation of 504 and 250 comments, respectively. Similar to the first stage, all comments associated with randomly selected projects were included and organized within conversations. The results of the IRR calculations for the first two rounds are presented in Table 10. The baserates given in the table reflect the frequency of occurrence in the final coding of each set, after social moderation. At the end of Round 2, sufficient levels of Cohen's kappa and Shaffer's rho values were obtained only for three codes, namely Supportive Feedback, Building a Following, and Role Playing. It was also observed that the low baserates of the remaining four codes—ranging from 0.02 to 0.10 in Round 2—could be limiting the generalizability of the kappa values obtained from the test sets.

Table 10

IRR Results for Stage 2, Rounds 1 and 2 (No Baserate Inflation)

	Round 1				Round 2			
Codes	Length	BR	Kappa	Rho	Length	BR	Kappa	Rho
Supportive Feedback	504	0.13	**0.82***	**< 0****	250	0.16	**0.82***	**< 0****
Building a Following	504	0.17	0.64	0.08	250	0.22	**0.72***	**0.02****
Personalized Tutoring	504	0.11	0.61	0.20	250	0.07	0.32	1
Agency in Learning	504	0.09	0.61	0.22	250	0.10	0.58	0.34
Cultural Competence	504	0.02	0.63	0.23	250	0.02	0.56	0.41
Conflict	504	0.09	0.52	0.53	250	0.02	0.26	1
Role Playing	504	0.10	**0.83***	**< 0****	250	0.07	**0.94***	**< 0****

Note. BR: Baserate
***** kappa > 0.65, ****** rho < 0.05

For this reason, baserate-inflated test sets were utilized in subsequent rounds. For the four remaining codes, test sets with a baserate of about 0.2 were created based on the annotation by the first rater. Just as in earlier rounds, the random selection process was applied at the project level in order to maintain the contextualized structure of the comments. For Round 3, three baserate-inflated sets were created from a coded set of 1,180 comments and passed onto the second rater for coding. As a result, sufficient levels of kappa and rho were achieved for three codes: Personalized Tutoring, Agency in Learning, and Conflict. The same procedure was repeated in Round 4—based on the annotation of 852 comments by the first rater—to obtain adequate levels of interrater reliability and generalizability for the Cultural Competence code (see Table 11).

Table 11

IRR Results for Stage 2, Rounds 3 and 4 (With Baserate Inflation)

	Round 3				Round 4			
	Length	BR	Kappa	Rho	Length	BR	Kappa	Rho
Personalized Tutoring	107	0.24	**0.84***	**< 0****				
Agency in Learning	107	0.21	**0.89***	**< 0****				
Cultural Competence	107	0.15	0.59	0.38	104	0.22	**0.76***	**0.03****
Conflict	101	0.21	**0.74***	**0.04****				

Note. BR: Baserate
***** kappa > 0.65, ****** rho < 0.05

Coding Results

The results of the coding process are provided in Table 12. Overall, the highest frequency was observed for Supportive Feedback, which was present in about 17% of all comments. This was followed by Building a Following and Role Playing, which occurred in around 16% and 11% of the messages, respectively. Agency in Learning and Personalized Tutoring were each present in approximately 10% of the comments, while Conflict was accounted for in about 5%. The lowest frequency was associated with Cultural Competence, appearing only in about 1.5% of the comments.

Disaggregating the comments by category, similar proportions of occurrences were found across all codes for the sent and received comments. This makes intuitive sense, given that whether a comment is classified as being sent or received is only dependent on the perspective from which the distinction is made. Therefore, there would not be any reason to expect a numerical difference between the comments sent and received by the sample users. The contextual comments, however, exhibit a slightly different pattern, in which relatively lower frequencies were observed for Supportive Feedback (9.31%) and Building a Following (8.58%)

while higher proportions were found for Conflict (10%) and Role Playing (13.56%). This may be partly attributed to the way the dataset was constructed, whereby only comments within conversations containing at least one comment sent or received by a sample user was selected. This means that relatively fewer contextual comments—messages sent and received between non-sample users—were included in the data overall, especially in shorter-length conversations where Supportive Feedback and Building a Following were more prevalent.

Table 12

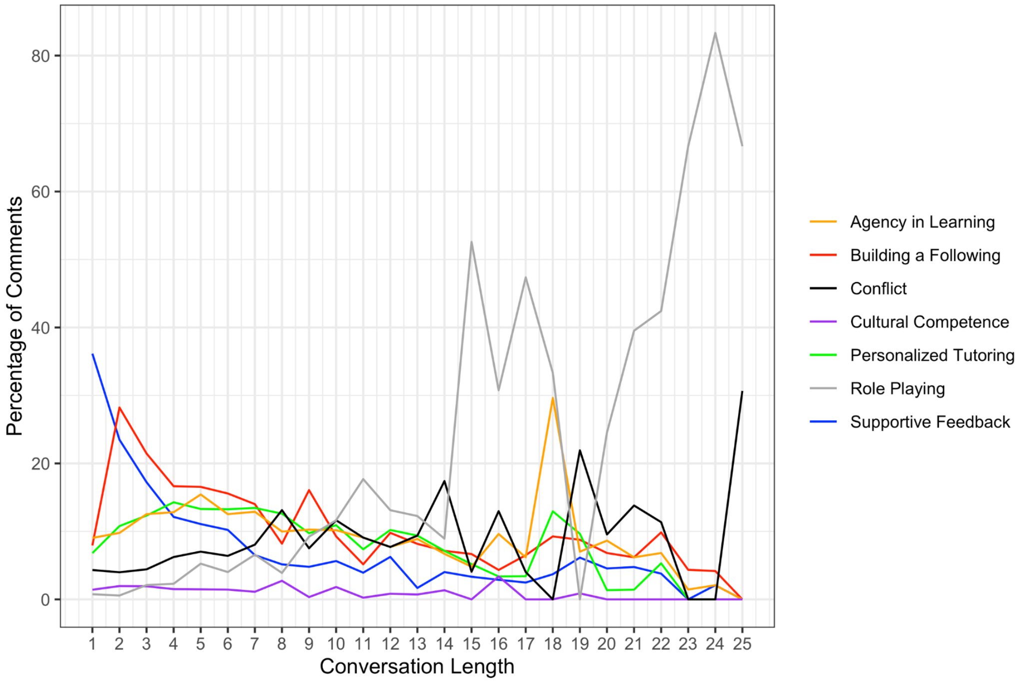
Code Frequencies of All Comments by Comment Type

	All Comments (45,617)		**Sent Comments** (20,178)		**Received Comments** (18,670)		**Contextual Comments** (7,752)	
Code	**Freq**	*%*	**Freq**	*%*	**Freq**	*%*	**Freq**	*%*
Supportive Feedback	7,947	*17.42*	3,910	*19.38*	3,365	*18.02*	722	*9.31*
Building a Following	7,143	*15.66*	3,437	*17.03*	3,286	*17.60*	665	*8.58*
Personalized Tutoring	4,483	*9.83*	2,165	*10.73*	1,675	*8.97*	775	*10.00*
Agency in Learning	4,608	*10.10*	2,121	*10.51*	1,858	*9.95*	691	*8.91*
Cultural Competence	676	*1.48*	343	*1.70*	241	*1.29*	108	*1.39*
Conflict	2,475	*5.43*	884	*4.38*	867	*4.64*	775	*10.00*
Role Playing	5,217	*11.44*	2,237	*11.09*	1,986	*10.64*	1,051	*13.56*

Beyond a discussion around contextual comments, patterns of code occurrences by conversation length provide key insights about the interaction of users in the Scratch community. Figure 4 presents the proportion of comments annotated for each code in conversations containing up to 25 comments. It can be seen that at the lower conversation lengths, Supportive Feedback and Building a Following account for a substantial percentage of comments. At conversation lengths above 14, Role Playing is shown to have large spikes, some of which make up well over 60 percent of all comments found at those conversation lengths. Several spikes of lower magnitude can be observed for Conflict, starting at length 14.

Figure 4

Proportion of Codes Occurring in Comments within Conversations of Varying Lengths



Epistemic Network Analysis

Using the *rENA* package in R (Marquart et al., 2019), ENA models were developed based on the coded comments. However, the comments annotated with the Role Playing code were excluded from the dataset used in ENA as well as in subsequent analyses of this study. While interesting in their own right, it was determined that the Role Playing comments constituted a mode of exchange among participants that was beyond the scope of this study. As a result, the remaining analyses in the study were focused on the remaining 40,400 comments.

Moving Window

ENA utilizes a moving window to capture the connections between elements of discourse co-occurring in a recent temporal context (Siebert-Evenstone et al., 2017). Prior to the organization of the comments into threaded conversations, an infinite moving window was considered for the ENA models in this study. The logic behind this determination was that the infinite window would take into account the asynchronous, text-based nature of interactions in the Scratch online community, where the all prior messages in a threaded conversation are readily visible to the user writing the comment. However, following the construction of the final dataset, it was noted that a number of conversations contained dozen of comments, some exceeding 100 comments. While it would not be impossible for a user to read and consider the previous 100 messages in crafting their response, it would also be highly unlikely. For this reason, it was determined that a moving window of a specified length would be needed to limit the connections between codes to those occurring within a certain number of comments in the conversation.

To this end, approaches outlined in Ruis et al. (2019) were applied to determine an appropriate window length to utilize in the ENA models. The methods were undertaken in two steps, involving both qualitative and quantitative analytical processes. First, a subsample of 400 randomly selected comments were annotated by two independent raters for all previous referents, or comments in the conversation that provide contextual information. For each comment, a window length was determined to be the number of lines from the comment to its furthest referent, inclusive. At each window length from one to nine, the agreement between the two raters was calculated using Cohen's kappa. The Shaffer's rho was computed to assess the generalizability of the kappa statistic. From the results presented in Table 13, it can be seen that a window length of five would capture 94% of the relevant connections made by comments in this subsample. This aligns closely with the proportion of the comments in the entire dataset, where slightly above 94% of all comments were contained within conversations of five comments or fewer. In addition, interrater agreement was found to be statistically significant at kappa > 0.65 and rho < 0.05 levels for window lengths up to five.

Table 13

Number and Percentage of Comments at Each Window Length

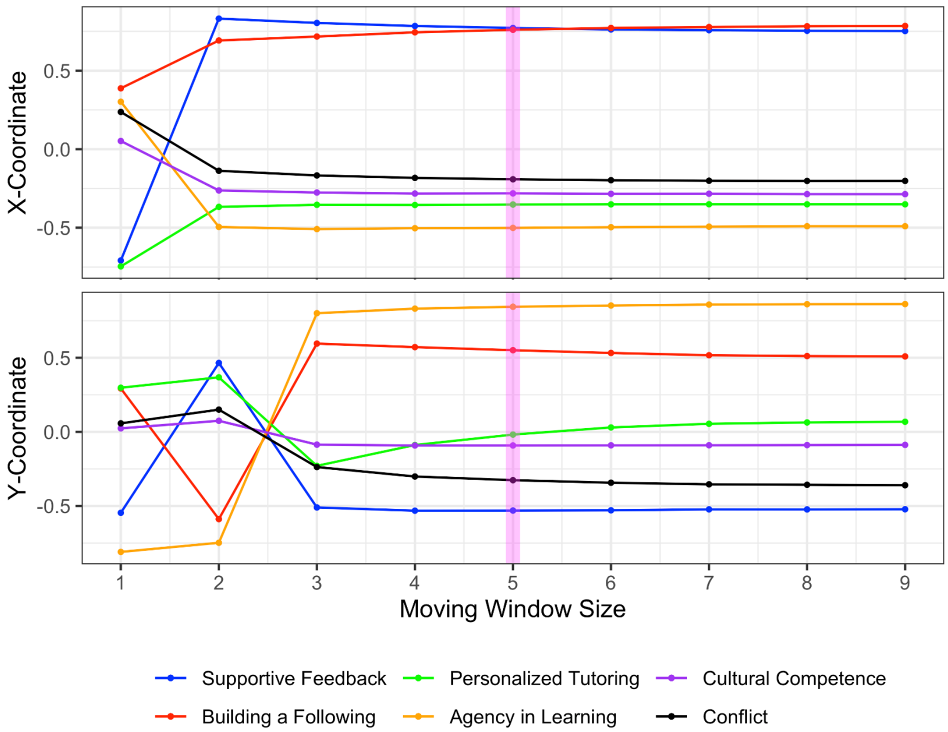
	Comments		Interrater Agreement	
Window Length	Cumulative Number	Cumulative Percentage	Cohen's Kappa	Shaffer's Rho
1	189	47.3%	**0.97***	**< 0.01****
2	296	74.0%	**0.94***	**< 0.01****
3	336	84.0%	**0.91***	**0.02****
4	361	90.3%	**0.94***	**< 0.01****
5	376	94.0%	**0.90***	**0.04****
6	385	96.3%	**0.84***	0.15
7	388	97.0%	**0.81***	0.22
8	393	98.3%	**0.82***	0.22
9	396	99.0%	**0.72***	0.37

***** kappa > 0.65, ****** rho < 0.05

The second step in the analysis consisted of evaluating the effect of varying window lengths on the stability of the ENA models produced. In particular, the analysis focused on the examining the positions of the network nodes for models utilizing different lengths of the moving window. Plots for the x- and y- dimensions for each node location are shown in Figure 5. The x-coordinates of most nodes appear to stabilize at length two. While a slight movement towards convergence can be observed for Supportive Feedback and Building a Following, it becomes negligible by length five. Similarly, the y-coordinates of all nodes except for Personalized Tutoring stay relatively constant after length three. The upward shift exhibited by the node for Personalized Tutoring beginning at length three also becomes more stable by around length five. Based on these results from both stages of analysis, it was determined that a moving window size of five would be optimal for the ENA models developed for this study.

Figure 5

Node positions of the ENA models at Each Length of the Moving Window



ENA Models

ENA models were developed to identify the patterns of interaction between Scratch users, as exhibited in the comments discourse. Each comment was defined as the basic unit of analysis for all models. The unit of conversation was made up of a threaded conversation, consisting of the first-level comment and associated second- and third-level comments, if any. Each conversational unit was grouped by the project with which it was related. For all models, the x- and y-axes of the ENA space was defined respectively by the first and second dimensions obtained through singular value decomposition.

For each analysis, two separate models were developed, one for the comments sent by the sample users and another for the comments that were received by them. As noted earlier, a number of sample users did not send or receive any comments during the study period. These users were not included in the ENA given that the dataset did not contain any comment associated with them. Therefore, the ENA results presented below are based on comments sent and received by 166 and 184 sample users, respectively.

Comments Sent and Received. The ENA models in Figure 6 display the average connections made by all comments that were sent (in red) and received (in blue) by sample users during the study period. The networks display similar connections, with the strongest linkage appearing between Supportive Feedback and Building a Following in both models. Moderate connections are also present between these two themes and Personalized Tutoring as well as Agency in Learning. Conflict and Cultural Competence show the weakest linkages to the other codes, indicating that that they constitute a relatively small component of the overall discourse. Looking at the subtracted model between the two networks in Figure 7 it can be seen that comments sent by sample users in general exhibited a stronger link between Personalized Tutoring and Supportive Feedback, while the comments they received contained greater focus on Building a Following and Supportive Feedback. A close-up of the group means of the sent and received comments projected onto the ENA space (square) along with their respective 95% confidence intervals (surrounding box) is shown in the inset. A two-sample t-test assuming unequal variance showed a statistically significant difference at the alpha = 0.5 level along the x-axis between the locations of the means for comments that were sent (mean = 0.0036, *SD* = 0.2798, *N* = 17,941) and received (mean = 0.0113, *SD* = 0.2885, *N* = 16,684; *t*(34,258.47) = -2.50, *p* = 0.0124, Cohen's *d* = 0.027).

The positions of the network nodes allow for the interpretation of the ENA space due to an optimization process that minimizes the distance between the projected points and network centroids of each unit of analysis (Shaffer et al., 2016). In this model, the x-axis is distinguished by Supportive Feedback and Building a Following on the right and Agency in Learning on the left. Likewise, the y-axis is defined mainly by Agency in Learning at the top and Supportive Feedback toward the bottom of the ENA space. Based on the statistically significant results of the t-test along the x-axis, it can therefore be concluded that the average comment received by sample users was relatively more associated with providing encouragement and building relationships, whereas the average comment sent by sample users was more focused on seeking information and collaboration related to project content.

Figure 6

Network Graphs of Comments Sent and Comments Received by Sample Users

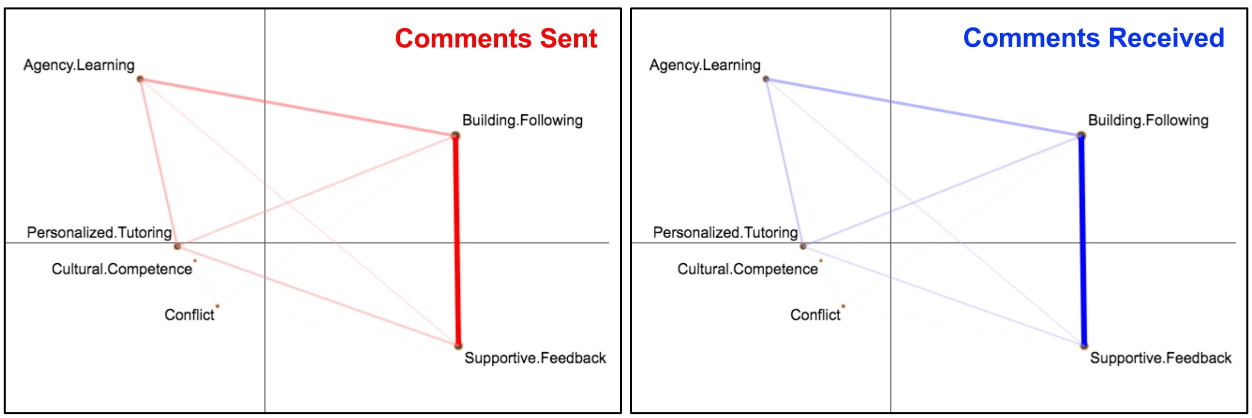
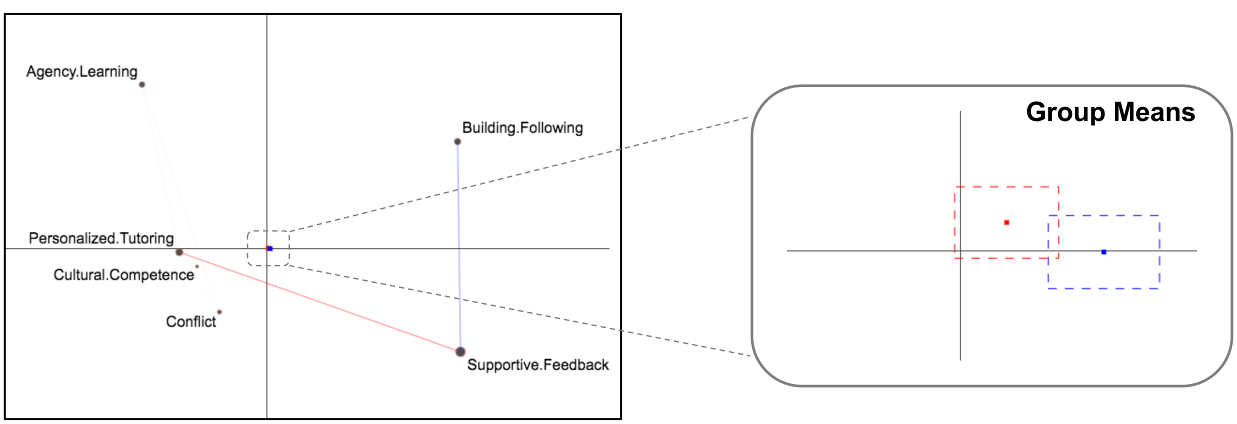


Figure 7

Subtracted Network Graph and Group Means for the Sent (red) and Received (blue) Comments



Level of Comment Posting. The second analysis focused on whether any differences existed in the discourse patterns linked to sample users depending on their level of posting comments. The sample was divided into two groups according to the total number of comments the user posted during the 3-month study period. Fifty-five sample user who had posted more than the sample mean of 100.89 comments were placed in the Higher Comments group, while the remaining sample users were placed in the Lower Comments group.

ENA models were developed for the comments sent and received by sample users in the two groups. For the sent comments (see Figure 8), thicker connections are visible for the Higher Comments Group between Agency in Learning and Building a Following. A statistically significant difference was found along the y-axis using a two-sample t-test assuming unequal variance between the Higher Comments group (mean = 0.0042, SD = 0.2097, N = 15,129) and the Lower Comments group (mean = -0.0073, SD = 0.1880, N = 2812; t(4,218.32) = 2.94, p = 0.003, Cohen's d = 0.058). This suggests that comments sent by the Higher Comments group included more connections to nodes located toward the top of the ENA space.

For the received comments, a more prominent connection can be seen in the Higher Comments Group between Building a Following and Agency in Learning as well as between Agency in Learning and Personalized Tutoring (see Figure 9). A statistically significant difference was observed for the received comments along the x-axis between the Higher Comments group (mean = -0.0126, SD = 0.2564, N = 8,814) and the Lower Comments group (mean = 0.0380, SD = 0.3185, N = 7,870; t(15,094.90) = -11.22, p < 0.001, Cohen's d = 0.175).

Figure 8

Network Graphs of Comments Sent by Higher Comments (dark red) and Lower Comments (dark blue) Groups

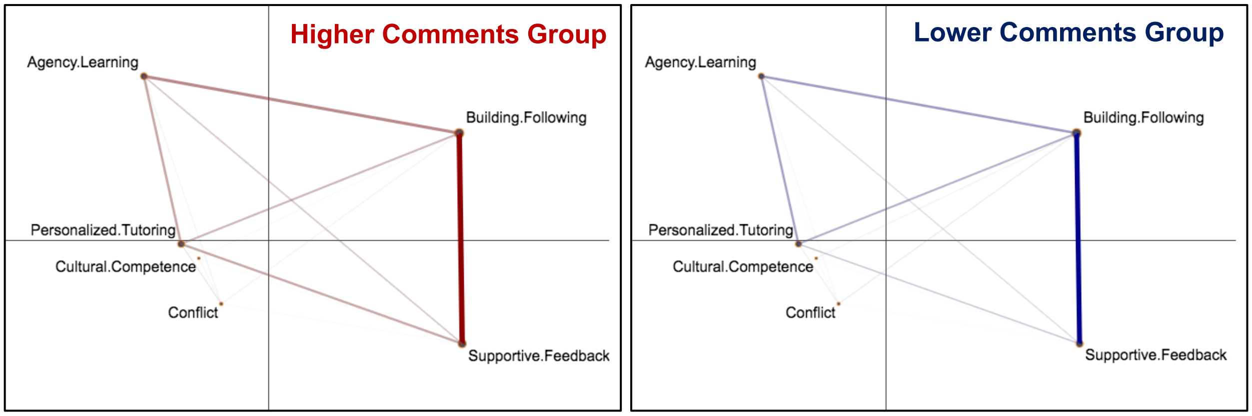
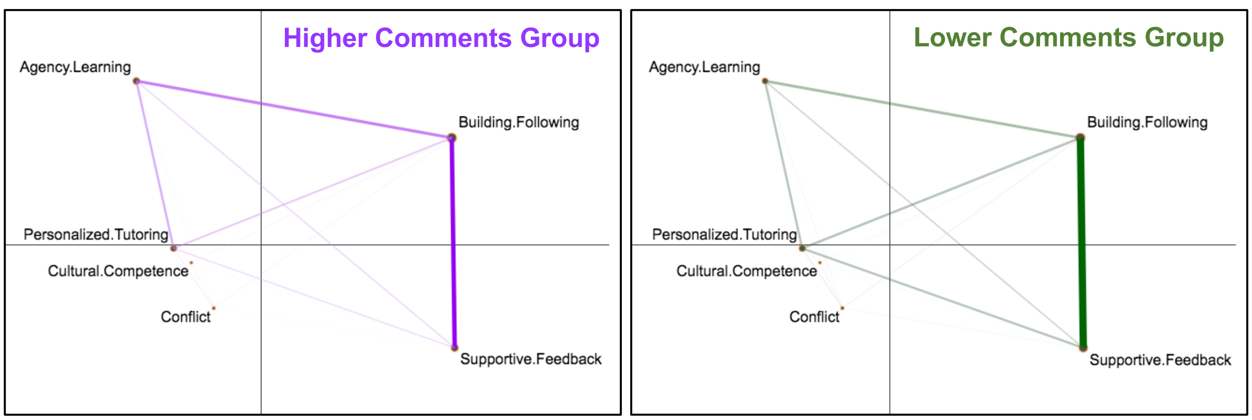


Figure 9

Network Graphs of Comments Received by Higher Comments (purple) and Lower Comments (dark green) Groups



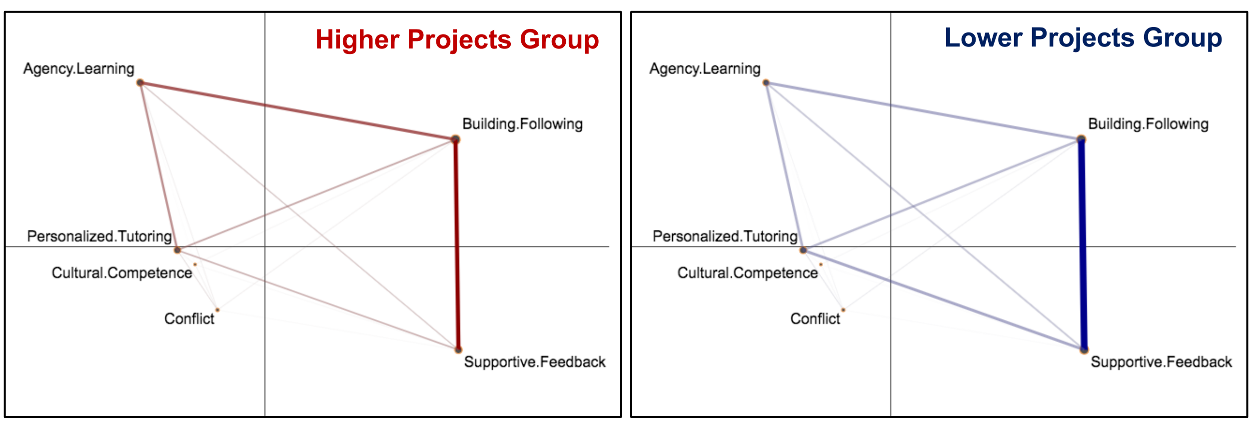
Level of Project Creation. The subsequent analysis revolved around the question of whether a different discourse pattern exists for comments associated with users who create more projects. For this analysis, two groups of sample users were formed based on the total number of projects created during the study period. Using the mean of 18.09 projects as the distinguishing criterion, 59 sample users were put into the Higher Projects group and 141 were placed in the

Lower Projects group. As noted earlier, users not having sent or received any comments during the study period were excluded from the respective datasets.

The networks graphs of comments sent by sample users in the two groups is shown in Figure 10. The strongest link is between Supportive Feedback and Building a Following for both groups. However, for those with higher levels of project creation, a prominent connection can also be observed between Building a Following and Agency in Learning. Due to this dominant association, the mean for the High Projects group is drawn toward to left side of the ENA space. The result is a statistically significant difference along the x-axis between the High Projects group (mean = -0.0173, *SD* = 0.2504, *N* = 9,588) and the Low Projects group (mean = 0.0277, *SD* = 0.3084, *N* = 8,353; $t(16{,}078.76) = -10.64$, $p < 0.001$, Cohen's $d = 0.16$). Similar discourse patterns were produced by the comments received by the two groups, also resulting in a statistically significant difference along the x-axis.

Figure 10

Network Graphs of Comments Sent by Higher Projects and Lower Projects Groups



Discourse Patterns of Comments Over Time

The next analysis was focused on the changes in discourse patterns of comments over time. However, rather than analyzing the change for all sample users, attention was given to those users who had exhibited either an increase or a decrease in project creation throughout the study period. For this analysis, a tally was made of the total number of projects created in each month for each user. Following this, the change between each month was notated. A plus (+) sign was used to indicate an increase in the project totals from one month to the next. Likewise, a minus (–) sign was used for a decrease and a zero (**0**) was applied when the figure stayed the same. Thirty-eight sample users exhibiting an overall increase in the number of projects without experiencing a reduction (notated with **+ +**, **0 +**, or **+ 0**) were placed in the Increasing Project Creation (IPC) group. Fifty-six users with an overall decrease in the number of projects created without showing a rise (notated with **– –**, **0 –**, or **– 0**) were categorized into the Decreasing Project Creation (DPC) group. The remaining 106 users (notated with **+ –**, **– +**, or **0 0**) were set aside due to the lack of a consistent trend in project creation over the 3-month period. Indeed, a calculation of the change in the number of projects created per user from January to March 2012 showed an average of a 7.97 project rise per user for the IPC group, a 6.48 project decline for those in the DPC group, and a 0.1 project increase for the uncategorized group. The final tally of users and associated comments for each group, accounting for sample users who did not send or receive any comments, is shown in Table 14. It is worth noting that the increasing or decreasing trend in project creation can also be seen in the number of comments that were sent and received by users in the respective groups.

Table 14

Number of Users and Comments Included in the Increasing/Decreasing Project Creation Groups

Group	Comment Type	Number of Users	Number of Comments		
			January	February	March
Increasing Project Creation (IPC)	Sent	25	617	688	982
	Received	31	502	492	947
Decreasing Project Creation (DPC)	Sent	51	2,648	1,986	1,865
	Received	52	2,626	1,781	1,621

A total of twelve ENA models were developed to examine the discourse patterns portrayed by the comments that were sent and received each month by sample users in the two groups. The network graphs and the plotting of the monthly group means for the IPC group are presented in Figure 11. For comments sent by the sample users, the January network displays strong and consistent connections between the four main codes of Supportive Feedback, Building a Following, Personalized Tutoring and Agency in Learning. The February and March networks show prominent linkages between Building a Following and Supportive Feedback as well as Agency in Learning. From the monthly group means, a general shift toward the right side of the ENA space can be observed. With no overlap in the 95% confidence interval along the x-axis, there is a statistically significant difference between January (mean = -0.0416, $SD = 0.2067$, $N = 617$) and February (mean = 0.0127, $SD = 0.2886$, $N = 688$; $t(1,243.81) = -3.94$, $p < 0.001$, Cohen's $d = 0.216$). No significant change was found between the group means for February and March.

Inspecting the networks for the comments received by the IPC group, no pronounced changes are visible across the three months. While some subtle shifts can be seen, including the emergence of a connection between Cultural Competence and Building a Following in February, the overall pattern of discourse remains stable. This is also confirmed in the plot of the group means, which are clustered together with substantial overlap in the confidence intervals. No statistically significant differences were found between the monthly means.

The network graphs and group means for the DPC group are provided in Figure 12. For the comments sent by the users in this group, the networks across the three months appear to remain relatively stable over the three months, with a strong association between Supportive Feedback and Building a Following and moderate connections making up the remaining edges between the four main codes. However, the plot of the group means indicates a gradual movement toward the left. A statistically significant difference was found along the x-axis between January (mean = 0.0154, *SD* = 0.2946, *N* = 2,648) and March (mean = -0.0114, *SD* = 0.2619, *N* = 1,865; *t*(4,275.85) = 3.21, *p* = 0.0014, Cohen's *d* = 0.096).

The network graphs for the comments received by the DPC group show a strengthening of the link between Building a Following and Supportive Feedback as well as Agency in Learning over the 3-month period. This is accompanied by a diminishing level of connection to Personalized Tutoring. This trend is confirmed in the plot of the group means, which reveals a clear diagonal shift toward the top right of the ENA space. Along the x-axis, statistically significant differences were found at the alpha = 0.05 level between all pairs of monthly group means. Along the y-axis, statistically meaningful differences were observed between January and February as well as between January and March. Table 15 presents a summary of the statistical comparisons of the monthly group means conducted using an independent two-sample t-test assuming unequal variances at an alpha level of 0.05.

Table 15

Statistical Comparisons of the Group Means for the Comments Received by the DPC Group

					Two Sample T-test (Welch)					
					February			March		
Axis	Month	*N*	Mean	SD	*t*	*p*	*d*	*t*	*p*	*d*
X	January	2,626	0.0081	0.2848	-2.59	0.0097**	0.080	-5.58	< .001***	0.180
	February	1,781	0.032	0.3120				-2.93	0.0034**	0.101
	March	1,621	0.0653	0.3473						
Y	January	2,626	-0.0106	0.1871	-2.18	0.029*	0.068	-3.58	< .001***	0.116
	February	1,781	0.0026	0.2030				-1.50	0.134	0.052
	March	1,621	0.0139	0.2320						

*$p < .05$, **$p < .01$, ***$p < .001$

Figure 11

Network Graphs and Group Means for Sample Users in the Increasing Project Creation Group

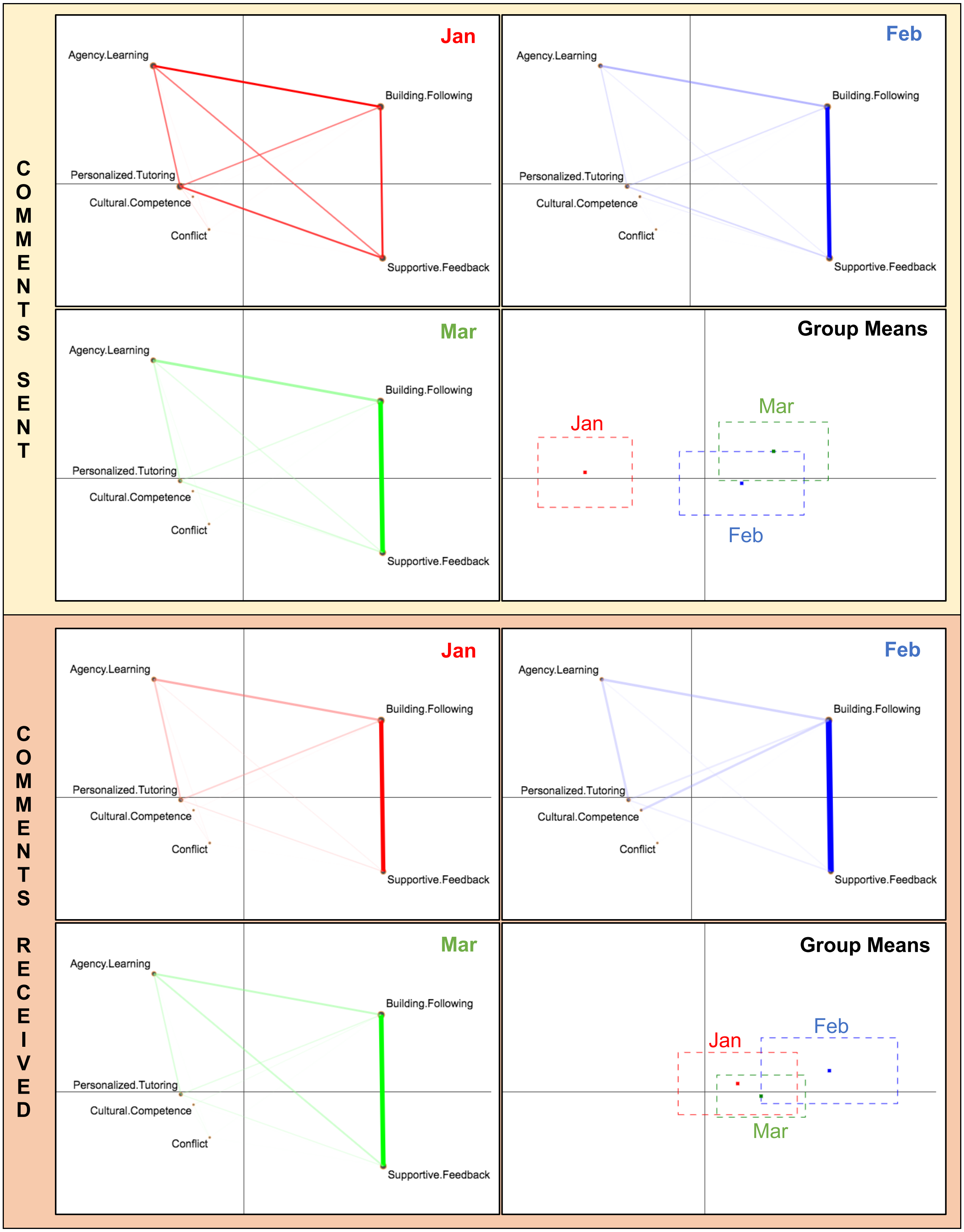
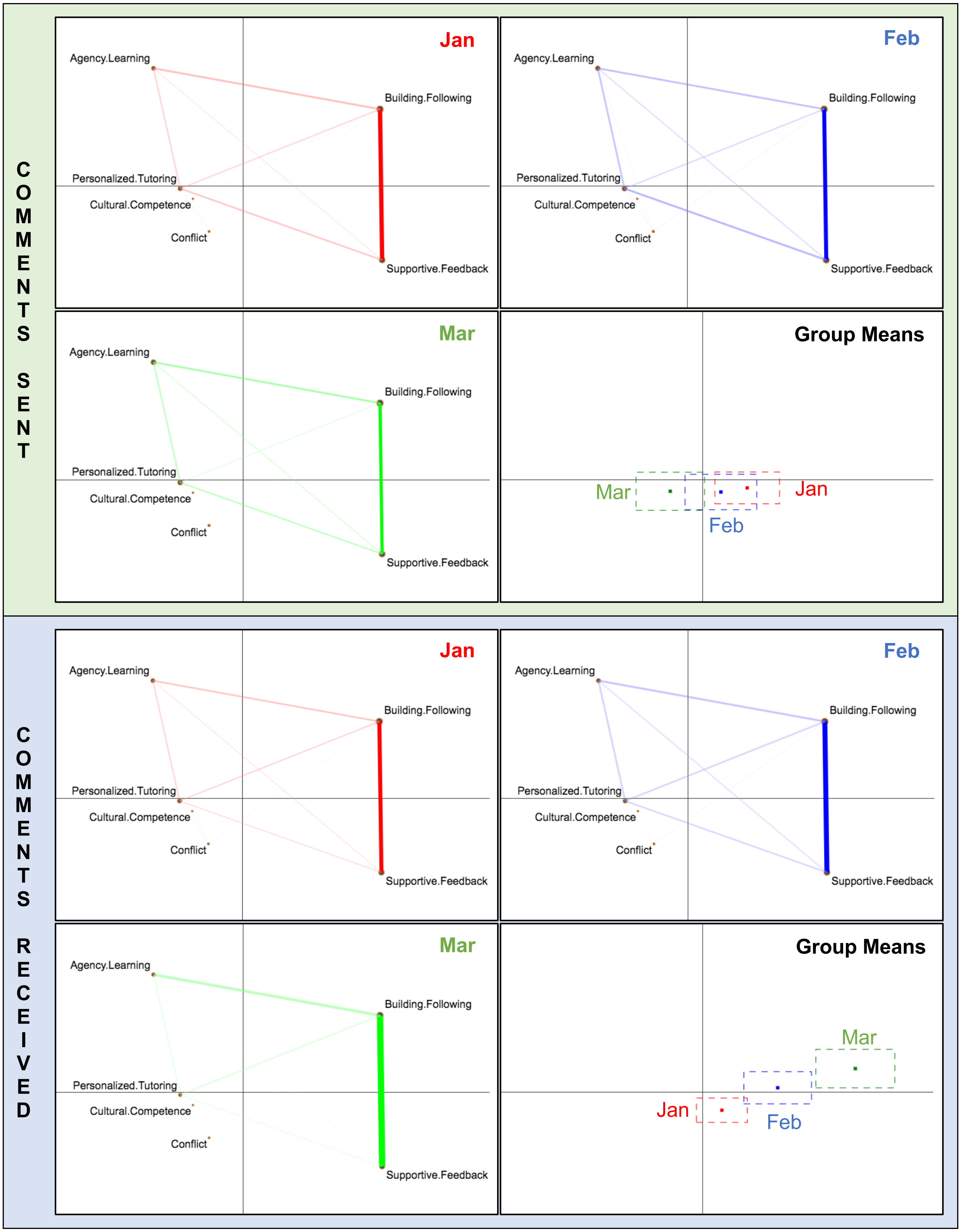


Figure 12

Network Graphs and Group Means for Sample Users in the Decreasing Project Creation Group



Generalized Linear Mixed Models

Generalized linear mixed models (GLMM) were developed to explore the contemporaneous associations between interaction and content creation behavior of users on the Scratch platform. Time series data was constructed at one-day intervals for each sample user. Counts of projects created during each calendar day served as the outcome variable. The predictor variables reflected the exchange of comments by sample users for the three different categories. First was comment activity, which represented the frequency of comments sent and received by the sample users. The frequency of occurrence of the six codes representing the functional focus of the user comments comprised the second level. The third category consisted of the network location of each comment on the ENA space, represented by its respective x- and y-coordinates of the projected point (corresponding to the first and second dimensions obtained from singular value decomposition). The counts of the variables at levels 1 and 2 were aggregated at one-day intervals, while the mean of the non-zero values were used for the network locations. The data for those users in the sample who had not sent or received any comments during the study period were excluded, resulting in 166 and 184 users in the sent and received datasets, respectively.

In the modeling process, both the Poisson and negative binomial models were considered. However, significant overdispersion was observed in the data, which resulted in the selection of the negative binomial model. The models were fit using the *glmer.nb* function from the *lme4* package in R, in which maximum likelihood estimates were obtained using Laplace approximation (Bates et al., 2015). Only random intercepts were modeled. Two models were developed for each category of the predictor variable, one for the comments sent and another for the comments received. The summary of the models for each predictor category generated using the *sjPlot* package in R (Lüdecke, 2020) is presented in Table 16.

For the fixed effects of the model, the estimated regression coefficients are reported in terms of incidence rate ratios, which represent the factor by which the expected outcome count increases for one unit increase in the predictor, given other variables in the model are held constant. For example, in the Category 1 Models, an additional comment sent by the sample user would result in the expected count of projects created to increase by a factor of 1.21. The proportion of variance explained by the random variable, in this case the user, is given by the intra-class correlation coefficient (ICC). The Akaike Information Criterion (AIC) and coefficient of determination (R^2) values are provided in relation to the overall fit of each model. The R^2 and ICC estimates for the GLMM have been calculated based on the methods outlined in Nakagawa et al. (2017). The marginal R^2 accounts only for the variance explained by the fixed effects, whereas the conditional R^2 considers the variance captured by both the fixed and random effects of the model.

The results indicate that nearly all of the interaction-related variables—for both sent and received comments—are statistically significant predictors of the number of projects created by sample users at the alpha = 0.05 level. The only exceptions are the code frequencies for Cultural Competence and the y-coordinates of the ENA projected points in comments sent by the users. In comparing the models, it can be observed that the Category 1 Models generally provide a better fit to the data.

Table 16

Summary of the GLMM Negative Binomial Models for Each Predictor Level

	Category 1 Models (Comment Activity)					
	Sent Comments			**Received Comments**		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	0.10	0.09 – 0.12	**<0.001*** **	0.10	0.09 – 0.12	**<0.001*** **
sent.comments	1.21	1.19 – 1.23	**<0.001*** **			
received.comments				1.23	1.21 – 1.25	**<0.001*** **
Random Effects						
σ^2	2.30			2.34		
τ_{00}	0.64 $_{user}$			0.73 $_{user}$		
ICC	0.22			0.24		
N / Observations	166 $_{user}$ / 15106			184 $_{user}$ / 16744		
AIC	14523.2			15768.6		
Marg. R^2 / Cond. R^2	0.114 / 0.308			0.122 / 0.331		

	Category 2 Models (Functional Focus)					
	Sent Comments			**Received Comments**		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	0.11	0.09 – 0.12	**<0.001*** **	0.10	0.09 – 0.12	**<0.001*** **
Supportive	1.14	1.08 – 1.20	**<0.001*** **	1.43	1.35 – 1.51	**<0.001*** **
Building	1.47	1.39 – 1.55	**<0.001*** **	1.23	1.16 – 1.31	**<0.001*** **
Personalized	1.12	1.04 – 1.22	**0.004** **	1.33	1.21 – 1.47	**<0.001*** **
Agency	1.33	1.23 – 1.43	**<0.001*** **	1.21	1.12 – 1.31	**<0.001*** **
Cultural	1.03	0.87 – 1.22	0.746	1.36	1.03 – 1.79	**0.031***
Conflict	1.20	1.07 – 1.34	**0.001** **	1.32	1.17 – 1.50	**<0.001*** **
Random Effects						
σ^2	2.29			2.33		
τ_{00}	0.67 $_{user}$			0.70 $_{user}$		
ICC	0.23			0.23		
N / Observations	166 $_{user}$ / 15106			184 $_{user}$ / 16744		
AIC	14565.9			15748.8		
Marg. R^2 / Cond. R^2	0.101 / 0.304			0.105 / 0.313		

	Category 3 Models (Network Location)					
	Sent Comments			**Received Comments**		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	0.15	0.13 – 0.17	**<0.001*** **	0.14	0.12 – 0.16	**<0.001*** **
ENA.x	1.47	1.02 – 2.12	**0.037***	0.44	0.30 – 0.63	**<0.001*** **
ENA.y	1.03	0.62 – 1.72	0.901	0.21	0.11 – 0.38	**<0.001*** **
Random Effects						
σ^2	2.36			2.39		
τ_{00}	0.68 $_{user}$			0.67 $_{user}$		
ICC	0.22			0.22		
N / Observations	166 $_{user}$ / 15106			184 $_{user}$ / 16744		
AIC	15205.1			16350.8		
Marg. R^2 / Cond. R^2	0.001 / 0.224			0.009 / 0.226		

**p* < .05, ***p* < .01, ****p* < .001

Vector Autoregression Models

Bivariate vector autoregression (VAR) models were developed to investigate the presence of time-lagged effects between the project creation and the interaction-related variables at the three levels. Time series data segmented at 1-day intervals was utilized in this analysis. VAR models were fitted using the *VAR* function included in the *vars* package in R (Pfaff, 2008) for each of the 9 predictor variables for the sent and received comments, resulting in a total of 18 VAR models for each sample user. Maximum lag length was designated at 7 to limit the temporal associations to a period of about one week. Lag selection for each VAR model was determined using the AIC.

For each VAR model, the Granger causality test was conducted using the *causality* function in the *vars* package, in which both variables in the bivariate function were considered to be the causal factor. Furthermore, several tests were conducted on the data as well as the VAR models to assess the viability of the results. An Augmented Dickey-Fuller test was carried out to confirm the stationarity of each time-series, while the stability of the model was checked by verifying that the absolute values of the eigenvalues of the coefficient matrices were less than 1. The Portmanteu test was also used to test the model for serially correlated errors.

Following this process, only the VAR models having satisfied all of the tests with a statistically significant Granger causality result at the alpha = 0.05 level were preserved. The frequencies of cases in which Granger causality was observed for each predictor is presented in the two tables below, one for comments sent (Table 17) and the other for the comments received (Table 18). Along with the number of cases with significant Granger causality found for each predictor variable in both the directions of causation, the average lag lengths for the respective set of cases are provided.

Examining the frequency of cases for the sent comments, it can be seen that predictors in the first (Comment Activity) and third (Network Location) categories generally performed better than those in the second (Functional Focus) category, although some exceptions are present. Overall, the Network Location predictors had the highest number of significant cases for both directions of causation. In over 17% of users, the ENA x-coordinate of the sent comments was found to Granger-cause the number of projects created. For the reverse causal direction, the number of projects created was found the Granger cause ENA y-coordinate values in nearly 26% of users. The average lag length for the cases with significant Granger causality mostly ranged between 2 and 3 for the sent comments.

For the received comments, the previous values of network location achieved the highest frequency of Granger causality for predicting current values of content creation. However, there were greater instances of Granger causality observed in the reverse direction. The highest frequencies were found for the number of comments received and for the Supportive Feedback codes, accounting for 22 and 24% of the users, respectively. This suggests that project creation in preceding segments were found to be associated with the numbers of comments received in the current period, especially for comments exhibiting Supportive Feedback.

Table 17

Cases of Significant Granger Causality for Comments Sent by Sample Users (N=166)

		X: Sent Comments / Y: Projects Created			
Category	**Predictor**	**Granger Causality X → Y**	**Avg. Lag Length**	**Granger Causality Y → X**	**Avg. Lag Length**
Comment Activity	Number of Comments	27 *(16.27%)*	2.33	27 *(16.27%)*	1.89
Functional Focus	Supportive Feedback	25 *(15.06%)*	2.60	24 *(14.46%)*	2.29
	Building a Following	25 *(15.06%)*	2.92	40 *(24.10%)*	2.53
	Personalized Tutoring	20 *(12.05%)*	2.35	20 *(12.05%)*	2.15
	Agency in Learning	10 *(6.02%)*	2.00	24 *(14.46%)*	1.96
	Cultural Competence	6 *(3.61%)*	2.33	7 *(4.22%)*	2.14
	Conflict	9 *(5.42%)*	2.78	20 *(12.05%)*	2.80
Network Location	ENA_x	29 *(17.47%)*	2.45	42 *(25.30%)*	2.38
	ENA_y	22 *(13.25%)*	2.36	43 *(25.90%)*	2.56

Table 18

Cases of Significant Granger Causality for Comments Received by Sample Users (N=184)

		X: Received Comments / Y: Projects Created			
Category	**Predictor**	**Granger Causality X → Y**	**Avg. Lag Length**	**Granger Causality Y → X**	**Avg. Lag Length**
Comment Activity	Number of Comments	22 (11.96%)	2.45	41 (22.28%)	2.05
Functional Focus	Supportive Feedback	19 (10.33%)	2.16	44 (23.91%)	2.11
	Building a Following	22 (11.96%)	2.50	21 (11.41%)	2.52
	Personalized Tutoring	16 (8.70%)	2.81	29 (15.76%)	2.24
	Agency in Learning	16 (8.79%)	2.50	28 (15.22%)	2.64
	Cultural Competence	1 (0.54%)	1.00	8 (4.35%)	2.50
	Conflict	11 (5.98%)	2.82	20 (10.87%)	2.10
Network Location	ENA_x	31 (16.85%)	2.06	36 (19.57%)	2.50
	ENA_y	28 (15.22%)	2.43	35 (19.02%)	2.31

Chapter 5: Discussion

Introduction

The purpose of this study was to explore how social interaction and collaboration are related to the creation of content by users in the Scratch online community. Through the examination of project-related comments exchanged by participants, this research aimed to provide insights on the activities of young participants in informal online learning environments, particularly on the patterns of discourse exhibited in the comments and the temporal associations between project creation and various dimensions of interaction. This study focused on the activities of a random sample of 200 Scratch users over a period of three months in 2012. Main methods of analysis included epistemic network analysis and time series analysis through generalized linear mixed modeling and vector autoregression techniques. In particular, the study aimed to address the following research questions:

- **RQ 1:** What patterns of discourse are observed in the comments of Scratch users? Are there differences in the discourse patterns among users with varying modes of participation?
- **RQ 2:** What relationships, if any, exist between the different elements of interaction and content creation?

Commenting in the Scratch Online Community

Scratch is a global community for young digital content creators. Established in 2007 to encourage and facilitate the development of computational thinking and programming skills among young people, the Scratch website has evolved into a platform where millions of users around the world come to interact and collaborate with one another in the creation of digital media content (Hill & Monroy-Hernández, 2017; Resnick et al., 2009). This study focused on the more active participants in the Scratch online community by sampling from a subset of users who had created at least one project in each month of the study period. Due to the application of this criterion, the 200 sample users included in this study were on average more experienced and demonstrated a higher level of project creation when compared against the overall population of users on the platform. For example, the average number of projects created by the sample users during the study period was just over 18, which was about 4 times higher than the total for all users who had created at least one project during the same period.

Project-related comments provide an important way for users to participate in the Scratch platform, especially as a means to communicate directly with fellow participants. The comments offer an opportunity for users to freely express their thoughts and opinions, going beyond the limited capacity to indicate an interest or appreciation for a project through the click of a love-it and favorite button. In this manner, comments allow users to connect, interact and collaborate with each other in various ways. At the same time, commenting is something that is completely optional to the user. Some users choose not to engage in the sharing of comments altogether. Examining the commenting behavior of the sample users, it was found that the average sample user posted just over 100 comments over the three-month period. This average value also included 34 users from the sample who had not posted a single comment during the same period.

Comments on the Scratch platform are diverse in terms of their form, content and function. Users are able to post messages in a new thread or as a nested response within an existing threaded message. This open structure creates additional possibilities in terms of how users utilize commenting in their interactions—not to mention added level of complexity for researchers trying to understand the patterns that exist within them. Conversations that are generated as the result of this process vary widely. Some consist of just one comment, while others include hundreds of messages. Analysis from the sample data, however, indicate that most of the conversations remain relatively brief, with over 90% of them containing 4 or fewer comments.

As for where the comments are posted, it was found that most users comment on projects that have been recently created. Over 94% of the comments in study dataset was related to projects created during the study period. There does not, however, appear to be any preference for commenting on one's own project versus another user's project. For the sample dataset, this ratio was found to be about 46 to 54. There was, however, significantly a higher number of comments posted on new projects as opposed to remixed projects.

The comments were just as varied in their function. In addition to the themes originally included in the coding scheme, three interesting phenomena were identified during the coding process. These trends highlight the myriad of different ways that comments enable users to participate in the Scratch community.

First was a role-playing activity that took place among users in the comments, often in real time with participants taking turns to create a story in a collaborative manner. Role Playing was coded in a total of 5,217 comments, accounting for about 11% of all comments in the dataset. A tendency for Role Playing to occur in extended exchanges among users was also observed, with participants often engaged in interactions which continued on for at least dozens, if not hundreds, of comments. It is difficult to conclude whether Role Playing was indeed a more widespread phenomenon across the entire Scratch community simply based on the sample data used for this study. However, it nevertheless appears that it is a significant phenomenon that merits additional attention. While beyond the scope of this research, further analysis focusing on these specific set of comments may shed valuable insights on the phenomenon of collaborative storytelling in online contexts.

Another trend involves contests and auditions, in which project creators invite participation from other users. For the contests, users are offered a prize, usually in the form of love it's or favorites given to projects designated by the winner. In the auditions, users are given the opportunity to try out for parts in animated videos and other types of projects that often involve voice recordings. In both cases, rules and timelines are established, and submissions are generally made through links provided in the comments. Such events help the project creator garner greater interest from others in the community and foster collaboration on future projects.

Lastly, users interact in with one another in the comments to engage in personal exchanges of creative digital content. These typically involve art trades and requests. In the former, two users agree to create artwork for each other, frequently of cartoon or animation characters that is specified by the receiver. The art requests, on the other hand, are unidirectional. The artist accepts requests submitted by other users through the comments and creates it for them, usually without asking for anything in return. In many ways, the art trades and requests seem to enable users to add a personal touch to the interactions they have in the Scratch community, while at the same time allowing them to share their own creative work with the outside world.

This study applied a coding scheme adapted from Fields et al. (2015), which was developed based on a grounded analysis of 2,273 randomly sampled project-comments from the Scratch online community. As both studies focused on comments from the same period (i.e. January to March 2012), the findings reported in the Fields et al. (2015) paper provide a suitable comparison group with which to assess the process and outcome of the qualitative coding carried out in this study. Two general observations are worth noting, keeping in mind that modification adopted for this study make more specific comparisons difficult.

First observation is that the majority of comments shared in the Scratch community was positive in nature. This was confirmed in both studies, with the proportion of negatively associated comments only ranging from 5 to 11% of all comments analyzed in the studies (Fields et al., 2015). Furthermore, the two results concur that encouraging feedback and efforts to build a community of one's own following account for a significant portion of all interactions. The second observation relates to the overall consistency with which content-focused codes appeared in the comments. In the two analyses, Personalized Tutoring accounted for about 9-10% of the comments while Agency in Learning made up between 5 and 10%. Given that the no modifications were made on these two codes, this may reflect the consistency with which the codebook was operationalized in the two studies. It also may point to the overall level of content-related discussions taking place within the comments. While they do not represent the most frequent codes, there nevertheless appears to be a consistent level of constructive interaction occurring between participants on the Scratch platform.

Patterns of Discourse

ENA was used in this study to generate graphical representations of the connections between the codes for each comment that was either sent or received by the sample users. These comment-level networks were then aggregated up, first by user and then by the specified group of users that were selected for each analysis conducted in the study. The network models therefore presented a snapshot of the discourse pattern for the average comment that was communicated by users from a certain group. In the network graphs, the edge weights of the model indicate a relative strength of connection between the codes. Therefore, the edge weights provide an indicator of the relative significance of each connection presented in the ENA model.

Figure 13 shows the weights of the edges of the ENA model for all comments that were sent and received by sample users (see Figure 6 for the corresponding network graphs). The edge weights signify the relative strength of connection between the codes in the model. It can be seen that the strongest link is between Supportive Feedback and Building a Following for both the sent and received comments. This is followed by the code pair Building a Following–Agency in Learning, and then by Personalized Tutoring–Agency in Learning. The discussion below focuses on these three most prominent connections in the discourse of users in the Scratch online community.

Figure 13

Heat Map of the ENA Edge Weights for All Comments Sent and Received by Sample Users

	Sent Comments	Received Comments
Supportive.Feedback & Building.Following	0.0954	0.1028
Supportive.Feedback & Personalized.Tutoring	0.0392	0.0315
Building.Following & Personalized.Tutoring	0.0360	0.0361
Supportive.Feedback & Agency.Learning	0.0267	0.0255
Building.Following & Agency.Learning	0.0522	0.0509
Personalized.Tutoring & Agency.Learning	0.0428	0.0402
Supportive.Feedback & Cultural.Competence	0.0043	0.0025
Building.Following & Cultural.Competence	0.0071	0.0077
Personalized.Tutoring & Cultural.Competence	0.0049	0.0049
Agency.Learning & Cultural.Competence	0.0025	0.0016
Supportive.Feedback & Conflict	0.0046	0.0046
Building.Following & Conflict	0.0114	0.0098
Personalized.Tutoring & Conflict	0.0154	0.0127
Agency.Learning & Conflict	0.0095	0.0066
Cultural.Competence & Conflict	0.0076	0.0075

Supportive Feedback–Building a Following

Supportive Feedback and Building a Following were the two codes that occurred most frequently in the dataset, being present in about 17% and 16% of all comments, respectively. As the higher value of the edge weight suggests, the two codes often occurred together in the same conversation, where an encouraging comment posted by a user visiting a project was followed by a response from the project creator acknowledging the message and expressing gratitude. As was shown in Figure 4 in Chapter 4, the frequency of occurrence is highest for conversations consisting of only 1-2 comments and then declines for longer conversations. These types of brief, positive exchanges between users can be found throughout the comments in the Scratch community, similar to the example below:

> User A: Impressive!!
> User B: Thanks and dont forget to click on love it!

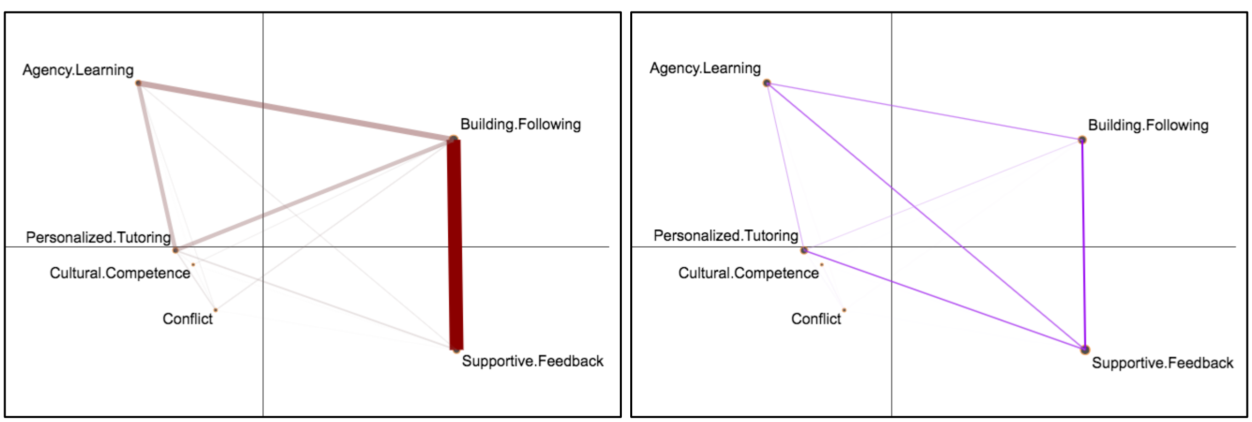
In other instances, the initial communication of support also includes other elements such as suggestions for improvement (Personalized Tutoring) or even an appeal to come view their own projects (Building a Following). The example below contains all three of these components:

> User C: I've seen chanmanpartyman's tutorial. I found a simpler version that a friend made, I used that, with better camera effects. Just see my Survival Pit 2, it's just a test. And for the lava that popping up out of nowhere, you may need a separate variable for that. chanmanpartyman's scripts are only for the scrolling floor. Anyway, cool test. Please see my projects.
> User D: thanks and i will. maybe youd like my real user better (nanosaurus) i have a bunch of cool games on there.

As shown in these exchanges, comments coded with these two codes mostly involve the project creator. Among received comments coded for Supportive Feedback, those that were directed at the project owner accounted for 92 percent. Similarly, Building a Following was mostly exhibited by the content creators, making up 71 percent of all sent comments coded for the theme. Further insights can be gleaned from the network graphs for the comments sent and received by sample users on their own projects, shown in Figure 14. For the comments sent by sample users on left, a dominant connection can be seen between Building a Following and Supportive Feedback along with moderate linkages to Agency in Learning and Personalized Tutoring. The opposite can be observed in the network to the right for the comments received by users on their projects. Distinct connections are visible linking Supportive Feedback with Building a Following, Agency in Learning and Personalized Tutoring. Similar to User C in the example above, many visitors appear to be combing other elements, such as suggestions, questions and requests, with their message of encouragement to the project creators. From this perspective, it seems feasible that Supportive Feedback may also be serving a key role as an entry point for other types of messages in the interactions among users in the Scratch community.

Figure 14

Comments Sent(left) and Received(right) by Sample Projects on Their Own Projects



Building a Following–Agency in Learning

Similar to the first pair of codes, the relationship between Building a Following and Agency in Learning also appears to be present in multiple dimensions. At the most basic level, the connection is representative of a request for resources, help or collaboration followed by a positive or polite response:

User E: I have a request.... can you get me the Kanto original or Johto original Champion music?
User F: As soon as I can get my microphone fixed, yeah! c:
User E: Awesome!Thanks.

In other situations, it involves the communication around a collaborative endeavor. In the example below, two participants engage in discussion in an effort to jointly identify a song for a collaboration on an animated music video:

User G: LOLZ. Hey, do you wanna do a collab project? The theme is AMVs.
User H: Sure! I've never done an AMV collab!
User G: Meh either. What song do you wanna do?
User H: I'm not sure. I don't want to do something too popular and overrated like Party Rock Anthem, though.

It is interesting to note that although they have agreed to undertake collaborative work, they are still in a preparatory phase, in which they have begun the process negotiation and search for common ground. In such cases, the association between Agency in Learning and Building a Following can be indicative of a state of willingness and readiness for collaboration among its participants.

At the same time, the connection between the two codes can occur in contexts in which Building a Following serves a moderating role in keeping a conversation cordial and constructive. In the example below, User I is given an answer that is not directly related to their question about website design. In reply, User I recognizes and expresses appreciation for the response by User J while providing an explanation for why it may be less relevant to their proposed need:

User I: Awesome! Just wondering, do you know anything about website design? Like maybe Weebly? I'm trying to find out how I could add a project to a website.

User J: well I know how. go to your project and click “embed” under link to this project and copy and paste a code into the “custom html” box
User I: I was looking for a way to do it without uploading to scratch, but that helps too. Thanks!

In this way, just as Supportive Feedback may be seen as providing an entry point for other types of interactions to occur, Building a Following may be serving an important role in allowing the discourse to remain friendly and civil.

Agency in Learning–Personalized Tutoring

Similar to the connection between Supportive Feedback and Building a Following, this pairing makes intuitive sense, especially when conceived in the form of a question and answer focusing on the topic of the project. This is exemplified in the example below, in which the second user provides a sequence of codes related to the question of how to save the high score of a game:

User K: OMG in high score it saves!!! HOW'D YOU DO THAT!
User L: I didn't make this but I'm pretty sure it's:
(when flag clicked)
(if (score) > (high score))
((set 'score' to 'high score)
Pretty simple but very effective.

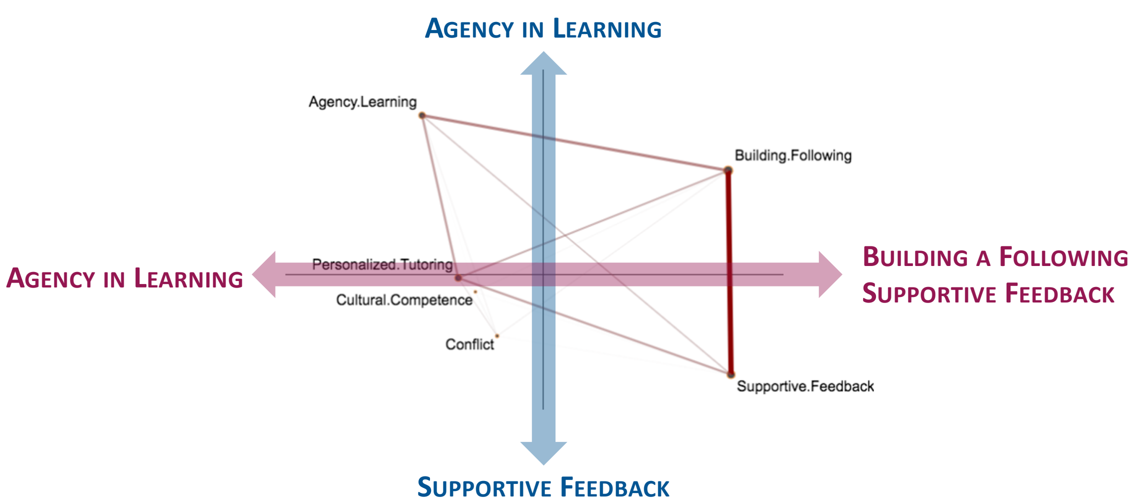
The dataset includes numerous instances capturing the sharing of content-related information and knowledge. In essence, the connection between Agency in Learning and Personalized Learning can be said to represent situations when relevant meaning or knowledge is provided to someone who is seeking it. Conversations containing such interactions provide evidence for the support that is given by certain users to facilitate the learning and development of others in the Scratch community. From this perspective, it can be argued that the connection between these two codes presents an interaction that could be considered most reflective of collaborative learning, in which the discourse contributes to the co-construction of meaning by members of the Scratch community (Stahl, 2004).

Comments Discourse and Content Creation

In order to facilitate the analysis of the relationship between content creation and the discourse patterns exhibited in the comments, the two-dimensional ENA space was interpreted based on the location of the network nodes, particularly those positioned toward the extreme edges of the network structure in the x and y dimensions. In the ENA model produced for this study, the x-axis was defined by Building a Following and Supportive Feedback to the right and Agency in Learning on the far left. The y-axis was distinguished primarily by Agency in Learning toward the top and Supportive Feedback toward the bottom of the ENA space. Figure 15 provides a visualization of the interpretive dimensions of the ENA space defined by the Scratch comments.

Figure 15

Interpretation of the X- and Y-dimensions of the ENA Space Defined by the Scratch Comments



The results of the ENA models were assessed based on this interpretative framework. Main avenues of comparative analysis among subsets of the sample users included: 1) higher versus lower number of total comments; 2) higher versus lower number of total projects; 3) increasing versus decreasing project creation over time.

The first analysis focused on examining whether significant differences existed in the discourse patterns of users who had posted a higher than average number of comments in the three-months of the study as compared to those who had not. The statistically significant differences between the location of the group means of the two group for the sent and received comments are shown in Table 19. In the table, upward and downward arrows indicate a statistically significant difference along the y-axis, while left and right arrows are used for a significant difference along the x-axis. The corresponding code or codes for each direction from the interpretative framework are listed below each arrow. Therefore, it can be seen that the Higher Comments Group are more closely associated with Agency in Learning in the comments that they sent and received, meaning that the discourse for this group was related more with seeking information and collaboration. On the other hand, the discourse for the Lower Comments Group demonstrated more linkages to giving encouragements and building relationships with other users.

Table 19

Statistically Significant Differences in the Group Means (Higher vs. Lower Comments)

	Sent Comments	**Received Comments**
Higher Comments Group	↑ Agency in Learning	← Agency in Learning
Lower Comments Group	↓ Supportive Feedback	→ Building a Following Supportive Feedback

Similarly, Table 20 presents the statistically significant differences in the position of the groups means of the Higher and Lower Project Groups. Users in the Higher Project Group had created more than the average of 18 projects during the 3-months from January to March 2020. Applying the interpretative framework to this comparison group, it can be observed that the discourse for the Higher Projects Group exhibited greater focus on seeking information and collaboration in both the comments they sent and received. In contrast, the discourse for the Lower Projects Group made more connections to providing encouraging feedback, gaining more attention and establishing relationships with other users.

Table 20

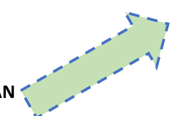
Statistically Significant Differences in the Group Means (Higher vs. Lower Projects)

	Sent Comments	**Received Comments**
Higher Projects Group	← Agency in Learning	← Agency in Learning
Lower Projects Group	→ Building a Following Supportive Feedback	→ Building a Following Supportive Feedback

Lastly, the changes in the groups means from January to March for the Increasing Project Creation (IPC) and Decreasing Project Creation (DPC) groups are shown in Table 21. These two groups were formed by identifying trends in the number of projects created by each user per month. Users in the IPC group demonstrated an overall increase in the monthly projects created from January to March without experiencing a decline in any subsequent months. The DPC group, on the other hand, displayed a downward trend over the three months without having an uptick in monthly project creations. Examining the change in the group means, it is possible to see that for the IPC, the discourse pattern for the sent messages shifted to have greater focus on providing encouragements and building a following of other users. No significant changes were observed for their received comments. For the DPC group, the discourse pattern for the sent comments shifted to have more connections with the seeking of information and collaboration. The group mean for the received comments changed meaningfully in both the x- and y-dimensions, thereby having more emphasis on Agency in Learning, Building a Following and Supportive Feedback.

Table 21

Statistically Significant Shifts in the Group Means from January to March (IDC vs. DPC)

	Sent Comments	**Received Comments**
Increasing Projects Creation (IPC)	JAN  MAR Building a Following Supportive Feedback	(No sig. change)
Decreasing Projects Creation (DPC)	MAR  JAN Agency in Learning	JAN  MAR Agency in Learning Building a Following Supportive Feedback

To summarize the results outlined above for the three types of analyses, it was found that at the level of the sample users, the comments involving users who exhibit a higher level of participation—through project creation or the sharing of comments—tend to have a greater focus on seeking information and collaboration. Those who demonstrate a lower level of participation in content creation and interaction through comments tend to make more connections to the more social aspects of gaining attention, building relationships and sharing encouraging messages with others.

For the subset of sample users demonstrating an increase or decrease in their project creation, it may appear that the findings are somewhat contradictory to the conclusions reached above. The changes in the group means of both groups appear to be going in the directions opposite to what would be expected. However, it should be noted that the change in their level of participation in the community may be having an effect on how they are interacting with others through the comments. Another key factor to consider is that along with the increase/decrease in the number of projects, there is a corresponding rise/decline in the number of comments sent and received by members in each group (see Chapter 4 for more details). Given that there is a higher prevalence of the Supportive Feedback and Building a Following codes in the overall dataset, an increase in comments would indicate a greater presence of those codes in the comments that are added each month to the IPC total. These results may also be associated with the particular users included in the groups or the projects that were created by them during the study period. However, it is difficult to ascertain the specific factors that may be affecting the discourse patterns for these two groups at the current level of analysis. Further research should be undertaken to examine this issue in greater detail.

Temporal Analyses of Discourse and Content Creation

The temporal relationship between different elements of interaction and content creation were examined through generalized linear mixed models (GLMM) and vector autoregression (VAR) models. Time-series data were constructed at 1-day intervals. The outcome variable consisted of the count of projects created, while the predictor variables consisted of the different dimensions of interaction in the comment activity, functional focus and network location categories. The predictor variables were also differentiated into two datasets by whether they were comments that were sent or received by the sample users.

To assess the contemporaneous effects of the predictor variable on project creation, three negative binomial GLMM models were developed for each dataset. The results found significant fixed effects for nearly all predictors in each dataset. This meant that the interaction-related variables were found to be significant predictors of project creation. For the random effects of the grouping variable, which in this analysis was the user, the intra-class correlation coefficient (ICC) was calculated. Similar values ranging from 0.22 to 0.23 were obtained for all models, indicating a consistent level in the proportion of the variance explained by the random variable. However, the variance explained by the fixed effects, as measured through the marginal R^2 values, varied across the models. The marginal R^2 value was particularly low for the Category 3 Model which had used the x- and y-coordinates of the ENA projected points for each comment. From these results, it was concluded that the Category 1 Models that used the counts of comments provided the best fit when considering the contemporaneous associations between interaction and content creation.

Several factors should be taken into account to better understand the result of the GLMM analysis. Given that the time-series datasets were constructed at 1-day intervals, there is a high likelihood that the results, especially for the Category 1 and 2 Models, are representative of the association between the comments sent and received by the user following the sharing of a project on the same day. Because the time-series data for comment activity and functional focus were simple aggregates of all comments or code counts occurring in the same calendar day, the model was not able to discriminate between comments posted before or after the project was shared. In fact, the incidence rate ratios, the factor by which the expected count of projects would rise per unit increase in the predictor, is generally higher for the received comments when compared against the models for the sent comments in the same category. This seems to further suggest that comments being directed at projects shared on the same day may be a key reason behind the significant fixed effects obtained in the models. While beyond the scope of this study, these observations raise several questions regarding the most appropriate segmentation for such analyses using GLMMs. Time segments that aggregate the data at smaller intervals might help alleviate the potential for reverse causality; however, creating segments that are too small can effectively remove the need for examining contemporaneous effects in the first place. Properly addressing this issue might require considering comment posting and project creation in terms of their sequential order of occurrence, rather than using a time variable to create aggregated counts at specified time intervals.

The second analysis examined the time-lagged effects of the predictor variables on the number of projects created. Using the 1-day interval time-series data used in the GLMMs, bivariate VAR models were constructed for each predictor variable as well as for each sample user. In this process, a total of 3,150 VAR models were fitted. Granger causality tests were carried out—along with a series of statistical tests to check the stationarity of the data, stability of the model and the absence of serially correlated errors—to assess the influence of the previous values of the predictor variables on the current value of the outcome variable. Significant results obtained for the Granger causality tests, including the lag length of used in the model, were tabulated to examine the overall trends across the dataset. The results showed that in general, higher frequencies of significant Granger causality findings were observed for the predictors in the first (Comment Activity) and third (Network Location) categories, for both sent and received comments. The average lag lengths for these models were mostly between 2 and 3, meaning that the lagged values of the interaction-related variables from the previous 1, 2, and 3 days helped to predict the number of projects created in the current day. While these results of the exploratory VAR models are preliminary in nature, they nevertheless suggest a potential for utilizing the x- and y-coordinate values of ENA projected points in conducting similar time-series analysis.

Methodological Considerations

Methodologically, this exploratory study aimed to bring together qualitative and quantitative methods to provide an in-depth analysis of a large set of data from the Scratch online community. The Scratch longitudinal dataset presents a valuable source of information on how young users engage in content creation and interact with one another in the online community. In this study, several new approaches were adopted, including the contextualization of the comments data into conversational threads grouped by the projects with which they were related. This organization of the data allowed for a more nuanced coding of the comments based on the conversational context present in each thread. Another contribution was the examination of the comments from the perspective of both sender and receiver. While this decision had the effect of doubling all of analyses conducted as part this study, it nevertheless shed light on some of the differences that exist in the patterns of discourse exhibited in the comments exchanged among users on the platform.

Despite these advances, several challenges remain that should be addressed in future research. First is the issue of single-comment conversations. In the current dataset, a total of 8,426 comments were identified as being part of single-comment conversations. In terms of the total number of comments, this accounted for just over 18%. However, it constituted 46% of all conversations in the dataset. If it were the case that all of such comments were indeed stand-alone message directed at the project or its creator, it would not pose a significant problem. However, in the qualitative coding of the data for this study, it was frequently observed that first-level comments were being used to respond to previous messages. For this reason, an exchange of several linked comments might occur as five single-comment conversations instead of one conversation with five interrelated comments. This creates a challenge not only in terms of contextualization of the data, but also in the development of analytical models such as ENA that take into account the co-occurrence of codes within a single conversational unit. While no easy fixes may be available, efforts should be made to come up with an effective and efficient means to address this issue.

Another challenge lies in the analysis of the substantial amount of data available in the Scratch longitudinal dataset, particularly in the coding of the comments. Given the exploratory nature of this study, manual coding was undertaken. One of the key reasons for this choice was that the comments were user-generated and therefore contained irregularities that may lead to problems for automated techniques. While more than 45 thousand lines were qualitatively annotated for this study, this only represented a small proportion of the entire dataset. In order to harness additional insights from this large dataset, a robust system and process for automated coding will need to be established.

Summary of Findings

This exploratory study aimed to examine the relationship between interaction, collaboration and content creation through the analysis of user-generated comments and log-data from the Scratch community. The research focused on more than 45 thousand comments associated with the online activity of 200 randomly selected sample users over a period of three months from January to March 2012. A mix of methodological techniques were applied in the analysis of the data. ENA was carried out to explore and identify patterns in the discourse of the comments. Negative binomial GLMMs and bivariate VAR models were used to investigate the temporal associations between interaction and content creation. The network location variables obtained from the ENA models were also utilized as predictor variables in the GLMMs and VAR models. A summary of the main findings of the study is presented in Table 22.

Table 22

Summary of the Study's Main Findings

Research Questions		Analytical Technique	Summary of Findings
1	What patterns of discourse are observed in the comments of Scratch users?	ENA	The ENA models show a higher prevalence of positive and socially-oriented messages within the discourse. Users who actively seek assistance or cooperation through the comments are often acknowledged and provided with pertinent information or knowledge. Such interactions are reflective of collaborative learning among users in the Scratch platform, in which the discourse contributes to the co-construction of meaning by members of the community.
	Are there differences in the discourse patterns among users with varying modes of participation?	ENA	ENA findings suggest that discourse involving sample users with a higher level of participation in the Scratch online community through project creation or the sharing of comments place greater focus on seeking information, help and collaboration. On the other hand, the discourse related to sample users with a lower level of participation in project creation or comment posting are more focused on socially-oriented exchanges such as the sharing of encouraging messages, expressing appreciation and seeking attention.
2	What relationships, if any, exist between the different elements of interaction and content creation?	Negative binomial GLMM	*[Contemporaneous Effects]* Results of the negative binomial GLMMs indicate that the number of comments sent and received by the sample users were most predictive of the number of projects created on the same day. The use of ENA network location values as predictors of content creation in the GLMMs was assessed; the models, however, were not selected due to their extremely low level of explained variance.
		Bivariate VAR	*[Time-lagged Effects]* Findings from the bivariate VARs show that the predictors using the number of comments or the ENA network location values were more frequently found to have Granger causal associations with project creation in both directions of causation. Among the functional codes, the predictors Supportive Feedback and Building a Following exhibited relatively higher incidences of Granger causality. Furthermore, project creation was more frequently found to precede interaction for most predictors, especially in the comments received by the creators.

Implications of the Study

The results of this study present several implications related to the learning and behavior of young people in informal online learning communities. Analysis of the comment data revealed that young users participate in the Scratch community in diverse ways that involve actively interacting and collaborating with others. Socially-oriented interaction, as exhibited through the sharing of positive feedback and efforts to develop relationships, was found to a significant part of the discourse. This confirms the importance of social dynamics in online learning communities such as the Scratch platform. Young people are seeking to connect with other users while they engage in other more cognitive aspects of the environments, such as creating projects and developing skills in programming. In this manner, the interplay between the social and cognitive engagement among users constitutes an important dimension in the comments discourse. Requests made by users for information, assistance and cooperation were frequently acknowledged and answered with specific guidance and suggestions. Such constructive interactions—exchanges leading to an improved understanding of relevant knowledge and concepts (Miyake, 1986)—create the basis on which collaborative learning can occur.

The findings from the temporal analyses also point to a mutually reinforcing relationship between social interaction and content creation. The number of comments sent or received were found to be most significantly linked to the number of projects created on the same day. Furthermore, project creation was more frequently found to precede interaction in several dimensions, especially in the comments received by the creators. For some users, the findings point to bidirectional Granger causality, indicating that the interaction and content creation variables may be jointly affecting one another. These results suggest that project creation and interaction may be part of an interconnected process, in which the social and cognitive dimensions each support and motivate the other. This is also reflective of the evolving nature of digital participation around learning and social engagement (Ito et al., 2009). For many users, the creation of projects often serves as a means for receiving comments, gaining attention and engaging in conversations with others in the community, which in turn drives additional efforts for content creation.

Given this interrelationship between social and cognitive dimensions in facilitating engagement and participation among users, there appears to be a need for improved integration of both components in online learning platforms and communities. Modes of interaction should go beyond linear, asynchronous discussions to one that is more dynamic in nature and can be carried out in real- or near-real-time. As observed from the Scratch data, learners have figured out ways to circumvent the limitations in the system, particularly in collaborating with one another in innovative ways. For example, during a role-playing session, a comments thread was transformed into a communication tool that was nearly synchronous. Therefore, efforts should be made in the design of online learning platforms to incorporate elements that promote and support greater real-time collaboration among participants in the joint creation of digital content.

REFERENCES

Ackermann, E. (2001). Piaget's constructivism, Papert's constructionism: What's the difference? *Constructivism: Uses and Perspectives in Education* (pp. 85-94). https://learning.media.mit.edu/content/publications/EA.Piaget%20_%20Papert.pdf

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), 357-376. https://doi.org/10.1037/0022-3514.45.2.357

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly, 50*(3), 367-403. https://doi.org/10.2189/asqu.2005.50.3.367

Amabile, T. M., & Pillemer, J. (2012). Perspectives on the social psychology of creativity. *The Journal of Creative Behavior, 46*(1), 3-15. https://doi.org/10.1002/jocb.001

Anderson, T., & Soden, R. (2001). Peer interaction and the learning of critical thinking skills. *Psychology Learning & Teaching, 1*(1), 37-40. https://doi.org/10.2304/plat.2001.1.1.37

Angeli, C., Valanides, N., & Bonk, C. J. (2003). Communication in a web-based conferencing system: The quality of computer-mediated interactions. *British Journal of Educational Technology, 34*(1), 31-43. https://doi.org/10.1111/1467-8535.d01-4

Aragon, C. R., Poon, S. S., Monroy-Hernández, A., & Aragon, D. (2009). A tale of two online communities: Fostering collaboration and creativity in scientists and children. *Proceedings of the 7th ACM Conference on Creativity and Cognition, Berkeley, CA, US* (pp. 9-18). https://doi.org/10.1145/1640233.1640239

Aragon, C. R., & Williams, A. (2011). Collaborative creativity: A complex systems model with distributed affect. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada* (pp. 1875-1884). https://doi.org/10.1145/1978942.1979214

Azmitia, M. (1988). Peer interaction and problem solving: When are two heads better than one? *Child Development, 59*(1), 87-96. https://doi.org/10.2307/1130391

Baker, M. J. (1999). Argumentation and constructive interaction. In P. Coirier & J. Andriessen (Eds.), *Studies in writing: Vol. 5. Foundations of argumentative text processing* (pp. 179-202). University of Amsterdam Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beals, L. M. (2010). Content creation in virtual worlds to support adolescent identity development. *New Directions for Youth Development, 2010*(128), 45-53. https://doi.org/10.1002/yd.374

Beghetto, R. A. (2010). Creativity in the classroom. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 447-463). Cambridge University Press.

Bell, N., Grossen, M., & Perret-Clermont, A.-N. (1985). Sociocognitive conflict and intellectual growth. *New Directions for Child and Adolescent Development, 1985*(29), 41-54. https://doi.org/10.1002/cd.23219852905

Blumer, H. (1986). *Symbolic interactionism: Perspective and method*. University of California Press.

Bose, E., Hravnak, M., & Sereika, S. M. (2017). Vector autoregressive models and Granger causality in time series analysis in nursing research: Dynamic changes among vital signs prior to cardiorespiratory instability events as an example. *Nursing Research, 66*(1), 12-19. https://doi.org/10.1097/nnr.0000000000000193

Brake, D. R. (2014). Are we all online content creators now? Web 2.0 and digital divides. *Journal of Computer-Mediated Communication, 19*(3), 591-609. https://doi.org/10.1111/jcc4.12042

Brennan, K. (2016). Audience in the service of learning: How kids negotiate attention in an online community of interactive media designers. *Learning, Media and Technology, 41*(2), 193-212. https://doi.org/10.1080/17439884.2014.939194

Brennan, K., Monroy-Hernández, A., & Resnick, M. (2010). Making projects, making friends: Online community as catalyst for interactive media creation. *New Directions for Youth Development, 2010*(128), 75-83. https://doi.org/10.1002/yd.377

Brennan, K., & Resnick, M. (2013). Imagining, creating, playing, sharing, reflecting: How online community supports young people as designers of interactive media. In C. Mouza & N. Lavigne (Eds.), *Emerging technologies for the classroom* (pp. 253-268). Springer.

Brennan, K., Valverde, A., Prempeh, J., Roque, R., & Chung, M. (2011). More than code: The significance of social interactions in young people's development as interactive media creators. *Preceedings of the 2011 EdMedia+ Innovate Learning Conference, Lisbon, Portugal* (pp. 2147-2156). https://www.learntechlib.org/p/38158

Brown, B. B., & Larson, J. (2009). Peer relationships in adolescence. In *Handbook of adolescent psychology: Contextual influences on adolescent development, Vol. 2, 3rd ed.* (pp. 74-103). John Wiley & Sons Inc.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. https://doi.org/10.3102/0013189x018001032

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science, 2*(1), 40-57.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes. *Small Group Research, 36*(5), 600-631. https://doi.org/10.1177/1046496405279309

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35. https://doi.org/10.3102/00346543064001001

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Coursey, L. E., Williams, B. C., Kenworthy, J. B., Paulus, P. B., & Doboli, S. (2018). Divergent and convergent group creativity in an asynchronous online environment. *The Journal of Creative Behavior*. https://doi.org/10.1002/jocb.363

Craft, A. (2003). The limits to creativity In education: Dilemmas for the educator. *British Journal of Educational Studies, 51*(2), 113-127. https://doi.org/10.1111/1467-8527.t01-1-00229

Craft, A. (2012). Childhood in a digital age: Creative challenges for educational futures. *London Review of Education, 10*(2), 173-190.

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 13*(4), 419-438. https://doi.org/10.1007/s11412-018-9292-z

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In *The nature of creativity: Contemporary psychological perspectives.* (pp. 325-339). Cambridge University Press.

Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention*. Harper Collins Publishers.

Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning, 9*(3), 247-281. https://doi.org/10.1007/s11412-014-9193-8

Dasgupta, S., Hale, W., Monroy-Hernández, A., & Hill, B. M. (2016). Remixing as a pathway to computational thinking. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, CA, US* (pp. 1438-1449). https://doi.org/10.1145/2818048.2819984

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education: A systematic literature review. *Thinking Skills and Creativity, 8*, 80-91. https://doi.org/10.1016/j.tsc.2012.07.004

De Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 87-103. https://doi.org/10.1007/s11412-007-9006-4

Dillenbourg, P., Baker, M. J., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189-211). Elsevier.

Drotner, K. (2020). Children's digital content creation: Towards a processual understanding of media production among Danish children. *Journal of Children and Media, 14*(2), 221-236. https://doi.org/10.1080/17482798.2019.1701056

Eagan, B. R., Rogers, B., Serlin, R., Ruis, A. R., Arastoopour Irgens, G., & Shaffer, D. W. (2017). Can we rely on IRR? Testing the assumptions of inter-rater reliability. *Proceedings of the 12th International Conference on Computer Supported Collaborative Learning, Philadelphia, PA, US* (pp. 529-532). https://repository.isls.org/handle/1/275

Erstad, O. (2012). The learning lives of digital youth—beyond the formal and informal. *Oxford Review of Education, 38*(1), 25-43. https://doi.org/10.1080/03054985.2011.577940

Falloon, G., & Khoo, E. (2014). Exploring young students' talk in iPad-supported collaborative learning environments. *Computers & Education, 77*, 13-28. https://doi.org/10.1016/j.compedu.2014.04.008

Fasko, D. (2001). Education and creativity. *Creativity Research Journal, 13*(3-4), 317-327. https://doi.org/10.1207/s15326934crj1334_09

Fields, D. A., Giang, M. T., & Kafai, Y. B. (2013). Understanding collaborative practices in the Scratch online community: Patterns of participation among youth designers. *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning, Madison, WI, US* (pp. 200-207). https://repository.isls.org//handle/1/1810

Fields, D. A., Giang, M. T., & Kafai, Y. B. (2014). Programming in the wild: Trends in youth computational participation in the online Scratch community. *Proceedings of the 9th Workshop in Primary and Secondary Computing Education, Berlin, Germany* (pp. 2-11). https://doi.org/10.1145/2670757.2670768

Fields, D. A., Kafai, Y. B., & Giang, M. T. (2016). Coding by choice: A transitional analysis of social participation patterns and programming contributions in the online Scratch community. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 209-240). Springer International Publishing. https://doi.org/10.1007/978-3-319-13536-6_11

Fields, D. A., Pantic, K., & Kafai, Y. B. (2015). “I have a tutorial for this”: The language of online peer support in the scratch programming community. *Proceedings of the 14th International Conference on Interaction Design and Children, Boston, MA, US* (pp. 229-238). https://doi.org/10.1145/2771839.2771863

Fischer, G. (2004). Social creativity: Turning barriers into opportunities for collaborative design. *Proceedings of the 8th Conference on Participatory Design: Artful integration: Interweaving media, materials and practices, Toronto, ON, Canada* (pp. 152-161). https://doi.org/10.1145/1011870.1011889

Fischer, G. (2007). Designing socio-technical environments in support of meta-design and social creativity. *Proceedings of the 2007 Conference on Computer Supported Collaborative Learning, New Brunswick, NJ, US* (pp. 2-11). https://dl.acm.org/doi/10.5555/1599600.1599602

Fischer, G. (2016). Exploring, understanding, and designing innovative socio-technical environments for fostering and supporting mass collaboration. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 43-63). Springer International Publishing. https://doi.org/10.1007/978-3-319-13536-6_3

Fosnot, C. T., & Perry, R. S. (2005). Constructivism: A psychological theory of learning. In C. T. Fosnot (Ed.), *Constructivism: Theory, perspectives, and practice* (2nd ed., pp. 8-38). Teachers College Press.

Frederiksen, J. R., Sipusic, M., Sherin, M., & Wolfe, E. W. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment, 5*(4), 225-297. https://doi.org/10.1207/s15326977ea0504_1

Garrison, D. R. (2011). *E-learning in the 21st century* (2nd ed.). Routledge. https://doi.org/10.4324/9780203838761

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education, 19*(3), 133-148. https://doi.org/10.1207/s15389286ajde1903_2

Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior, 92*, 562-577. https://doi.org/10.1016/j.chb.2018.07.003

Gee, J. P. (2005). *An introduction to discourse analysis: Theory and method* (2nd ed.). Routledge.

Gee, J. P. (2018). *Introducing discourse analysis: From grammar to society*. Routledge.

Gee, J. P., & Green, J. L. (1998). Discourse analysis, learning, and social practice: A methodological study. *Review of Research in Education, 23*, 119-169. https://doi.org/10.2307/1167289

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The Interpretation of Cultures* (pp. 3-30). Basic Books.

Gilster, P. (1997). *Digital literacy*. Wiley Computer Pub.

Goggins, S. P., Laffey, J., & Gallagher, M. (2011). Completely online group formation and development: Small groups as socio-technical systems. *Information Technology & People, 24*(2), 104-133. https://doi.org/10.1108/09593841111137322

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424. https://doi.org/10.2307/1912791

Greeno, J. G., & Van De Sande, C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Educational Psychologist, 42*(1), 9-23. https://doi.org/10.1080/00461520709336915

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*(9), 444-454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications, 1*(2), 147-166.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4), 397-431. https://doi.org/10.2190/7mqv-x9uj-c7q3-nrag

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1998). Transcript analysis of computer-mediated conferences as a tool for testing constructivist and social-constructivist learning theories. *Proceedings of the 14th Annual Conference on Distance Teaching & Learning, Madison, WI, US* (pp. 139-146). https://eric.ed.gov/?id=ED422835

Gutierrez, F. J., Simmonds, J., Hitschfeld, N., Casanova, C., Sotomayor, C., & Peña-Araya, V. (2018). Assessing software development skills among K-6 learners in a project-based workshop with Scratch. *Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training, Gothenburg, Sweden* (pp. 98-107). https://doi.org/10.1145/3183377.3183396

Halverson, E. R., & Sheridan, K. (2014). The maker movement in education. *Harvard Educational Review, 84*(4), 495-504. https://doi.org/10.17763/haer.84.4.34j1g68140382063

Harel, I., & Papert, S. (1991). *Constructionism*. Ablex Publishing.

Harlan, M. A., Bruce, C., & Lupton, M. (2012). Teen content creators: Experiences of using information to learn. *Library Trends, 60*(3), 569-587.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*(1), 569-598. https://doi.org/10.1146/annurev.psych.093008.100416

Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *Journal of the Learning Sciences, 22*(3), 413-461. https://doi.org/10.1080/10508406.2013.799475

Hill, B. M., & Monroy-Hernández, A. (2017). A longitudinal dataset of five years of public activity in the Scratch online community. *Scientific Data, 4*(1), 170002. https://doi.org/10.1038/sdata.2017.2

Hoadley, C. (2010). Roles, design, and the nature of CSCL. *Computers in Human Behavior, 26*(4), 551-555. https://doi.org/10.1016/j.chb.2009.08.012

Hoffmann, C. P., Lutz, C., & Meckel, M. (2015). Content creation on the Internet: a social cognitive perspective on the participation divide. *Information, Communication & Society, 18*(6), 696-716. https://doi.org/10.1080/1369118x.2014.991343

Hu, S., Cao, Y., Zhang, J., Kong, W., Yang, K., Zhang, Y., & Li, X. (2012). More discussions for granger causality and new causality measures. *Cognitive Neurodynamics, 6*(1), 33-42. https://doi.org/10.1007/s11571-011-9175-8

Isen, A. M. (1999). On the relationship between affect and creative problem solving. In S. W. Russ (Ed.), *Affect, creative experience, and psychological adjustment* (pp. 3-18). Routledge.

Ito, M., Baumer, S., Bittanti, M., Boyd, D., Cody, R., Herr Stephenson, B., Horst, H. A., Lange, P. G., Mahendran, D., Martínez, K. Z., Pascoe, C. J., Perkel, D., Robinson, L., Sims, C., & Tripp, L. (2009). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. The MIT Press. https://doi.org/10.7551/mitpress/8402.001.0001

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction, 43*, 39-51. https://doi.org/10.1016/j.learninstruc.2016.01.005

Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 263-280. https://doi.org/10.1007/s11412-016-9238-2

Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive–situative divide and combining individual and social processes. *Educational Psychologist, 45*(1), 15-27. https://doi.org/10.1080/00461520903433539

Jeong, H., Cress, U., Moskaliuk, J., & Kimmerle, J. (2017). Joint interactions in large online knowledge communities: The A3C framework. *International Journal of Computer-Supported Collaborative Learning, 12*(2), 133-151. https://doi.org/10.1007/s11412-017-9256-8

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist, 51*(2), 247-265. https://doi.org/10.1080/00461520.2016.1158654

Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *International Journal of Computer-Supported Collaborative Learning, 9*(3), 305-334. https://doi.org/10.1007/s11412-014-9198-3

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist, 31*(3-4), 191-206. https://doi.org/10.1080/00461520.1996.9653266

Kafai, Y. B. (1996). Learning design by making games. In Y. B. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking, and learning in a digital world*. Lawrence Erlbaum Assoc.

Kafai, Y. B., Fields, D. A., Roque, R., Burke, Q., & Monroy-Hernandez, A. (2012). Collaborative agency in youth online and offline creative production in Scratch. *Research and Practice in Technology Enhanced, 7*(2), 63-87.

Kafai, Y. B., & Peppler, K. A. (2012). Developing gaming fluencies with Scratch: Realizing game design as an artistic process. *Games, learning, and society: Learning and meaning in the digital age*, 355-380.

Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education, 97*, 116-128. https://doi.org/10.1016/j.compedu.2016.03.002

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123-1134. https://doi.org/10.1037/0003-066X.39.10.1123

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*(4), 664-687. https://doi.org/10.3102/00028312027004664

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory Into Practice, 41*(1), 33-39. https://doi.org/10.1207/s15430421tip4101_6

Koschmann, T. (2002). Dewey's contribution to the foundations of CSCL research. *Proceedings of the 4th International Conference on Computer Supported Collaborative Learning, Boulder, CO, US* (pp. 17-22). https://dl.acm.org/doi/10.5555/1658616.1658618

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*(3), 335-353. https://doi.org/10.1016/s0747-5632(02)00057-2

Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2004). Measuring perceived quality of social space in distributed learning groups. *Computers in Human Behavior, 20*(5), 607-632.

Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social aspects of CSCL environments: A research framework. *Educational Psychologist, 48*(4), 229-242. https://doi.org/10.1080/00461520.2012.750225

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*(3), 287-315. https://doi.org/10.1207/s1532690xci1503_1

Kutaka-Kennedy, J. (2015). A proposed model to increase creativity, collaboration and accountability in the online classroom. *International Journal of Information and Education Technology, 5*(11), 873.

Lave, J. (1991). Situating learning in communities of practice. In *Perspectives on socially shared cognition.* (pp. 63-82). American Psychological Association. https://doi.org/10.1037/10096-003

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* [doi:10.1017/CBO9780511815355]. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lenhart, A., Fallows, D., & Horrigan, J. B. (2004). *Content creation online*. Pew Research Center. https://www.pewresearch.org/internet/2004/02/29/content-creation-online-2/

Literat, I., & Glaveanu, V. P. (2018). Distributed creativity on the Internet: A Theoretical foundation for online creative participation. *International Journal of Communication, 12*. https://ijoc.org/index.php/ijoc/article/view/7621

Livingstone, S. (2008). Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media & Society, 10*(3), 393-411. https://doi.org/10.1177/1461444808089415

Livingstone, S. (2010). Digital learning and participation among youth: Critical reflections on future research priorities. *International Journal of Learning and Media, 2*(2-3), 1-13.

Lowenthal, P. R. (2010). The evolution and influence of social presence theory on online learning. In T. Kidd (Ed.), *Online Education and Adult Learning: New Frontiers for Teaching Practices* (pp. 124-139). IGI Global. https://doi.org/10.4018/978-1-60566-830-7.ch010

Lu, J., Chiu, M. M., & Law, N. W. (2011). Collaborative argumentation and justifications: A statistical discourse analysis of online discussions. *Computers in Human Behavior, 27*(2), 946-955. https://doi.org/10.1016/j.chb.2010.11.021

Lüdecke, D. (2020). sjPlot: Data visualization for statistics in social science (R Package version 2.8.4). https://CRAN.R-project.org/package=sjPlot

Lutz, C., Hoffmann, C. P., & Meckel, M. (2014). Beyond just politics: A systematic literature review of online participation. *First Monday, 19*(7), 1-36.

Malinen, S. (2015). Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in Human Behavior, 46*, 228-238. https://doi.org/10.1016/j.chb.2015.01.004

Mamykina, L., Candy, L., & Edmonds, E. (2002). Collaborative creativity. *Communications of the ACM, 45*(10), 96–99. https://doi.org/10.1145/570907.570940

Marquart, C. L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., & Shaffer, D. W. (2019). rENA: Epistemic Network Analysis (R package version 0.2.0.1). https://cran.r-project.org/web/packages/rENA

Martin, L. L., Ward, D. W., Achee, J. W., & Wyer, R. S. (1993). Mood as input: People have to interpret the motivational implications of their moods. *Journal of Personality and Social Psychology, 64*(3), 317-326. https://doi.org/10.1037/0022-3514.64.3.317

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology, 14*(1), 6-23.

Mead, G. H. (1934). *Mind, self and society: From the standpoint of a social behaviorist*. The University of Chicago Press.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10*(2), 151-177. https://doi.org/10.1207/s15516709cog1002_2

Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning. In *The Cambridge handbook of the learning sciences, 2nd ed.* (pp. 418-438). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.026

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning, 9*(2), 137-160. https://doi.org/10.1007/s11409-013-9105-8

Monroy-Hernández, A. (2009). Designing a website for creative learning [Poster paper]. *Proceedings of the WebSci09 Conference: Society On-Line, Athens, Greece.*

Mumford, M. D. (2003, 2003/07/01). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal, 15*(2-3), 107-120. https://doi.org/10.1080/10400419.2003.9651403

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface, 14*(134), 20170213. https://doi.org/doi:10.1098/rsif.2017.0213

Organisation for Economic Cooperation and Development. (2007). *Participative web and user-created content*. https://doi.org/10.1787/9789264037472-en

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49*(1), 345-375. https://doi.org/10.1146/annurev.psych.49.1.345

Papert, S. (1993). *Mindstorms: Children, computers, and powerful ideas* (2nd ed.). Basic Books, Inc.

Peppler, K. A., & Solomou, M. (2011). Building creativity: Collaborative learning and creativity in social media environments. *On the Horizon, 19*(1), 13-23. https://doi.org/10.1108/10748121111107672

Pew Research Center. (2018). *Teens, social media & technology 2018*. www.pewresearch.org

Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software, 27*(4). http://www.jstatsoft.org/v27/i04/

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Viking Press.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. University of Chicago Press.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. Basic Books.

Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology, 20*(5), 347-356. https://doi.org/10.1080/01449290110084683

Preece, J., Abras, C., & Maloney-Krichmar, D. (2004). Designing and evaluating online communities: Research speaks to emerging practice. *International Journal of Web Based Communities, 1*(1), 2–18. https://doi.org/10.1504/ijwbc.2004.004795

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239-257. https://doi.org/10.1007/s11412-009-9070-z

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. *Communications of the ACM, 52*(11), 60–67. https://doi.org/10.1145/1592761.1592779

Reychav, I., & Wu, D. (2015). Mobile collaborative learning: The role of individual learning in groups through text and video content delivery in tablets. *Computers in Human Behavior, 50*, 520-534. https://doi.org/10.1016/j.chb.2015.04.019

Reynolds, R. (2016). Defining, designing for, and measuring “social constructivist digital literacy” development in learners: A proposed framework. *Educational Technology Research and Development, 64*(4), 735-762. https://doi.org/10.1007/s11423-015-9423-4

Richard, G. T., & Kafai, Y. B. (2016). Blind spots in youth DIY programming: Examining diversity in creators, content, and comments within the Scratch online community. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, US* (pp. 1473-1485). https://doi.org/10.1145/2858036.2858590

Robinson, K. (2011). *Out of our minds: Learning to be creative*. John Wiley & Sons.

Rogoff, B. (1998). Cognition as a collaborative process. In *Handbook of child psychology: Volume 2: Cognition, perception, and language.* (pp. 679-744). John Wiley & Sons Inc.

Rolim, V., Ferreira, R., Lins, R. D., & Găsević, D. (2019). A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education, 42*, 53-65. https://doi.org/10.1016/j.iheduc.2019.05.001

Roque, R., Dasgupta, S., & Costanza-Chock, S. (2016). Children's civic engagement in the scratch online community. *Social Sciences, 5*(4), 55.

Roque, R., Rusk, N., & Blanton, A. (2013). Youth roles and leadership in an online creative community. *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning, Madison, WI, US* (pp. 399-405). https://repository.isls.org//handle/1/1932

Roque, R., Rusk, N., & Resnick, M. (2016). Supporting diverse and creative collaboration in the Scratch online community. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 241-256). Springer International Publishing. https://doi.org/10.1007/978-3-319-13536-6_12

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences, 2*(3), 235-276. https://doi.org/10.1207/s15327809jls0203_1

Roschelle, J., Tatar, D., Chaudhury, S. R., Dimitriadis, Y., Patton, C., & DiGiano, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. *Computer, 40*(9), 42-48. https://doi.org/10.1109/MC.2007.321

Rovai, A. P. (2002). Building sense of community at a distance. *The International Review of Research in Open and Distributed Learning, 3*(1). https://doi.org/10.19173/irrodl.v3i1.79

Ruis, A. R., Siebert-Evenstone, A., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking. *Proceedings of the 13th International Conference on Computer Supported Collaborative Learning, Lyon, France* (pp. 136-143). https://repository.isls.org//handle/1/4395

Runco, M. A. (2010). Divergent thinking, creativity, and ideation. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 413-446). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.026

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92-96. https://doi.org/10.1080/10400419.2012.650092

Sarmiento, J. W., & Stahl, G. (2007). Group creativity in virtual math teams: Interactional mechanisms for referencing, remembering and bridging. *Proceedings of the 6th ACM Conference on Creativity and Cognition, Washington, DC, US* (pp. 37-44). https://doi.org/10.1145/1254960.1254966

Sawyer, R. K. (2003). *Group creativity: Music, theater, collaboration*. Lawrence Erlbaum Associates Publishers.

Sawyer, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of music, 34*(2), 148-165.

Sawyer, R. K., & DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts, 3*(2), 81-92. https://doi.org/10.1037/a0013282

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9-45.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media, 1*(2).

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 175-187). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.015

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics, 4*(3), 123-139.

Simonton, D. K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist, 55*(1), 151-158. https://doi.org/10.1037/0003-066X.55.1.151

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*(3), 273-307. https://doi.org/10.1007/s11412-015-9218-y

Stahl, G. (2000). A model of collaborative knowledge-building. *Proceedings of the Fourth International Conference of the Learning Sciences, Ann Arbor, MI, US* (pp. 70-77). http://www.umich.edu/~icls/proceedings/abstracts/ab70.html

Stahl, G. (2004). Building collaborative knowing. In J.-W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *What we know about CSCL: And implementing it in higher education* (pp. 53-85). Springer. https://doi.org/10.1007/1-4020-7921-4_3

Strijbos, J. W., & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. *Computers in Human Behavior, 26*(4), 495-505. https://doi.org/10.1016/j.chb.2009.08.014

Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*(4), 491–494. https://doi.org/10.1016/j.chb.2009.08.006

Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior, 38*, 110-117. https://doi.org/10.1016/j.chb.2014.05.022

Suthers, D. D. (2012). Computer-supported collaborative learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 719-722). Springer. https://doi.org/10.1007/978-1-4419-1428-6_389

Tan, J. P.-L., Caleon, I. S., Jonathan, C. R., & Koh, E. (2014). A dialogic framework for assessing collective creativity in computer-supported collaborative problem-solving tasks. *Research & Practice in Technology Enhanced Learning, 9*(3), 411-437.

Terras, M. M., Ramsay, J., & Boyle, E. A. (2015). Digital media production and identity: Insights from a psychological perspective. *E-Learning and Digital Media, 12*(2), 128-146. https://doi.org/10.1177/2042753014568179

Torrance, E. P. (1965). Scientific views of creativity and factors affecting its growth. *Daedalus, 94*(3), 663-681.

Tu, C.-H. (2000). On-line learning migration: From social learning theory to social presence theory in a CMC environment. *Journal of Network and Computer Applications, 23*(1), 27-37. https://doi.org/10.1006/jnca.1999.0099

Turvey, K. (2006). Towards deeper learning through creativity within online communities in primary education. *Computers & Education, 46*(3), 309-321. https://doi.org/10.1016/j.compedu.2005.11.004

United Nations Educational, Scientific, and Cultural Organisation. (2013). *Media and information literacy: Policy and strategy guidelines*. https://unesdoc.unesco.org/ark:/48223/pf0000225606

Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media effects: Theory and research. *Annual Review of Psychology, 67*(1), 315-338. https://doi.org/10.1146/annurev-psych-122414-033608

Velasquez, N. F., Fields, D. A., Olsen, D., Martin, T., Shepherd, M. C., Strommer, A., & Kafai, Y. B. (2014). Novice programmers talking about projects: What automated text analysis reveals about online Scratch users' comments. *Proceedings of the 47th Hawaii International Conference on System Sciences, Waikoloa, HI, US* (pp. 1635-1644). https://doi.org/10.1109/HICSS.2014.209

Vrasidas, C., & McIsaac, M. S. (1999). Factors influencing interaction in an online course. *American Journal of Distance Education, 13*(3), 22-36. https://doi.org/10.1080/08923649909527033

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Vygotsky, L. S. (1994). *The Vygotsky reader*. Basil Blackwell.

Vygotsky, L. S. (2004). Imagination and creativity in childhood. *Journal of Russian & East European Psychology, 42*(1), 7-97. https://doi.org/10.1080/10610405.2004.11059210

Wagner, E. D. (1994). In support of a functional definition of interaction. *American Journal of Distance Education, 8*(2), 6-29. https://doi.org/10.1080/08923649409526852

Wallas, G. (1926). *The art of thought*. Harcourt, Brace and Co.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research, 19*(1), 52-90. https://doi.org/10.1177/009365092019001003

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3-43. https://doi.org/10.1177/009365096023001001

Webb, N. M. (1987). Peer interaction and learning with computers in small groups. *Computers in Human Behavior, 3*(3), 193-209. https://doi.org/10.1016/0747-5632(87)90023-9

Wheeler, S., Yeomans, P., & Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology, 39*(6), 987-995. https://doi.org/10.1111/j.1467-8535.2007.00799.x

Wise, A. F., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 445-470. https://doi.org/10.1007/s11412-011-9120-1

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A social constructivist interpretation. *The Internet and Higher Education, 10*(1), 15-25. https://doi.org/10.1016/j.iheduc.2006.10.005

APPENDIX

IRB Notice of Approval for Human Research

PEPPERDINE

Pepperdine University
24255 Pacific Coast Highway
Malibu, CA 90263
TEL: 310-506-4000

NOTICE OF APPROVAL FOR HUMAN RESEARCH

Date: May 28, 2020

Protocol Investigator Name: Seung Lee

Protocol #: 20-04-1324

Project Title: Interaction, collaboration and content creation in informal online learning environments: Multidimensional analyses of longitudinal data from the Scratch coding community

School: Graduate School of Education and Psychology

Dear Seung Lee:

Thank you for submitting your application for exempt review to Pepperdine University's Institutional Review Board (IRB). We appreciate the work you have done on your proposal. The IRB has reviewed your submitted IRB application and all ancillary materials. Upon review, the IRB has determined that the above entitled project meets the requirements for exemption under the federal regulations 45 CFR 46.101 that govern the protections of human subjects.

Your research must be conducted according to the proposal that was submitted to the IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an amendment to the IRB. Since your study falls under exemption, there is no requirement for continuing IRB review of your project. Please be aware that changes to your protocol may prevent the research from qualifying for exemption from 45 CFR 46.101 and require submission of a new IRB application or other materials to the IRB.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite the best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the IRB as soon as possible. We will ask for a complete written explanation of the event and your written response. Other actions also may be required depending on the nature of the event. Details regarding the timeframe in which adverse events must be reported to the IRB and documenting the adverse event can be found in the *Pepperdine University Protection of Human Participants in Research: Policies and Procedures Manual* at community.pepperdine.edu/irb.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval. Should you have additional questions or require clarification of the contents of this letter, please contact the IRB Office. On behalf of the IRB, I wish you success in this scholarly pursuit.

Sincerely,

Judy Ho, Ph.D., IRB Chair

cc: Mrs. Katy Carr, Assistant Provost for Research