Changes in Positioning:
An Alternative Perspective on Learning in Massive Open Online Courses

by

Yi Cui

M.A. (Applied & Foreign Linguistics), Beijing Foreign Studies University, 2006
B.A. (English), Northeast Normal University, 2001

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Educational Technology & Learning Design Program
Faculty of Education

© Yi Cui 2019
SIMON FRASER UNIVERSITY
Summer 2019

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name:	**Yi Cui**
Degree:	**Doctor of Philosophy**
Title:	**Changes in Positioning: An Alternative Perspective on Learning in Massive Open Online Courses**
Examining Committee:	**Chair: Laura D'Amico** Adjunct Professor
	Alyssa Wise Senior Supervisor Associate Professor
	Phil Winne Supervisor Professor
	Marek Hatala Internal Examiner Professor School of Interactive Art & Technology
	Sean Goggins External Examiner Associate Professor College of Engineering University of Missouri
Date Defended/Approved:	August 7, 2019

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Discussion forums are widely provided in Massive Open Online Courses for learners to interact and exchange learning support. Developing one's forum participation pattern to interact substantively about the course content can be an important form of MOOC learning. This mixed-method study examines learners' forum participation patterns as positions that can be characterized by characteristics related to their contributions and social relations. The series of positions that a learner takes over time form their position trajectory. This study analyzed learners' positions in the beginning, middle, and end periods in a statistics MOOC and a writing MOOC. Through performing content analysis and social network analysis on the discussions, five kinds of participation characteristics were extracted for each learner: quantity of content-related contributions, input seeking and providing activities, deep consideration of the discussion content, connectedness in the social network, and strength of social connections. Positions in each time period were identified through clustering groups of learners who had similar participation characteristics. The identified positions fell into six primary types: enthusiastic central providers, enthusiastic central reciprocators, moderate providers, moderate reciprocators, moderate deep thinkers, and minimal peripheral contributors. The forum at any time point usually contained a small group of enthusiastic contributors, a big proportion of moderate contributors, and a majority of minimal contributors. This study further examined the position trajectories for learners who participated in multiple periods, and performed case studies on learners who showed the frequent trajectories. In both MOOCs only 17% of the multi-period learners showed constructive development in participation pattern and changes in language and participation focus that suggested identity development. This study is the first effort among MOOC research to examine changes in participation pattern using multiple contribution and social characteristics. The identified positions provided a critical ground for studying content-related interaction and learning community in MOOC forums. The moderate contributor groups are under-researched in the MOOC literature and promising for expanding the understanding of MOOC learners. The findings in this study also demonstrate the usefulness of the position perspective for understanding MOOC learning and both the need and potential avenues to help MOOC learners become more competent forum participants.

Keywords: Massive Open Online Courses; online discussion; forum participation; role; learning analytics; mixed methods research

Acknowledgements

Foremost, I would like to thank my senior supervisor, Dr. Alyssa Friend Wise for her motivation, patience, and support. It has been a real privilege to learn from a supervisor who is not only so knowledgeable and rigorous as a researcher but also so genuine and fun as a person. Thank you for helping me to grow academically and personally. You are a best role model.

I am also grateful to my supervisor, Dr. Phil Winne who offered me invaluable guidance and advice. His insightful comments and suggestions greatly helped me to make this paper stronger. I am most lucky to have been inspired by his erudition, intellectual curiosity, and great sense of humor.

A sincere thank you to Dr. Marek Hatala, Dr. Sean Goggins, and Dr. Laura D’Amico for being my examining committee. Their thought-provoking questions and comments made the defense an enriching and unforgettable experience for me.

I’m thankful to my colleagues and friends at Simon Fraser University and New York University, especially Christopher Jeschelnik and Yeonji Jung. Their generous help, valuable suggestions, and warm encouragement are truly appreciated.

I gratefully acknowledge the funding and working opportunities provided by Simon Fraser University and New York University during my doctoral study. I also want to thank the Center for Advanced Research through Online Learning, Stanford University for sharing the MOOC data that made this study possible. A special thank you to Kathy Mirzaei for her kind help to facilitate the data sharing process.

Finally, my wholehearted thanks go to my dear family and friends. To my parents, my brother, and my lovely niece Tiantian, thank you for your unconditional love, and for always supporting me to pursuit my dreams and taking pride in what I do. To Bonnie and Jessica, thank you for your heartful care and encouragement. To A’Mao, BB, Fei, and Yezi, thank you for always being there for me and making hard times easier. To Ang, thank you for your continuous and unparalleled love, support, and understanding. This journey would not have been so perfect without you. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction and Problem Statement

Massive Open Online Courses (MOOCs) are large-scale learning environments characterized by usually freely accessible learning materials and large numbers of participants (Deboer, Ho, Stump, & Breslow, 2014). MOOC is depicted as offering access to affordable education opportunities, especially to the underprivileged (Breslow et al., 2013). As MOOCs pool learners with diverse academic, professional, and cultural backgrounds from all over the world, they are also expected to provide the experience of learning in a world classroom (Dillahunt, Wang, & Teasley, 2014). In the past several years, MOOCs have attracted unprecedented public involvement and enormous governmental and institutional input (Gaebel, 2013; Hollands & Tirthali, 2014; Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2015). Millions of learners invest time and efforts to reap the learning opportunities brought forth by MOOCs. It is thus worthwhile to study how learning occurs and can be supported in these environments (Kovanovic, 2017; Ross, Sinclair, Knox, Bayne, & Macleod, 2014).

Research on MOOC learning usually adopts course performance measures (e.g., grade and certificate) as metrics for learning and examines their connections with course participation to identify useful factors for learning improvement (Almeda, 2018; Houston, Brady, Narasimham, & Fisher, 2017). Although course performance measures are straightforward and widely used in MOOCs for learning assessment, its insufficiency as the sole perspective to MOOC learning has been noted (Wise & Cui, 2018b; Wang, 2017). For one thing, course performance does not well capture development in learners' identity in relation to peers and the community of practice (Nelimarkka & Vihavainen, 2015; Wang, 2017). Moreover, it is not viable for learners without performance-orientation (e.g., auditors who do not take course assessments, see Kizilcec, Piech, and Schneider, 2013) and MOOCs without formal performance assessments.

A small number of MOOC studies have explored alternative perspectives on MOOC learning. For instance, some research investigated learners' post-course activities for indicators of learning, such as career advance in course-related fields

(Wang, 2017) and participation in course-related online open communities (Chen, Davis, Lin, Hauff, & Houben, 2016). These perspectives may set overly high standards for learning for a large proportion of MOOC takers.

MOOC research has also delved into during-course social interactions to investigate learning. Research on connectivist MOOCs (cMOOCs) has looked into social media activities. For instance, in a study conducted on a cMOOC *Connectivism and Connective Knowledge (CCK08)*, Joksimović, Dowell et al. (2015) measured learning outcomes as connections that learners made on Facebook, Twitter, and blogs. In another study conducted on the same course, Joksimović, Kovanović et al. (2015) investigated learning by examining (1) over time changes in the topics discussed on social media by course takers and (2) the alignment between discussed topics and topics introduced in learning materials. As social media are used as primary venues for connectivist learning (Danial, 2012), activities on these platforms can be useful for understanding learning in cMOOCs.

Research on social interaction in xMOOCs usually focuses on activities in course forums. Forum discussion can be useful for understanding xMOOC learning for several reasons. First, discussion forums are hosted internally on course platforms and easily accessible to all course takers. It is estimated that a substantial proportion of committed learners participate in forums through varied forms, such as posting and reading (Wise & Cui, 2018b). Second, forum discussion provides opportunities for responsive interpersonal interaction which is valued by learners as a useful source of learning support. Breslow et al. (2013) found that discussion forum was the most frequently used learning resource when MOOC learners worked on assignments. Third, fine grained clickstream data and text data generated from forum discussion provide the opportunity to study MOOC learning in process, which can complement the end-of-course perspectives. A small number of MOOC studies have probed discussion forums for indicators of learning but have not yet established a viable conceptual perspective. For one thing, they generally perceived discussion content as approximation for learning outcomes without attending to the process of participation. Furthermore, they mostly assessed discussion content from the social constructivist perspective, while most xMOOCs are designed and administered based on the instructivist approach (Kellogg, Booth, & Oliver, 2014; Rodriguez, 2012; Tawfik et al., 2017).

Positioning theory can be useful for studying MOOC learning from a participation perspective. According to positioning theory, learners take positions dynamically when taking part in conversations; positioning is the result of social negotiation among conversation participants (Davies & Harré, 1990; Dennen, 2011; Harré & van Langenhove, 1998). In MOOC contexts, learning can be conceptualized as developing one's participation pattern to interact substantively in a content domain (Barab & Duffy, 2000; Sfard, 1998). When participating in discussions, MOOC learners take positions in relation to other participants and can change their positions over time. Learning is indicated by constructive development in their participation patterns.

The MOOC literature indicates characteristics of learners' forum contributions and the relations they develop with others can be useful for understanding positions, but the two perspectives have rarely been combined. Moreover, there is not a refined content analysis model for assessing position-related contribution characteristics. This study investigates learning in MOOC discussion forums from the position perspective by conducting a mixed method study on two courses. The goal is to (1) develop a theoretical framework that encompasses contribution characteristics and social relationships for conceptualizing positions, (2) compile a content analysis rubric for position-related contribution characteristics, and (3) identify positions and position trajectories to improve the understanding of MOOC learning and forum interaction.

Chapter 2 of this dissertation first introduces MOOCs, then reviews prior work on MOOC learning and positions in MOOC forums. Chapter 3 raises the research questions and explains important decisions in research design. After research methods are introduced in Chapter 4, the analysis results are reported in Chapter 5. Finally, Chapter 6 discusses the research findings, the implications for MOOC research and practice, and offers suggestions for future work.

Chapter 2.

Literature Review

This chapter first introduces MOOCs and discusses the importance of understanding MOOC learning. Then it reviews prior perspectives adopted in MOOC research for understanding learning. Finally, a position perspective is proposed as an alternative perspective on MOOC learning and prior work useful for applying this perspective in MOOC contexts is reviewed.

2.1. MOOCs

2.1.1. cMOOC and xMOOC

MOOCs first appeared in 2008 when Athabasca University made an online course Connectivism and Connective Knowledge (CCK08) freely open to not-for-credit participation and attracted more than 2,200 registrations. This course adopted a connectivist approach to learning and instruction; learners participated by creating content related to course topics and connecting with each other through a variety of web tools, such as personal blogs, online discussion forums, and synchronous online meetings; the instructors facilitated learning by curating and disseminating participant-generated content through RSS feeds. CCK08 and its successors offered in this format were later labelled as cMOOCs, i.e. connectivist MOOCs (Daniel, 2012). The innovative pedagogy and massive scale of participation previously unseen in online learning attracted attention from researchers and course designers, but the impact largely remained in academia (Morrison, 2013).

In 2011, Stanford University offered a free artificial intelligence course online and received over 160,000 registrations (Daniel, 2012). Shortly, other elite North American universities like MIT and Harvard also started to offer free open online courses. In contrast to cMOOCs, these courses and most of the MOOCs offered since then are characterized by a behaviorist / cognitivist approach to learning and teaching; the courses follow curricula similar to those for traditional on-campus courses; learners participate by using pre-prepared learning materials (e.g., video lectures and text-based reading materials), participating in discussion forums, doing assignments, and taking quizzes and exams. Due to the size of learner population, learning assessment primarily relies on auto-grading and peer review. Courses offered in this format are labelled as xMOOCs (Daniel, 2012).

2.1.2. MOOC hype and recent trends

xMOOCs quickly ignited the interest of mainstream media and the public because of the participating universities' fame, enormous number of participants, and the notion of free access to top quality education. MOOC was portrayed as an educational innovation with the power to make quality education more accessible, drive down the rocketing cost of higher education, and force traditional mortar-and-brick institutions to reform to stay relevant (Horn & Christensen, 2013). As MOOC quickly became a media buzzword, many universities joined to offer courses for varied motivations, such as expanding the institution's influence, exploring new arena for learning and teaching, and the fear of being left behind (Daniel, 2012).

In the next several years, MOOCs proliferated not only in North America but also other parts of the world, such as Europe and Asia. The number of participating institutions, courses, and registered learners all grew with strong momentum. At the end of 2017, more than 9,400 courses were provided by over 800 universities; the number of learners exceeded 81 million (Shah, 2017). Along with the growth in numbers, the development of MOOCs showed several noteworthy trends. First, course platforms started to emphasize monetizing MOOCs. Although most platforms still allow free auditing for some courses, they are becoming increasingly focused on making profit from course credentials, specializations, and degree programs (Shah, 2016). Second, the subjects being offered have shifted partly due to adjustment in the business model. The earliest xMOOCs were mostly computer science and engineering courses, and were later outnumbered by humanities courses in 2013; more recently partly due to the platforms' drive for profitability, courses that focus on applied skills in business and technology are gaining ground (Shah, 2014, 2016, 2017).

2.1.3. MOOC controversy and the need to understand learning in MOOCs

Ever since the beginning of the MOOC hype, there have been heated debates over MOOC's value for education and learning. In contrast to course providers' enthusiasm, some universities refrained from joining the club due to concerns over incompatibility of the MOOC format with their teaching and learning philosophy (Kolowich, 2013). In addition, when some universities attempted to substitute domestically developed course content with MOOC content designed by other universities, they were rejected and criticized by faculty members who contended that those courses did not sufficiently address the specific learning characteristics of their students (Parry, 2013).

More recently, findings from empirical research indicated MOOCs' value and potential for facilitating education and providing educational opportunities. First, research on the user population revealed MOOCs are providing learning opportunities for people with diverse motivations. For instance, Macleod, Haywood, Woodgate, and Alkhatnai (2015) found many full-time employees took MOOCs for career development, either for learning new skills or gaining credentials. For these learners, self-paced MOOCs and courses that set flexible window period for assignments and exams allow them to arrange their study more flexibly. Furthermore, Dillahunt et al. (2014) found many users with limited economic means took MOOCs as a trial before enrolling in formal college courses and were more likely to pass MOOCs than others. In addition, some universities offer graduate level MOOCs and allow learners to transfer their MOOC credentials to the universities' on-campus Master programs so that they can get the full degree faster (MicroMasters). In this sense, MOOCs offer value in providing flexible and accessible learning opportunities to under-served population. Moreover, another important motivation for taking MOOCs is to enjoy learning new things, reported by participants of diverse age groups and educational backgrounds (Macleod et al., 2015). By providing learners the access to courses on an array of subjects offered by diverse institutions, MOOCs can contribute to life-long learning, which is a noble mission of education (Aspin & Chapman, 2000; Sharples, 2000).

Second, MOOCs provide unprecedented opportunities for understanding learning in online environments (Reich, 2015). Research on learning in conventional online environments has developed insightful understanding and theories, but mostly focuses on small scale contexts. The scale and diversity of MOOC learner population provide conditions for research design and analysis methods that were often not feasible in the past. For instance, the enormous number of participants in the same course can provide abundant sample size for experiments that compare multiple approaches or tools in parallel, which is not always possible in conventional online learning environment. Moreover, the diversity of learners who participate in the same learning environment provides the chance to account for the influence of different factors on learning, such as professional experience, motivation, and cultural background (Gillani & Eynon, 2014).

With that said, empirical studies also presented reasons for concerns, with a major one being low completion rate and less than satisfactory learning experience which has been a prevalent problem since the first MOOCs (Hew & Cheung, 2014; Khalil & Ebner, 2013) and remains a focal area in MOOC research and design (Gasevic, Kovanovic, Joksimovic, & Siemens, 2014; Veletsianos & Shepherdson, 2016). Although it has been argued that a large proportion of people who register for a MOOC do not intend to finish the course, thus render completion rate an unfitting criterion for evaluating the quality of MOOCs (Deboer et al., 2014), MOOC literature on learner motivation, behavior, and perceived learning experience shows learners who intend to finish the course also often drop out due to learning difficulty and compromised learning experience (Belanger & Thornton, 2013; Hone & El Said, 2016; Yang, Sinha, Adamson, & Rosé, 2013). These problems call for further efforts to understand how learning occurs and can be better supported in these large-scale learning environments.

2.2. Understanding learning in MOOCs

Research on MOOC learning varies with respect to what is learning, what indicates the occurrence of learning, and how to measure learning. This section reviews perspectives on learning adopted in prior work.

2.2.1. The course performance perspective (measure learning at the end of the course)

Most MOOC studies measure learning outcomes with performance metrics used by the course / instructor, such as grades and certificates (Barba, Kennedy, & Ainley, 2016; Gillani & Eynon, 2014; Houston et al., 2017). Course performance seems to be the most straightforward approach to measuring learning. Research conducted from this perspective has shed light on the connections between course performance and participation. For instance, Gillani and Eynon (2014) found that forum contributors received higher final grades than non-contributors in a business MOOC. Houston et al. (2017) found that the quantity of forum contribution was useful for predicting final grades in a MOOC on programming and two offerings of a MOOC on innovation. Wise and Cui (2018b) found that learners who posted in discussions forums in a statistics MOOC were more likely to pass the course, and that the quantity of learning-related contributions was the most useful predictor for final grades in comparison to the quantity of learning-unrelated contributions and undistinguished contributions.

Useful as it can be, limitations of the course performance perspective are also noteworthy. First, some forms of learning outcomes are not well captured by this perspective. For instance, for alumni MOOC learners who engage substantially in networking with peers and assisting others to learn (Nelimarkka & Vihavainen, 2015), the development in their identity and their relationship with peers can be a meaningful form of learning (Wenger, 1998), but not one well captured by grades and certificates. Second, the course performance perspective is viable in contexts where formal assessment is considered important, but many MOOC learners are not performance-oriented and do not take assessment seriously (Kizilcec et al., 2013). Thus grades and certificates are not suitable metrics for their learning. In addition, this perspective is not applicable in contexts without formal assessments, such as cMOOCs (Daniel, 2014).

2.2.2. The post-course perspectives (measure learning after the course)

A small number of MOOC studies examined learners' post-course activities related to the MOOC they had taken for signs of learning. For instance, Wang (2017) used learners' post-course career advance as an alternative criterion for learning success in a MOOC on big data. Career advancers were defined as those who joined academic societies or submitted papers to conferences related to the course within two years after taking the course. The career advancers were found to have used learning resources more frequently during the course and gained better course assessment results than the non-advancers. In another exploratory study, Chen et al. (2016) looked for changes in MOOC learners' Github and Stackflow activities after they had taken a programming MOOC. It was found that the course takers asked less questions on Stackflow after taking the course; the quantity of their programming activities on Github peaked during the course. But the mechanism for connections between these changes and learning was not established in the study.

The post-course perspectives can be potentially useful for assessing MOOC success and understanding learning for the subpopulation that continue to participate in post course activities. However, this may not align with most MOOC takers' learning goals and thus can pose an overly high standard for what is considered learning.

2.2.3. The social interaction perspectives (measure learning during the course)

Some MOOC studies examined during-course social interaction for signs of learning. Social interaction is an important element for the quality of online learning with its positive impact on motivation, retention and learning performance documented in the literature (Coomey & Stephenson, 2001; Moore, 1989; Swan, 2002). MOOC research has investigated social interaction in social media and discussion forums.

Learning in social media

Multiple types of social media are used by MOOC learners for social interaction, such as Facebook, Twitter, and Google Hangout. In particular in cMOOC contexts these tools are used as the primary venues for connectivist learning (Daniel, 2012). Learners participate in learning by contributing content to and making connections on social media. Some MOOC studies have looked into these activities to investigate learning. For instance, Joksimović, Dowell et al. (2015) conceptualized learning outcomes in a connectivist MOOC on connectivism as accumulated social capital (i.e., connections that learners established via interaction) on Twitter, blog, and Facebook. They compared characteristics of contributions made by learners who gained different amount of social capital. It was found that learners with more social capital used more narrative and formal language, presented well-connected ideas, and discussed new ideas frequently. In another study on the same course, Joksimović, Kovanović et al. (2015) investigated learning by examining (1) over time changes in topics that learners discussed on social media and (2) alignment between topics emerged in discussions and topics introduced in recommended readings. It was found that throughout the course, learners tended to continue having conversations on fundamental concepts introduced early in the course, rather than taking up new topics in the recommended readings. Social interactions in social media can be a rich source for understanding learning in cMOOCs given their special roles in connectivist contexts.

Learning in discussion forum

Discussion forums are widely provided on xMOOC course platforms as venues for responsive social interaction. They are used for varied purposes, such as questions and answers about course content, discussion of complex concepts, sharing resources, raising / addressing technical and logistical questions, and connecting with others (Stump, DeBoer, Whttinghill, & Breslow, 2013; Wise, Cui, Jin, & Vytasek, 2017). Participation in forum discussion is usually not required but encouraged. In some courses, instructional team members also participate in forum interaction.

Forum discussion can be useful for understanding MOOC learning for several reasons. First, discussion forums are hosted internally on course platforms and easily accessible to course takers. It is estimated that a substantial proportion of committed learners participate in forums through varied forms, such as posting and reading (Wise & Cui, 2018b). Second, forum discussion provides opportunities for responsive interpersonal interaction, which is valued by learners as a useful source of learning support. Breslow et al. (2013) found that discussion forum was the most frequently used learning resource when MOOC learners worked on assignments. Third, fine grained clickstream data and text data generated from forum discussion provide the opportunity to study MOOC learning in process, which can complement the end-of-course perspectives.

A small number of MOOC studies have probed learning in discussion forum, but have not yet established a viable conceptual perspective due to two limitations. First, prior studies generally perceived discussion content as approximation for learning outcomes, without attending to how students participate in discussion processes. For instance, Wong, Pursel, Divinsky, and Jansen (2016) assessed discussion content using Anderson and Krathwohl’s cognitive thinking taxonomy (2001) that consists of six levels: remembering, understanding, applying, analyzing, evaluating, and creating. Wong et al.’s (2016) study was conducted on a MOOC on art and two MOOCs on innovation.

Words related to cognitive levels were used as approximation for the level of cognitive thinking involved in discussion. It was found that cognitive activities generally increased over time in all courses while the increase in the art MOOC tended to be the most prominent; discussions involved all levels of cognitive thinking, with creating activities being the most common (20%-28%) and analyzing activities the least common (7%-16%) in all courses; the proportion of different levels tended to change over time. While the analysis of cognitive thinking characteristics provided understanding about the discussion content, Wong et al.'s approach did not account for how individual learners engaged in these activities and interacted with others.

Second, except for Wong et al. (2016), other studies mostly examined forum learning from the social constructivist perspective and assessed discussion content with the Interaction Analysis Model for Examining Social Construction of Knowledge (IAM, see Gunawardena, Lowe, & Anderson, 1997) and the cognitive presence model (Garrison, Anderson, & Archer, 1999; Garrison, Anderson, & Archer, 2001). Although these models are widely applied for assessing learning in text-based online discussion in small group collaboration and non-MOOC contexts (Gašević, Adesope, Joksimović, & Kovanović, 2015; Goggins, Galyen, Petakovic, & Laffey, 2016; Kovanović et al., 2016; Wise & Chiu, 2011), they are limited in comparability with the actual pedagogical and interaction contexts in most xMOOCs. For example, Kellogg et al. (2014) examined discussions in a MOOC on digital learning and a MOOC on teaching mathematics using the IAM models. They assessed discussions for five phases of knowledge construction: sharing information, exploring dissonance, negotiating meaning, testing and modifying, and summarizing and applying. It was found that the vast majority of discussions involved sharing information and exploring dissonance; few discussions moved beyond phase two. Tawfik et al. (2017) used the same model and reported similar findings for discussions in a chemistry MOOC: 96% of the discussions remained at phase one and only 4% reached phase two. A major problem with applying the knowledge construction models for assessing MOOC forum learning is that the majority of MOOCs adopt instructivist approaches to learning and are not designed to promote knowledge construction (Daniel, 2012). The guidance and facilitation that are considered to be indispensable for achieving advanced knowledge construction phases are absent in MOOC discussions (Kovanović, 2017), even in the small number of constructivist MOOCs (e.g., Kellogg et al., 2014; Tawfik et al., 2017). As a result, MOOC discussions rarely exceeded the basic phases of knowledge construction. Thus it is important to assess discussion content using models that are compatible with the pedagogical and interactional contexts in MOOCs.

2.3. A new perspective: Learning as changes in positioning in forum discussion

According to the participation perspective to learning, learning occurs in the process of participating in social activities; the way in which an individual participates in social processes determines and reflects their identity; learning outcomes are seen as changes in the way of participation (Barab & Duffy, 2000; Sfard, 1998). In the context of MOOC forum discussion, learning can be conceptualized as developing one’s way of participation to interact substantively in a content domain. Changes in a learner’s participation pattern that are constructive in respect to substantive interaction can be considered as indicators of learning.

2.3.1. Theoretical constructs for learning as participating in discussion: Role and position

The role concept is frequently used for characterizing participation patterns in group discussion (Cesareni, Cacciamani, & Fujita, 2016; De Wever, Van Keer, Schellens, & Valcke, 2009; Wise, Saghafian, & Padmanabhan, 2012). Roles are defined based on discussants’ similarities in attitude, motivation, function, and social relationship in forum interaction (Strijbos & De Laat, 2010). Roles in group discussion can be pre-defined and assigned (e.g., De Wever et al., 2009; Wise et al., 2012), or emergent and taken up spontaneously in the process of interpersonal interaction (Herrmann, Jahnke, & Loser, 2004; Strijbos & De Laat, 2010).

The position concept in positioning theory shares similarities with emergent role, and can be better adapted for studying changes in learners’ forum participation patterns. Originating from social psychology, positioning theory focuses on understanding and characterizing how conversation participants interact with and relate to each other; participants can take up different positions when coproducing the interactions, such as expert and novice, powerful and powerless, superior and inferior (Davies & Harre, 1990; Dennen, 2007; Harré & van Langenhove, 1998). Positions are similar to emergent roles in that they both are products of specific interaction contexts and processes, and associate with certain behavioral patterns and relationships (Davies & Harré, 1990; Dennen, 2007; Dowell, Nixon, & Graesser, 2018; Herrmann et al., 2004). Yet emergent roles in discussion forums are usually characterized as continuous and rigid constructs formed over (and often identified at the end of) interaction process; in contrast, positions are more dynamic and transitory states that discussion participants can take up and modify swiftly, accommodating to the need of identifying dynamic and fluid development in discussion participation (Harré & Van Langenhove, 1998). Thus this study proposes to use position as the conceptual construct for learners participation patterns while drawing upon the literature on emergent roles in MOOC discussions. Positioning is defined as the action of taking a position.

2.3.2. Understanding position and positioning based on contribution characteristics and social relationships

MOOC learner's position in forum interaction can be understood from two perspectives. Learners participate in discussions through making contributions. Learners' contributions can reveal their participation focus, interest, and approaches to learning and thus can be fundamental for understanding the positions they take (Davies & Harré, 1990). Social relationships that learners form with others can also be useful for understanding positions in that they can influence how learners take positions in forum interactions (Strijbos & Weinberger, 2010). For instance, compared to isolated learners, well-connected learners may be more motivated to seek help, as they can reach a bigger audience and thus have advantages in getting responses. Moreover, social contexts can also influence what position a learner inclines to take in interaction. For instance, Wise and Cui (2018a) found that a cohesive community in a statistics MOOC showed more evident social bonding signs than the less connected communities. It is possible that members in the cohesive community are more motivated to assist each other in learning.

Contribution characteristics and social relations have been examined separately but rarely combined in the MOOC literature on roles in discussion forums. The remainder of this section reviews prior works that are helpful for understanding position and positioning in MOOC discussion forums.

2.3.3. Prior work on contribution characteristics

Content analysis methods are widely used for understanding characteristics of learning and interaction processes in forum discussion (Henri, 1992). While many MOOC studies involve analyzing the content of discussion contributions, only a small number have done so for the purpose of characterizing roles.

Shortly after the first xMOOCs appeared in 2012, Stump et al. (2013) used a grounded theory method to analyze discussions in a MOOC on circuits and electronics. They developed a two-dimension content analysis framework for classifying posts based on (a) the discussion topic and (b) poster's roles in discussion (see Table 2.1). This study found help seekers and help givers were the primary roles in the forum. This study also found that in addition to the discussions related to the learning of circuits and electronics, there were also some discussions that involved topics not directly related to learning, such as technical issues and course policy. Subsequent MOOC studies confirmed this distinction and found learning-related and unrelated discussions not only differ in interaction purposes, but also involve distinct communication techniques and social relations patterns (Cui & Wise, 2015; Wise, Cui, & Jin, 2017; Wise & Cui, 2018a).

Table 2.1 Content analysis scheme for MOOC forum contributions (Stump et al., 2013)

	Code	Description
Topic of Post	Content	Posts specifically addressing circuits and electronics material
	Other coursework	Posts discussing courses other than circuits and electronics
	Social/affective	Posts addressing social, emotional, or community- building aspects of the class
	Course website/technology	Posts that addressed the online interface
	Course structure/policies	Posts regarding the course organization, guidelines, or requirements
	Other	Posts conveying anything not related to class content, other courses, social aspects of the class, course website or technology, or course requirements
	Missing data	Posts in which data had been censored by course staff (in these cases, particular thread numbers could not be located)
	Non-English	Posts written in other languages
Role of Poster	Help-seeker	Posts in which the poster asked for help, information, etc.
	Help-giver	Posts in which the poster provided help, insight, or information
	Other	Posts in which the poster was not explicitly seeking, declaring, or providing information, such as an opinion

These findings indicate the importance of differentiating learning-related and unrelated discussions when studying MOOC learning and interaction.

The distinction between learning-related and unrelated discussions was made in Liu, Kidziński, and Dillenbourg's (2016) research on roles in a programming MOOC. Discussion posts were first categorized into six learning-related categories (question, answer, clarification request, clarification, positive feedback, negative feedback) and an off-topic category, based on Sridhar, Getoor, and Walker's (2014) scheme for online debate stances. The content analysis process was facilitated by semi-automated classification using a classifier built with features extracted from 100 hand-coded posts (e.g., course-related keywords, post length and grammatical quality, number of question marks). Based on the classified posts, a contribution characteristics profile was built for each learner. Two roles were defined based only on the quantity of answers and questions in learners' posts: *answer seekers* who mainly asked questions (> 80% of postings were questions) and *answer providers* who mainly answered questions (> 80% of postings were answers). It was further found that *answer providers* achieved higher course grades than *answer seekers*. These roles are useful for understanding information exchange behaviors, but are inadequate for understanding positions in relation to learning. For instance, answers that provide factual information and those that discuss the applicability of concepts in different circumstances can reflect important differences in the answer providers' positions. Thus it is important that the content of contributions is assessed in relation to learning. This has not yet been done in the literature on roles and positions in MOOC forum discussion, but relevant work on roles in non-MOOC discussion can provide useful inspiration.

Zhu (1996) presented a scheme for categorizing discussion contributions for characterizing how learners participated in learning interaction (see Table 2.2). Adopting social constructivist and cognitive learning perspectives, Zhu used a two-dimension content analysis scheme that incorporated Hatano and Inagaki's (1991) theory of group interaction and Graesser and Person's (1994) theory of question analysis. The first dimension examines whether the message involves vertical interaction (concentrating on seeking "the answer" from more capable members) or horizontal interaction (concentrating on expressing ideas and contributing to constructing knowledge with each other). The second dimension examines in what ways learners contribute to discussions, such as asking questions, providing factual answers, and scaffolding (see Table 2.2).

Using this scheme, Zhu (1996) analyzed discussion contributions generated in two weeks in a graduate course on learning technologies. Based on contribution characteristics learners were characterized as *contributor*, *wanderer*, *seeker*, and *mentor*. *Contributors* included all learners who ever made any number and type of messages. *Wanders* were those who usually discussed teaching and learning in a general way and seldom addressed specific educational issues presented in weekly readings. They seemed to be lost in the readings and discussions, but still struggled and strived to understand the issues. *Seekers* were those who sought specific information to understand the issues. Finally, *mentors* were those who helped others to understand the readings as well as develop understanding of learning and teaching in a general way. This study found learners often played multiple roles in discussion.

Table 2.2 Content analysis scheme for online discussions (Zhu, 1996)

Interaction Types	Categories	Characteristics
Vertical	Type I Question	Ask for information or requesting an answer
Horizontal	Type II Question	Inquire, start a dialogue
Horizontal	Answer	Provide answers to information-seeking questions
Horizontal	Information sharing	Share information
Horizontal	Discussion	Elaborate, exchange, and express ideas or thoughts
Horizontal	Comment	Judgmental
Horizontal	Reflection	Evaluation, self-appraisal of learning
Horizontal	Scaffolding	Provide guidance and suggestions to others

Fahy, Crawford, and Ally (2001) modified Zhu's (1996) scheme and tested the reliability of the new scheme (see Table 2.3): *answer*, *information sharing* and *comment* were combined as *statement*; *discussion* and *reflection* were combined as *reflection*; *scaffolding* was split into *scaffolding and engaging*; a new category *references and authorities* was added. Three types of coding reliability were reported, including (a) intra-rater agreement for the primary coder (86% with a 10-day elapse), (b) inter-rater agreement for two pairs of coders (60% and 70% respectively, the two pairs coded the same data), and (c) inter-rater kappa for two pairs of coders (0.45 and 0.65 respectively, the two pairs coded the same data). Using this tool, Fahy et al. coded 2,558 sentences from online discussions in a graduate course on distant education. It was found that *statements* (52%) and *reflections* (21%) were most common, followed by *references* (10%) and *scaffolding and engaging* (10%); *vertical questions* (1%) and *horizontal questions* (2%) only took a minimal portion. Zhu (1996) and Fahy et al. (2001) showed learners' roles can be usefully characterized based on their approaches to the discussion content and how these approaches relate to learning.

Table 2.3 Content analysis scheme for online discussions (Fahy et al., 2001)

Primary categories	Secondary categories	Characteristics
T1 - Questioning	T1(a): vertical	Assume a "correct" answer exists, and the question can be answered if the right answer can be found.
	T1(b): horizontal	There may not be one right answer, and others are invited to help provide a plausible or alternate "answer", or to help shed light on the question.
T2 - Statements	T2(a): direct	Contain little self-revelation and usually do not invite response or dialogue. The main intent is to impart facts or information.
	T2 (b): answers or comments	Direct answers to questions, or comments referring to specific preceding statements.
T3 – Reflections		Express thoughts, judgments, opinions or information which are personal and are usually guarded or private. The speaker may also reveal personal values, beliefs, doubts, convictions, and ideas acknowledged as personal.
T4 – Scaffolding and engaging		Intended to initiate, continue or acknowledge interpersonal interaction, and to "warm" and personalize the discussion by greeting or welcoming.
T5 – References and authorities	T5(a): references, quotations, paraphrases	References to, and quotations or paraphrases of other sources.
	T5(b): citations or attributions	Citations or attributions of quotations or paraphrases.

It should be noted that learning and interaction contexts examined in Zhu (1996) and Fahy et al. (2001) are different to those in MOOC forums which may result in differences in contribution characteristics. For instance, forum discussions in formal learning contexts often involve pedagogical designs, such as anchoring group projects and readings, designated learner roles, instructor facilitation, and requirement for participation (Fahy et al., 2001; Zhu, 1996). In contrast, MOOC forum discussions are usually not pedagogically structured, and learners participate voluntarily for varied purposes without a clearly-defined common goal (Margaryan, Bianco, & Littlejohn, 2015). Furthermore, learners in formal learning contexts often interact with the same small group of peers for an extended period whereas MOOC forum participants are often exposed to a larger number of peers who may not participate consistently. These differences can impact the relationships learners form with each other and the positions they take up in interaction (Golder & Donath, 2004). Therefore it is important to adapt the existing schemes to the particular learning and interaction contexts in MOOC forums. Additionally it can be worthwhile to develop a more compact framework to further boost the coding reliability.

2.3.4. Prior work on social relations

Social relations in MOOC discussion forums are usually examined using social network analysis (SNA) methods. SNA research often conceptualizes relationships in discussion forums as a network with participants as nodes and reply structures between them as ties (Dowell et al., 2015; Jiang, Fitzhugh, & Warschauer, 2014; Joksimović et al., 2016; Poquet & Dawson, 2016).

MOOC research has characterized discussant roles in information exchange based on structural similarity (sharing similar connections with similar others in the social network). For instance, Kellogg et al. (2014) identified learners who participated in discussions with similar patterns in a digital learning MOOC and a mathematics MOOC. For each course, a directed weighted social network was constructed based on the Direct Reply tie definition that constructs ties between people who have directly replied to each other's posts. Regular equivalence algorithm was used to partition forum participants into groups based on the similarity of their ties to others with similar ties. Four participation patterns were identified: *reciprocators* who participated in at least one mutual exchange; *networkers* who gave and received responses but with different peers; *broadcasters* who only initiated discussions; and *the invisible* who responded to others' posts but did not receive any response. *Reciprocators* made up the largest proportion of learners in both courses. This approach to identifying roles in information exchange is particularly useful for mega networks. However, the reply structure in the social network is a high level structural proxy of the actual information flow pattern. In comparison analyzing the content of discussion messages can capture learners' roles in information exchange more directly (e.g., Hecking, Chounta, & Hoppe, 2017).

Other MOOC research that adopted the SNA methods has found learner's connectedness in the social network and the strength of their social connections can provide understanding of discussant roles that complement the content analysis approach.

Connectedness in the social network

Connectedness indicates how well-connected a node is in a social network. Learner’s connectedness can be quantified by node-level centrality properties, such as degree (number of direct connections a node has), betweenness (the number of times a node is part of the shortest path between two other nodes in the network), and closeness (average of the shortest path from a node to all other nodes in the network, see Jiang, Fitzhugh, & Warschauer, 2014; Joksimović et al., 2016). Centrality properties are often associated with assumptions about influence, power, and privilege. For instance, high in-degree (number of connections pointing to a node) indicates a prestige status and high out-degree (number of connections pointing from a node) indicates a hub status; high betweenness indicates a broker status; high closeness indicates easy access to resources (Jiang, Fitzhugh, & Warschauer, 2014; Joksimovic et al., 2016). Learner’s connectedness can also be presented graphically by a core–periphery structure. Social graphs for large social networks often consist of a small number of highly connected nodes at the core and a large number of less connected and isolated nodes in the periphery (Goggins et al., 2016).

Both degree and betweenness and the associated core-periphery structure have been used for characterizing roles in MOOC forums (Brinton et al., 2014; Dowell et al., 2015; Joksimovic et al., 2016; Poquet & Dawson, 2016). For instance, Poquet and Dawson (2016) investigated influential learners in an undirected weighted network (constructed based on thread copresence) for discussion forums in a solar energy MOOC. They clustered forum participants through k-means clustering using betweennenss (indexing the quantity of participated conversations) and clustering coefficient (number of triangles a node is in divided by number of triangles it could be in, indexing a learner’s level of embeddedness in different conversations). Two of the four clusters they detected demonstrated characteristics of influential members. One cluster consisted of 8 highly influential members who had very high betweenness and low clustering coefficient. They were likely to be community TAs. The other cluster consisted of 82 moderately influential members with moderate betweenness and low clustering coefficient. They were likely to be learners who participated in social interaction moderately actively. In the same study, Poquet and Dawson (2016) also identified information brokers among frequent contributors (who posted in at least three weeks) using the core-peripheral approach. An undirected weighted social network was constructed for frequent contributors and showed a core-peripheral structure. At the core were a small group of learners who exchanged interaction with each other frequently and connected with more peers, including those in the periphery of the network. These core members were considered important information brokers between less active members.

A small number of MOOC studies examined changes that occurred over time in individual learners' connectedness. Tawfik et al. (2017) found in a chemistry MOOC that degree and betweenness of the top 25 participants ranked by degree centrality showed little change over time. Yang et al. (2013) found in a literature MOOC that the few high centrality participants were mainly from the earliest cohort that participated in the forums; late starters often remained in the periphery, had trouble getting integrated into discussions, and tended to make less contributions.

Strength of social connections

The strength of social connections between learners can be measured with weighted edges. Repeated interactions between two learners result in higher edge weight which indicates potentially stronger connections between the two. Yang, Wen, and Rosé (2014a) found that MOOC learners who developed a higher number of stronger social ties in the discussion forum were more likely to continue participating in the forum. Furthermore, Wise and Cui (2018a) examined a statistics MOOC and found forums participants in communities with strong inter-learner social connections tended to revisit discussion to offer help to others, after they had received help in these discussions. Focusing on individual learners, each learner in the social network forms an ego network with others that are directly connected to the learner themselves (Scott, 2000). Average edge weight in the ego network indicates the overall strength of a learner's social connections with those who have interacted directly with them. Connection strength in ego networks have been found to impact many aspects of social interaction, such as willingness to collaborate and share resources with others (Arnaboldi, Conti, Passarella, & Dunbar, 2017). Thus average edge weight in the ego network can be promising for understanding learner positions in MOOC discussions.

2.3.5. A prior study that integrated contribution characteristics and social characteristics

Contribution characteristics and social characteristics facilitate understanding of positions from complementary perspectives. Content analysis has strength in generating in-depth understanding of interaction at a more granular level (such as the topics involved in interaction and the effectiveness of communication) while SNA methods make high level measurement of interaction patterns (Forestier, Stavrianou, Velcin, & Zighed, 2012; Gleave, Welser, Lento, & Smith, 2009). Combining the two perspectives is promising for providing a holistic understanding of positions.

In the MOOC literature, such integration has been done only in Hecking et al. (2017). The two approaches were combined through (1) identifying information-seeking and information providing activities in forum contributions through content analysis, (2) restoring connections and direction of connections between forum participants based on the information-seeking and giving relationships identified through content analysis, then constructing a social network, and (3) identifying learners with structural similarity (sharing similar connections with similar others in the social network) in the social network. The study was conducted on sub-forums dedicated to learning-related discussions in two MOOCs on corporate finance and global warming. First, contribution characteristics were identified through categorizing discussion posts as information-seeking (questions, clarification requests, reports of technical issues) and information-providing (answers, issue resolutions, hints, recommendations). Posts that did not involve information-seeking or information-providing activities were filtered out. The content analysis process was facilitated by using a message classifier built with learning-related features (question words, question / exclamation marks, post length, help seeking and giving phrases) and thread structural features (start / reply post, number of votes) extracted from 500 hand-coded posts. Second, a directed weighted social network was constructed based on the reply structures between information seeking messages and information providing messages identified through the content analysis process. Finally, learners were partitioned into groups through blockmodeling based on their information seeking / providing patterns. Three roles were identified in each course: *core users*, *peripheral information seekers*, and *peripheral information givers*. *Core users* were the dominant role and characterized learners who both seek and gave information; *peripheral information seekers* were learners who only received information from other groups; and *information providers* were those who only gave information to other groups. Using both content analysis and social network analysis methods, this study led to improved understanding of how learners participated in information exchange in discussions. However, as with Stump et al. (2012) and Liu et al. (2016), the content analysis framework they used did not allow for identifying roles in association with learning.

2.4. Chapter summary

This literature review reveals that the course performance perspective alone is inadequate for understanding MOOC learning. Learning as development in participation patterns characterized by position trajectory offers a promising alternative perspective. While prior MOOC research on role-related contribution characteristics and social relations provides useful grounding for understanding positions, three major research gaps remain to be addressed. First, prior research mainly analyzed contribution and social characteristics as static end-of-phase constructs. It is necessary to investigate changes in these characteristics over time to understand development in forum participation patterns. Second, contribution and social characteristics can provide complementary understanding of positions, but there is not yet a comprehensive framework that encompasses the two perspectives. Third, there is not yet an appropriate content analysis scheme for contribution characteristics. The schemes used in MOOC contexts mainly focused on identifying information seekers and providers without substantially associating positions with learning. The complex schemes developed in non-MOOC contexts need to be adapted to MOOC contexts.

Chapter 3.

Research Questions and Study Framing

3.1. Research questions

The overarching goal of this study is to improve understanding of MOOC learning defined as development in the way that one interacts with others in forum discussions. Such development is indicated by changes in a learner's participation pattern that are constructive with respect to substantive interaction about the course content. This study uses position as a conceptual construct for learner's way of participation that can be characterized by characteristics related to their forum contributions and social relationships in the forum. Positioning is defined as the action of position taking. The research goal is addressed through answering the following research questions:

RQ1a: What common participation positions are found in MOOC forums?

RQ1b: Which of the positions are found across different courses and different time periods in the same course?

RQ2: What changes occurred in characteristics of individual learner's position?

RQ3: How did the changes manifest as common trajectories and represent learning?

3.2. Study framing

3.2.1. Conceptualizing positions in MOOC discussions

Answering the research questions first calls for a conceptual framework that provides a holistic perspective on positions in MOOC discussions. Based on the literature review, the current study presents a two-dimension framework that encompasses contribution characteristics and social characteristics which accounts for the nature of a learner's contribution content and their social relationships respectively (see Table 3.1).

Table 3.1 Framework for position characteristics

Dimensions	Aspects
Contribution characteristics	Related / unrelated to the course content Input seeking / input providing Involving / not involving deep consideration of the discussion content
Social characteristics	Individual connectedness in the social network Strength of social connections

For the first dimension in the framework, contribution characteristics are assessed by three aspects about a contribution: does it (1) relate to the course content; (2) seek or provide input; and (3) contain deep consideration of the discussion content. First, MOOC discussions related and unrelated to the course content have been found to serve distinct purposes and differ in both interaction characteristics and social network properties (Wise, Cui, & Jin, 2017; Wise & Cui, 2018a). Therefore differentiating the two kinds of contributions is important for studying learner positions in the forum and MOOC learning. Content discussion involves interactions related to subject matter knowledge specified in the course syllabus and domain knowledge related to the course subject. Non-content discussion involves interactions not directly related to either form of knowledge, such as technical and logistical issues.

Second, input seeking and input providing activities differ in interaction purposes and the way they contribute to forum interactions. Learners who seek input request help or resource from others to meet their learning needs; at the same time their requests invite participation from others and potentially generate interaction opportunities (Liu et al., 2016; Stump et al., 2014). Learners who provide input contribute content and resources to forum discussion; their knowledge and expertise can be useful for facilitating other's learning, and is crucial for interactions in MOOC forums to sustain (Liu et al., 2016; Yang, Adamson, Rosé, 2014). Input seeking and input providing activities have been differentiated in MOOC studies that examined learners' roles in discussion; input providers and input seekers have been found to differ in their learning outcomes measures by course grades (Hecking et al., 2017; Liu et al., 2016).

Finally, prior studies on MOOC and non-MOOC forum discussions revealed learners' approaches to discussion content often associate with different levels of effort and learning that can be differentiated hierarchically (Anderson & Krathwohl, 2001; Wong et al., 2016). In this study the differentiation is made based on presence / absence of deep consideration of the discussion content. Deep consideration involves efforts to understand and make sense of the content, such as explaining, elaborating, comparing, justifying, and reasoning whereas a non-deep contribution generally involves plain and straightforward information exchange through naming, listing, and describing content.

Social characteristics are assessed by two aspects: (1) individual connectedness in the social network and (2) strength of social connections with adjacent neighbors in the social network. Both aspects have been found useful for characterizing roles and understanding interactions in MOOC and non-MOOC contexts (Arnaboldi et al., 2017; Joksimovic et al., 2016; Poquet & Dawson, 2016; Wise & Cui, 2018a). In summary, the conceptual framework for position consists of two dimensions (five aspects): contribution characteristics (content-relatedness, input seeking / input providing, deep consideration) and social characteristics (connectedness in the social network, strength of social connections).

3.2.2. Important decisions in operationalizing the conceptual framework to understand positions and position trajectories

From an empirical perspective, addressing the research goal of understanding MOOC learning by investigating positions and position trajectories involves multiple operational decisions. First, to investigate changes in a learner's position during a MOOC, the course needs to be divided into multiple time segments. In this study a course is divided into three segments each with a similar number of weeks. This decision was made based on multiple considerations. For one thing, the level of participation in MOOC forum is highly variant. Segmenting the course on a weekly basis can produce a lot of time segments without participation. Multi-week segments can alleviate this problem. Moreover, the three-segment solution allows for the potential for examining trajectories with multiple stages and the three segments respectively correspond to the beginning, middle, and end phases in a course.

Second, to characterize learners' positions in discussions related to the learning of course content, a participation profile needs to be built for each learner using their contribution characteristics and social characteristics. Contribution characteristics were extracted through performing content analysis on discussion contributions. As documented in the literature review, a content analysis scheme for discussions in the MOOC context needs to be developed based on the contribution dimension in the conceptual framework (see Table 3.1). Content-related discussions have been found to be more useful for understanding interaction about the course content and predicting learners' course grades, in comparison to content-unrelated discussions and undifferentiated discussions (Cui & Wise, 2015; Wise, Cui, Jin, & Vytasek, 2017; Wise & Cui, 2018a). Therefore this study extracts contribution characteristics from content-related discussions. Content-relatedness of a discussion is determined at thread level given that MOOC discussions are presented to learners as threaded conversations. Threads with a content-related initiating message are considered content-related as the initiating message largely sets the direction and scope of discussion. Considering that discussions with a non-content initiating message may change direction as new participants join (Stump et al., 2013), for threads with a non-content initiating message, the decision is made based on whether or not the thread contains a substantial proportion of content-related replies.

As for social characteristics, extracting these characteristics first requires constructing the social network. Prior research had indicated content and non-content discussions in MOOCs are participated by substantially different people (Wise & Cui, 2018a). Since this study aims to understand forum participation that occurs around the course content, the social networks are constructed for content-related discussions. Choosing a tie definition for constructing the social network is also an important decision. In MOOC forums, a learner can establish social connections through replying to other's messages. They can also establish connections through accessing and being informed by multiple posts before making their own. Wise and Cui (2018a) presented the Limited Copresence tie definition to operationalize this assumption about social connections in MOOC social networks. Limited Copresence defines a tie as being present in the same part of a discussion based on the assumption that a participant in a thread with a small number of replies has ties with all others in the same thread. However the number of messages MOOC learners read before posting has not yet been empirically verified. Without setting a limit, the copresence assumption becomes problematic when size of thread is large and when many distinct people are involved in the same thread, which are often the case in MOOCs. An alternative tie definition is the Direct Reply definition that constructs a tie only if there is a direct reply relationship between two nodes in the same thread, without making any assumption about others who may have been informed by a post but not replied directly to it. This tie definition is used widely in MOOC research. In comparison to the Limited Copresence definition, it results in social networks with similar overall structures but lower degree and edge weight (Fincham, Gaševic, & Pardo, 2018; Wise & Cui, 2018a). The current study uses the Direct Reply definition for constructing the social networks. The implications of this operational decision will be discussed in Chapter 6.

Third, to gain a big picture understanding of the common participation patterns in MOOC forums, this study needs to identify groups of learners that have the similar positions in discussion, thus participation profiles for learners in the same time period need to be grouped based on similarity in contribution and social characteristics. Cluster analysis is used for this task. Cluster analysis is an unsupervised machine learning method useful for discovering structures in data, and has been widely adopted for discovering groups of learners based on similarity in multiple types of participation characteristics in MOOC and non-MOOC forums (Eynon, Hjorth, Yasseri, & Gillani, 2016; Poquet & Dawson, 2016; Wise, Speer, Marbouti, & Hsiao, 2013).

Fourth, learners who participated in multiple time periods can be considered to have one position in each period; movement between these positions forms a learner's position trajectory. Changes that occurred over time in learners' positions are identified through comparing characteristics of the start position and the end position in the trajectory. Given that each position profile consists of multiple characteristics, for each identified position this study assigns a label that represents each characteristic, in addition to a profile name that highlights the key characteristics of the position. To understand common changes in position, the most frequent trajectories are examined both quantitatively through identifying the direction of changes and qualitatively through case studies.

Finally, to understand how variations in learning context may relate to position and positioning, two MOOCs offered in different subjects were selected for the study. To control for other differences in the learning context, the two selected courses were from the same course platform and had similar course length, number of forum participants, quantity of forum activities, and course policy for forum participation.

Chapter 4.

Methods

4.1. Data source

This study used data from StatLearn and SciWrite, two completed MOOCs offered on the Stanford open-source platform Lagunita. StatLearn, offered in 2015, was an introductory course in supervised learning with a focus on regression and classification methods. Learners participated by watching lecture videos and answering quiz questions. SciWrite, offered in 2014, was a course on effective writing in science disciplines. Learners participated by watching lecture videos, responding to quiz questions and assignments, taking a final exam and had the choice to participate in two optional essay writing / peer grading assignments. Passing grades for StatLearn and SciWrite were respectively 50% and 60% of the total score. Both courses required a minimal of 90% score to pass with distinction.

Both courses provided a discussion forum as a supplementary venue for participation. Learners were invited to post questions and comments about the course for response by other learners and the instructional team. Participation in forum was optional and not graded. The forums used a three-level threaded interface with threads listed on the left and the expanded messages shown on the right. Data available included demographics and discussion forum logs.

4.2. Data

4.2.1. Demographics

The demographic data contained user id and self-reported sex, year of birth, level of education, and country. There were 76,311 entries in the demographic data table for StatLearn and 42,683 entries for SciWrite. For both courses, some user ids appeared multiple times in the table and contained identical information for gender, year of birth, and level of education, but differed in country. The multiple country values for the same user id were produced if a user accessed the platform from different IP addresses which each generated its own row in the table. These duplicated ids were removed and yielded 55,957 user ids for StatLearn and 31,558 for SciWrite.

The learner population in StatLearn consisted of a majority of males (67%) and a lower proportion of females (21%); in SciWrite the proportion of male and female learners was similar (48% vs. 44%, see Figure 4.1)

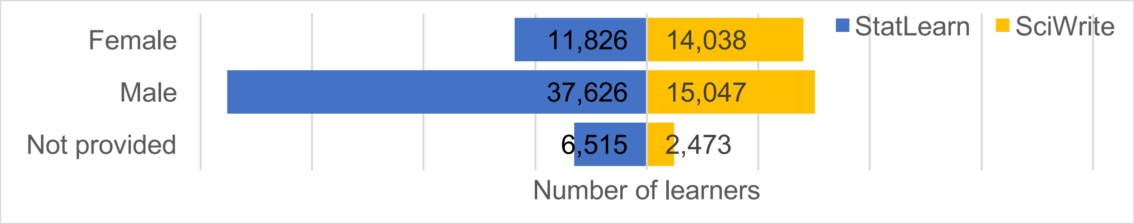


Figure 4.1 **Proportion of female and male learners in StatLearn and SciWrite**

Learners in the two courses were of similar ages, ranging from 0 to 120 (calculated from self-reported year of birth). It appears some learners provided false data but the results cannot be verified. Age distribution is given here to provide a sense of the learner population's age (see Figure 4.2).

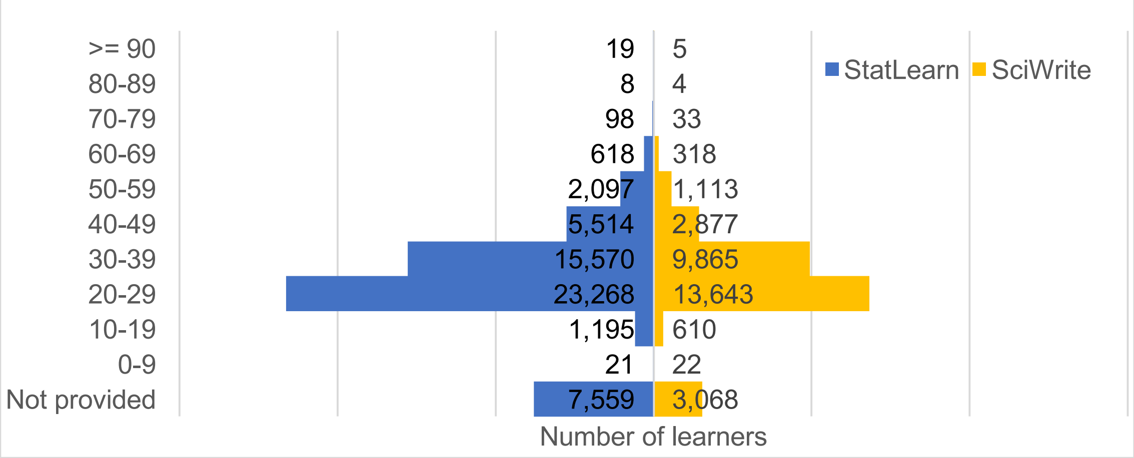


Figure 4.2 **Learner age in StatLearn and SciWrite**

The learner populations of both courses were well-educated with more than 80% having a bachelor's or more advanced degree (see Figure 4.3).

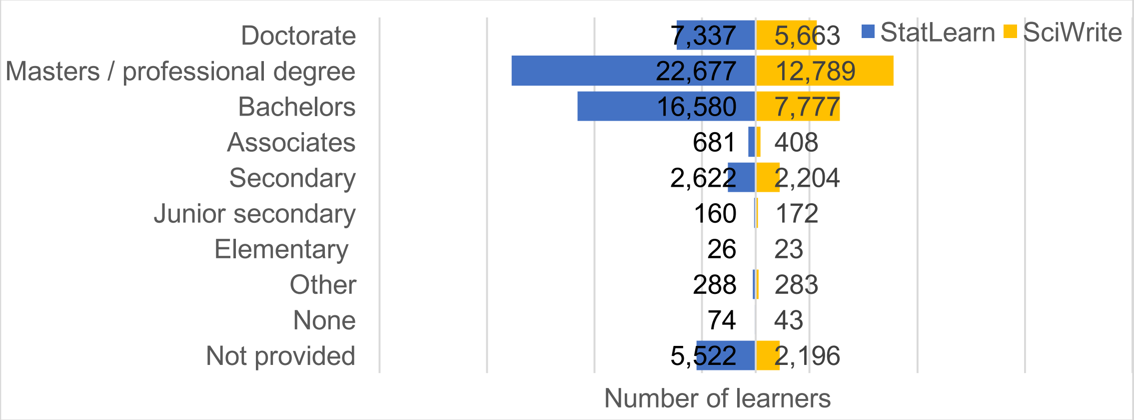


Figure 4.3 **Learners' education level in StatLearn and SciWrite**

Learners in StatLearn and SciWrite were respectively located in 233 and 188 countries. Due to the multiple country values for some user ids, the country data were not broken down to describe the geographical distribution of the learner population.

4.2.2. Discussion forum logs

Forum information provided in the data set included: thread id, post id, user id, post position in thread (starter, reply, reply to reply), hierarchical post relationship (parent post and child post), post text, post creation date and time, and number of votes post received. Thread titles were not included in the data set.

StatLearn and SciWrite included 9 and 8 weeks of lectures respectively. An inspection of forum activities showed in both courses a substantial number of users continued participating in discussion forum for another two weeks after conclusion of the course lectures. Thus for the two courses, a total of 11 weeks and 10 weeks of forum activities respectively were included for analysis. To examine participation patterns over time, the examined weeks in both courses were divided into beginning, middle, and end periods (hereafter T1, T2, T3), each time period containing ~1/3 of the examined weeks (see Table 4.1).

Table 4.1 Weeks in course periods

	StatLearn	SciWrite
Beginning period (T1)	Week 1-3	Week 1-3
Middle period (T2)	Week 4-7	Week 4-6
End period (T3)	Week 8-11	Week 7-10

During the examined weeks, a total of 1,229 users generated 850 threads with 4,201 messages in the StatLearn forum; a total of 1,194 users generated 1,189 threads with 4,385 messages in the SciWrite forum. In cleaning the data, for both courses a small number of threads and subthreads that were missing the initiating messages for unidentified causes were removed from the data set. Additionally messages containing foreign language or only punctuation were removed. After removing messages with these characteristics, for StatLearn 97% threads (n = 821), 96% messages (n = 4,016), and 97% users (n = 1,187) remained for analysis; for SciWrite 99.5% threads (n = 1,182), 99.5% messages (n = 4,360), and 99.7% users (n = 1,191) users remained for analysis.

4.3. Content analysis and extracting contribution characteristics

4.3.1. Content analysis

Study data in the two courses were hand coded for characteristics in three aspects. Each message was first coded for whether it related to learning of the course content. Then each content-related message was coded for two subsequent aspects: (1) whether it sought input on the course content, provided input, or did both and (2) whether it contained deep consideration of the discussion content. Although content analysis was conducted at the message level, messages were sampled and analyzed in intact threads to preserve interaction context. A detailed coding guide with description of characteristics and examples was used to facilitate coding (see Appendix).

Coder training was conducted on data from successive offerings of StatLearn and SciWrite until reliability, as indexed by Cohen's kappa, was stable at an acceptable level (> 0.7). To verify the conventional practice of double coding 20% of total data, Cantor's (1996) and Donner's (1999) methods were used to estimate the minimum sample size needed to infer the desired reliability level for the entire coded corpus (kappa > 0.7, alpha = 0.05, power = 0.8), based on coding reliability and results achieved for initial double coding conducted on a sample set from the study data (see Table 4.2). The estimation was conducted in R using the N.cohen.kappa function in the irr package and the Power3Cats function in the kappaSize package. Results showed for StatLearn minimally 24, 154, and 203 messages (of the entire corpus of 4,016) were needed for the three aspects respectively. For SciWrite minimally 36, 105, and 105 messages (of the entire corpus of 4,360) were needed. This indicated double coding 20% of the data (803 messages for StatLearn and 872 messages for SciWrite) is more than acceptable. As data sampling needed to be conducted at thread level for the reasons described above, 20% of the total threads were selected for double coding through stratified sampling based on time of creation for the thread starter post and overall thread length. The remaining 80% of threads were single coded. To check for coder drift, three rounds of double coding were conducted before, between, and after two rounds of single coding. All differences in double coding were discussed and reconciled before the next round of coding.

All rounds of double coding reached acceptable inter-rater reliability (kappa > 0.7). Over time some small drift was observed but reliability remained at acceptable levels (see Table 4.2). These results indicated that the content analysis scheme developed in this study was of good reliability.

Table 4.2 Inter-rater reliability for double coding

Rounds in double coding	# of threads/ messages	Aspect 1 Content/Non-content *K**/%***	Aspect 2 Seek/Provide/Both *K**/%***	Aspect 3 Deep/Non-deep *K**/%***
StatLearn Initial*	11 / 35	0.941 / 0.971	0.877 / 0.952	0.809 / 0.905
1	52 / 209	0.894 / 0.947	0.879 / 0.940	0.795 / 0.914
2	52 / 234	0.796 / 0.902	0.854 / 0.928	0.761 / 0.884
3	51 / 228	0.815 / 0.908	0.861 / 0.932	0.714 / 0.864
Overall	166 / 706	0.840 / 0.921	0.865 / 0.934	0.763 / 0.889
SciWrite Initial*	16 / 56	0.958 / 0.982	0.866 / 0.941	0.866 / 0.941
1	76 / 259	0.906 / 0.958	0.834 / 0.929	0.799 / 0.917
2	76 / 265	0.926 / 0.970	0.910 / 0.961	0.717 / 0.870
3	76 / 268	0.935 / 0.970	0.951 / 0.979	0.701 / 0.887
Overall	244 / 848	0.925 / 0.967	0.898 / 0.956	0.747 / 0.895

* The initial round was conducted to estimate necessary sample size for double coding.

**K = Kappa, % = percentage agreement

4.3.2. Distinguishing content and non-content discussions

Each thread was classified as either content-related or non-content based on two rules. First, threads with a content-related starter post were labeled as content-related. Second, for threads with a non-content starter, the decision was made based on whether the thread contained at least 50% content-related replies (see Section 5.1.1).

4.3.3. Overview of discussion activities at the course level

For each course and each of the three time periods in the same course, the number of content threads and non-content threads generated in the forum was calculated. Then focusing within the content threads, the number and proportion of messages with different characteristics were calculated, including content messages, non-content messages, input seeking messages, input providing messages, both-seeking-and-providing messages, messages with deep consideration (hereafter deep messages), and messages without deep consideration (hereafter non-deep messages).

4.3.4. Extracting contribution characteristics

Learners who posted to only content threads, only non-content threads, and both kinds of threads were identified. For every learner who posted to any content threads, the number of course periods they participated in was calculated. The following contribution features were computed for each course period they participated in: number of content messages, number of input seeking messages, number of input providing messages, number of both-seeking-and-providing messages, number of deep messages, number of non-deep messages, and number of deep messages added to ongoing non-deep discussions (to measure the frequency that a learner steered the depth of a discussion, hereafter depth-switching messages).

4.4. Network analysis and extracting social characteristics

Directed weighted social networks were constructed for content discussions in each of the three time periods (T1, T2, T3) in StatLearn and in SciWrite. The nodelists were extracted from forum data using the user id of messages. Edgelists were extracted using the Direct Reply tie definition: the author of each post was considered the source node and connected with the author of its parent message as target node. Forum participants who started a thread but did not receive any replies were considered isolated nodes. Network graphs were visualized for each time period. Social network properties were computed at the global and community levels to explore the overall learning context. Global properties included number of nodes, number of edges, average degree, average edge weight, and number of communities. Community detection was performed through modularity maximization using the Louvain method (De Meo, Ferrara, Fiumara, & Provetti, 2011). For each community the number of nodes was calculated.

Social features computed for each individual node included degree, betweenness, and average edge weight in the ego network (hereafter ego edge weight). Network construction and property computation were performed in R using the igraph package. Network graphs were visualized using the Force Atlas layout algorithm in Gephi for Mac.

4.5. Cluster analysis: Identifying positions in the forum

A participation profile was built for each learner in each of the three time periods in the two courses, using contribution features and social features described in Sections 4.3 and 4.4. Cluster analysis was implemented on all profiles in the same time period to detect clusters of similar profiles. Learners in the same cluster were considered to have the same overall position in discussions.

4.5.1. Clustering features

Contribution features included three measures for quantity of input seeking/providing activities (number of input seeking messages, number of input providing messages, number of both-seeking-and-providing messages) and three measures for quantity of activities with/without deep consideration (number of deep messages, number of non-deep messages, number of depth-switching messages). Social features included two measures for learner’s connectedness (degree, betweenness), and one measure for strength of social connections in the ego network (ego edge weight). All features were continuous variables.

Histogram for each of the variable was examined for variance in the variable. The variables were then checked in pairs for correlation, using Pearson's correlation coefficient when both variables were normally distributed (|Skewness| <2, |Kurtosis| <2). Spearman's correlation coefficient was calculated if distributions were non-normal.

4.5.2. Cluster analysis procedures

Pairwise scatterplots for variables were examined for non-convex distribution of the data and outliers that can affect performance of a cluster analysis. K-means clustering was implemented for cluster analysis, using Euclidean distance to determine the distances between clusters. Before clustering all variables were scaled to account for differences in metrics. To select the optimal number of clusters, scree plots were examined for the leveling-off point which suggests additional groups would not have meaningful differences between them. Cluster analysis and visualization was implemented in R using the kmeans function from the stats package, and the fviz_nbclust function and the fviz_cluster function from the factoextra package.

Each cluster identified through the cluster analysis represented a position in discussion. Profiles for the identified positions were summarized based on the mean scores of all clustering variables for all cluster members. A summary was compiled for positions found in each time period. Each position was assigned a label that represented all position characteristic determined based on the actual data described in Section 5.3.3. A profile name that represented the key characteristics was also assigned to each position. To identify positions that were found across courses and time periods, positions from different time periods in the two courses were compared and summarised based on the key characteristics.

4.6. Extracting and characterizing position trajectories

Learners who participated in multiple time periods were considered to have one position in each period. Shifts across these positions formed a learner's position trajectory. As both courses examined in this study only contained a small number of learners who participated in three periods (hereafter 3-period learners, *n* = 34 in StatLearn, *n* = 15 in SciWrite) and a larger number of learners who participated in two periods (hereafter 2-period learners, *n* = 116 in StatLearn, *n* = 58 in SciWrite) this study examined trajectories as consisting of two positions (a start position and an end position). The 3-period learners were considered to have two trajectories (see Figure 4.4).

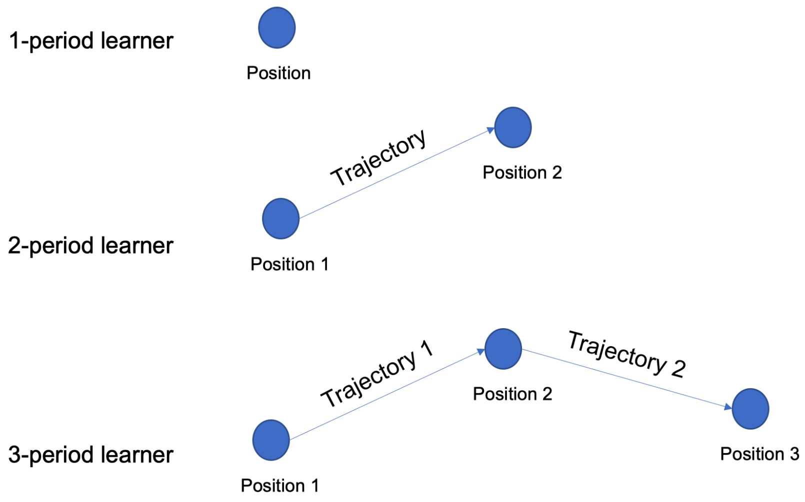


Figure 4.4 Postion and position trajectory

After trajectories were extracted from the two courses, changes in a learner's position were identified by comparing characteristics of the start position and the end position in the trajectory. The proportions of learners who did and did not show changes over time were calculated. Changes in each position characteristic were summarized for frequency. Position trajectories that occurred frequently within each course and across both courses were examined for changes in position.

4.7. Case studies

To develop a contextualized understanding of changes in position and in what ways such changes may indicate learning, case studies were performed on learners who typified frequent trajectories in discussions. Learners' contribution characteristics and social characteristics in the two time periods involved in the trajectory were compared. In addition, for the threads in which these learners participated, the content and interaction processes were described, including discussion topics, process of exchanges among thread participants, and characteristics of messages posted by the examined learners.

Chapter 5.

Results

5.1. Results for content analysis and contribution characteristics

5.1.1. Differentiating content and non-content discussions

In StatLearn a total of 1,187 forum users generated 821 threads (containing 4,016 messages) during the 11 examined weeks. Of the 4,016 messages, 2,048 were labeled as content messages; 1,968 were labeled as non-content messages. Of the total 821 threads in this course, 498 threads had a content message as the starter and thus were considered content threads. Among the 323 threads that had a non-content starter, 304 did not contain any content replies and were characterized as non-content threads; the remaining 19 threads had 6% to 71% content messages (see Figure 5.1). Manual examination of the thread content showed threads containing 50% or more content messages did have substantial interaction related to the course content (while those below did not), and thus these 9 threads were additionally labeled as content threads. In summary 507 threads (62% of the total) containing 2,391 messages were labeled as content threads; 314 threads (38% of the total) containing 1,625 messages were labeled as non-content threads.

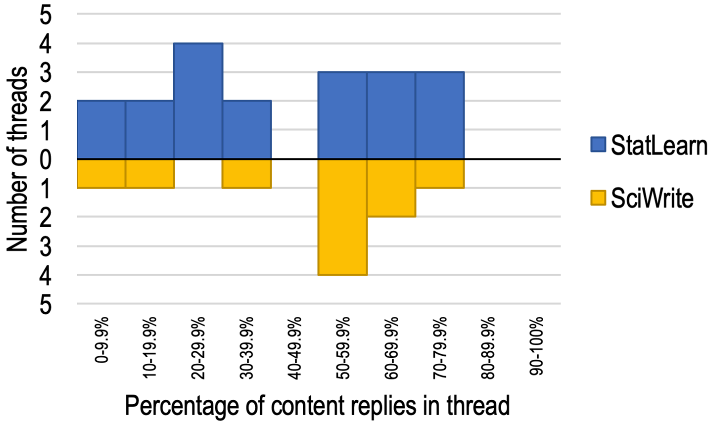


Figure 5.1 Threads with a non-content starter and >0 content replies

In SciWrite a total of 1,191 forum users generated 1,182 threads (containing 4,360 messages) during the 10 examined weeks. Of the 4,360 messages, 1,395 were labeled as content messages; 2,965 were labeled as non-content messages. Of the total 1,182 threads, 436 threads had a content-related starter and thus were considered content threads. Among the 746 threads that had a non-content starter, 736 threads did not contain any content replies and were characterized as non-content threads; the remaining 10 threads had 8% to 74% content messages (see Figure 5.1). Manual examination showed threads that contained 50% or more content messages did have substantial content-related interaction (while those below did not), and thus these 7 threads were additionally labeled as content threads. In summary 443 threads (37% of the total) containing 1,556 messages were labeled as content threads; 739 threads (63% of the total) containing 2,804 messages were labeled as non-content threads.

5.1.2. Overview of content discussions in the two courses

The StatLearn forum contained a higher proportion of content threads (62%, *n* = 507) than SciWrite forum (37%, *n* = 443). The number of content and non-content threads generated over time in the two courses are summarized in Figure 5.2. In Statlearn, the number of content threads was similar in T1 and T2 (*n* = 210 vs. *n* = 188), then dropped dramatically in T3 (*n* = 109). In SciWrite, the quantity of content threads started high in T1 (*n* = 339), dropped dramatically in T2 (*n* = 63) and remained low in T3 (*n* = 41).

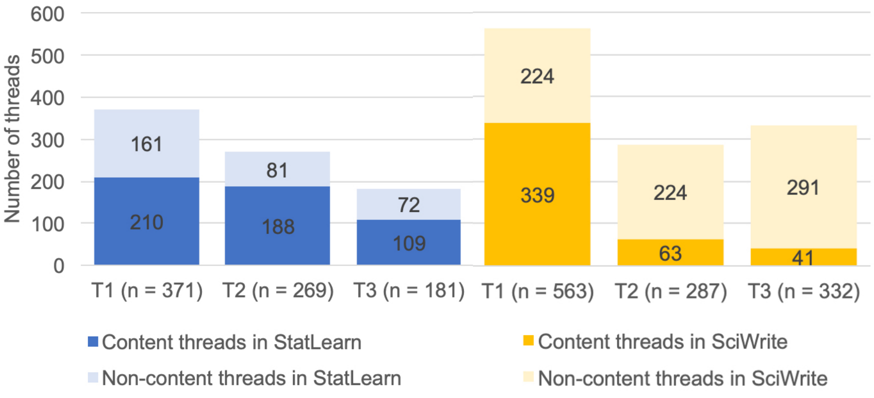


Figure 5.2 Number of content and non-content threads generated over time

5.1.3. Focusing within content threads: Content-relatedness, input seeking and input providing activities, and deep consideration of the discussion content

In StatLearn content threads contained 2,391 messages, of which 85% (*n* = 2,025) were content-related. In SciWrite content threads contained 1,556 messages, of which 90% (*n* = 1, 396) were content-related. In both courses about 1/3 of the content messages sought input and 2/3 provided input. Only 1-2% did both. These ratios were relatively stable across time periods (see Table 5.1).

Table 5.1 Proportion of input seeking, input providing, and seeking+providing messages within content threads

	StatLearn			**SciWrite**		
	Seek	**Provide**	**Both**	**Seek**	**Provide**	**Both**
T1	37% (*n* = 311)	62% (*n* = 515)	1% (*n* = 10)	29% (*n* = 318)	70% (*n* = 754)	1% (*n* = 7)
T2	37%* (*n* = 288)	61%* (*n* = 476)	3%* (*n* = 20)	40% (*n* = 83)	59% (*n* = 123)	1% (*n* = 1)
T3	38% (*n* = 152)	61% (*n* = 248)	1% (*n* = 5)	33% (*n* = 36)	67% (*n* = 74)	0% (*n* = 0)
Total	37% (*n* = 751)	61% (*n* = 1239)	2% (*n* = 35)	31% (*n* = 437)	68% (*n* = 951)	1% (*n* = 8)

* Percentages added up to 101% because of rounding.

The overall proportion of deep message was slightly higher in StatLearn than in SciWrite (37% vs. 27%). In StatLearn during the three time periods the proportion of deep messages gradually dropped from 41% to 27%. In SciWrite, the proportion of deep messages started off at 30% in T1, dropped to 16% in T2, and resurged to 24% in T3 (see Table 5.2).

Table 5.2 Proportion of deep and non-deep messages in content threads

	StatLearn		**SciWrite**	
	Deep	**Non-deep**	**Deep**	**Non-deep**
T1	41% (*n* = 346)	59% (*n* = 490)	30% (*n* = 319)	70% (*n* = 760)
T2	36% (*n* = 286)	64% (*n* = 498)	16% (*n* = 33)	84% (*n* = 174)
T3	27% (*n* = 110)	73% (*n* = 295)	24% (*n* = 26)	76% (*n* = 84)
Total	37% (*n* = 742)	63% (*n* = 1283)	27% (*n* = 378)	73% (*n* = 1018)

5.1.4. Participants in content and non-content discussions

Which type of discussion did learners participate in?

Of the total 1,187 participants in StatLearn forum, 34% (n = 407) only posted to content discussion; 24% (n = 286) posted to both content and non-content discussion; 42% (n = 494) only posted to non-content discussion. Of the total 1,191 participants in SciWrite forum, 24% (n = 283) only posted to content discussion; 23% (n = 269) posted to both content and non-content discussion; 53% (n = 639) only posted to non-content discussion. These results indicate content and non-content discussions in the two courses engaged substantially different people. This pattern was also found in each time period in the two courses (see Figure 5.3).

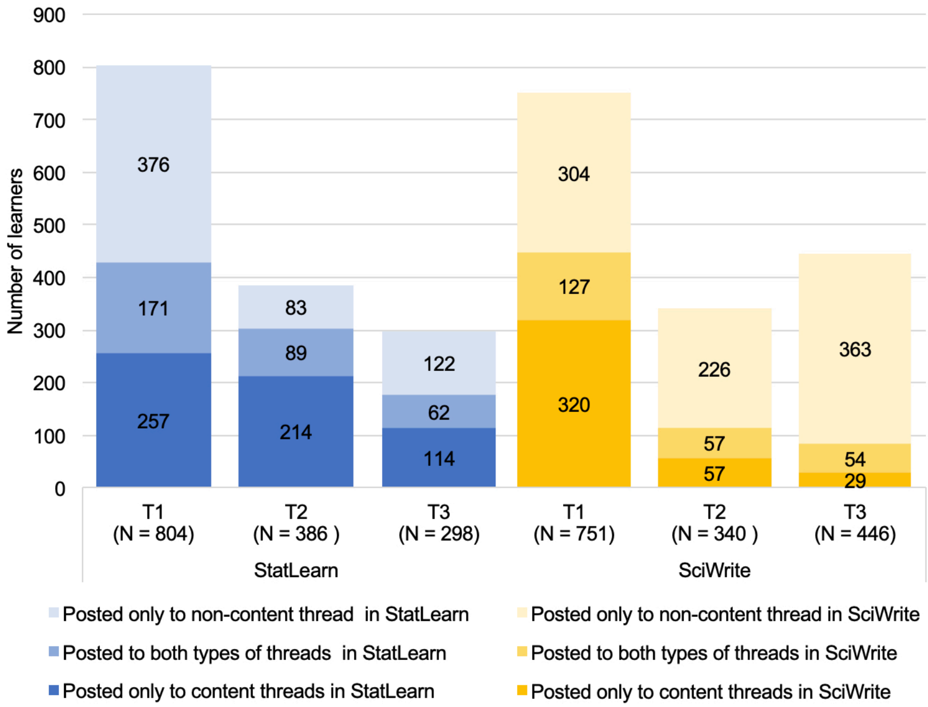
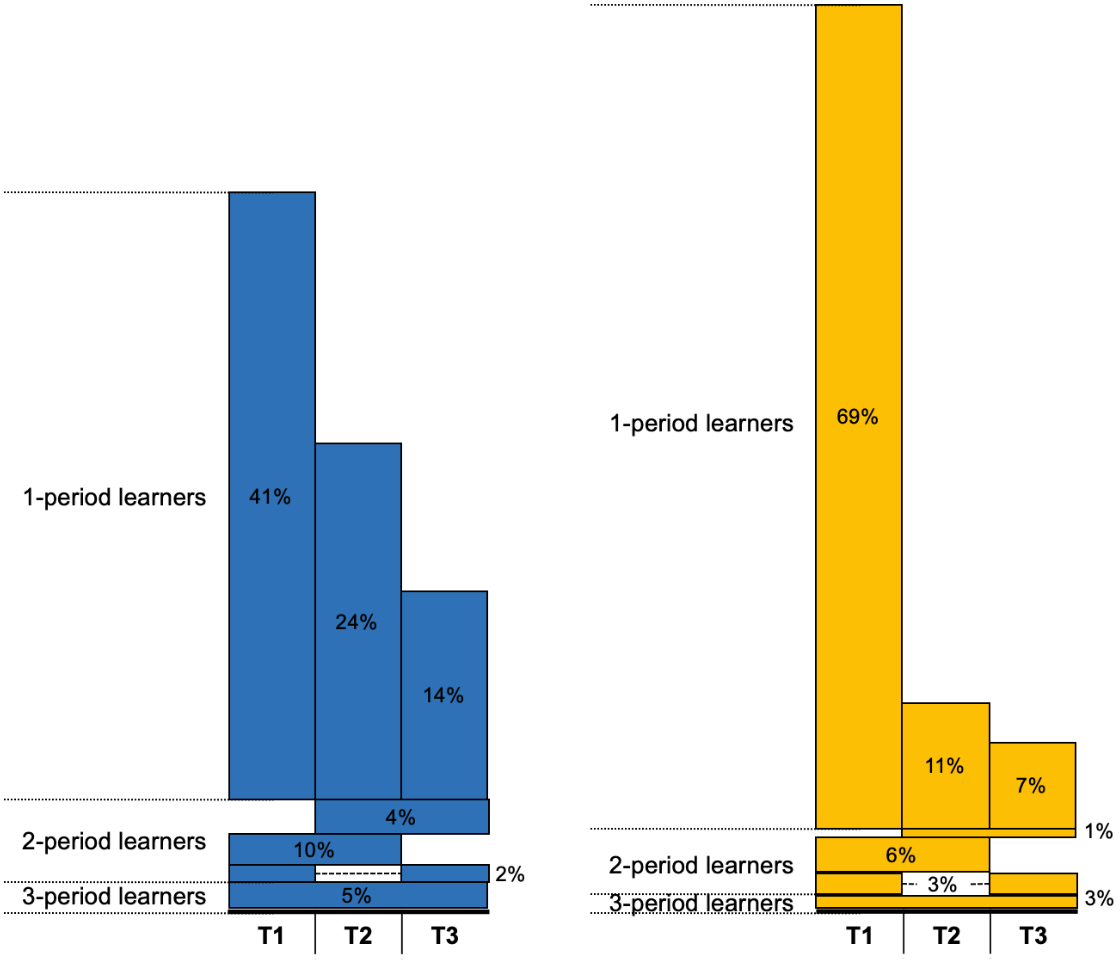


Figure 5.3 Number of learners who posted to different types of threads

How many time periods did learners participate in?

In StatLearn among the 693 participants in content discussions (407 only in content discussions and 286 in both content and non-content discussions), 689 were learners; 4 were instructional team members (1 instructor and 3 TAs). Seventy-nine percent (n = 539) of the 689 learners only posted in one of the three course periods; 16% (n = 116) posted in two course periods; only 5% (n = 34) posted in all three course periods (see Figure 5.4). Three instructional team members posted in two course periods; one posted in all time periods. In SciWrite among the 552 participants (283+269) in content discussion, 550 were learners and 2 were instructional team members (1 instructor and 1 TA). Among the 550 learners 87% (*n* = 477), 10% (*n* = 58), and 3% (*n* = 15) learners respectively posted in one, two, and three time periods (see Figure 5.4). The two instructional team members both posted in all time periods.



a. StatLearn

b. SciWrite

Figure 5.4 Proportion of 1-period, 2-period, 3-period learners and the time periods they participated in

5.1.5. Contribution characteristics

Contribution characteristics for learners in each time period in the two courses were extracted to use in cluster analysis: number of input seeking messages, number of input providing messages, number of both seeking and providing messages, number of deep messages, number of non-deep messages, and number of depth-switching messages (see Table 5.4 for a summary of all clustering variables).

5.2. Results for network analysis and social characteristics

5.2.1. Comparing social networks in the two courses

Social network properties for the three periods in the two courses were compared to account for differences in the learning context. First, in both courses the social networks decreased in size over time, but with different patterns (see Table 5.3). For StatLearn, the number of nodes and edges dropped slightly in T2 then dramatically in T3 indicating that major decrease in network size occurred in T3. In SciWrite the number of nodes and edges dropped dramatically in T2, then dropped further in T3, indicating that major decrease in network size started in T2. Changes in network size showed the same pattern as changes in the quantity of discussion quantity described in Section 5.1.2.

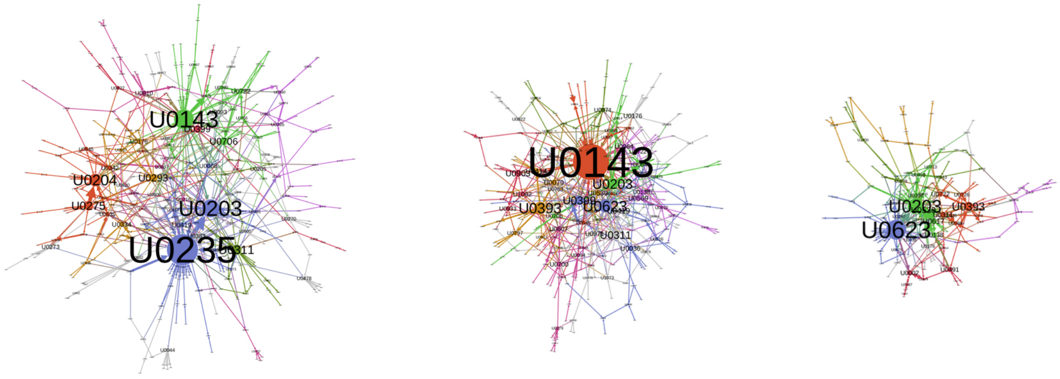
Table 5.3 Global and community network properties

	StatLearn T1	StatLearn T2	StatLearn T3	SciWrite T1	SciWrite T2	SciWrite T3
Number of nodes	403	327	204	447	144	87
Number of isolates	13	17	17	19	6	34
Avg degree with isolates (SD)	3.35 (5.61)	3.52 (6.22)	2.92 (4.59)	3.38 (6.92)	2.60 (4.20)	1.08 (1.30)
Avg degree without isolates (SD)	3.46 (5.66)	3.72 (6.34)	3.19 (4.71)	3.53 (7.03)	2.71 (4.26)	1.77 (1.24)
Number of edges	674	576	298	756	187	47
Avg edge weight (SD)	1.09 (0.38)	1.10 (0.35)	1.15 (0.53)	1.08 (0.37)	1.06 (0.33)	1.06 (0.24)
Number of communities*	20	18	24	23	22	14
Avg size of community*	19.50 (17.64)	17.22 (12.38)	7.79 (9.63)	18.61 (19.79)	6.27 (6.25)	3.79 (2.04)

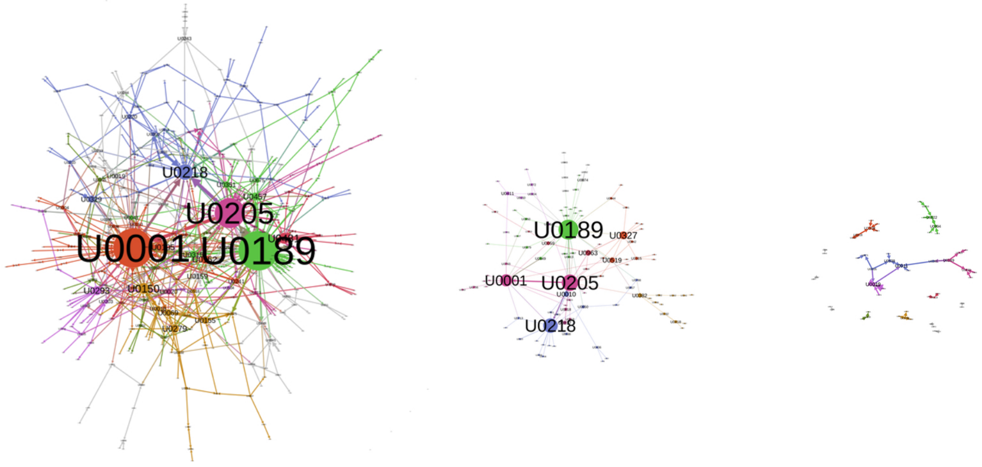
* Excluding one-member communities

Second, in StatLearn the average degree calculated with and without the isolated nodes was relatively stable over time indicating learners in the three time periods had connections with similar numbers of people (see Table 5.3). In SciWrite the average degree in T3 was lower than in T1 and T2, indicating that learners in T3 interacted with fewer people than those in the two preceding time periods. Average edge weight in all networks was slightly over 1 indicating forum participants infrequently had repeated interactions with the same peers.

Finally, the StatLearn T1 network contained multiple central nodes and many inter-community connections indicating that a big number of learners participated across discussion threads. These characteristics largely held in the T2 and T3 networks (see Figures 5.5a to c). In contrast, while the SciWrite T1 network also contained multiple central nodes and many inter-community connections, in the T2 network there were fewer inter-community connections; in the T3 network there were few central nodes and the network became highly disconnected (see Figures 5.5d to f).



a. StatLearn T1 b. StatLearn T2 c. StatLearn T3



d. SciWrite T1 e. SciWrite T2 f. SciWrite T3

Figure 5.5 Network graphs for time periods in the two courses
Node size represents degree. Color indicates community.

5.2.2. Social characteristics

Social characteristics of learners in the social networks of each time period in the two courses were extracted to use in cluster analysis: degree, betweenness, and ego edge weight (see Table 5.4 for a summary of all clustering variables).

5.3. Results for cluster analysis: Identifying positions in discussions

5.3.1. Clustering features

Six contribution features and three social features were extracted for each learner in each of the three time periods in the two courses. Contribution features included number of input seeking messages, number of input providing messages, number of both-seeking-and-providing messages, number of deep messages, number of non-deep messages, and number of depth-switching messages. Social features included degree, betweenness, and ego edge weight.

Histograms of the variables showed the number of both-seeking-and-providing messages had little variance (0 for most people). Thus each of these messages was counted as a seeking message and as a providing message, and added to values of the two variables. The number of depth-switching messages also showed low variance (0 for most people) but was kept for clustering as it cannot be substituted by other variables in characterizing this particular kind of posting behavior.

In SciWrite T3 degree and ego edge weight were highly correlated (Spearman's coefficient = 0.94) but both were kept as they measure different aspects of social connection (breadth vs. strength). This also allowed for using the same variable set for all time periods. Correlation coefficients for other variable pairs were all below 0.8.

Eventually 8 variables were used for cluster analysis and are summarized in Table 5.4. Non-convex distributions were not identified in any of the six data sets. A small number of outliers were identified in each data set and were taken out from the data (see Table 5.5). After outlier removals, the six data sets had 80 to 443 cases, resulting in 10 to 55 cases per variable.

Table 5.4 Clustering variables

		Clustering variables							
Period		Seek	Provide	Deep	Non-deep	Depth switch	Degree	Betweenness	Edge weight
StatLearn T1	Mean	0.8	1.21	0.83	1.15	0.09	3.13	336	1.11
(n = 399)	SD	1.4	2.52	1.82	1.73	0.31	4.36	1252	0.36
StatLearn T2	Mean	1.02	1.62	0.95	1.62	0.07	3.65	329	1.1
(n = 299)	SD	1.71	3.96	1.88	3.02	0.3	6.47	1457	0.44
StatLearn T3	Mean	0.9	1.43	0.62	1.68	0.09	3.13	117	1.06
(n = 175)	SD	1.76	3.48	1.32	3.27	0.3	4.92	396	0.47
SciWrite T1	Mean	0.72	1.45	0.66	1.5	0.05	3.03	245	1.08
(n = 445)	SD	1.1	2.81	1.61	2.03	0.25	4.45	1069	0.38
SciWrite T2	Mean	0.72	0.89	0.29	1.32	0.03	2.53	39	1.09
(n = 112)	SD	0.76	2.01	0.89	1.51	0.16	3.49	177	0.42
SciWrite T3	Mean	0.43	0.84	0.28	0.99	0.03	1.05	0	0.68
(n = 81)	SD	0.63	0.96	0.84	0.8	0.16	1.31	1	0.69

Table 5.5 Data sets for cluster analysis

	Number of cases for cluster analysis	Number of outliers
StatLearn T1	398	1
StatLearn T2	298	1
StatLearn T3	173	2
SciWrite T1	443	2
SciWrite T2	110	2
SciWrite T3	80	1

5.3.2. Number of clusters

The scree plot for StatLearn T1 showed flattening after the 3-cluster solution and the 5-cluster solution (see Figure 5.6a) suggesting an optimal number of clusters could range from 3 to 5. A comparison of the 3-cluster and 4-cluster solutions showed three of the clusters in the 4-cluster solution had similar characteristics to the clusters in the 3-cluster solution while the additional cluster made a substantially higher number of depth switching messages than other clusters. An examination of the 5-cluster solution showed two of the clusters showed negligible differences for all variables. Thus the 4-cluster solution was selected. It accounted for 56% of the total variance.

The scree plot for StatLearn T2 showed flattening after the 6-cluster solution (see Figure 5.6b) suggesting an optimal number of clusters could be 6. An examination of cluster centroids revealed two of the six clusters posted a minimal number of messages and showed negligible differences for all variables except members in one cluster had substantially lower average edge weight in their ego network. These members were

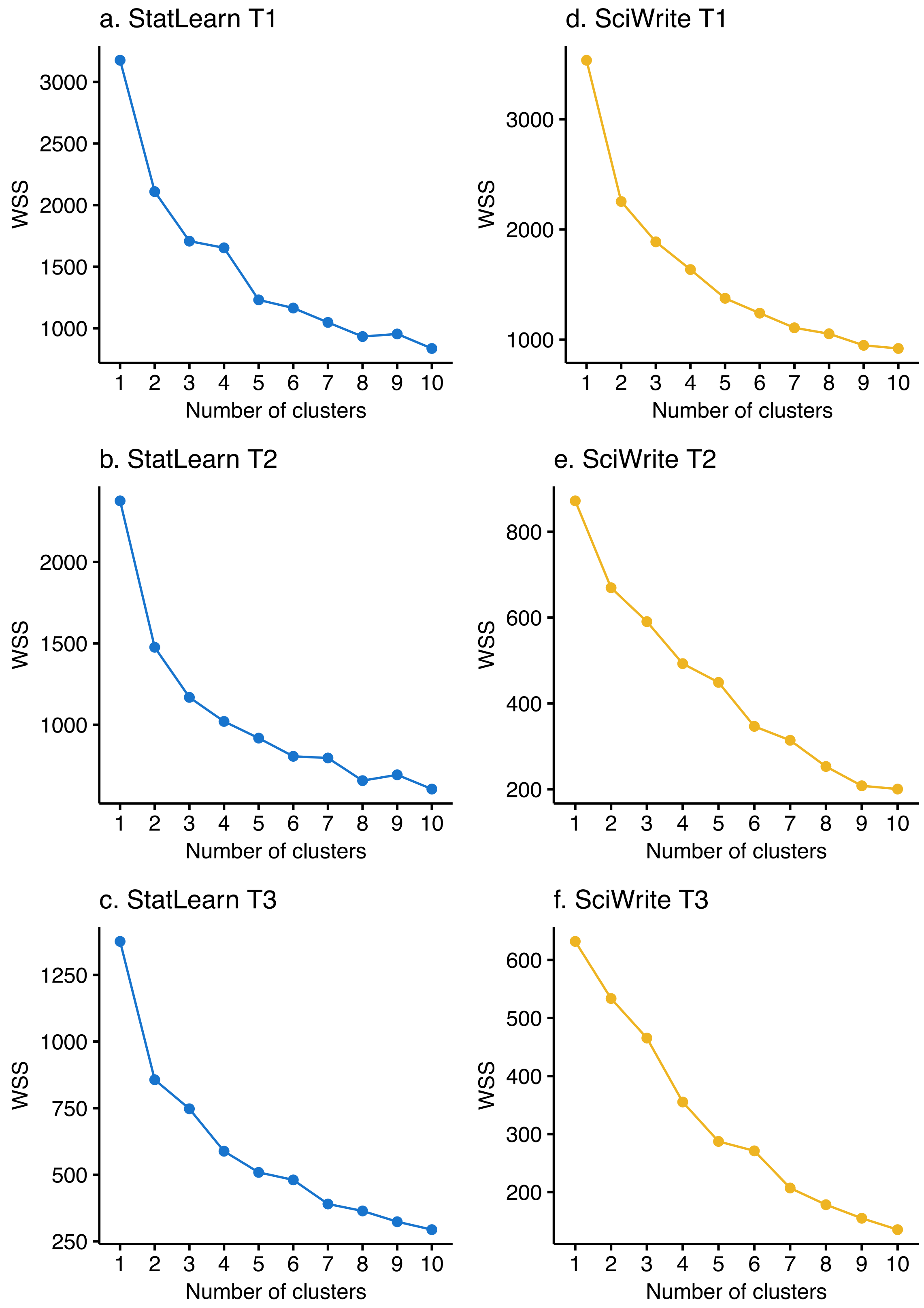


Figure 5.6 **Scree plots for the cluster analysis**

isolated nodes in the social network. Given that isolates existed in all time periods but did not always appear as a cluster in solutions with a low number of clusters, a decision was made not to investigate them as a separate cluster in this time period. An examination of the 5-cluster solution revealed the clusters all had distinct characteristics. Thus the 5-cluster solution was selected. It accounted for 64% of the total variance.

The Scree plot for StatLearn T3 showed flattening after the 5-cluster solution (see Figure 5.6c) suggesting an optimal number of clusters could be 5. An examination of cluster centroids revealed two clusters posted a minimal number of messages and showed negligible differences for all variables except one cluster had substantially lower average edge weight in the ego network and were found out to be isolates. For the same reason discussed above, these isolates were not investigated as a separate cluster. An examination of the 4-cluster solution revealed the clusters all had distinct characteristics. Thus the 4-cluster solution was selected. It accounted for 59% of the total variance.

The scree plot for SciWrite T1 showed flattening after the 5-cluster solution (see Figure 5.6d) suggesting an optimal number of clusters could be 5. An examination of cluster centroids revealed two of the clusters showed negligible differences for all variables. An examination of the 4-cluster solution revealed the clusters all had distinct characteristics. Thus the 4-cluster solution was selected. It accounted for 56% of the total variance.

The scree plot for SciWrite T2 showed flattening after the 4-cluster solution (see Figure 5.6e) suggesting an optimal number of clusters could be 4. An examination of the 4-cluster solution revealed the clusters all had distinct characteristics. Thus the 4-cluster solution was selected. It accounted for 49% of the total variance.

The scree plot for SciWrite T3 showed flattening after the 5-cluster solution (see Figure 5.6f) suggesting an optimal number of clusters could be 5. An examination of cluster centroids revealed the clusters all had distinct characteristics. Thus the 5-cluster solution was selected. It accounted for 57% of the total variance.

The distribution of cluster members on the first two principal components (automatically computed by fviz_cluster function) is shown in Figure 5.7 to illustrate the segregation between clusters. In all time periods the identified clusters varied substantially in size and compactness. Clusters with more members were more compact

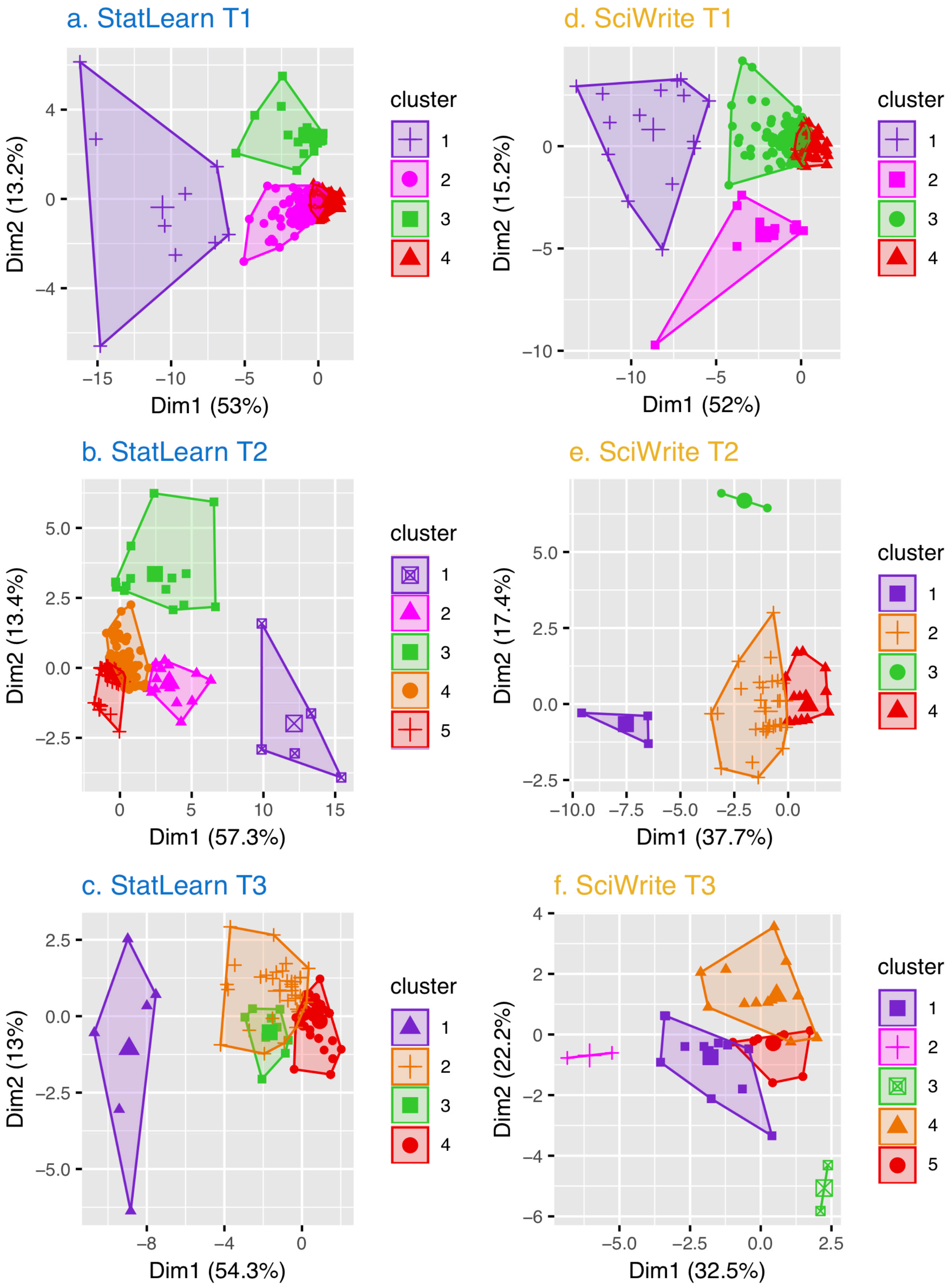


Figure 5.7 Distribution of cluster members on two principal component coordinates

than those with less members. In addition, clusters with more members tended to overlap in the projection while the smaller clusters tended to be disjointed.

5.3.3. Summarizing position characteristics

The number of clusters identified in each time period, size of each cluster, and centroid variable values for each cluster are summarized in Table 5.6. Cluster profiles are visualized in Figure 5.8. Clustering results for the six time periods showed the identified positions can be summarized based on characteristics in five aspects: total content contribution quantity (indicated by the sum of deep messages and non-deep messages, see Table 5.6), input providing and input seeking activities, deep consideration of the discussion content, connectedness in the social network, and strength of social connections. Specific characteristics of the five aspects are summarized below. A five-letter profile label is assigned for each position with each letter representing one characteristic. Profile labels for the identified positions in the six time periods are summarized in Table 5.6.

Aspect 1: Total content contribution quantity

Based on the total number of content messages posted, positions identified in the six time periods can be summarized as having high (H), medium (M), or low (L) contribution quantity. StatLearn T1, StatLearn T2, StatLearn T3, SciWrite T1, and SciWrite T2 each contained a cluster that posted less than 1.5 messages per person indicating a majority of the cluster members were single-message contributors (see Table 5.6). These single-message posters clusters were considered to have low contribution quantity in comparison to other clusters in the same time period. In contrast to the low quantity cluster, each of the five time periods also contained a cluster that contributed substantially more messages than other clusters (contribution quantity 2.7 to 4.4 standard deviations higher than the second highest value in the same time period, see Figure 5.8). These clusters were considered to have high contribution quantity in comparison to other clusters in the same time period. The rest of the clusters in each time period had similar contribution quantity (z-scores for contribution quantity ranged between 0.10 and 1.68, see Figure 5.8) and thus were considered to have medium contribution quantity. In SciWrite T3 none of the five clusters contributed a substantially higher number of messages than other clusters. The two clusters with the lowest contribution quantity were single-message posters clusters and thus were considered to have low contribution quantity. The other three clusters posted similar numbers of

Table 5.6 Clustering results

Time period	Clusters	# of learners (%)	Positions	T* Mean (SD)	S* Mean (SD)	P* Mean (SD)	%P	D* Mean (SD)	ND* Mean (SD)	%D	DS* Mean (SD)	DE* Mean (SD)	B* Mean (SD)	E* Mean (SD)
StatLearn T1	Cluster 1	9 (2%)	HPNHR	14.56	5.22	9.56	65%	6.33	8.22	44%	0.56	20.78	5228	1.37
			(Enthusiastic central providers)	(5.32)	(4.49)	(6.33)		(3.94)	(3.67)		(0.73)	(9.95)	(2740)	(0.23)
	Cluster 2	88 (22%)	MENLR	3.28	1.50	1.84	55%	1.31	1.98	40%	0.00	4.98	575	1.53
			(Moderate reciprocators)	(1.77)	(1.57)	(1.45)		(1.48)	(1.51)		(0)	(2.75)	(682)	(0.39)
	Cluster 3	28 (7%)	MPDLR	2.43	0.86	1.61	65%	1.68	0.75	69%	1.07	2.93	190	1.20
			(Deep thinkers)	(2.13)	(1.15)	(2.08)		(1.19)	(1.35)		(0.26)	(2.23)	(407)	(0.3)
	Cluster 4	273 (69%)	LENLN	0.98	0.41	0.57	58%	0.33	0.65	34%	0.00	1.82	54	0.96
			(Minimal reciprocators)	(0.67)	(0.58)	(0.68)		(0.56)	(0.63)		(0)	(1.43)	(202)	(0.22)
StatLearn T2	Cluster 1	5 (2%)	HPNHR	21.40	8.80	14.40	62%	7.00	14.40	33%	0.20	25.80	4281	1.35
			(Enthusiastic central providers)	(5.37)	(7.36)	(8.38)		(3.39)	(5.46)		(0.45)	(4.09)	(889)	(0.11)
	Cluster 2	16 (5%)	MPNLR	8.06	2.44	5.75	70%	3.06	5.00	38%	0.00	10.56	1240	1.27
			(Moderate providers)	(3.11)	(2.45)	(2.93)		(1.91)	(3.18)		(0)	(2.66)	(884)	(0.27)
	Cluster 3	15 (5%)	MPDLR	4.93	1.33	3.93	75%	2.60	2.33	53%	1.07	6.93	716	1.44
			(Deep thinkers)	(3.83)	(1.54)	(3.45)		(1.35)	(2.87)		(0.26)	(5.26)	(925)	(0.78)
	Cluster 4	61 (20%)	MENLR	2.72	1.31	1.44	52%	1.05	1.67	39%	0.00	4.48	294	1.53
			(Moderate reciprocators)	(1.29)	(1.07)	(1.2)		(1.04)	(1.19)		(0)	(2.7)	(373)	(0.46)
	Cluster 5	201 (67%)	LENLN	1.18	0.59	0.60	50%	0.37	0.81	31%	0.00	1.63	30	0.92
			(Minimal reciprocators)	(0.75)	(0.63)	(0.69)		(0.52)	(0.8)		(0)	(1.2)	(91)	(0.28)
StatLearn T3	Cluster 1	6 (3%)	HENHR	12.33	5.83	6.83	54%	4.17	8.17	34%	0.67	13.33	947	1.50
			(Enthusiastic central reciprocators)	(1.86)	(4.58)	(3.97)		(2.32)	(2.4)		(0.52)	(3.83)	(286)	(0.3)
	Cluster 2	37 (21%)	MPNLR	3.03	1.03	2.00	66%	0.65	2.38	21%	0.00	4.86	174	1.54
			(Moderate providers)	(1.62)	(1.38)	(1.7)		(0.89)	(1.57)		(0)	(2.81)	(248)	(0.4)
	Cluster 3	9 (5%)	MPDLR	2.89	0.89	2.00	69%	1.44	1.44	50%	1.11	4.00	108	1.16
			(Deep thinkers)	(0.78)	(0.78)	(0.71)		(0.73)	(1.13)		(0.33)	(2.29)	(144)	(0.19)
	Cluster 4	121 (70%)	LENLN	1.00	0.50	0.51	50%	0.26	0.74	26%	0.00	1.44	9	0.87
			(Minimal reciprocators)	(0.7)	(0.59)	(0.62)		(0.47)	(0.69)		(0)	(1.18)	(38)	(0.37)

*T = total content messages, S = seeking messages, P = providing messages, %P = percentage of providing messages, D = deep messages, ND = non-deep messages, %D = percentage of deep messages, DS = depth-switching messages, DE= degree, B = betweenness, E = ego edge weight

Table 5.6 Clustering results (continued)

Time period	Clusters	# of learners (%)	Positions	C* Mean (SD)	S* Mean (SD)	P* Mean (SD)	%P	D* Mean (SD)	ND* Mean (SD)	%D	DS* Mean (SD)	DE* Mean (SD)	B* Mean (SD)	E* Mean (SD)
SciWrite T1	Cluster 1	15 (3%)	HPNHR	12.13	2.93	9.33	76%	3.80	8.33	31%	0.07	15.67	2270	1.37
			(Enthusiastic central providers)	(3.14)	(2.25)	(3.98)		(2.18)	(3.06)		(0.26)	(3.13)	(1047)	(0.23)
	Cluster 2	14 (3%)	MPDLR	3.79	0.43	3.36	89%	2.07	1.71	55%	1.07	5.29	452	1.26
			(Deep thinkers)	(2.83)	(0.76)	(2.41)		(2.43)	(1.2)		(0.27)	(2.89)	(546)	(0.34)
	Cluster 3	103 (23%)	MENLR	2.87	1.15	1.75	60%	0.98	1.89	34%	0.00	4.40	360	1.42
			(Moderate reciprocators)	(1.53)	(1.16)	(1.51)		(1.08)	(1.35)		(0)	(2.13)	(561)	(0.48)
	Cluster 4	311 (70%)	LENLN	1.17	0.45	0.72	62%	0.24	0.93	20%	0.00	1.58	16	0.94
			(Minimal reciprocators)	(0.52)	(0.56)	(0.71)		(0.46)	(0.62)		(0)	(1.04)	(70)	(0.24)
SciWrite T2	Cluster 1	3 (3%)	HENHR	4.67	2.00	3.00	60%	1.00	3.67	21%	0.00	9.33	327	1.54
			(Enthusiastic central reciprocators)	(2.89)	(1.73)	(3)		(1)	(2.08)		(0)	(2.31)	(183)	(0.32)
	Cluster 2	2 (2%)	MPDLR	2.00	0.50	1.50	75%	1.50	0.50	75%	1.00	3.00	50	1.25
			(Deep thinkers)	(1.41)	(0.71)	(0.71)		(0.71)	(0.71)		(0)	(0)	(59)	(0.35)
	Cluster 3	36 (33%)	MENLR	2.00	0.92	1.08	54%	0.33	1.67	17%	0.00	3.08	18	1.35
			(Moderate reciprocators)	(0.72)	(0.87)	(0.84)		(0.53)	(0.79)		(0)	(1.4)	(33)	(0.47)
	Cluster 4	69 (63%)	LENLN	0.88	0.54	0.35	39%	0.06	0.83	7%	0.00	1.29	2	0.91
			(Minimal reciprocators)	(0.32)	(0.5)	(0.48)		(0.24)	(0.38)		(0)	(0.79)	(11)	(0.28)
SciWrite T3	Cluster 1	13 (16%)	MPNLR	2.15	0.38	1.77	82%	0.54	1.62	25%	0.00	2.31	0	1.49
			(Moderate providers)	(0.9)	(0.51)	(0.6)		(0.97)	(1.12)		(0)	(1.11)	(1)	(0.84)
	Cluster 2	2 (3%)	MPNHR	2.00	0.00	2.00	100%	0.00	2.00	0%	0.00	5.50	8	1.53
			(Moderate central providers)	(0)	(0)	(0)		(0)	(0)		(0)	(0.71)	(2)	(0.39)
	Cluster 3	2 (3%)	MPDLN	1.50	0.00	1.50	100%	1.50	0.00	100%	1.00	0.00	0	0.00
			(Deep thinkers)	(0.71)	(0)	(0.71)		(0.71)	(0)		(0)	(0)	(0)	(0)
	Cluster 4	25 (31%)	LSNLN	1.28	1.20	0.08	6%	0.08	1.20	6%	0.00	0.80	0	0.39
			(Minimal seekers)	(0.54)	(0.5)	(0.28)		(0.28)	(0.58)		(0)	(1.29)	(1)	(0.53)
	Cluster 5	38 (48%)	LPNLN	0.79	0.00	0.79	100%	0.13	0.66	16%	0.00	0.63	0	0.61
			(Minimal providers)	(0.53)	(0)	(0.53)		(0.34)	(0.58)		(0)	(0.54)	(0)	(0.5)

*T = total messages, S = seeking messages, P = providing messages, P% = percentage of providing messages, D = deep messages, ND = non-deep messages, %D = percentage of deep messages, DS = depth-switching messages, DE= degree, B = betweenness, E = ego edge weight

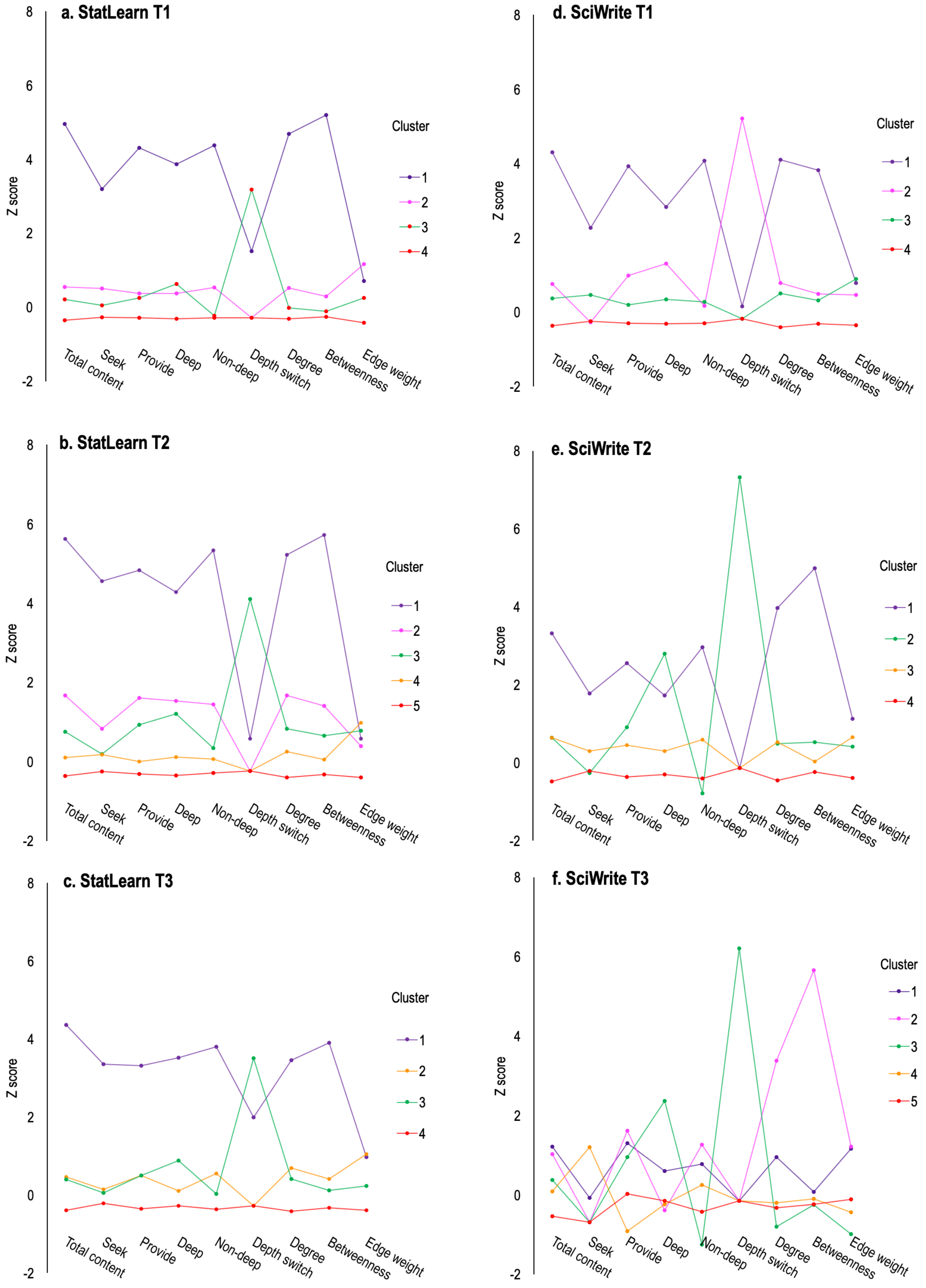


Figure 5.8 Cluster profiles

Total number of content messages was computed by summing deep and non-deep messages; it was included in the profile to assist cluster interpretation. All values are shown in z-score to account for differences in measurement units.

messages (z-scores for contribution quantity ranged between 0.37 and 1.22, see Figure 5.8) and were considered to have medium contribution quantity.

In all time periods except SciWrite T3, there was a small cluster of learners with high contribution quantity (2%-3%). A higher proportion of learners (22%-35%) were in clusters with medium contribution quantity. The majority of learners (63%-70%) were in clusters with low contribution quantity (see Table 5.6). In SciWrite T3 approximately half learners were in clusters with medium and low contribution quantity respectively (see Table 5.6).

Aspect 2: Input seeking and input providing activities

Based on the proportion of input seeking and input providing activities, positions identified in the six time periods can be summarized as input providers (P), input seekers (S), or reciprocators (E). StatLearn T1, StatLearn T2, StatLearn T3, SciWrite T1, and SciWrite T2 each contained some clusters distinguished by the high proportion of input providing messages they posted (62%-89%) compared to other clusters in the same period (see Table 5.6). Thus, these clusters were considered to be input providers and other clusters in the same time period (proportion of input providing messages 39%-62%, see Table 5.6) were considered to be reciprocators. In SciWrite T3, four of the five clusters were input providers; in contrast, Cluster 4 posted an extremely low proportion of input providing messages (6%) thus was considered to be input seekers.

In all time periods except SciWrite T3, the identified clusters showed either the input provider characteristic or the reciprocator characteristic. The input provider characteristic was identified in a smaller proportion of learners (2%-26%) in comparison to the reciprocator characteristic (see Table 5.6). In SciWrite T3, 69% learners showed the input provider characteristic and the other 31% showed the input seeker characteristic.

Aspect 3: Deep consideration of the discussion content

Based on deep consideration of the discussion content contained in contributions, positions identified in the six time periods can be summarized as deep thinkers (D) and non-deep thinkers (N). In each of the six time periods one of the clusters not only posted a high proportion of deep messages (50% to100%) but also posted at least one depth switching message (1 to 1.11 per person, see Table 5.6).

These clusters were considered to have the characteristics of deep thinkers. In contrast, other clusters all posted a lower proportion of deep messages (0% to 44%) and less than one depth switching message (0 to 0.67 per person, see Table 5.6); thus they were considered to not have such characteristics. The deep thinker characteristics were identified in a small proportion of learners (2% to 5%, see Table 5.6).

Aspect 4: Connectedness in the social network

Based on degree and betweenness, positions identified in the six time periods can be summarized as having high (H) or low (L) social centrality. In each of the six time periods there was a cluster that had substantially higher degree than other clusters (degree 2.4 to 4.2 standard deviations higher than the second highest value in the same time period, see Figure 5.8). These clusters were considered to have high degree centrality in comparison to other clusters in the same time period. Degree centrality for the remaining clusters in each time period were similar (z-scores for degree ranged between -0.81 to 1.67, see Figure 5.8). Thus these clusters were considered to have low degree centrality. In addition, for all clusters in all time periods the distribution of betweenness centrality showed the same pattern as the degree centrality. The clusters with high degree centrality all had high betweenness centrality (betweenness 3.3 to 5.6 standard deviations higher than the second highest value in the same time period, see Figure 5.8). Those with low degree centrality all had low betweenness centrality (z-scores for betweenness ranged between -0.33 to 1.40, see Figure 5.8). Thus, the two centrality indices were interpreted together as overall centrality. High centrality was only identified in a small proportion of learners (2-3%, see Table 5.6).

Aspect 5: Strength of social connections

Based on average edge weight in the ego network, positions identified in the six time periods can be summarized as having repeated interactions (R) or one-off interactions (N) with peers. Each of the six time periods contained at least one cluster whose average edge weight in ego network did not exceed 1 (see Table 5.6) indicating these learners did not have repeated interactions with the same peers in their ego network. Thus, these clusters were considered to have one-off interactions in discussion. For the rest of the clusters, the average edge weight in the ego network ranged between 1.16 and 1.54 (see Table 5.6) indicating learners in these clusters had some (yet still infrequent) repeated interactions with the same peers. The characteristic of having repeated interactions with peers was identified in a smaller proportion of learners (19% to 37%) compared to the one-off interaction characteristic (see Table 5.6).

5.3.4. What positions were identified in each time period?

Characteristics of positions identified in each of the six time periods are described below. A five-letter label and a profile name that highlighted the key characteristics were assigned to each cluster.

Positions in StatLearn T1 (see Figure 5.8a and Table 5.6)

Cluster 1: HPNHR (Enthusiastic central providers)

Two percent of learners were grouped into Cluster 1. They were distinctive for their substantially higher total content contribution quantity (n = 14.50) indicated by the sum of deep messages and non-deep messages and social centrality indicated by degree and betweenness. Their total contribution quantity, degree, and betweenness were respectively 4.4, 4.2, and 4.9 standard deviations higher than the second highest values in this time period. Thus, they were considered to be enthusiastic central contributors. In addition, their contributions contained a higher proportion of input providing messages (65%) than Cluster 2 (55%) and Cluster 4 (58%) in this time period. Thus, they were considered to be primarily input providers. These learners posted slightly less deep messages than non-deep messages, and posted less than one depth-switching message in discussion. They had some repeated interaction with the same peers (ego edge weight = 1.37).

Cluster 2: MENLR (Moderate reciprocators)

Twenty-two percent of learners were grouped into Cluster 2. These learners' contribution quantity (n = 3.28) was substantially lower than Cluster 1 but considerably higher than the lowest value in this time period (Cluster 5, n = 0.98). Thus they were considered to be moderate contributors. Their messages distributed relatively evenly between input providing and seeking (55% vs. 45%), thus they were considered to be reciprocators in discussion. They posted fewer deep messages than non-deep messages, and did not post any depth-switching message. They were considered to be peripheral in the social network as their degree and betweenness were both substantially lower than Cluster 1 and were not distinguishable from other clusters in this time period. They had some repeated interactions with the same peers.

Cluster 3: MPDLR (Moderate deep thinkers)

Seven percent of learners were grouped into Cluster 3. Their total contribution quantity (n = 2.43) was similar to Cluster 2 and so they were also considered to be moderate contributors. They were the only group that made a substantially higher proportion of deep messages than non-deep messages (69% vs. 31%) and posted more than one depth-switching message. They were thus considered to be deep thinkers in discussion. Their contributions contained a higher proportion of input providing messages (65%) than Cluster 2 (55%) and Cluster 4 (58%) in this time period. Thus, they were considered to be primarily input providers. Like Clusters 2, they were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 4: LENLN (Minimal peripheral reciprocators)

Sixty-nine percent of learners were grouped into Cluster 4. These learners were distinctive in that the majority of them only posted a single message (n = 0.98) during this time period. Thus, they were considered to be minimal contributors in discussions. Another distinctive characteristic of these learners was they did not interact repeatedly with the same peers (ego edge weight = 0.96) and thus were considered to have one-off interactions in discussions. The proportion of input providing and input seeking messages posted by members of this group were relatively even (58% vs. 42%). These learners were considered to be reciprocators. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 2 and 3, they were peripheral in the social network.

Positions in StatLearn T2 (see Figure 5.8b and Table 5.6)

Cluster 1: HPNHR (Enthusiastic central providers)

Two percent of learners were grouped into Cluster 1. They were distinctive for their substantially higher total contribution quantity (n = 21.40) and central status in the social network. Their total contribution quantity, degree, and betweenness were respectively 3.9, 3.5, and 4.3 standard deviations higher than the second highest values in this time period. In addition, their contributions contained a higher proportion of input providing messages (62%) than Cluster 4 (52%) and cluster 5 (50%) in this time period

(61% providing messages). They posted fewer deep messages than non-deep messages and made less than one depth-switching message during this time period. They had some repeated interactions with the same peers.

Cluster 2: MPNLR (Moderate providers)

Five percent of learners were grouped into Cluster 2. They had the second highest contribution quantity (*n* = 8.06) in this time period but substantially lower than Cluster 1. A distinctive characteristic of this cluster was the high proportion of input providing messages they made (70%). These learners posted fewer deep messages than non-deep messages, and did not post any depth-switching message. They were peripheral in the social network and had some repeated interactions with the same peers.

Cluster 3: MPDLR (Moderate deep thinkers)

Five percent of learners were grouped into Cluster 3. These learners were distinctive in that they posted a high proportion of deep messages (53%) and made more than one depth-switching message. In addition, their contribution contained a higher proportion of input providing messages (75%) than other clusters in this time period; thus they were considered to be primarily input providers. Their total contribution quantity (*n* = 4.93) was lower than Cluster 2. They were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 4: MENLR (Moderate reciprocators)

Twenty percent of learners were grouped into Cluster 4. These learners were similar to Cluster 2 except that their messages were evenly distributed between input providing and seeking (52% vs.48%) thus these learners were considered to be reciprocators. In addition, their contribution quantity (*n* = 2.72) was also lower than Cluster 2.

Cluster 5: LENLN (Minimal peripheral reciprocators)

Sixty-seven percent of learners were grouped into Cluster 5. These learners were distinctive for being single-message contributors and having one-off interaction. The proportions of input providing and input seeking messages posted by this group were even (50% vs.50%); thus these learners were considered to be reciprocators. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 2, 3, and 4, they were peripheral in the social network.

Positions in StatLearn T3 (see Figure 5.8c and Table 5.6)

Cluster 1: HENHR (Enthusiastic central reciprocators)

Three percent of learners were grouped into Cluster 1. They were distinctive for their substantially higher total contribution quantity (*n* = 12.33) and central status in the social network; their total contribution quantity, degree, and betweenness were respectively 3.9, 2.8, and 3.5 standard deviations higher than the second highest values in this time period. They posted similar proportions of input providing and seeking messages (54% vs.46%) and were considered to be reciprocators. They posted fewer deep messages than non-deep messages, and made less than one depth-switching message during this time period. They had some repeated interaction with the same peers.

Cluster 2: MPNLR (Moderate providers)

Twenty-one percent of learners were grouped into Cluster 2. They had the second highest contribution quantity (n =3.03) in this time period, but still substantially lower than Cluster 1. Their contribution contained a higher proportion of input providing messages (66%) than Cluster 1 (54%) and Cluster 4 (50%) in this time period, thus they were considered to be primarily input providers. These learners posted fewer deep message than non-deep messages, and did not post any depth-switching message. They were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 3: MPDLR (Moderate deep thinkers)

Five percent of learners were grouped into Cluster 3. These learners were distinctive in that they posted a high proportion of deep messages (50%) and made more than one depth-switching message; thus they were considered to be deep thinkers. Their contribution contained a higher proportion of input providing messages (69%) than other clusters in this time period, thus they were considered to be primarily input providers. Similar to Cluster 2, they made a medium amount of contribution (*n* = 2.89), had peripheral social status, and had some repeated interaction with peers.

Cluster 4: LENLN (Minimal peripheral reciprocators)

Seventy percent of learners were grouped into Cluster 4. These learners were distinctive for being single-message contributors and having one-off interaction. The proportions of input providing and input seeking messages posted by this group was even (50% vs.50%); thus these learners were considered to be reciprocators. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 2 and 3, they were peripheral in the social network.

Positions in SciWrite T1 (see Figure 5.8d and Table 5.6)

Cluster 1: HPNHR (Enthusiastic central providers)

Three percent of learners were grouped into Cluster 1. They were distinctive for their substantially higher total contribution quantity (*n* =12.13) and central status in the social network; their total contribution quantity, degree, and betweenness were respectively 3.6, 3.3, and 3.3 standard deviations higher than the second highest values in this time period. In addition their contribution contained a higher proportion of input providing messages (76%) than Cluster 3 (60%) and Cluster 4 (62%) in this time period and thus were considered to be primarily input providers. They posted fewer deep messages than non-deep messages, and made less than one depth-switching message. They had some repeated interaction with the same peers.

Cluster 2: MPDLR (Moderate deep thinkers)

Three percent of learners were grouped into cluster 2. These learners posted a high proportion of deep messages (55%) and made at least one depth-switching message; thus they were considered to be deep thinkers. In addition, their contribution contained a higher proportion of input providing messages (89%) than other clusters in this time period and thus were considered to be primarily input providers. They had the second highest total contribution quantity (*n* = 3.79) in this period, but still substantially lower than Cluster 1. Cluster 2 learners were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 3: MENLR (Moderate reciprocators)

Twenty-three percent of learners were grouped into Cluster 3. Their contribution quantity (*n* = 2.87) was slightly lower than Cluster 2. They posted a lower proportion of input providing messages (60%) than Clusters 1 and 2, thus were considered to be reciprocators. They posted fewer deep messages than non-deep messages, and did not post any depth switching message. They had peripheral social status and some repeated interaction with the same peers.

Cluster 4: LENLN (Minimal peripheral reciprocators)

Seventy percent of learners were grouped into Cluster 4. These learners were distinctive for being single-message contributors and having one-off interaction. They posted a lower proportion of input providing messages (62%) than Clusters 1 and 2, and were considered to be reciprocators. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 2 and 3, they were peripheral in the social network.

Positions in SciWrite T2 (see Figure 5.8e and Table 5.6)

Cluster 1: HPNHR (Enthusiastic central reciprocators)

Three percent of learners were grouped into Cluster 1. They were distinctive for their substantially high total contribution quantity (*n* = 4.67) and central status in the social network; their total contribution quantity, degree, and betweenness were respectively 2.7, 3.5, and 4.5 standard deviations higher than the second highest values in this time period. Their contribution contained a lower proportion of input providing messages (60%) than Cluster 2 (75%) and thus were considered to be reciprocators. They posted fewer deep messages than non-deep messages, and did not post any depth-switching message. They had some repeated interaction with the same peers.

Cluster 2: MPDLR (Moderate deep thinkers)

Two percent of learners were grouped into Cluster 2. These learners were distinctive in that they posted a high proportion of deep messages (75%) and made at least one depth-switching message; thus they were considered to be deep thinkers. In addition, they posted a higher proportion of input providing messages (75%) than other clusters in this time period and thus were considered to be primarily input providers. They had the second highest total contribution quantity (*n* = 2.00) in this period, but substantially lower than Cluster 1. Cluster 2 learners were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 3: MENLR (Moderate reciprocators)

Thirty-three percent of learners were grouped into Cluster 3. These learners they posted similar proportions of input providing and input seeking messages (54% vs.46%) and thus were considered to be reciprocators. They made the same amount of contribution (*n* = 2.00) as Cluster 2. They made fewer deep messages than non-deep messages, and did not post any depth-switching message. They were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 4: LENLN (Minimal peripheral reciprocators)

Sixty-three percent of learners were grouped into Cluster 4. These learners were distinctive for being single-message contributors and having one-off interaction. They posted a lower proportion of input providing messages (39%) than other clusters in this period and thus were considered to be reciprocators. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 2 and 3, they were peripheral in the social network.

Positions in SciWrite T3 (see Figure 5.8f and Table 5.6)

Cluster 1: MPNLR (Moderate providers)

Sixteen percent of learners were grouped into Cluster 1. They had the highest contribution quantity (n = 2.15) in this time period, but only slightly higher than the cluster with the second highest value (Cluster 2, *n* = 2.00). They posted a higher proportion of input providing messages (82%) than Cluster 4 (6%) in this time period and thus were considered to be primarily input providers. These learners posted fewer deep messages than non-deep messages, and did not post any depth-switching message. They were peripheral in the social network and had some repeated interaction with the same peers.

Cluster 2: MPNHR (Moderate providers with high centrality)

Three percent of learners were grouped into Cluster 2. They only posted input providing messages and thus were considered to be input providers. These learners posted fewer deep messages than non-deep messages, and did not post any depth-switching message. They had repeated interaction with the same peers. They were similar to Cluster 1 except for their central social status; their degree and betweenness were respectively 2.4 and 5.6 standard deviations higher than the second highest values in this time period.

Cluster 3: MPDLN (Moderate deep thinkers)

Three percent of learners were grouped into Cluster 3. Their total contribution (*n* = 1.50) was lower than Cluster 2. They were distinctive in that they only posted deep messages and made one depth-switching message in discussion; thus they were considered to be deep thinkers. All of their messages were posted to provide input thus they were considered to be input providers. Learners in this cluster were isolates in the social network and did not have connection with any peer.

Cluster 4: LSNLN (Minimal peripheral seekers)

Thirty-one percent of learners were grouped into Cluster 4. These learners were distinctive for being single-message contributors and having one-off interaction. As a group they posted 6% input providing messages and 94% input seeking messages; thus they were considered to be primarily input seekers. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. They were peripheral in the social network.

Cluster 5: LPNLN (Minimal peripheral providers)

Forty-eight percent of learners were grouped into Cluster 5. These learners were distinctive for being single-message contributors and having one-off interaction. They only posted input providing messages and thus were considered to be input providers. They posted fewer deep messages than non-deep messages and did not post any depth-switching message. Like Clusters 1, 3 and 4, they were peripheral in the social network.

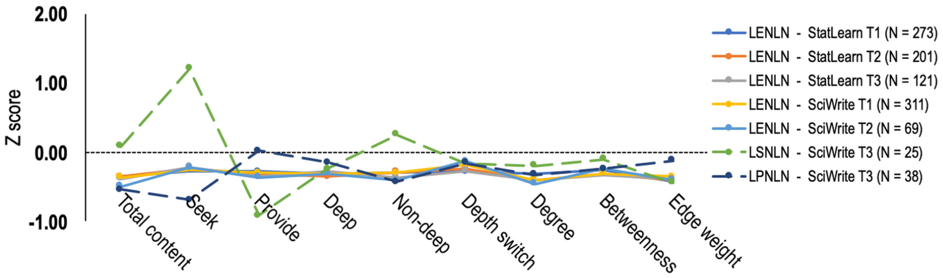
5.3.5. Answering research question 1: (a) What common learner positions are found in MOOC forums? (b) Which of the positions are found across courses and time periods?

To present an overview of positions in discussion forums, the positions identified in the six time periods were summarized into six primary types based on the key characteristics. Figure 5.9 displays visualized position profiles.

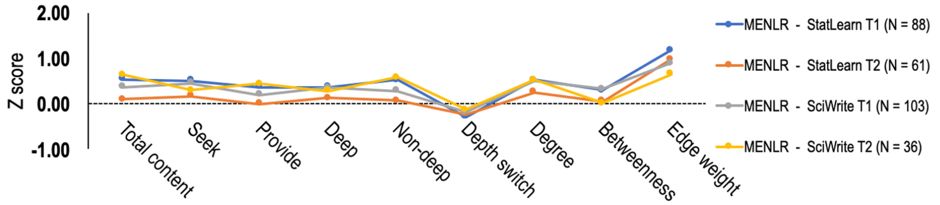
Minimal peripheral contributors (reciprocators, seekers, providers) – LENLN / LSNLN/ LPNLN

Minimal peripheral contributors were identified in all time periods (see Figure 5.

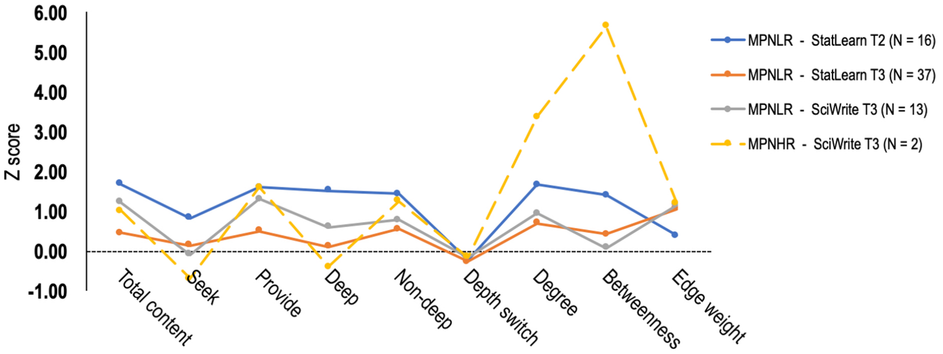
a. Minimal peripheral contributors



b. Moderate pheripheral reciprocators



c. Moderate peripheral providers



d. Moderate peripheral deep thinkers

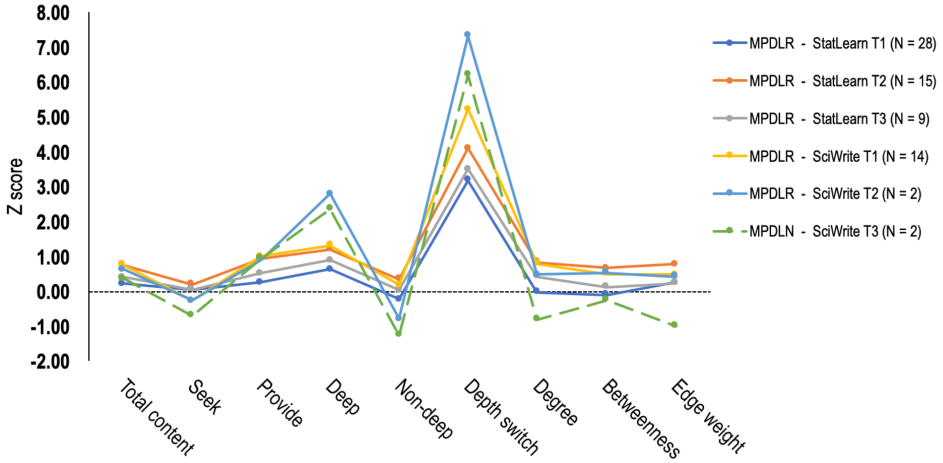
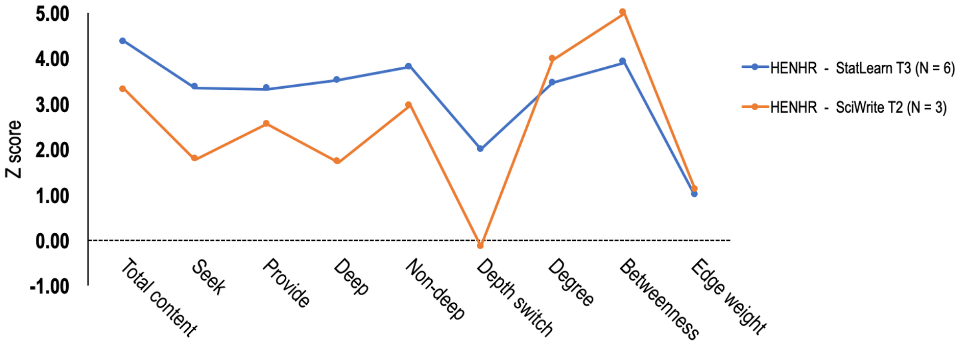


Figure 5.9 Postion profiles grouped based on key characteristics
Total number of content messages was computed by summing deep and non-deep messages; it was included in the profile to assist cluster interpretation. All values are shown in z-score to account for differences in measurement units.

e. Enthusiastic central reciprocators



f. Enthusiastic central providers

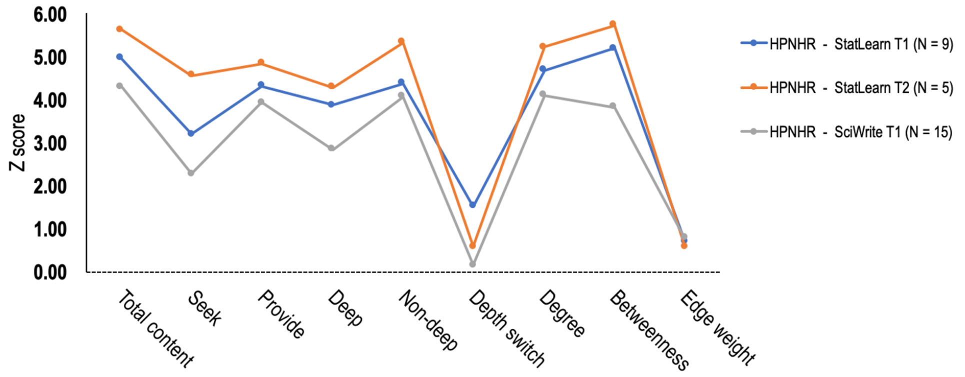


Figure 5.9 Postion profiles grouped based on key` characteristics (continued)

9a). They were non-deep thinkers, peripheral in social networks, and had one-off interactions with peers. The minimal contributors identified in all time periods except SciWrite T3 were reciprocators. In SciWrite T3, two minimal contributor clusters were identified showing the characteristics of providers and seekers.

Moderate reciprocators - MENLR

Moderate reciprocators were identified in four of the six time periods (see Figure 5.9b). They had medium contribution quantity, and sought and provided input relatively evenly. They were non-deep thinkers, peripheral in social networks, and had some repeated interactions with the same peers.

Moderate providers – MPNLR / MPNHR

Moderate providers were identified in three of the six time periods (see Figure 5.9c, two of the four groups were from SciWrite T3). They had medium contribution quantity, and mainly provided input in discussions. They were non-deep thinkers and had some repeated interactions with the same peers. Moderate providers were generally peripheral in social networks, except the group identified in SciWrite T3 that had high centrality.

Moderate deep thinkers – MPDLR / MPDLN

Moderate deep thinkers were identified in every time period (see Figure 5.9d). They posted a high proportion of deep messages and added deep messages to on-going non-deep discussions. Deep thinkers had medium contribution quantity and peripheral status in social networks. They generally had some repeated interactions with the same peers, except that the group in SciWrite T3 were isolates.

Enthusiastic central reciprocators – HENHR

Enthusiastic central reciprocators were identified in two of the six time periods (see Figure 5.9e). They had high contribution quantity, provided and sought input evenly, and did not show deep thinker characteristics. They were central in the social networks and had some repeated interactions with the same peers.

Enthusiastic central providers – HPNHR

Enthusiastic central providers were identified in three of the six time periods (see Figure 5.9f). These learners had high contribution quantity. They mainly provided input in discussion and did not show deep thinker characteristics. They were central in the social networks and had some repeated interactions with the same peers.

5.4. Identifying changes in position and indication of learning

5.4.1. Answering research question 2: What changes occurred in characteristics of individual learner’s position?

Position trajectories and changes in position

In StatLearn among learners examined in the cluster analysis there were 116 2-period learners and 31 3-period learners (three of the 34 3-period learners identified in content discussions were removed as outliers before cluster analysis). In SciWrite among learners examined in cluster analysis there were 57 2-period learners and 13 3-period learners (two of the 15 3-period learners identified in content discussions were removed as outliers before cluster analysis). One trajectory was extracted for each 2-period learner. Two trajectories were extracted for each 3-period learner. This resulted in 178 trajectories from StatLearn and 84 from SciWrite. Among these 262 trajectories a total of 35 unique trajectories were found (see Table 5.7).

The start position and the end position in the same trajectory were compared to identify changes in the five position characteristics: total content contribution quantity (up / same / down), connectedness in the social network (up / down), providing / seeking activities (providing up / same / seeking up), deep consideration of the discussion content (up / same / down), and strength of social connections (up / same / down). Among the 35 unique trajectories, 30 showed changes in one or more aspects while 5 did not show change in any aspect. Trajectory frequency and the changes identified in each aspect are summarized in Table 5.7.

Did learners take different positions over time?

In StatLearn 46% of the total 2-period learners (with 1 trajectory) changed positions while 54% of them did not. In SciWrite 78% of the 2-period learners changed positions while 22% did not (see Table 5.8). Among the small number of 3-period learners (with 2 trajectories) in StatLearn, 58% changed positions in one of the two trajectories; 32% changed positions in both trajectories; 10% did not change position in either trajectory (see Table 5.9). In SciWrite 38% of the 3-period learners changed positions in one of the two trajectories; 62% of them changed positions in both trajectories (see Table 5.9). These results indicate a substantial proportion of MOOC learners in these forums did take different positions over time. In both courses the proportion of learners that showed changes was higher for 3-period learners than for 2-period learners. Looking across 2-period and 3-period learners in the same course, in StatLearn 55% (81 out of 147) showed changes while 45% did not. In SciWrite, 82% (58 out of 71) showed changes while 18% did not.

Table 5.7 Position trajectories and changes in position

	Changes in position					StatLearn Total (n = 178)	SciWrite Total (n = 84)
Trajectories	Quantity	Seek /Provide	Depth	Centrality	Edge weight		
1. LENLN>>LENLN	**Same**	**Same**	**Same**	**Same**	**Same**	**38%**	**15%**
2. MENLR>>LENLN	**Down**	**Same**	**Same**	**Same**	**Down**	**15%**	**7%**
3. LENLN>>MENLR	**Up**	**Same**	**Same**	**Same**	**Up**	**6%**	**10%**
4. MENLR>>MPNLR	**Same**	**P_Up***	**Same**	**Same**	**Same**	**5%**	**7%**
5. LENLN>>MPNLR	**Up**	**P_Up***	**Same**	**Same**	**Up**	**5%**	4%
6. MENLR>>MENLR	**Same**	**Same**	**Same**	**Same**	**Same**	**4%**	**6%**
7. LENLN>>MPDLR	**Up**	**P_Up***	**Up**	**Same**	**Up**	**3%**	1%
8. MPDLR>>MPNLR	**Same**	**Same**	**Down**	**Same**	**Same**	**3%**	1%
9. MPNLR>>MPNLR	**Same**	**Same**	**Same**	**Same**	**Same**	**3%**	**-**
10. HPNHR>>LENLN	**Down**	**S_Up***	**Same**	**Down**	**Down**	1%	**5%**
11. LENLN>>LPNLN	**Same**	**P_Up***	**Same**	**Same**	**Same**	**-**	**11%**
12. MENLR>>LPNLN	**Down**	**P_Up***	**Same**	**Same**	**Down**	**-**	**6%**
13. MENLR>>LSNLN	**Down**	**S_Up***	**Same**	**Same**	**Down**	**-**	**5%**
14. MPDLR>>MPDLR	Same	Same	Same	Same	Same	2%	-
15. MENLR>>MPDLR	Same	P_Up*	Up	Same	Same	2%	1%
16. MPDLR>>MENLR	Same	S_Up*	Down	Same	Same	2%	1%
17. MPDLR>>LENLN	Down	S_Up*	Down	Same	Down	2%	1%
18. MPNLR>>LENLN	Down	S_Up*	Same	Same	Down	2%	-
19. HPNHR>>MPNLR	Down	Same	Same	Down	Same	2%	-
20. MENLR>>HENHR	Up	Same	Same	Up	Same	1%	2%
21. HPNHR>>HPNHR	Same	Same	Same	Same	Same	1%	-
22. MENLR>>HPNHR	Up	P_Up*	Same	Up	Same	1%	-
23. LENLN>>HENHR	Up	Same	Same	Up	Up	1%	-
24. MPNLR>>HENHR	Up	S_Up*	Same	Up	Same	1%	-
25. HPNHR>>HENHR	Same	S_Up*	Same	Same	Same	1%	-
26. HPNHR>>MPDLR	Down	Same	Up	Down	Same	1%	-
27. LENLN>>LSNLN	Same	S_Up*	Same	Same	Same	-	4%
28. HPNHR>>MENLR	Down	S_Up*	Same	Down	Same	-	4%
29. MPDLR>>LPNLN	Down	Same	Down	Same	Down	-	2%
30. LENLN>>MPDLN	Up	P_Up*	Up	Same	Same	-	1%
31. MENLR>>MPDLN	Same	P_Up*	Up	Same	Down	-	1%
32. HENHR>>LPNLN	Down	P_Up*	Same	Down	Down	-	1%
33. HENHR>>LSNLN	Down	S_Up*	Same	Down	Down	-	1%
34. HPNHR>>LSNLN	Down	S_Up*	Same	Down	Down	-	1%
35. HPNHR>>MPNHR	Down	Same	Same	Same	Same	-	1%

P_Up = providing_up, S_Up = seeking_up, boldness indicates top trajectories in the course, "-" indicates trajectory was not found in the course.

Table 5.8 Presence / absence of changes in position (2-period learners)

	StatLearn 2-period learners (n=116)	SciWrite 2-period learners (n=58)
Change in trajectory	46%	78%
No change in trajectory	54%	22%

Table 5.9 Presence / absence of changes in position (3-period learners)

	StatLearn 3-period learners (n=31)	SciWrite 3-period learners (n=13)
Change in one trajectory	58%	38%
Change in both trajectories	32%	62%
No change in either trajectory	10%	0%

Frequency of changes

Looking at frequency of changes in different characteristics, changes in contribution quantity, input seeking / providing activities, and strength of social connections occurred sometimes while changes in deep consideration of content and connectedness in the social network were relatively infrequent (see Table 5.10).

Table 5.10 Frequency of changes in position

Characteristics	Changes	StatLearn trajectories (n = 178)	SciWrite trajectories (n = 84)
Total content messages	Up	16%	18%
	Down	21%	35%
	No change	62%	48%
Seek / provide	Provide_up	16%	33%
	Seek_up	8%	21%
	No change	75%	45%
Depth	Non-deep to deep	6%	5%
	Deep to non-deep	7%	6%
	No change	86%	88%
Centrality	Low to high	2%	2%
	High to low	3%	12%
	No change	95%	86%
Interaction	One-off to repeat	15%	14%
	Repeat to one-off	19%	31%
	No change	66%	55%

Changes in contribution quantity occurred in approximately one third of the trajectories in StatLean and half of the trajectories in SciWrite (see Table 5.10). In StatLearn, the frequency of quantity decreases and quantity increases did not differ dramatically (21% vs. 16%). In SciWrite, quantity decreases occurred twice frequently as quantity increases (35% vs. 18%).

Changes in input seeking and input providing activities occurred in one quarter of the trajectories in StatLearn and approximately half of the trajectories in SciWrite (see Table 5.10). In both courses, increases in input providing activities (16% in StatLearn, 33% in SciWrite) occurred more frequently than increases in input seeking activities (8% in StatLearn, 21% in SciWrite). Specifically, in StatLearn and SciWrite 16% and 32% of the trajectories showed the changes from reciprocators to input providers; 9% and 11% of the trajectories showed the change from input providers to reciprocators. In SciWrite 11% of the trajectories showed the change from input seekers to reciprocators and another 11% showed the change from input providers and reciprocators to input seekers (see Table 5.7).

Changes in strength of social connections occurred in approximately one third of the trajectories in StatLearn and half of the trajectories in SciWrite (see Table 5.10). In StatLearn, the frequency of the change from having repeated interactions to having one-off interaction was similar to the frequency of the changes in the other direction (19% vs. 15%); in SciWrite, the former type of change occurred more frequently than the latter (31% vs. 14%).

Changes in deep consideration of the discussion content occurred in 14% and 12% of the trajectories in StatLearn and SciWrite respectively. In both courses the frequency of increases and decreases in deep consideration were similar (see Table 5.10).

Changes in connectedness in the social network occurred in 5% and 14% of trajectories in StatLearn and SciWrite respectively (see Table 5.10). In StatLearn the movement from central to peripheral status and the movement in the other direction occurred with similar frequency (2% vs. 3%). In SciWrite increases in social centrality occurred dramatically less frequently than decreases (2% vs. 12%).

5.4.2. Answering research question 3: How did the changes in position manifest as common trajectories and represent learning defined as development in one's forum participation pattern?

Frequency rank was used as the criterion for selecting frequent trajectories as the distribution of trajectory frequency by percentage in the two course did not show a natural breaking point. In both courses, when frequency rank dropped below 9 there were multiple trajectories with equivalent rank, each of them only representing 1%-4% of all trajectories in the course. Thus trajectories with top 9 frequency in each course were selected for further examination (see bolded in Table 5.7). The top 9 trajectories in StatLearn occurred 5 to 67 times. They made up 82% of the trajectories in the course. The top 9 trajectories in SciWrite occurred 4 to 13 times. They made up 71% of the trajectories in the course. Considering frequent trajectories from both courses, a total of 13 unique trajectories were selected from the two courses for further examination (see Table 5.7). Examination of the position changes in the frequent trajectories showed they can be grouped into four categories.

First, three of the 13 trajectories showed constructive change combinations in respect to substantive interaction: Trajectory 3 LENLN>>MENLR (Up – Same – Same – Same – Up), Trajectory 5 LENLN>>MPNLR (Up – P_Up – Same – Same – Up), and Trajectory 7 LENLN>>MPDLR (Up – P_Up – Up – Same – Up). For all of these trajectories, contribution quantity increased from low (single message) to medium indicating that learners with these trajectories changed from making minimal contribution to moderate contributions. The increase in contribution quantity was always accompanied by changes in other aspects. For instance, learners with these trajectories all changed from having one-off interaction to repeated interactions with the same peers indicating while posting more messages, they also started to develop stronger social connections. On top of these changes, learners with Trajectory 7 also changed from non-deep thinkers to deep thinkers indicating they switched from mainly engaging in simple information exchange to approaching the discussion content in more sophisticated ways. Together these three trajectories made up 14% of the trajectories in StatLearn and 14% in SciWrite. They were found in 17% (*n* = 35) of the 147 multi-period learners StatLearn and 17% (*n* = 12) of the 71 multi-period learners in SciWrite. It should be noted learners with Trajectories 5 and 7 also changed from reciprocators to input providers indicating that they changed from exchanging knowledge and resources with others to mainly supplying such content. Such changes are considered to be neutral as both positions can contribute to interaction (see an in-depth discussion in Section 6.2.2).

Second, five of the 13 trajectories showed unconstructive change combinations: Trajectory 2 MENLR>>LENLN (Down – Same – Same – Same – Down), Trajectory 13 MENLR>>LSNLN (Down – S_UP – Same – Same – Down), Trajectory 12

MENLR>>LPNLN (Down – P_UP – Same – Same – Down), Trajectory 10 HPNHR>>LENLN (Down – S_UP – Same – Down – Down), and Trajectory 8 MPDLR>>MPNLR (Same – Same – Down – Same – Same). For Trajectory 2, Trajectory 13, Trajectory 12, and Trajectory 10, the changes included decreases in contribution quantity from medium or high to low level and the switch from having repeated interaction with the same peers to having one-off interaction. These changes indicated learners with these trajectories reduced their contributions to the single message level and stopped reinforcing social connections through repeating interactions with the same peers. In addition to these changes, learners with Trajectory 13 also changed from reciprocators to input seekers indicating they switched from engaging in two-way exchange of knowledge and resources to mainly requesting such content without giving back. As for learners with Trajectory 8, they stopped showing deep thinker characteristics indicating they switched from approaching the discussion content in ways associated with deeper learning to mainly engaging in simple exchange in information. In summary, the changes found in these trajectories can be considered unconstructive for having substantive interaction in MOOC discussions. These trajectories made up 19% of the trajectories in StatLearn and 24% in SciWrite.

Third, two of the 13 trajectories contained neutral change combinations: Trajectory 4 MENLR>>MPNLR (Same – P_UP – Same – Same – Same) and Trajectory 11 LENLN>>LPNLN (Same – P_UP – Same – Same – Same). Learners with these trajectories switched from reciprocators to input providers and did not change in other aspects. Together they made up 5% of the trajectories in StatLearn and 18% in SciWrite.

Finally, three of the 13 trajectories did not show changes in any of the five aspects, including Trajectory 1 LENLN>>LENLN (Same – Same – Same – Same – Same), Trajectory 6 MENLR>>MENLR (Same – Same – Same – Same – Same), and Trajectory 9 MPNLR>>MPNLR (Same – Same – Same – Same – Same). These trajectories together made up 46% of the trajectories in StatLearn and 21% in SciWrite.

In summary, only a small number of frequent trajectories showed position changes that potentially indicated development in participation patterns. They made up 14% of the trajectories and 17% of the learners in both courses. The vast majority of the frequent trajectories showed unconstructive changes, neutral changes, or no change at all.

5.4.3. Case studies

To deepen and contextualize understanding of position trajectories and how changes in position may indicate learning, case studies were conducted for 4 trajectories. These included Trajectory 7 that typified trajectories with constructive changes, Trajectory 2 that typified trajectories with unconstructive changes, Trajectory 4 that typified trajectories with neutral changes, and Trajectory 1 that typified trajectories without change.

Case study 1 Trajectories with constructive changes

StatLearn U0644: Trajectory 7 LENLN>>MPDLR (Up – P_Up – Up – Same – Up) from a minimal peripheral reciprocator to a moderate deep thinker

Learner U0644 showed this position trajectory in StatLearn T1 (week 1-3) to T2 (week 4-7). This learner changed participation characteristics in several aspects, including from a minimal contributor to a moderate contributor, from a reciprocator to a provider, from a non-deep thinker to a deep thinker, and from having one-off interaction to having some repeated interactions with the same peers.

In T1 U0644 posted one content message in one thread. In week 2, he/she participated in a thread about the *t*-test analysis method and the rule for rounding decimals for solving a quiz question. After a peer learner and an instructional team member provided answers to the original question, U0644 raised a non-deep follow up question about the rule for rounding, but received no response.

> “Does that mean rounded to one decimal? i.e. if the result is 0.29 (which it isn't!), should we say 0.3 or trim it to 0.2? ... What would be the appropriate response in this case?”

In T1 this learner had direct connections with four people. He/she did not repeat interactions with any of them.

In T2, U0644 not only posted more messages than in T1, but also participated across different threads. In total, this learner posted 6 messages in five threads. He/she had direct connections with 10 people and repeated interactions with 2 of them. Most of this learner’s messages were posted to provide knowledge and resources. For instance, in week 4 he/she participated in a thread initiated by a peer learner seeking input on R programming solutions for a quiz question. U0644 started a subthread with a non-deep message to share his/her solution. He/she also expressed uncertainty about parameter setting in the solution.

> "…I specified R=10, I=100, sim='fixed', n.sim=1000.... I submitted a somewhat central value and got the answer right. However, **I'm wondering if** the parameters I chose are correct or not."

In this thread, U0644 had repeated interactions with a peer learner (U0203) who responded to this post and expressed confusion about getting different results when using U0644's method compared to results of a second method. U0644 responded with a non-deep message that further introduced his/her solution and sought more detail about the peer's method. In response U0203 provided the code for the method. Four other learners joined this subthread and continued to discuss with U0203 the coding solution and the parameter setting. But U0644 did not post to this discussion again.

In T2 U0644 also showed the characteristics of deep thinkers. Half of his/her messages in this time period contained deep consideration of the discussion content, including a depth-switching messages. For instance, in week 7 a peer learner raised a non-deep question on which log (natural, base-2, base-10) should be used for calculating cross-entropy. After five other peers posted non-deep responses, U0644 posted a depth-switching message in which he/she provided detailed explanations of using log in calculating entropy and using log operations in the R programing language. In addition to addressing the peer's question about the specific learning task, U0644 also showed his/her connection with the general domain knowledge by summarizing different research fields' preference for log bases. Notably in this post and other messages posted in this week U0644 used more assertive and confident language compared to earlier weeks. For instance,

> "**Remember**, the use of the entropy is to select how to partition a dataset.… **The fact is** that in R, log is natural, log10 is 10 based and log2 is 2 based. **Simply** doing ?log would tell you that... As a last note, the base used depicts the unit of the result. If you use log2, the result is in bits.... If you use the natural log, the result is in nats.… **From doing many courses**, it seems statistics oriented people tend to use the natural log, while in many big data and computational scenarios, log2 is common."

Furthermore, it is notable this learner often used social presence indicators (e.g., greetings, addressing people by name, see Wise et al. 2004) in this time period. In week

7, he/she also shared personal learning experience indicating a sense of closeness in social interaction.

> "…. when I first came across it, I learned it to be Shannon Entropy…. the base of the log confused me too. Later, I found that the base of the log doesn't matter! - as long as you're consistent."

Case study 2 Trajectories with unconstructive changes

SciWrite U0155: Trajectory 2 HPNHR>>LENLN (Down – S_Up – Same – Down – Down) from an enthusiastic central provider to a minimal peripheral reciprocator

Learner U0155 showed this position trajectory in SciWrite T1 (week 1-3) to T2 (week 4-6). This learner changed participation characteristics in several aspects, including from an enthusiastic contributor to a minimal contributor, from a provider to a reciprocator, from central to peripheral social status, and from having repeated interactions with the same peers to having one-off interaction.

In T1 this learner posted 9 messages in 5 different threads. The focuses of these discussions were similar. They all related to sharing and discussing solutions for sentence editing exercises. Four of U0155's messages were posted to share his/her editing solutions without deep explanation / justification for the editing choices. For instance,

> "I have tried**; any comments for the below**? 1) Anti-inflammatory drugs may prevent Alzheimer's disease. 2) About 0.5% to 2.3% of newborns have clinical seizure.…"

Usually these messages received general and shallow responses from the peers, such as "Your edit is really good". Occasionally when others raised questions about his/her editing decisions, this learner provided explanations. For instance,

> "[Instructor's name], thanks for the comments. Yes. I see your point. On the other hand, the original sentence did not imply they are not related. I put in 'very' only because of the 'specialized knowledge' in the original sentence. Not sure if there is another way to put it to retain the meaning of 'specialized knowledge' or maybe just 'technical' is enough? Cheers."

U0155 also commented on others' editing solutions, but tended to provide general comments (e.g., "I think the last one can be improved") and used inconclusive and uncertain language. For instance,

> “**I don't know if** 'severe' is a better word to replace 'devastating'? The latter sounds better for large scale events, just a matter or preference **I guess**. I agree with [peer learner’s name] about 'kinetic activity'.”

In T1 U0155 often had multiple rounds of exchanges with the same peers in the same discussion. He/she also interacted with the same peers across different threads. He/she had direct connections with 9 people, and repeated interactions with 6 of them.

In T2, this learner only posted one message. He/she initiated a thread to seek input on the use of the present tense. An instructional team member responded to this question.

Case study 3 Trajectories with neutral changes

SciWrite U0418: Trajectory 4 MENLR>>MPNLR (Same – P_UP – Same – Same – Same) from a moderate reciprocator to a moderate provider

U0418 showed this trajectory in SciWrite T2 (week4-6) to T3 (week7-10). Over the two time periods, this learner changed from a reciprocator to a provider. He/she remained as a moderate contributor, non-deep thinker, and peripheral member in the social network. He/she had some repeated interactions with the same peers in both time periods.

This learner’s participation focus and language style did not differ noticeably over time. In both time periods, U0418 posted three messages in two threads to seek and provide input about editing solutions. He/she consistently used social presence indicators in posting (e.g., addressing others by name, expressing gratitude). For instance,

> “For your interest, [TA’s name], here is my edit: ‘We defined the athlete's first injury each season as initial and all other injuries….’”

In T2 this learner had connections with four people, and repeated interaction with one of them. Similarly in T3 this learner had connections with three people, and repeated interaction with one of them.

It is notable that one of this learner’s input seeking messages showed a distinct characteristic rarely found in other input seeking messages in this course. In T3 week 8 this learner initiated a thread to seek feedback for his/her writing. Instead of making a general request for comments, this learner explicitly asked for suggestions on four specific challenges he/she had when drafting this piece of writing.

> “I would greatly appreciate your feedback on an introduction I am writing for the third assignment…. **the particular issues I'm grappling with are**:
> - How to state why the case is important without turning the introduction into a discussion…. My introduction so far: I had just begun treating a 52-year-old alcoholic man…. To anyone that takes time to read it, thank you very much!”

Case study 4 Trajectories with no change

SciWrite U0384: Trajectory 1 LENLN>>LENLN (Same – Same – Same – Same – Same) remaining as a minimal peripheral reciprocator

Learner U0384 showed this trajectory in SciWrite T1 (week 1-3) to T2 (week 4-6). Over the two time periods, this learner showed no change in any of the five aspects. He/she remained as a minimal reciprocator and non-deep thinker with peripheral status in the social network.

In T1 week 1 U0384 participated in a thread initiated by a peer learner for discussing and sharing editing for a sentence. U0384 posted a non-deep message to provide his/her opinion on this topic.

> “…I have always been taught to refer to the figures directly, not by way of a parenthetical reference….”

In T2 week 5, U0384 participated in a thread initiated by a peer for sharing usual practice in writing.

> “… I draw a big diagram that shows the structure, with little bullet points for each part of the sequence. It's like a road map.”

In both time periods, this learner had connection with two people and did not repeat interaction with anyone. He/she did not show changes in language style and participation focus.

5.5. Chapter summary

Analysis in this chapter resulted in several important findings for understanding positions, position trajectories, and learning in MOOC forum discussions. First, positions found in the two MOOCs fell into six primary types based on the key characteristics, including enthusiastic central providers, enthusiastic central reciprocators, moderate deep thinkers, moderate reciprocators, moderate providers, and minimal contributors.

These positions were all found across courses. The discussion forums usually contained a small group of learners with high contribution quantity and central social status, two or three bigger groups of learners with moderate contribution quantity and peripheral social status, and a vast majority of learners with low contribution quantity and peripheral social status.

Second, 55% of the StatLearn learners and 82% of the SciWrite learners who posted in content discussions in multiple time periods changed positions over time. Changes in contribution quantity, seeking / providing activities, and strength of social connections occurred more frequently than changes in deep thinker characteristics and connectedness in the social network.

Finally, only a small number of the frequent trajectories in the two courses showed constructive position changes that indicated potential development participation pattern. They made up 14% of the trajectories and were found in 17% of the multi-period learners in both courses. The majority of the frequent trajectories showed unconstructive position changes, neutral changes, or no change at all.

Chapter 6.

Discussion and Conclusions

This study investigated MOOC learning defined as development in the way that one interacts with others in forum discussions. Such development is indicated by changes in a learner's participation pattern that are constructive with respect to substantive interaction about the course content. This study conceptualizes different forum participation patterns as positions in the discussion that learners take in relation to each other. The series of positions that a learner takes over time form their participation trajectory. In this study participants' positions in discussion forums were analyzed for the beginning, middle, and end periods in two MOOCs: StatLearn and SciWrite. Positions were characterized using five types of characteristics related to learners' forum contributions and social relations: quantity of content contributions, input seeking and providing activities, deep consideration of the discussion content, connectedness in the social network, and strength of social connections. In the first phase of analysis, positions in each time period were identified through clustering groups of learners who had similar participation characteristics. Based on the key characteristics, the positions found in the two MOOCs fell into six primary types: enthusiastic central providers, enthusiastic central reciprocators, moderate providers, moderate reciprocators, moderate deep thinkers, and minimal peripheral contributors. The forum for each time period usually consisted of an enthusiastic central group (either providers or reciprocators) that made up a small proportion of the forum population (2% to 3%), a minimal peripheral contributor group that made up a majority of the population (63% to 79%), two or three moderate contributor groups (including a moderate deep thinker group, and at least one of the moderate provider group and the moderate reciprocator group) that together made up a large proportion of the population (26% to 35%).

In the second phase of analysis, for learners who participated in multiple time periods in the same course, their positions in different time periods were compared to determine whether and how their participation patterns changed over time. Looking across all position trajectories found in the two courses, changes occurred in the five aspects with different frequency: contribution quantity, input seeking and providing activities, and strength of social connections occurred sometimes whereas changes in deep consideration of the discussion content and social connectedness occurred infrequently. Fifty-five percent of the learners in StatLearn and 82% of the learners in SciWrite changed positions over time. Thirteen frequent trajectories in the two courses fell into four categories based on whether and how changes occurred: trajectories with development in participation patterns, trajectories with unconstructive changes, trajectories with neutral changes, and trajectories with no change. In both courses, less than 20% of multi-period learners showed development in participation patterns over time whereas the majority of multi-period learners did not show such changes. Case studies on the four trajectory categories revealed the learner who showed constructive changes also displayed changes in language and participation focus that suggested potential identity development. Learners with the other three trajectory categories did not show such changes in language and participation focus.

These results are discussed in this chapter. Section 6.1 reviews the positions found and their usefulness for understanding forum interaction. Section 6.2 discusses changes in each aspect of participation characteristics and explores possible explanations for the changes. Section 6.3 discusses position trajectories and their usefulness for understanding development in participation patterns. Sections 6.4 and 6.5 discuss the implications for MOOC research and MOOC practice. Limitations of the current study and suggestions for future research are discussed in Section 6.6. Finally, Section 6.7 concludes this paper.

6.1. Participation positions in MOOC forums

Prior MOOC studies have examined forum participation from a contribution perspective (e.g., Liu et al., 2016) or a social relations perspective (e.g., Poquet & Dawson, 2016). The contribution perspective accounts for a leaner’s effort level and approach to interaction. The social relations perspective accounts for the characteristics of a learner’s social relationships. In this study the two perspectives were combined for characterizing participation positions. The identified positions are useful for understanding both the big picture of MOOC forum interactions and different participation groups.

6.1.1. Participation positions: Understanding the big picture of MOOC forum interactions

Prior MOOC studies have investigated the composition of forum populations and the role of different participant groups from different perspectives. Studies that focused on contribution quantity generally have reported that discussion forums are populated by a small number of learners contributing a large proportion of the messages and a large group of learners each posting a few messages (Clow, 2013; Cohen, Shimony, Nachmias, & Soffer, 2019). Studies that focused on social relations usually have noted a core-periphery structure in which several central individuals had a disproportionately large number of connections and a large number of peripheral participants had connections with a small number of peers (Boroujeni, Hecking, Hoppe, & Dillenbourg, 2017; Kellogg et al., 2014; Sinha, 2014). Based on the positive correlation between contribution quantity and social centrality reported for MOOC forums (Houston et al., 2017), MOOC studies generally can be said to have conceptualized forum populations as consisting of a small group of core members who participate quite frequently, and a large group of peripheral members who rarely participate (Boroujeni et al., 2017; Kellogg et al., 2014; Sinha, 2014). However researchers have also speculated that like other online learning communities, MOOC discussion forums may also contain a middle group who participate regularly but not as frequently as the core group (Kellogg et al., 2014; Poquet & Dawson, 2016; Wenger, 1999). Poquet and Dawson (2016) clustered forum participants in a solar energy MOOC based on betweenness and clustering coefficient, and found in addition to a small group of highly influential learners and a majority of uninfluential learners, there was a large proportion of learners who were moderately influential. Poquet and Dawson noted that this group was likely to be learners who participated in social interactions moderately actively. But this speculation had not been verified. The positions found in the current study validates the basic core-periphery structure: the enthusiastic central contributors (making up 2% to 3% of the forum population) can be seen as the core members and the minimal peripheral contributors (making up 63% to 79% of the forum population) can be seen as the peripheral members. Furthermore the moderate contributors (together making up 26% to 35% of the forum population) can be seen as the regular members who participated moderately actively. These findings offer a critical ground for studying MOOC forum interaction.

6.1.2. Participation positions: Understanding different participant groups

This study characterized positions using five contribution and social characteristics. The resultant positions provided rich illustrations of how learners participated in forum interactions and deepened the understanding of participant groups identified from a single perspective in prior research.

Enthusiastic central providers and enthusiastic central reciprocators

Most of the examined forums contained either a group of enthusiastic central providers or enthusiastic central reciprocators. Members in these groups were distinctive from other groups in making a substantially higher number of content contributions, having social connections with a large number of people, and being important connecting points between people in different conversations. They were not deep thinkers in discussion. They sometimes had repeated interactions with the same peers. The two groups each made up 2% to 3% of the total forum population.

Prolific forum contributors have been examined in several MOOC studies. A widely-cited study was conducted by Huang, Dasgupta, Ghosh, Manning, and Sanders (2014) on superposters in 44 MOOCs. Superposters were learners whose weekly contribution quantity ranked in the top 5% of all forum participants. Huang et al. investigated the association between these learners' prolific posting behaviours and final grades, the conventional metric for learning outcomes. Huang et al. found superposters obtained better final grades than the average forum participants (which does not necessarily imply a causal relationship). Other MOOC studies have also reported positive association between contribution quantity and grades / certificates (Wang, Yang, Wen, Koedinger, & Rosé, 2015; Wise & Cui, 2018b). The two groups of enthusiastic central contributors found in the current study can help explain this correlation. For enthusiastic central providers, their active input providing activities suggested they had more knowledge, expertise, and resources related to the course content than the average learners (Liu et al., 2016; Yang, Adamson, & Rosé, 2014), and thus were more likely to obtain better grades. The high scores obtained by the prolific contributors may also be attributed to their social centrality. Being the most well-connected members in the social network, they had direct connections with many people and participated across many conversations. Thus they were more privileged than the average learners in accessing knowledge and resources (Kellogg et al., 2014), which may have contributed to their final grades.

Interestingly in the same study Huang et al. (2014) also found as a group the prolific learners' final grades were lower than the learners whose forum reputation (computed based on both the contribution quantity a learner made and the number of upvotes they received) ranked top 5% in the course. Huang et al. raised the point that the reputation scores can be seen as a combined indicator of contribution quantity and quality, thus may better account for the final grades. This finding also suggested that the most prolific learners may not have always had high contribution quality. The current study offers evidence for this speculation. The prolific contributor groups did not show deep thinker characteristics. This may partially explain why the superposters in Huang et al.'s study did not get the highest final grades.

Moreover Huang et al. (2014) reported that the superposters' high contribution quantity positively correlated with the total contribution quantity made by other participants. This may relate to the high degree and betweenness of the enthusiastic central contributors found in the current study. Low responsiveness is a major discouraging factor for MOOC forum participation (Almatrafi, Johri, & Rangwala, 2018; Gütl, Rizzardini, Chang, & Morales, 2014; Kizilcec & Halawa, 2015). The enthusiastic central contributors not only posted a large number of messages, but also had direct connections with many peers and participated widely across different conversations. Their activities may have attended to the emergent learning needs of a wide range of learners, and thus encouraged participation.

From the social relations perspective, MOOC research have paid substantial attention to the most well-connected participants, especially the positive association between their central social status and learning outcomes measured by grades and certificates (Houston et al., 2017; Wise & Cui, 2018b). This association is usually attributed to these learners' access to a large number of people and topics. It has also been attributed to their high effort level based on the positive correlation between social centrality and contribution quantity (Houston et al., 2017; Wise & Cui, 2018b). The enthusiastic central providers found in the current study offer an additional conjecture. As discussed earlier, their active input providing activities may suggest they had more knowledge about the course content than the average learners and thus were more likely to get better learning performance than others (Liu et al., 2016; Yang, Adamson, & Rosé, 2014). Considering the strength of social connections, prior research found the central participants in MOOC social networks often had strong connections with each other (Poquet & Dawson, 2016; Wise & Cui, 2018a). The current study found the average edge weight in the ego networks of the enthusiastic central contributors was not high suggesting although these learners had interactions with a large number of peers, they only interacted frequently with a small proportion of them. This finding is useful for understanding the prolific and well-connected learners' roles in MOOC forums. While they exchange information broadly in the forum, they may only develop meaningful social bonding with a small number of peers.

Moderate deep thinkers, moderate providers, and moderate reciprocators

The current study identified three moderate contributor groups in the examined forums: providers who were deep thinkers, providers who were non-deep thinkers, and reciprocators who were non-deep thinkers. The three groups were similar in that their members all posted a medium number of content contributions, had direct connections with a small number of people, and had some repeated interactions with the same peers. Together they made up 26% to 35% of the forum population. The moderate contributor groups have not received as much attention as have the prolific learners and well-connected learners in MOOC research.

The moderate deep thinker group was found in every time period, making up 2% to 7% of the forum population. They were distinctive for adopting sophisticated approaches to the discussion content. They not only posted a high proportion of deep messages but also added deep messages to ongoing non-deep discussions. Dowell et al.'s (2015) study on MOOC learners' language and discourse characteristics, learning performance measured using the course grade, and social centrality provided some useful insights for understanding the deep thinkers. They found forum participants who used more abstract, explanatory, and referential language in discussions (which seem to align with the deep thinkers' sophisticated approaches) obtained higher final grades in a MOOC on infrastructure. Moreover, Dowell et al. found these learners were not as well connected in the social networks as the learners who used more concrete, narrative, simple, and less referential languages. This may help understand the deep thinkers' peripheral social status and infrequent repeated interactions with the same peers.

Whereas the deep thinkers' medium contribution quantity may partially account for their peripheral social status, their language style may have also obstructed their interactions with other learners. Understanding these characteristics can offer insights for assisting these learners in building social connections.

Looking at the moderate providers, in both courses this group either were not found in T1 and T2 forums or only made up a small proportion of the forum population (up to 5%). In the T3 forum though the moderate providers made up approximately 20% of the population. Nelimarkka and Vihavainen's (2015) study on alumni learners provides some pointers for understanding this group. Alumni learners were MOOC learners who continued to participate in discussion forums for courses they had passed. Nelimarkka and Vihavainen found alumni learners were not the most prolific contributors but usually devoted their contributions to help others and took on the role of mentors. Interviews with these learners revealed they offered help to others for varied motivations, such as contributing to the learning community, returning the favor for the help they received as new learners, observing how others learn to deepen their own understanding of the course subject, and networking with people who had similar interests. Understanding the motivations related to input providing activities in MOOCs can be useful for designing incentives to better encourage and engage the moderate provider group.

Finally, in both courses the moderate reciprocator group was the second largest group in the T1 and T2 forums (20% to 33%) but was not found in the T3 forum. These learners engaged in input providing and seeking activities in a balanced manner. A notable characteristic of the moderate reciprocators is they seemed to have slightly higher average edge weight in the ego network than the moderate providers and the deep thinkers in the same time periods, which may have to do with their two-way interaction patterns. One speculation is the reciprocators may have engaged in back and forth exchanges with the same peers in the same discussions. It is also possible that the versatile interaction mode (both seeking and providing instead of a single mode) may have resulted in more opportunities for the reciprocators to interact with the same peers across discussions. Verifying these connections may provide useful insights for designing discussion activities to promote stronger social connections between learners.

Minimal peripheral contributors

Minimal peripheral contributors formed the largest group in the examined forums, making up 63% to 79% of the forum population at all points in time. Members in this group were mostly single-message contributors. They were non-deep thinkers and usually made the lowest proportion of deep messages among all participant groups. Most of these learners had direct interactions with few people and tended to have one-off interactions; some of them were isolates in the social network. These findings align with the MOOC literature on learners with low participation level in discussion forums, who have been found to have low learning performance (measured by grades and certificates), peripheral social status, and high dropout rate (Houston et al., 2017; Poquet & Dawson, 2016; Yang et al., 2013). As the minimal peripheral contributors generated limited data in the discussion forums, further understanding of this group calls for using other data collecting methods, such as surveys and interviews.

6.2. Changes in position characteristics

Examination of position trajectories revealed changes occurred in five aspects of participation characteristics. This section discusses the changes in each aspect and their connections with development in learner's participation patterns, as well as the factors useful for understanding the changes.

6.2.1. Changes in contribution quantity

Given that MOOC discussions related and unrelated to the course content have been found to differed dramatically in interaction purposes and interaction characteristics (Stump et al., 2013; Wise & Cui, 2018a; Wise & Cui, 2018b), the current study differentiated the two types of discussions and only focused on the content-related contributions. In the current study 16% of StatLearn trajectories and 18% of SciWrite trajectories showed increases in quantity of content-related contribution while 21% and 35% of the trajectories in the two courses showed decreases respectively. It is worth mentioning that the current study filtered out non-content contributions, thus contribution quantity more directly indicates the effort that a learner invests in learning-related discussions. Effort level reflects commitment to the learning goal and often associates with learning orientation and persistence (Rosé et al. 2014; Strijbos & De Laat, 2010;

Tang, Xing, & Pei, 2018). For instance, Rosé et al. (2014) found MOOC learners who had a high forum participation level from the beginning of the course were more likely to continue participating. Considering forum participants as members of a group, an increase in the amount of effort they invest to the group activity reflects stronger commitment to the group goal; it also often associates with playing more important roles and creating greater value for the group (Strijbos & De Laat, 2010; Wenger, Trayner, & De Laat, 2011). For instance, an increase in contribution quantity can be aligned with a learner's change from being a peripheral member to a core member in the learning community. In the MOOC context learners' contributions to discussion forums are crucial for sustaining interpersonal interaction and social learning in the asynchronous course format. An increase in contribution quantity indicates the learner produced more learning artefacts that can be received by others and thus may have created more value for the group. Furthermore, as discussed in Section 6.1.1, becoming a highly prolific contributor indicates the learner may be positively influencing other learners and motivating them to participate more actively (Huang et al., 2014). Thus increases in contribution quantity are a constructive change for learning-related interaction whereas decreases in the contribution quantity are unconstructive.

In the current study a substantial proportion of learners in the two courses decreased contribution quantity over time. In most cases, this was a decrease to the level of a single message signifying a high likelihood of disengagement and dropout. This finding underlines the importance of understanding the cause of such changes. The changes in individual learner's contribution quantity can be accounted for from various perspectives. First, motivation can influence the persistence of effort. Tang et al. (2018) found learners in a project management MOOC fell into three groups based on how their weekly forum participation quantity changed over time (participation quantity included both posting and viewing activities): one group sustained low participation quantity; one group sustained higher participation quantity; one group had the highest participation quantity at the beginning of the course but gradually disengaged. Tang et al. found learners with intrinsic motivations (e.g., interest in learning) were more likely to maintain a higher engagement level whereas learners with extrinsic motivations (e.g., career advancement) were more likely to have a low engagement level. They also found that a large proportion of the gradually disengaged learners expressed intrinsic motivations at the beginning of the course, hence they argued that intrinsic motivation is not constant but needs continuous support and scaffolding during the learning process. Second, Poquet and Dawson (2016) found increases in learner's contribution quantity can be triggered by external incentives. In a solar energy MOOC, after the course team declared an incentive policy to encourage forum participation, learners who had not contributed actively in the forum generated a large number of messages. However Poquet and Dawson further found most of these messages were shallow in content. Finally, MOOC studies also noted the quantity of forum activities usually increase before milestone events in the course, such as assignment deadlines and final exams (Boroujeni et al., 2017; Poquet & Dawson, 2016; Tang et al., 2018). Understanding these factors can be useful for promoting participation level in the forums.

6.2.2. Changes in input seeking and input providing activities

In StatLearn and SciWrite 9% and 22% of the trajectories showed increase in *input seeking* activities respectively while 16% and 32% showed increase in *input providing* activities. These two types of activities involve different knowledge and skills, and contribute to forum interaction in different ways. Providing input requires knowledge and expertise related to the course subject. By contributing content and resources, learners play an crucial role in maintaining dynamic and sustained interactions in MOOC forums (Liu et al., 2016; Yang, Adamson, Rosé, 2014). Input providing can also benefit the providers themselves. Through helping and mentoring the novice learners, the experienced learners can enrich their perspectives on learning the course content (Nelimarkka & Vihavainen, 2015).

Seeking input also can benefit both the seekers and the group. Help-seeking is an important resource management strategy in self-regulated learning. It involves the capability to identify the problem and the help needed, request help, and evaluate and apply the received help to the problem domain (Karabenick & Dembo, 2011; Zimmerman, 2008). Through requesting knowledge, assistance, and resources from instructors and peers, help seekers are more likely to succeed in addressing the challenges and stay engaged in the course (Corrin, de Barba, & Bakharia, 2017; Najafi, Rolheiser, Harrison, & Heikoop, 2018). From the group perspective, help seeking can create interaction opportunities and mobilize other learners to participate. The processes that help seekers work out the problems can be useful for other learners (Jiang et al., 2015). The MOOC community has noted these useful characteristics of help-seeking and designed technologies to tap this resource. For instance, Yang, Adamson, & Rosé (2014) built an automated tool that analyzes the questions raised in forums and helps the seekers send invitations to learners whose participation history indicated they would be able to provide the needed help.

Changes in input seeking and input providing roles in MOOC forums has only been investigated in Boroujeni et al. (2017). They examined bi-weekly time periods in two MOOCs and found MOOC learners' roles as help seekers, help givers, and reciprocators (who exchanged information with the same people) changed over time. But their study did not report frequency of the changes. The current study revealed the change from reciprocators to input providers was most frequent in both StatLearn and SciWrite (respectively 16% and 32% of the trajectories), followed by the change from input providers to reciprocators (respectively 9% and 11% of the trajectories). In SciWrite another 11% of the trajectories showed the change from input seekers to reciprocators. Given the contributive characteristics of input seeking and input providing activities, shifts between the input providers and reciprocators do not necessarily associate with a single value judgement related to forum interaction. In SciWrite 11% trajectories showed changes from input providers and reciprocators to input seekers. Such shift is unconstructive as altruistic input providing activities are crucial for sustained forum interactions in MOOC discussion forums (Zhang, Skryabin, & Song, 2016).

The factors that associate with changes in input seeking and providing activities have not been directly examined in the MOOC literature. One possible factor affecting input seeking and providing behaviors is a change in social relations. As learners develop stronger social bonding and a sense of community, they may be more willing to help others and contribute to the community (Nelimarkka & Vihavainen's, 2015); they may also perceive the discussion forum as a safe environment and become more motivated to seek help (Howley, Tomar, Yang, Ferschke, & Rosé, 2015). It is also possible that as the learners become more knowledgeable about the course subject they felt more confident in their ability to help others. Finally, prior knowledge and interest about the discussion topics may also influence input seeking and providing behaviors. Learners can be more motivated to provide help on topics about which they have more interest and expertise (Yang, Adamson, & Rosé, 2014). As for help seeking, they may seek help more frequently on topics about which they have some prior knowledge other than topics that they are unfamiliar with (Bartholomé, Stahl, Pieschl, & Bromme, 2006).

6.2.3. Changes in deep consideration of the discussion content

Only 6% of the StatLearn trajectories and 4% of the SciWrite trajectories showed a switch from non-deep thinkers to deep thinkers. The current study used deep versus non-deep consideration of the discussion content to characterize differences in the ways that learners approached the discussion content. Deep consideration involves engaging with the learning content in complex and sophisticated ways, such as giving explanation, elaboration, interpretation, or comparisons. Contributions without deep consideration generally focused on simple and straightforward exchanges of information. The change from non-deep thinkers to deep thinkers indicates the learner not only adopted deep approaches more frequently in posting, but also started to post deep messages in on-going non-deep discussions indicating the ability to alter the depth of discussion. Such change is constructive for multiple reasons. First, prior MOOC studies often found a positive association between the complexity of learners' approaches to discussion content and their course performance (e.g., course grades and certificates). For instance, Wang et al. (2016) found learners who displayed constructive and interactive behaviors in discussions in a psychology MOOC obtained higher scores than those who did not display such behaviors. Looking across weeks, Wang et al. further found learners obtained better scores in the weeks when they showed higher order thinking behaviors. In another study, Gillani, Eynon et al. (2014) found learners who showed more advanced knowledge construction levels in forum discussions in a business MOOC were more successful in passing the course than those who showed less advanced levels. Second, from the participation perspective, the change from non-deep thinkers to deep thinkers can indicate changes in the learners' commitment, roles, and identity. Wen, Yang, and Rosé (2014b) found MOOCs learners who displayed higher cognitive engagement levels in the discussions were less likely to drop out. As deep thinkers' engagements with the discussion content can serve as a model for other learners in the conversation (McKendree, Stenning, Mayes, Lee, & Cox, 1998), becoming a deep thinker means a learner can potentially cast positive influence on their peers' way of thinking. Such change can also associate with identity development as a more knowledgeable and confident participant who is capable of steering the depth of discussion.

In the current study the vast majority of the trajectories in the two courses remained as non-deep thinkers over time. The lack of change in this aspect can be explained from several perspectives. First, deep thinkers were a small group in the forums and had limited social impact on other learners. The examined forums only contained 2% to 7% of deep thinkers at any point in time. These learners were moderate contributors and not well-connected. Thus only a small proportion of forum participants had directed interactions with them and were exposed to their potential influence. Second, changing one's learning approach is hard and usually requires the careful planning and scaffolding on the part of instructors (Leflay & Groves, 2013). Instructor's facilitation and encouragement are considered important factors associated with learner's cognitive engagement in learning (Budsankom, Sawangboon, Damrongpanit, & Chuensirimongkol, 2015; Zhu, 2006). For instance, Christopher, Thomas, and Tallent-Runnels (2004) studied the unfacilitated online discussion of ten graduate students in an education course. Over ten weeks of discussion the students showed medium level of thinking (high level = synthesis and evaluation, medium level = application and analysis, low level = knowledge and comprehension) and did not change over time. In the current study it is unlikely that even learners who had direct interaction with the deep thinkers could learn to adopt their thinking approaches without careful guidance and scaffolding. Third, the characteristics of the learning content can also be an explanatory factor for the lack of change from non-deep thinkers to deep thinkers. Wang et al. (2016) examined weekly lecture topics in a biology MOOC and found topics with everyday words and phenomenon that learners were familiar with were more likely to trigger higher order thinking whereas technical topics that learners were unfamiliar with were less likely to trigger higher order thinking. Considering that the examined MOOCs were both on technical topics (statistics and writing in the sciences), the learning content may have not helped the learners to develop into deep thinkers. Finally, it is possible that changes in thinking approaches take longer than the examined time periods in the current study. As MOOC specializations (a series of courses related to a subject) become more popular, it could be interesting to trace cross-course development in learners. Finally, 7% of the StatLearn trajectories and 5% of the SciWrite trajectories showed a switch from deep-thinkers to non-deep thinkers signifying the need for a learning environment that supports the deep thinkers to continue adopting the complex approaches in discussions.

6.2.4. Changes in connectedness in the social network

This study used degree and betweenness as indicators of a learner's connectedness in the social network. In StatLearn and SciWrite 3% and 2% of the trajectories showed a change from peripheral to central social status respectively. At the same time 2% and 12% of the trajectories in the two courses showed changes in the other direction respectively. An increase in social centrality indicated the learner switched from the majority group of less-connected people to the small group of disproportionately well-connected people. For individual learners, an increase in degree means they changed from having direct connections with a small number of people to a large number of people, and thus likely had access to more information sources and perspectives related to the course content. An increase in betweenness means the learner participated across a greater number of conversations and thus likely had access to more topics related to the course content. From the group perspective, a larger number of members with high degree centrality was interpreted to mean the questions raised in the discussion forum were more likely to get responses. A larger number of members with high betweenness centrality means there were more learners who facilitated flow of information and people across conversations. For example, learners who have participated in multiple conversations often refer to content in other conversations relevant to the current one. They can also direct input seekers to existing threads relevant to the seekers' questions. Moreover, a social network with a bigger number of well-connected members is more resilient. Communities in MOOC forums are usually connected by the central members. A network with more central members is less likely to fragment when a few of them stop participating (Callaway, Newman, Strogatz, & Watts, 2000; Gillani, Yasseri, Eyon, & Hjorth, 2014). This was seen in the social networks in the current study. All networks except the SciWrite T3 network had multiple central nodes and contained a large connected component. In contrast, the SciWrite T3 network had few central nodes and was highly disconnected.

This study revealed that the less-connected learners rarely made their way into the small well-connected group. The infrequent change from peripheral to central status has been noted in the MOOC literature (Tawfik et al., 2017; Yang et al., 2013). Tawfik et al. built weekly social networks for a chemistry MOOC and found little change in the top 25 learners ranked by degree and betweenness centrality. Yang et al. (2013) examined the weekly social networks in a literature MOOC and found most of the few central participants joined the discussion forum in the early weeks and continued interacting with each other. Learners who joined the discussion forum later in the course were more likely to remain in the periphery and appeared to have trouble getting integrated into the discussion. The rich-get-richer phenomenon indicated the need to understand how MOOC forum participants form bonds in interaction and how discussion activities can be structured to help latecomers to get integrated. While little empirical work has been published on this topic, the work of Gillani, Yasseri et al. (2014) on social networks in a business MOOC can provide useful pointers. They examined forums designated for different topics and found two forums that contained more iterative and deep conversations had proportionately larger numbers of highly connected members than other forums. Furthermore, in the second offering of the same course, the proportion of highly connected members in these two forums were even higher than in the first offering. Gillani, Yasseri et al. (2014) noted this may have been the result of an additional participation incentive (the total number of upvotes that learners received on their messages were counted as up to 8% of the final scores). Thus it is possible that in the current study the infrequent change from peripheral to central status was partially due to lack of inclusive and deep discussions as well as effective incentives for the peripheral learners to connect with more peers.

6.2.5. Changes in strength of social connections

This study used the average edge weight in the ego network to measure the strength of a learner's social connections with the peers who directly interacted with them. Based on the Direct Reply tie definition, the edge weight between two learners increases when they respond to each other multiple times either in the same conversation or across different conversations. An increase in the average edge weight in the ego network indicated a learner changed from only having one-off ties to having some repeated interactions with the same peers, which may signal stronger social bonding. This study revealed in StatLearn and SciWrite respectively 14% and 15% of the trajectories showed the change from having one-off ties to repeated interactions with the same peers. Comparatively 20% and 30% of the trajectories in the two courses showed changes in the other direction respectively.

An increase in repeated interaction with the same peers and the potential for stronger social bonding to emerge can be constructive to forum interaction in several ways. First, having repeated exchanges with the same peers often associates with more extended and in-depth discussions of the learning content. Chung and Paredes (2015) found learners who had stronger ties between each other exchanged deeper insights and had richer discussion about the learning content. Second, stronger social ties often associate with a greater sense of closeness and trust between learners, which can motivate help seeking and providing activities (Krackhardt, 1992). Finally, strong social bonding can lead to improved sense of community and satisfaction with the learning experience (Thomas, 2000). Rosé et al. (2014) found learners who had stronger social ties in MOOC forums were more likely to continue participating than those who only had weak social ties.

It should be noted that although about 15% of the trajectories in the two courses showed the change from having one-off ties to having repeated interactions with the same peers, none of the groups in the examined forums had very high average edge weight in the ego network suggesting repeated interactions still occurred relatively infrequently. This finding aligns with prior work reporting weak social ties in MOOCs on other subjects and underscores the need to structure discussion activities to promote stronger social bonding in MOOC forums. For example Kellogg et al. (2014) examined the discussion forums in a digital learning MOOC and a mathematics MOOC and found the social networks had low average edge weight with most ties consisted of a single communication. Partly due to this reason MOOC forums are often considered to be crowds of discussants loosely connected by topics rather than a community of members bonded by social connections. MOOC research has proposed different strategies to address this issue, such as designing small group discussion to promote repeated interactions within a small number of learners (Wen, Yang, & Rosé, 2015; Zheng, Vogelsang, & Pinkwart, 2015), designing learning tasks that require extended discussions (Gillani, Yasseri et al., 2014), and using facilitation techniques that guide the learners to work out the problems together instead of providing them with straightforward answers (Wise & Cui, 2018a).

6.3. Position trajectories as manifestation of changes

Position trajectories are combinations of the changes reviewed in the prior section. This section discusses the importance of examining the concurrent changes together to understand MOOC learners' participation process. It also discusses four trajectory categories and their conceptual connections with development in learners' participation patterns.

6.3.1. Changes in a learner’s position often occurred in multiple aspects, thus call for a comprehensive diagnosis.

Only a small number of MOOC studies have investigated chronological changes in forum participation patterns. Those that have mostly focused on a single aspect of such change, such as contribution quantity (Tang et al., 2018), help-seeking and providing activities (Hecking et al., 2017), or social centrality (Tawfik et al., 2017). The current study combined multiple perspectives and examined learners’ position trajectories from five aspects. Results showed the changes in position often involved changes in multiple aspects. Among the 30 trajectories that included any change, 24 showed changes in multiple aspects while only 6 trajectories showed changes in just one. The complex nature of changes in position underscores the need for a comprehensive diagnosis that considers multiple aspects.

The small number of studies that looked across changes in multiple aspects had only focused on changes within the *contribution characteristics*. Nelimarkka and Vihavainen (2015) looked at both contribution quantity and input seeking and providing activities when examining alumni learners’ participation behaviors before and after they became alumni learners. Results showed while these learners’ contribution quantity decreased noticeably when they took the same course again, the focus of their participation switched from serving their own learning needs to helping others to learn. Combining the two perspectives allowed for a more accurate account for these learners’ participation characteristics and estimation of their influence on forum interaction. Poquet and Dawson (2016) found although an incentive policy for forum participation resulted in a brief surge in contribution quantity, most of these contributions were shallow in content. This indicated looking at both effort level and effort quality allows for more accurate evaluation of the behavioral changes. The current study verified the benefit in integrating the examination of changes across different contribution characteristics. For instance, learners with Trajectory MPDLR>>MENLR showed no change in contribution quantity, but they did change from input providers to reciprocators and stopped showing deep thinker characteristics. This change combination indicates as the course proceeded these learners remained motivated to participate in forum interaction, but they may not have as much knowledge about the course topics being discussed in the second time period as in the first time period and thus may need more support for understanding the course content.

Furthermore, the current study built on the prior research to simultaneously examine the changes in both *contribution characteristics* and *social characteristics*. This further expanded the understanding of changes in participation patterns. For instance, examination of the learners with Trajectory LENLN>>MENLR showed they only increased contribution quantity from low to medium level, and remained as reciprocators and non-deep thinkers over time. However examination of the social characteristics revealed while these learners increased their effort level, they also switched from having one-off interactions to having repeated interactions with the same peers. This is a quite different change than if the increased contribution quantity occurred only in one-off interaction with different peer. The changes found in these learners indicated they were more likely to develop stronger social bonding than they were in the first time period.

6.3.2. Some trajectories contained constructive change combinations and showed signs of potential identity development.

The 13 position trajectories that occurred with high frequency varied in the specific change combinations they contained. Generally there were four categories of trajectories: trajectories with constructive changes, trajectories with unconstructive changes, trajectories with neutral changes, and trajectories with no change.

Three of the 13 frequent trajectories showed constructive changes in participation patterns. Together they made up 14% of the StatLearn trajectories and 14% of the SciWrite trajectories. Learners with these trajectories all changed from making minimal contributions to medium contributions and from having one-off interactions to having repeated interactions with the same peers. They all remained peripheral in the social networks. The differences among the trajectories lay in whether and how input seeking and providing activities and deep consideration of the discussion content changed over time.

Trajectory LENLN>>MPDLR (Up – P_Up – Up – Same – Up) showed changes in four of the five aspects, the most among the three trajectories. It is also the only frequent trajectory that showed increase in deep thinker characteristics, making it the most desirable in this category. This trajectory only made up 3% of the trajectories in StatLearn and 1% of the trajectories in SciWrite. Learners with this trajectory changed from minimal reciprocators and non-deep thinkers to moderate providers and deep thinkers. These changes indicated increased learning effort and complexity of their approaches to the discussion content. The increase in input providing activities coupled with the emerging deep thinker characteristics suggested positive development in the learner's relationship with the learning content. The case study of StatLearn learner U0644 further revealed changes in his/her posting language and participation focus. Compared to T1, this learner used more certain and confident language in T2 (especially in the last week of T2), such as "***Remember**, the use of the entropy is to select how to partition a dataset*" and "***The fact is** that in R, log is natural, log10 is 10 based and log2 is 2 based*". Moreover, in T2 this learner not only continued exchanging information about specific quiz questions, but also started to contribute general domain knowledge related to the course. Similar changes in the use of language and participation focus have been reported for MOOC learners who took up a role as mentors for peer learners (Nelimarkka & Vihavainen, 2015). It is possible that U0644 also developed his/her identity as a more knowledgeable and competent discussant as suggested by his/her use of more certain and confident language. In addition to changes in contribution characteristics, U0644 also started to have repeated interactions with the same peers in discussions which appears to contribute to developing stronger social ties.

Promoting constructive changes such as those observed for U0644 requires identifying factors for the changes. Some hypotheses can be proposed for future investigation. For instance, the observed changes may be attributed to the improved perception of the social learning environment. In T2, he/she used social presence indicators (e.g., addressing peers by name) and referred to personal learning experience in his/her postings: "…when I first came across it… the base of the log confused me too. Later, I found…". These characteristics indicated a sense of closeness that may motive the learner to commit more effort and help others more actively. It is also possible that as the course proceeded the learner's increased knowledge about the subject allowed him/her to make connections with their prior knowledge. For example in T2 they said "From doing many courses, it seems statistics oriented people tend to use the natural log, while in many big data and computational scenarios, log2 is common". This may have helped the learner to think about the discussion content with more depth and made him/her more confident in helping others.

Finally, analysis of learner U0644's posting patterns revealed opportunities for further improving participation through developing stronger social ties. This learner seldom revisited the threads he/she had participated in which resulted in several unfinished dialogues and missed bonding opportunities. This might be addressed through encouraging learners to subscribe discussions they joined so that they can be notified of new posts added to the conversations. Moreover, some of this learner's input providing messages were made several days after the requests were posted which may have limited his/her chances to engage in ongoing discussions. This can be addressed by analyzing the learner's participation history and automatically sending him/her help requests that fit their expertise and interest (Yang, Adamson, & Rosé., 2014).

6.3.3. Trajectories that showed unconstructive changes signified the need for learning support.

Five of the frequent trajectories contained unconstructive change combinations. Together they made up 19% of the StatLearn trajectories and 24% of the SciWrite trajectories. These trajectories varied in the specific changes. One of the trajectories only showed the change from deep thinkers to non-deep thinkers. Three trajectories showed changes from medium to minimal contribution quantity and from having repeated interactions to one-off interactions. They differed in whether and how changes occurred in input seeking and providing activities. The final trajectory (HPNHR>>LENLN) showed changes in four of the five aspects, the most among the five trajectories. The case study on SciWrite learner U0155 with this trajectory resulted in improved understanding of the changes and signified the need for learning support. From T1 to T2 learner U0155 changed from an enthusiastic provider to a minimal reciprocator while remaining a non-deep thinker. These changes are interpreted to reflect a decrease in the effort and input providing activities. In this process the learner stopped having direct interactions with multiple people and participating in multiple threads. Thus, in T2 he/she was likely to have exposed to fewer information sources and perspectives, and accessed fewer discussion topics. His/her impact on forum interaction also may have reduced dramatically. Moreover, this learner stopped having repeated interactions with the same peers with the result he/she no longer had extended discussions with others. Together these changes signified the learner was becoming disengaged. Analysis of U0155's contribution content and interaction processes showed he/she mostly participated in discussions about sentence editing solutions. When sharing editing solutions to seek input, this learner did not explicitly state what kind of help he/she expected (e.g., "I have tried; any comments for the below?") and sometimes received very general comments from other learners (e.g., “Your edit is really good”). This may have discouraged his/her further engagement. Moreover this learner’s comments about others’ editing solutions also tended to be general (e.g., “I think the last one can be improved. Good try!”) and seldom involved deep consideration. These non-deep exchanges on similar topics may have contributed to the learner being intellectually under-challenged and resulted in disengagement (Schussler, 2009). This issue might be addressed by guiding the learner to improve help-seeking skills, such as explicitly stating the challenges they have and the kind of help they want to receive so as to induce constructive responses from others. For instance, another learner wrote a more sophisticated post when seeking suggestions for his/her writing: “I welcome any comments, but the particular issues I'm grappling with are: how to state why the case is important without turning the introduction into a discussion…”. At the same time, it can also be useful to guide and model for the learners how to provide constructive editing suggestions. Another plausible factor for U0155’s disengagement may be confusion. This learner sometimes used indecisive and uncertain language (e.g., “I have the same problem with other exercises and this as well. Can't know for certain when to stop cutting.”). Unaddressed confusion can have negative impact on learning engagement and perceived learning experience (Yang, Wen, Howley, Kraut, & Rosé, 2015). Identifying and addressing confusion in MOOC discussion forums can be challenging for intervention due to the volume of activities, yet automated solutions have been proposed. For instance, confusion detection can be done through identifying learners whose language shows linguistic features associated with confusion and frustration (Yang, Wen, Howley, Kraut, & Rosé, 2015). It also can be done through identifying unsolved questions in conversations through analyzing sentence structures and sequences (Almatrafi et al., 2018; Yang, Wen, & Rosé, 2014b). Once confusion is detected, the support can be provided through automatically recommending relevant learning resources (Agrawal, Venkatraman, Leonard, & Paepcke, 2015) or selecting and engaging learners with relevant expertise to provide help (Yang, Adamson, & Rosé, 2014).

6.3.4. Trajectories with neutral changes and trajectories without changes

Two of the frequent trajectories showed only the change from reciprocators to input providers. One of them featured non-deep thinkers who had medium contribution quantity, peripheral social status, and repeated interactions with peers. The other featured non-deep thinkers who had low contribution quantity, peripheral social status, and one-off interaction. Together they made up 5% of the trajectories in StatLearn and 18% of the trajectories in SciWrite. Three of the frequent trajectories did not show change in any of the five aspects. Learners with these trajectories were all non-deep thinkers with peripheral social status. One trajectory featured moderate reciprocators who had repeated interactions with the same peers. One trajectory featured moderate providers who had repeated interactions with the same peers. The other trajectory featured minimal reciprocators who had one-off ties. Together they made up 45% of the trajectories in StatLearn and 21% of the trajectories in SciWrite.

Case studies for the trajectories with neutral changes and the trajectories without change did not yield useful explanation for the lack of constructive changes. To better understand the learners with these trajectories and their needs for learning support, future research can investigate their motivations, learning goals, and perception of the learning experience.

6.4. Implications for MOOC research

6.4.1. Multi-aspect approaches to understanding position and position trajectory

MOOC discussion forum participation is a complex process. This study showed the importance and usefulness of examining forum participation across multiple aspects together. First, positions identified based on multiple contribution and social characteristics greatly improved the understanding of participant groups. For instance, prior MOOC studies identified the provider group and reported they generally obtained higher final scores than other learners (Hecking et al., 2017; Liu et al., 2016). The current study further found that learners with the provider characteristics are heterogeneous in other aspects such as deep consideration of the discussion content and connectedness in the social network. Unveiling these differences are critical for understanding different provider groups' impact on forum interaction and their need for learning support. For example, deep thinking providers and non-deep thinking providers contributed different proportions of deep contributions to forum interaction and modeled different ways of thinking for other learners (McKendree et al., 1998). The non-deep thinking providers may need guidance to adopt more complex way of thinking so that they can achieve better learning outcomes (Anderson & Krathwohl, 2001). Similarly, central providers and peripheral providers differed in the number of people they interacted with and the number of conversations they participated in. This disparity indicates the two groups differed dramatically in how broadly they impacted forum interaction. Peripheral providers may need guidance to enrich their learning by accessing more people and topics.

Second, position trajectories that characterize concurrent changes in different aspects provided more well-rounded understanding of forum participation processes. MOOC research has noted the complexity of participation process (Bergner, Kerr, & Pritchard, 2015; Goggins & Xing, 2016; Tang et al., 2016), but changes in participation have mostly been investigated from a single aspect (Hecking et al., 2017; Tang et al., 2018; Yang, Wen, & Rosé, 2014a). The current study revealed changes in position usually occurred in multiple aspects. Adopting a multi-aspect approach can result in more thorough understanding of MOOC learners' participation processes, challenges, and need for support. For instance, a majority of MOOC learners decrease or terminate forum participation as the course proceeds, but disengagement is a process that can be curbed if addressed in a timely fashion (Tang et al., 2018; Yang et al., 2013). While a decrease in contribution quantity can signal the risk of disengagement (Tang et al., 2018), simultaneous changes in deep consideration, providing activities, and social connections can offer insights into its potential causes, such as inadequate prior knowledge, unaddressed confusion, or lack of support for a social learning environment (see Section 6.3). Thus examining position trajectories across multiple aspects allows for both detecting pivotal points for intervention (even before contribution quantity drops) and designing well-targeted learning support (Yang et al., 2013; Zhu et al., 2016).

To analyze the contribution characteristics from multiple aspects the current study developed a content analysis scheme that examined discussion messages for relatedness to the course content, input seeking and providing activities, and deep consideration of the discussion content. This is a valuable addition to the content analysis repertoire for MOOC discussion. The content analysis achieved good coding agreement across the two courses and different time periods which offers evidence for the reliability of the scheme.

6.4.2. The moderate contributors: An overlooked participant group

Prior MOOC studies have paid much attention to the prolific learners and the learners with high centrality in discussion forums (Huang et al., 2014; Tawfik et al., 2017; Wise, Cui, & Jin, 2017). Learners who had low forum participation levels and peripheral social status have also been investigated in various MOOC studies, especially those focusing on attrition and disengagement (Sinha, Li, Jermann, & Dillenbourg, 2014; Yang et al., 2013; Zheng, Rosson, Shih, & Carroll, 2015). The current study identified three moderate contributor groups that are under-researched in the MOOC literature: moderate deep thinkers, moderate reciprocators, and moderate providers. Members of these groups made a medium amount of contribution in a small number of conversations, interacted with a small number of people, and had some repeated interactions with the same peers. The three groups made up a considerable proportion of the forum population in each time period (26% to 35% in the current study). Understanding the three moderate contributor groups is important in that they showed unique characteristics that were not found in the prolific and the minimal contributor groups. For instance, deep thinkers were the only group that not only frequently adopted complex approaches to learning but also added deep messages to on-going non-deep conversations. This group can be useful for understanding connections between contribution quantity, contribution quality, and learning performance (Dowell et al., 2016; Huang et al., 2014). Moderate reciprocators appeared to have stronger social ties in their ego networks than other moderate contributors, which may relate to their balanced information exchange behaviors. Studying this group can help understand the constructive posting patterns for promoting social bonding (Gillani, Yasseri et al., 2014). Moderate providers are also interesting in that this group seemed to make up a bigger proportion of the forum population in the final time period than in the earlier time periods. They can be useful for understanding the factors that associate with changes in input providing activities (Howley et al., 2015).

6.4.3. Changes in positioning: An alternative perspective on MOOC learning

Most work on MOOC learning has adopted a course performance perspective that measures learning outcomes with grades and certificates (Gardner & Brooks, 2018). Course performance is a straightforward metric, but is inadequate for capturing learning in other forms, such as development in learner's relationship with the domain knowledge and the peers (Wenger, 1998). It is also not suitable for studying MOOC learners who are not performance-oriented (Kizilcec et al., 2013). Other MOOC work has adopted a post-course perspective to study MOOC learning. This perspective examines MOOC learners' post-course activities related to the course they had taken for signs of learning, such as career advance (Wang, 2017) and changes in activities in online learning communities related to the course (Chen et al., 2016). While the post-course perspective has the potential to capture learning from the participation perspective, the learning outcomes are assessed after the course has concluded. Thus it can be limited in its usefulness for understanding and supporting the learning process.

The position perspective taken in the current work provides a useful complement to the existing perspectives. By studying learners' participation process in the discussion forums, this perspective examines learning as development in their way of participation, which is not necessarily related to the performance orientation. Adopting this perspective, the current study revealed a different kind of learning (admittedly for a small percentage of learners) than what have been captured from the performance perspective and the post-course perspective. By examining the changes in learners' connections with the domain knowledge and people around them, this perspective also resulted in useful insights about their challenges and learning needs. Given the diversity of reasons people enroll in MOOCs this offers a useful complement for understanding and supporting MOOC learning.

6.5. Implications for MOOC teaching and learning

This study has important implications for MOOC teaching and learning. First, the fact that most forum participants did not improve their interaction over time raises a critical question about the value of MOOC forums and how they can be redesigned to better support learning. Using discussion forums to support learning has been a focus of MOOC design efforts over the past several years. One strategy is to optimize forum usability and assist learners to locate useful information and people. These efforts range from the design of basic interface features such as keyword search and post ranking system to more complex designs, such as content-based discussion classifiers (e.g., Cui, Jin, & Wise, 2016) and question recommendation systems (e.g., Yang, Adamson, & Rosé, 2014). These efforts can potentially mitigate some challenges that hamper MOOC learners from successfully accessing the knowledge and perspectives shared by other participants. Another strategy draws from the experience of group discussion in formal educational environments and introduces small group and collaborative discussion into MOOCs (Wen et al., 2015; Wichmann et al., 2016; Zheng, Vogelsang, & Pinkwart, 2015). This approach integrates forum discussion activities into the overall learning design. By designing group projects, tasks, and roles, it anchors forum interaction with clearly defined group goals and group boundaries. This strategy can potentially alleviate the problems causes by information overload and constantly changing discussants. Small group discussion environments also have the potential to help group members build stronger social bonding through repeated interactions. In complement to these existing approaches, the current study leads to the proposal of a different strategy. That is one that focuses on the development of MOOC learners to become more competent forum participants who can make efficient use of the resources in the massive and open discussion forums while consciously overcoming the challenges. As discussed in the earlier sections in this chapter, this can be done through diagnosing learners' position trajectories and providing customized learning support, such as guiding them to adopt deeper approaches to the discussion content, participate across multiple discussions, and engage in extended interactions with the same peers. Moreover, the information about learners' positions and position trajectories can be provided to the learners to assist them understanding and regulating their own learning (Wise et al., 2013). Through innovatively aggregating these approaches, it may be possible for the MOOC community to be more successful in supporting MOOC learners to reap the benefit of studying in the world classroom.

A second implication of this work for MOOC teaching and learning comes from the finding that learners mostly had weak ties in their ego network, indicating the need to design forum activities to strengthen social connections. Strong social connections and cohesive learning communities are considered promising for alleviating the high dropout rate and the lack of social learning environment in MOOCs (Brown et al., 2015; Rosé et al., 2014). The active participants in MOOC forums (e.g., instructors, TAs, active learners) are often considered important figures in the social networks (Jiang et al., 2015; Poquet & Dawson, 2016). But as seen in the enthusiastic central contributors, even the active and well-connected learners only developed strong ties with a small proportion of people that they interacted with. Thus their impact on facilitating broad social bonding among forum participants should be estimated accordingly. In contrast, learning design for forum discussion can be promising for promoting stronger social ties more broadly (Kellogg et al., 2014; Wise & Cui, 2018a). Discussion activities can be designed to encourage repeated interactions with the same peers through various means, such as quiz questions and assignments that can incur in-depth and extended conversations, designated discussions on different topics that lead learners to participate with the same people across conversations, and small group projects that require learners to interact intensively with the same peers for a prolonged period of time (Wen et al., 2015; Zheng, Vogelsang, & Pinkwart, 2015). In addition to learning design, the instructional team can purposefully guide learners to have extended interactions with each other. Wise and Cui (2018a) found the strength of learner-learner connections in a content-oriented discussion network was associated with the instructor's facilitation style: stronger learner-learner connections were found in the community with an instructor who guided learner to work out the answers themselves than in the community with an instructor who usually provided straightforward answers.

A third implication of the work is that participation positions identified in this study can inform forum facilitation and learning design. Learners with different positions can be engaged to help support different types of emergent learning needs in discussion forums. For instance, the enthusiastic central providers can transmit knowledge and resources to a broad range of peers. Engaging these learners as "community TAs" may improve responsiveness in the discussion forum. At the same time, there is also value in engaging the deep thinkers to help facilitate forum interactions. While they model the deep and complex approaches to the discussion content, the instructional team can focus on guiding and scaffolding other learners to pick up these approaches. Moreover, learners' positions can be useful for organizing discussion activities. For instance, small group collaboration is considered to be a promising learning design format for MOOCs (Goggins et al., 2016; Tawfik et al., 2017; Wen e al, 2015). Zheng, Vogelsang, and Pinkwart (2015) noted grouping solutions designed based on learner characteristics were more likely to lead to better retention rate and learning performance than random grouping solutions in small group collaboration in MOOCs. As learner's position profiles include multiple participation characteristics related to how learners interact in relation to each other, they can provide useful information for the grouping solutions.

6.6. Limitations and future work

The scope of this study was limited to examining the positions and position trajectories in a statistics MOOC and a writing MOOC. Additional work on MOOCs on other subjects, using other pedagogies (e.g., constructivism), and involving other forms of discussion designs (e.g., structured small group discussion) is needed to determine the extent to which the current findings generalize more broadly. At the same time, it is also important to further validate and refine the observed patterns. Future work can replicate this study using larger datasets so that the positions identified through cluster analysis contain enough cases for post hoc analysis (e.g., Wise et al., 2013; Kovanović et al., 2019). It can also be worthwhile to study position trajectories in other time spans, such as across milestone events within a course (e.g., segmenting a course based on assignment deadlines) and across courses (e.g., MOOC specialization courses). In addition, this study was only able to examine changes in *posting* activities and construct social networks based on *posting* activities alone, as the examined discussion forums did not support tracing *reading* activities at the post level. *Reading* is an important component of forum participation (Wise & Chiu, 2011). Learners who only read the discussion content without posting any message are common in MOOC forums (Mustafaraj & Bu, 2015). Future work on the changes between readers and contributors can further expand the understanding of positions and position trajectories. Moreover, the Direct Reply social networks built in the current study only reflected the responding relations between learners, but social connections in discussion forums can also be formed and strengthened through *reading* activities (Wise, Cui, & Jin, 2017). Future work can include post-level reading data to build social networks and verify to what extent the patterns observed in the current study still hold.

Finally, the observations about positions and position trajectories were made based on forum participation data alone. Future work can deepen the understanding of the identified positions and position trajectories by using other data collection methods such as survey and interview to investigate different groups' motivation, goal, and learning experience. This can be particularly useful for understanding the minimal and moderate contributor groups who generate limited amount of data in discussion forums (Adamopoulos, 2013; Kizilcec et al., 2013; Zheng, Rosson et al., 2015).

Looking forward, this study is the first effort among MOOC research to examine changes in multiple contribution and social characteristics over time. To use the resultant insights to support MOOC practice, future studies can consider automating the position and trajectory identification processes through integrating real-time content analysis and social network analysis. The real-time information about positions and position trajectories in the discussion forum can be used in teaching dashboards to assist MOOC instructors' facilitation / intervention decisions (e.g., Cobos, Gil, Lareo, & Vargas, 2016; Jiang et al., 2015). Such information can also be provided in learning dashboards together with customized learning support and recommendations to help learners better understand and regulate their own learning (e.g., Davis et al., 2017; Pardos, Tang, Davis, & Le, 2017).

6.7. Conclusion

Discussion forums provide the primary venue for MOOC learners to interact and exchange learning support. Interacting substantively in discussions is critical for MOOC learners to tap these resources and become more successful in learning. This study examined MOOC learners' participation patterns in discussions as positions that can be characterized by their contribution and social characteristics. It was found learners in the same forum took several distinctive participation positions that differed in contribution quantity, input seeking and providing activities, deep consideration of the discussion content, social centrality, and strength of social connections. This study further found a substantial proportion of forum participants who participated in multiple time periods and changed positions over time. A small number of these participants showed development in their way of participation while a majority of them did not show constructive changes. Additionally, this work identified three moderate contributor groups that are under-researched in the MOOC literature. Studying these groups can expand the knowledge about the MOOC learner population and forum interaction. These findings demonstrate the usefulness of the position perspective for understanding MOOC learning and both the need and potential avenues to help MOOC learners become more competent forum participants. This new line of work is promising for making progress on tapping the potential of MOOC forums that is not yet realized.

References

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In *Proceedings of the 34th International Conference on Information Systems* (vol. 5, pp. 4720-4740).

Agrawal, A., Venkatraman, J., Leonard, S., & Paepcke, A. (2015). YouEDU: Addressing confusion in MOOC discussion forums by recommending instructional video clips. In *Proceedings of the 8th International Conference on Education Data Mining* (pp. 297-304). New York, NY, USA: ACM.

Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education, 118*, 1-9.

Almeda, M. (2018). Comparing the factors that predict completion and grades among for-credit and open/MOOC students in online learning. *Online Learning, 22*(1), 1-18.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Arnaboldi, V., Conti, M., Passarella, A., & Dunbar, R. I. (2017). Online social networks and information diffusion: The role of ego networks. *Online Social Networks and Media, 1*, 44-55.

Aspin, D. N., & Chapman, J. D. (2000). Lifelong learning: Concepts and conceptions. *International Journal of lifelong education*, *19*(1), 2-19.

Barab, S. A., & Duffy, T. (2000). From practice fields to communities of practice. *Theoretical foundations of learning environments, 1*(1), 25-55.

Barba, P. D., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning, 32*(3), 218-231.

Bartholomé, T., Stahl, E., Pieschl, S., & Bromme, R. (2006). What matters in help-seeking? A study of help effectiveness and learner-related factors. *Computers in Human Behavior, 22*(1), 113-129.

Belanger, Y. & Thornton, J. (2013). Bioelectricity: A quantitative approach - Duke University’s first MOOC. Retrieved September 14, 2018 from http://hdl.handle.net/10161/6216.

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 234–241). New York, NY, USA: ACM.

Boroujeni, M. S., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. In *Proceedings of the 7th International Learning Analytics & Knowledge Conferenc*e (pp. 128-137). New York, NY, USA: ACM.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment, 8*, 13-25.

Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. M. F. (2014). Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE transactions on Learning Technologies, 7*(4), 346-359.

Brown, R., Lynch, C., Wang, Y., Eagle, M., Albert, J., Barnes, T., ... McNamara, D. (2015). Communities of performance & communities of preference. In *Proceedings for CEUR Workshop,* 1446. Retrieved September 21, 2018 from http://www.upenn.edu/learninganalytics/ryanbaker/GEDM_2015_Submission_2.pdf.

Budsankom, P., Sawangboon, T., Damrongpanit, S., & Chuensirimongkol, J. (2015). Factors affecting higher order thinking skills of students: A meta-analytic structural equation modeling study. *Educational Research and Reviews*, *10*(19), 2639-2652.

Callaway, D. S., Newman, M. E., Strogatz, S. H., & Watts, D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical review letters*, *85*(25), 5468-5471.

Cantor, A. B. (1996) Sample-size calculation for Cohen's kappa. *Psychological Methods*, *1*, 150-153.

Cesareni, D., Cacciamani, S., & Fujita, N. (2016). Role taking and knowledge building in a blended university course. *International Journal of Computer-Supported Collaborative Learning*, *11*(1), 9-39.

Chen, G., Davis, D., Lin, J., Hauff, C., & Houben, G. J. (2016). Beyond the MOOC platform: Gaining insights about learners from the social web. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 15-24). New York, NY, USA: ACM.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

Christopher, M. M., Thomas, J. A., & Tallent- Runnels, M. K. (2004). Raising the bar: Encouraging high level thinking in online discussion forums. *Roeper Review, 26*(3), 166-171.

Chung, K. S. K., & Paredes, W. C. (2015). Towards a social networks model for online learning & performance. *Journal of Educational Technology & Society, 18*(3), 240-253.

Clow, D. (2013). MOOCs and the funnel of participation. In *Proceedings of the 3rd International Conference on Learning Analytics & Knowledge* (pp. 185-189). New York, NY, USA: ACM.

Cobos, R., Gil, S., Lareo, A., & Vargas, F. A. (2016). Open-DLAs: An open dashboard for learning analytics. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 265-268). New York, NY, USA: ACM.

Cohen, A., Shimony, U., Nachmias, R., & Soffer, T. (2019). Active learners' characterization in MOOC forums and their generated knowledge. *British Journal of Educational Technology, 50*(1), 177-198.

Coomey, M., & Stephenson, J. (2001). Online learning: It is all about dialogue, involvement, support and control-according to the research. In Stephenson, J. (Ed.) *Teaching and Learning Online: Pedagogies for New Technologies* (pp. 37-52). London : Kogan Page; Sterling, VA: Stylus Pub.

Corrin, L., de Barba, P. G., & Bakharia, A. (2017). Using learning analytics to explore help-seeking learner profiles in MOOCs. *In Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 424-428). New York, NY, USA: ACM.

Cui, Y., & Wise, A. F. (2015). Identifying content-related threads in MOOC discussion forums. In *Proceedings of the 2nd ACM Conference on Learning@ Scale* (pp. 299-303). New York, NY, USA: ACM.

Cui, Y., Jin, W. Q., & Wise, A. F. (2017). Humans and machines together: Improving characterization of large scale online discussions through Dynamic Interrelated Post and Thread Categorization (DIPTiC). In *Proceedings of the 4th ACM Conference on Learning@ Scale* (pp. 217-219). New York, NY, USA: ACM.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education, 2012*(3). doi: 10.5334/2012-18.

Daniel, J. (2014). Foreword to the special section on Massive Open Online Courses MOOCs – Evolution or revolution? *Journal of Online Learning and Teaching*, *10*(1), i–iv.

Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the theory of social behaviour, 20*(1), 43-63.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G. J. (2017). Follow the successful crowd: Raising MOOC completion rates through social comparison at scale. In *Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 454-463). New York, NY, USA: ACM.

De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. In *Proceedings for the 11th International Conference on Intelligent Systems Design and Applications* (pp. 88-93). IEEE.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2009). Structuring asynchronous discussion groups: The impact of role assignment and self-assessment on students' levels of knowledge construction through social negotiation. *Journal of Computer Assisted Learning, 25*(2), 177-188.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course": Reconceptualizing educational variables for Massive Open Online Courses. *Educational Researcher, 43*(2), 74-84.

Dennen, V. P. (2007). Presence and positioning as components of online instructor persona. *Journal of Research on Technology in Education, 40*(1), 95-108.

Dennen, V. P. (2011). Facilitator presence and identity in online discourse: Use of positioning theory as an analytic framework. *Instructional Science, 39*(4), 527-541.

Dillahunt, T. R., Wang, B. Z., & Teasley, S. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *The International Review of Research in Open and Distributed Learning, 15*(5). doi: 10.19173/irrodl.v15i5.1841.

Donner A. (1999). Sample size requirements for interval estimation of the intraclass kappa statistic. *Communication in Statistics 28*, 415-429.

Dowell, N. M., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gašević, D., Hennis, T. A., Vries, P., & Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 205-257).

Dowell, N., Nixon, T., & Graesser, A. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multi-party interactions. *Behavior research methods, 2018*, 1-35.

Eynon, R., Hjorth, I., Yasseri, T., & Gillani, N. (2016). Understanding communication patterns in MOOCs: Combining data mining and qualitative methods. In S. ElAtia, D. Ipperciel, & O. R. Zaïane (Eds.), *Data Mining and Learning Analytics: Applications in Educational Research* (pp.207–221). New Jersey: John Wiley & Sons.

Fahy, P. J., Crawford, G., & Ally, M. (2001). Patterns of interaction in a computer conference transcript. *The International Review of Research in Open and Distributed Learning, 2*(1).

Fincham, E., Gaševic, D., & Pardo, A. (2018). From social ties to network processes: Do tie definitions matter?. *Journal of Learning Analytics, 5*(2), 9-28.

Forestier, M., Stavrianou, A., Velcin, J., & Zighed, D. A. (2012). Roles in social networks: Methodologies and research issues. *Web Intelligence and Agent Systems: An International Journal, 10*(1), 117-133.

Gaebel, M. (2013). MOOCs: Massive Open Online Courses. *EUA Occasional papers.* Retrieved June 21, 2018, from http://supporthere.org/sites/default/files/eua_occasional_papers_moocs_4.pdf.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2), 127-203.

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*(2-3), 87-105.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of distance education, 15*(1), 7-23.

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education, 24*, 53-65.

Gašević, D., Kovanović, V., Joksimović, S., & Siemens, G. (2014). Where is research on Massive Open Online Courses headed? A data analysis of the MOOC Research Initiative. *The International Review of Research in Open and Distributed Learning, 15*(5), 134-176.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education, 23*, 18-26.

Gillani, N., Eynon, R., Osborne, M., Hjorth, I., & Roberts, S. (2014). Communication communities in MOOCs. Retrieved March 18, 2017, from https://arxiv.org/pdf/1403.4640.pdf?.

Gillani, N., Yasseri, T., Eynon, R., & Hjorth, I. (2014). Structural limitations of learning in a crowd: Communication vulnerability and information diffusion in MOOCs. *Nature Scientific Reports, 4,* 6447 (2014). doi: 10.1038/srep06447.

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). A conceptual and operational definition of 'social role' in online community. In *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1-11). IEEE. doi: 10.1109/HICSS.2009.6

Goggins, S. P., Galyen, K. D., Petakovic, E., & Laffey, J. M. (2016). Connecting performance to social structure and pedagogy as a pathway to scaling learning analytics in MOOCs: An exploratory study. *Journal of Computer Assisted Learning, 32*(3), 244-266.

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education, 94*, 241-251.

Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. *Internet Research, 5*, 19-22.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104-137.

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.

Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014). Attrition in MOOC: Lessons learned from drop-out students. In *Proceedings for International Workshop on Learning Technology for Education in Cloud* (pp. 37-48). Springer, Cham.

Harré, R., & Van Langenhove, L. (Eds.). (1998). *Positioning theory: Moral contexts of international action*. Wiley-Blackwell.

Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. B.Resnick, J. M.Levine, & S. D.Teasley (Eds.), *Perspectives On Socially Shared Cognition* (pp. 331–348). Washington, DC: American Psychological Association.

Hecking, T., Chounta, I. A., & Hoppe, H. U. (2017). Role modelling in MOOC discussion forums. *Journal of Learning Analytics, 4*(1), 85-116.

Henri, F. (1992). Computer conferencing and content analysis. In Kaye, A (Ed.), *Collaborative Learning Through Computer Conferencing* (pp. 117-136). London: Springer-Verlag.

Herrmann, T., Jahnke, I., & Loser, K. U. (2004). The role concept as a basis for designing community systems. In Darses, F (Ed.), *Cooperative systems design: Scenario-based design of collaborative systems.* In *Proceedings of the International Conference on the Design of Cooperative Systems* (pp. 163-178).

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of Massive Open Online Courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45-58.

Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and reality*. Full report. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University, New York, NY.

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education, 98*, 157-168.

Horn, M., & Christensen, C. (2013). Beyond the buzz: Where are MOOCs really going? Wired. Retrieved May 5, 2018, from http://www. wired. com/opinion/2013/02/beyond-the-mooc-buzz-where-are-they-going-really.

Houston, S. L., Brady, K., Narasimham, G., and Fisher, D. (2017). Pass the idea please: The relationship between network position, direct engagement, and course performance in MOOCs. In *Proceedings of the 4th ACM Conference on Learning@ Scale* (pp. 295-298). New York, NY, USA: ACM.

Howley, I., Tomar, G., Yang, D., Ferschke, O., & Rosé, C. P. (2015). Alleviating the negative effect of up and downvoting on help seeking in MOOC discussion forums. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 629-632). Springer, Cham.

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. In *Proceedings of the 1st ACM Conference on Learning@ Scale Conference* (pp. 117-126). New York, NY, USA: ACM.

Jiang, S., Fitzhugh, S. M., & Warschauer, M. (2014). Social positioning and performance in MOOCs. In *Proceedings of Educational Data Mining 2014 Workshops at the 7th International Conference on Educational Data Mining* (pp.55-58).

Jiang, S., Williams, A., Schenke, K., Warschauer, M., & O'Dowd, D. (2014). Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 273-275).

Jiang, Z., Zhang, Y., Liu, C., & Li, X. (2015). Influence analysis by heterogeneous network in MOOC forums: What can we discover?. In *Proceedings of the 8th International Conference on Education Data Mining* (pp. 242-249).

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). How do you connect?: Analysis of social capital accumulation in connectivist MOOCs. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (pp. 64-68). New York, NY, USA: ACM.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media?: A topic analysis of discourse in a cMOOC. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (pp. 156-165). New York, NY, USA: ACM.

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & De Kereki, I. F. (2016). Translating network position into performance: Importance of centrality in different network configurations. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (pp. 314-323). New York, NY, USA: ACM.

Karabenick, S. A., & Dembo, M. H. (2011). Understanding and facilitating self-regulated help seeking. *New Directions for Teaching and Learning, 126*, 33-43.

Kellogg, S., Booth, S., & Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *The International Review of Research in Open and Distributed Learning, 15*(5), 263-289.

Khalil, H., & Ebner, M. (2013). “How satisfied are you with your MOOC?”-A research study on interaction in huge online courses. In Proceedings of the World Conference on Educational Media and Technology (pp. 830-839). EdMedia.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In Proceedings of the 2nd ACM Conference on Learning@ Scale (pp. 57-66). New York, NY, USA: ACM.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in Massive Open Online Courses. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 170-179). New York, NY, USA: ACM.

Kolowich, S. (2013). Why some colleges are saying no to MOOC deals, at least for now. *The Chronicle of Higher Education*. Retrieved November 28, 2018, from http://www.chronicle.com/article/Why-Some-Colleges-Are-Saying/138863.

Kovanović, V. (2017). Assessing cognitive presence using automated learning analytics methods (PhD thesis). School of Informatics, University of Edinburgh, Edinburgh, UK.

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology, 46*(3), 510-527.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., ... & Gašević, D. (2019). Examining communities of inquiry in Massive Open Online Courses: The role of study strategies. *The Internet and Higher Education, 40*, 20-43.

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the 6th International Conference On Learning Analytics & Knowledge* (pp. 15-24). New York, NY, USA: ACM.

Krackhardt, D. (1992). The strength of strong ties: The importance of philo in organizations. In N. Nohria & R. Eccles (Eds.), *Networks and Organizations: Structures, Form and Action* (pp. 216-239). Boston, MA: Harvard Business School Press.

Leflay, K., & Groves, M. (2013). Using online forums for encouraging higher order thinking and 'deep' learning in an undergraduate Sports Sociology module. *Journal of Hospitality, Leisure, Sport & Tourism Education, 13*, 226-232.

Liu, W., Kidziński, Ł., & Dillenbourg, P. (2016). Semiautomatic annotation of MOOC forum posts. In Li, Y., Chang, M., Kravcik, M., Popescu, E., Huang, R., Kinshuk, . . . Chen, N (Eds.), *State-of-the-art and future directions of smart learning* (pp. 399-408). Springer Singapore.

Macleod, H., Haywood, J., Woodgate, A., & Alkhatnai, M. (2015). Emerging patterns in MOOCs: Learners, course designs and directions. *TechTrends, 59*(1), 56-63.

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of Massive Open Online Courses (MOOCs). *Computers & Education, 80*, 77-83.

McKendree, J., Stenning, K., Mayes, T., Lee, J., & Cox, R. (1998). Why observing a dialogue may benefit learning. *Journal of Computer Assisted Learning, 14*(2), 110-119.

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education (3)*2, 1-7.

Morrison, D. (2013). The ultimate student guide to xMOOCs and cMOOCs. *MOOC News and Reviews*. Retrieved March 11, 2018, from http://moocnewsandreviews.com/ultimate-guide-to-xmoocs-and-cmoocso.

Mustafaraj, E., & Bu, J. (2015). The visible and invisible in a MOOC discussion forum. In *Proceedings of the 2nd ACM Conference on Learning@ Scale* (pp. 351-354). New York, NY, USA: ACM.

Najafi, H., Rolheiser, C., Harrison, L., & Heikoop, W. (2018). Help-seekers and help-providers in MOOCs. In *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 1476-1485). Association for the Advancement of Computing in Education.

Nelimarkka, M., & Vihavainen, A. (2015). Alumni & tenured participants in MOOCs: Analysis of two years of MOOC discussion channel activity. In *Proceedings of the 2nd ACM Conference on Learning@ Scale* (pp. 85-93). New York, NY, USA: ACM.

Pardos, Z. A., Tang, S., Davis, D., & Le, C. V. (2017). Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework. In *Proceedings of the 4th ACM Conference on Learning@ Scale* (pp. 23-32). New York, NY, USA: ACM.

Parry, M. (2013). A star MOOC professor defects—at least for now. *The Chronicle of Higher Education*. Retrieved March 11, 2018, from http://www.chronicle.com/article/A-MOOC-Star-Defects-at-Least/141331.

Poquet, O., & Dawson, S. (2016). Untangling MOOC learner networks. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (pp. 208-212). New York, NY, USA: ACM.

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34-35.

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for Massive Open Online Courses. *European Journal of Open, Distance and E-Learning*, *2012*(2), 1-13.

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014). Social factors that contribute to attrition in MOOCs. In *Proceedings of the 1st ACM Conference on Learning@ Scale Conference* (pp. 197-198). New York, NY, USA: ACM.

Ross, J., Sinclair, C., Knox, J., Bayne, S., & Macleod, H. (2014). Teacher experiences and academic identity: The missing components of MOOC pedagogy. *Journal of Online Learning and Teaching*, *10*(1), 57–69.

Schussler, D. L. (2009). Beyond content: How teachers manage classrooms to facilitate intellectual engagement for disengaged students. *Theory into Practice, 48*(2), 114-121.

Scott, J. (2000). *Social Networks Analysis: A Handbook.* London, UK: SAGE Publications.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4-13.

Shah, D. (2014). MOOCs in 2014: Breaking down the numbers. *EdSurge News.* Retrieved March 3, 2016, from https://www.edsurge.com/n/2014-12-26-moocs-in-2014-breaking-down-the-numbers.

Shah, D. (2016). By the numbers: MOOCs in 2016. Retrieved April 13, 2017, from https://www.class-central.com/report/mooc-stats-2016/.

Shah, D. (2017). By the numbers: MOOCs in 2017. Retrieved June 21, 2018, from https://www.class-central.com/report/mooc-stats-2017/.

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers & Education, 34*(3), 177-193.

Sinha, T. (2014). Supporting MOOC instruction with social network analysis. Retrieved March 18, 2019, from https://arxiv.org/pdf/1401.5175.pdf.

Sinha, T., Li, N., Jermann, P., & Dillenbourg, P. (2014). Capturing ‘attrition intensifying’ structural traits from didactic interaction sequences of MOOC learners. In *Proceedings of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*. Retrieved March 6, 2019, from https://arxiv.org/pdf/1409.5887.pdf.

Sridhar, D., Getoor, L., & Walker, M. (2014). Collective stance classification of posts in online debate forums. In *Proceedings of ACL 2014 Joint Workshop on Social Dynamics and Personal Attributes in Social Media* (pp.109-117). ACL.

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.

Strijbos, J. W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*(4), 491-494.

Stump, G. S., DeBoer, J., Whittinghill, J., & Breslow, L. (2013). *Development of a framework to classify MOOC discussion forum posts: Methodology and challenges.* Retrieved June 21, 2018 from https://tll.mit.edu/sites/default/files/library/Coding_a_MOOC_Discussion_Forum.pdf.

Swan, K. (2002). Building learning communities in online courses: The importance of interaction. *Education, Communication & Information, 2*(1), 23-49.

Tang, H., Xing, W., & Pei, B. (2018). Exploring the temporal dimension of forum participation in MOOCs. *Distance Education, 39*(3), 353-372.

Tawfik, A. A., Reeves, T. D., Stich, A. E., Gill, A., Hong, C., McDade, J., ... & Giabbanelli, P. J. (2017). The nature and level of learner–learner interaction in a chemistry massive open online course (MOOC). Journal of Computing in Higher Education, 29(3), 411-431.

Thomas, S. L. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The journal of Higher Education, 71*(5), 591-615.

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *The International Review of Research in Open and Distributed Learning, 17*(2).

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *Proceedings of the 8th International Conference on Educational Data Mining*. Retrieved June 12, 2018, from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_89.pdf

Wang, Y. (2017). *Demystifying learner success: Before, during, and after a Massive Open Online Course* (PhD thesis). Teachers College, Columbia University, New York City, USA.

Wen, M., Yang, D., & Rosé, C. P. (2014a). Linguistic reflections of student engagement in Massive Open Online Courses. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. Retrieved March 3, 2019, from https://www.cs.cmu.edu/~diyiy/docs/icwsm14.pdf.

Wen, M., Yang, D., & Rosé, C.P. (2014b). Sentiment analysis in MOOC discussion forums: What does it tell us?. In *Proceedings of the 7th International Conference on Educational data mining*. Retrieved April 24, 2018, from http://www.cs.cmu.edu/~mwen/papers/edm2014-camera-ready.pdf.

Wen, M., Yang, D., & Rosé, C. P. (2015). Virtual teams in Massive Open Online Courses. In *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 820-824). Springer, Cham.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press, Cambridge, UK.

Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework*. Heerlen, The Netherlands: Open University of the Netherlands, Ruud de Moor Centrum. Retrieved April 25, 2019 from https://wenger-trayner.com/documents/Wenger_Trayner_DeLaat_Value_creation.pdf

Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., & Hoppe, H. U. (2016). Group formation for small-group learning: are heterogeneous groups more productive?. In *Proceedings of the 12th International Symposium on Open Collaboration* (pp. 14-17). New York, NY, USA: ACM.

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.

Wise, A. F., & Cui, Y. (2018a). Learning communities in the crowd: Characteristics of content related interactions and social relationships in MOOC discussion forums. *Computers & Education, 122*, 221-242.

Wise, A. F., & Cui, Y. (2018b). Unpacking the relationship between discussion forum participation and learning in MOOCs: Content is key. In *Proceedings of the 8th International Learning Analytics & Knowledge Conference* (pp. 330-339). New York, NY, USA: ACM.

Wise, A. F., Cui, Y., & Jin, W. Q. (2017). Honing in on social learning networks in MOOC forums: Examining critical network definition decisions. In *Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 383-392). New York, NY, USA: ACM.

Wise, A. F., Cui, Y., Jin, W. Q., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education, 32*, 11-28.

Wise, A. F., Saghafian, M., & Padmanabhan, P. (2012). Towards more precise design guidance: Specifying and testing the functions of assigned student roles in online discussions. *Educational Technology Research and Development, 60*(1), 55-82.

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y. T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science, 41*(2), 323-343.

Wong, J. S., Pursel, B., Divinsky, A., & Jansen, B. J. (2016). An analysis of cognitive learning context in MOOC forum messages. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1315-1321). New York, NY, USA: ACM.

Yang, D., Adamson, D., & Rosé, C. P. (2014). Question recommendation with constraints for massive open online courses. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 49-56). New York, NY, USA: ACM.

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop*. Retrieved May 17, 2019, from https://www.cs.cmu.edu/~diyiy/docs/nips13.pdf.

Yang, D., Wen, M., & Rosé, C. (2014a). Peer influence on attrition in massive open online courses. In *Proceedings of the 7th International Conference of Educational Data Mining* (pp. 405-406).

Yang, D., Wen, M., & Rosé, C. (2014b). Towards identifying the resolvability of threads in MOOCs. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 21-31).

Yang, D., Wen, M., Howley, I., Kraut, R., & Rosé, C. (2015). Exploring the effect of confusion in discussion forums of Massive Open Online Courses. In *Proceedings of the 2nd ACM Conference on Learning@ Scale (pp. 121-130)*. New York, NY, USA: ACM.

Zhang, J., Skryabin, M., & Song, X. (2016). Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis. *Distance Education, 37*, 270–286.

Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). Understanding student motivation, behaviors and perceptions in MOOCs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1882-1895). New York, NY, USA: ACM.

Zheng, Z., Vogelsang, T., & Pinkwart, N. (2015). The impact of small learning group composition on student engagement and success in a MOOC. In *Proceedings of the 8th International Conference of Educational Data Mining* (pp. 500-503).

Zhu, E. (1996). Meaning negotiation, knowledge construction, and mentoring in a distance learning course. In *Proceedings of Selected Research and Development Presentations at the 1996 National Convention of the Association for Educational Communications and Technology*. ERIC documents: ED397849.

Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science, 34*(6), 451-480.

Zhu, M., Bergner, Y., Zhang, Y., Baker, R., Wang, Y., & Paquette, L. (2016). Longitudinal engagement, performance, and social connectivity: a MOOC case study using exponential random graph models. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (pp. 223-230). New York, NY, USA: ACM.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45(*1), 166-183.

Appendix. Coding Guide

Each message is coded first on the dimension of whether it is related to the course content; then each content-related message is coded on two subsequent dimensions: whether it seeks input, provides input, or both; and whether it contains deep consideration of the content involved (see Figure A1).

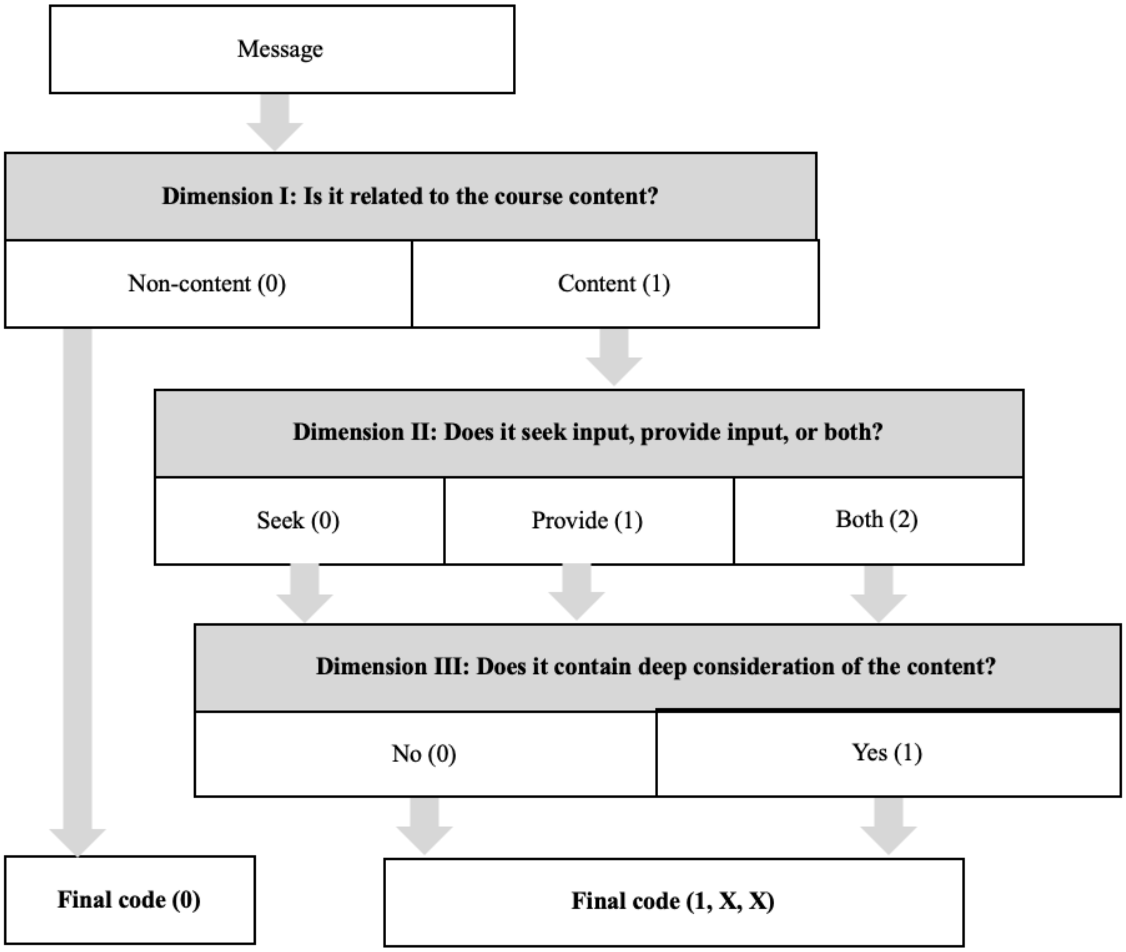


Figure A1. Content Analysis Dimensions and Categories

Dimension 1: Is the message related to the course content?

Each message is labeled as either non-content or content. Non-content messages involve interactions not directly related to the course or domain knowledge. These messages generally address technical, social / affective, and logistical issues. Technical issues often involve use of course site (e.g., assignment interface, browser compatibility) and software (e.g., downloading, configuration). Social and affective issues usually involve self-introduction, study group, and emotions. Logistical issues usually regard availability and quality of learning materials, course design, course policy and management (e.g., deadlines, forum regulation, assignment requirement, grade and credential). Content messages involve learning of subject matter knowledge specified in the course syllabus and domain knowledge related to the course subject. Common foci of content messages include Q&A, discussion, deliberation, and exchange of learning resources.

Table A1.

	Foci	Examples	
Non-content	Technical issues	1.	*How do I install RStudio?*
		2.	*How do we use the deletion feature?*
	Socializing / affection	3.	*Is anyone interested in forming a study group?*
		4.	*Thanks to all the staff for the great course!*
	Logistical issues	5.	*Where can I download the slides?*
		6.	*The due date for homework has been changed.*
		7.	*How can I post my answer without violating the Honor Code?*
		8.	*I haven't received the essays for peer review. What's going on?*
Content	Q&A	9.	*What is the difference between mean and median?*
		10.	*Can I use "examine" instead of "analyze" in question 2.4?*
	Discussion & deliberation	11.	*The goal of a supervised learning algorithm can be "inference" rather than "prediction", therefore to say that the goal of "any" supervised learning algorithm is prediction is problematic.*
		12.	*A passive voice attaches real meaning to these relationships, whereas an active voice sounds more assertive when the relationship is not.*
	Learning resource	13.	*I've found this link with introductions to statistics concepts: www.statsoft.com/textbook/elementary-statistics-concepts/.*
		14.	*I feel challenged in transforming negative words with verbs into positive words. Can anyone suggest some resources?*

Dimension 2: Does the message seek input, provide input, or do both?

This dimension addresses whether a message seeks input from others (e.g., answers, comments, scaffolding, resources), provides such input, or does both. Each message is labeled as seeking, providing, or both.

Table A2.

	Example	
Seeking*	*15.*	*Can I use "examine" instead of "analyze" in question 2.4?*
	16.	*I find the distinction between inference and prediction confusing.*
Providing	*17.*	*That's the method to use when you want to compare two groups of normally distributed data. Then why do we use non-parametric methods for the SNA data? Hint: think about what type of distribution is common in SNA data.*
	18.	*Steven Pinker wrote a great article on why academics write so poorly. It's worth a read. http://chronicle.com/article/Why-Academics-Writing-Stinks/<zipRedac>48989/.*
Both	*19.*	*I would say for these data log transformation is needed before you do the regression. Can anyone explain the difference between classification and clustering?*
	20.	*Hi Dave, you didn't get it wrong. I got the same result as yours by running the code suggested by Lilian. I'm wondering if there is a theoretical reason the three methods perform so closely.*

* When a learner provides their opinion in order to set up for a question, the message is labeled as seeking input.

* Rhetorical questions are labeled as seeking input.

Dimension 3: Does the message contain deep consideration of the content?

A message is labeled as either non-deep or deep based on the author's approach to the content. Non-deep messages generally involve exchange of simple and straightforward information, knowledge, and resource through seeking, naming, listing, and describing of such content. Messages containing deep consideration generally involve Q&A, discussion, deliberation, and exchange of learning resources related to abstract / complicated concepts, principles, theories, and processes. These messages demonstrate efforts to understand and make sense of such content through discussion, explanation, elaboration, interpretation, comparison, justification, and reasoning.

Table A3.

	Examples	
Non-deep	*21.*	*Is the mean of a left-skewed distribution smaller or greater than its median?*
	22.	*You can use lm() to build a linear regression model in R.*
	23.	*I used formula 12.7 and got the correct answer.*
	24.	*The correct model for predicting total sales is sales = f(TV, Radio, Newspaper), not f(TV) + f(Radio) + f(Newspaper).*
	25.	*You should avoid overusing nouns. Try using verbs instead.*
	26.	*The plural form of "information" should be "pieces of information".*
	27.	*How did you edit "Multiple mechanisms play only a small role or work by impacting one of the three primary mechanisms"? I don't have a clue.*
Deep	*28.*	*Why is the mean of a left-skewed distribution smaller than its median?*
	29.	*Here's what the R codes do step by step. First, x[y==1,] selects all rows of x where the corresponding row of y equals 1. The right hand side of the equation shifts these values by one …. Hope that helps.*
	30.	*This is how I derived formula 12.7. First, ….*
	31.	*Sales = f(TV) + f(Radio) + f(Newspaper) is having each of the predictor on a different graph and summing the values for each Xi. But the predictors (TV, Radio and Newspaper) should be used together and considered acting "simultaneously" to predict the final sales value. This is why we use f(TV, Radio, Newspaper) instead of three independent functions.*
	32.	*You present a fascinating view on using nouns instead of verbs. Perhaps it's a case of wanting to give gravitas to the text. The authors may use "stuffy" words to give a sense of authority to the text rather than to obscure.*
	33.	*"Information" is a mass noun, which always takes the singular verb. It's true that it's a "collection" of information. So if you wanted to single out one particular item from that collection of information, you would say "piece of information".*
	34.	*I don't understand "Multiple mechanisms play only a small role or work by impacting one of the three primary mechanisms". Mechanisms always play a role whatever the context is, so if they play a role they must work. This is redundant.*