Looking at MOOC discussion data to uncover the relationship between discussion paces, learners' cognitive presence and learning achievements

Zhi Liu^{1,2} · Xi Kong¹ · Sannyuya Liu^{1,2} · Zongkai Yang^{1,2} · Cuishuang Zhang²

Received: 21 May 2021 / Accepted: 6 February 2022 / Published online: 3 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The MOOCs (Massive Open Online Courses) forum carries rich discussion data that contains multi-level cognition-related behavior patterns, which brings the potential for an in-depth investigation into the development trend of the group and individual cognitive presence in discourse interaction. This paper describes a study conducted in the context of an introductory astronomy course on the Chinese MOOCs platform, examining the relationship between discussion paces (i.e., instructor-paced or learner-paced discussion), cognitive presence, and learning achievements. Using content analysis, lag sequential analysis, logistic regression, and grouped regression approaches, the study analysed the online discussion data collected from the Astronomy Talk course involving 2603 participants who contributed 24,018 posts. The findings of the study demonstrated the significant cognitive sequential patterns, and revealed the significant differences in the distribution of cognitive presence with different discussion paces and learning achievement groups, respectively. Moreover, we found that the high-achieving learners were mostly in the exploration, integration, and resolution phase, and learner-paced discussion had a greater moderating effect on the relationship between cognitive presence and learning achievements. Based on the findings and discussion, suggestions for improving the learners' cognitive presence and learning achievements in the MOOC environment are discussed.

Keywords MOOCs · Discussion forum · Cognitive presence · Discussion paces · Learning achievements

---

✉ Sannyuya Liu
lsy5918@mail.ccnu.edu.cn

Extended author information available on the last page of the article

1 Introduction

As a large-scale learning community widely adopted during the COVID-19 epidemic, MOOCs provides remote learners with a wealth of learning resources, interactive contexts, and flexible learning time, meeting the self-regulated learning needs for diversified learners around the world and providing an important way for informal learning (Hood et al., 2015; Fidalgo et al., 2014). The pedagogical design and learning regularities in MOOCs have drawn increasing attention in the interdisciplinary field of educational technology. Due to the openness and autonomy of MOOCs, however, students' low completion rates in MOOCs have been primary concerns for educators and institutions (Alraimi et al., 2015; Gütl et al., 2014; Clow, 2013; Yang et al., 2013). The lack of active interaction, teachers' inadequate guidance, and especially low cognitive level may be the key factors leading to low completion rates in MOOCs, while interactive discussion in the forum plays an important role in facilitating interaction and cognition of learners (Khalil & Ebner, 2014; Ferrari et al., 2020; Xiong et al., 2015). Flanders (1970) has found that discourse and interactive discussion accounts for about 80% of all teaching and learning behaviors through considerable empirical studies. MOOCs discussion forums provide an important communication channel allowing for group knowledge construction and individual cognitive processing, and learners also generate a large amount of discussion data in the forum. The investigation and guidance of discussion activities in a MOOC forum can help teachers better adjust the teaching process (Ramesh et al., 2014). At the same time, the participation and quality of learners' interactive discussion in online learning are closely related to the cognitive level and final learning achievements.

In a MOOC forum, large-scale discussion behavioral data can be fully recorded such as asking questions, answering questions, self-reflecting, quoting knowledge, opinion arguing and exchange (Wang et al., 2015). These data contain rich and multi-level cognition-related behavior patterns, which brings the potential for an in-depth investigation into the varying trend of the group and individual cognition processing. On the other hand, to analyze the cognitive levels in interactive discussion data, Garrison et al. (2001) also provide an efficient instrument based on the CoI (Community of Inquiry) framework, which has been very widely used to analyze discussions data, especially in MOOCs (Hu et al., 2020). Furthermore, discussion paces (i.e., instructor-paced or learner-paced) in a MOOC forum typically have a key effect on learning performance and cognitive processing (Moore et al., 2021). The instructor-paced discussion means that after completing a lesson, the instructor will post a discussion task related to course content for promoting learners' interaction and reflection. Also, the instructor will help to solve questions raised by learners. The learner-paced discussion refers to a discussion thread organized by learners, where learners can launch topic posts on interesting questions, and then other peers provide their insights to solve the issues. The instructor-paced discussion was often conducted with the focus of learning tasks, unit tests, and key knowledge, and the instructor typically requires approximately several hours per week responding to questions from learners. The

previous study also showed that weekly instructor intervention leads to higher rates of forum discussion completion in MOOC learning environments (Safitri et al., 2019). Instead, the learner-paced discussion might mainly focus on learners' needs such as confusion and curiosity in the learning process. According to the previous study (Moore et al., 2021), the different discussion pacing types might result in different cognitive levels which is related to learners' high order thinking and learning outcomes. Thus, to better engage learners and improve their cognitive levels and learning achievements in MOOCs, discussion paces need to be designed and considered (Lee & Recker, 2021). However, little literature has shown that how discussion paces affect learners' cognitive levels and their learning achievements in a MOOC discussion forum. Therefore, it is necessary to leverage the MOOC discussion data to investigate the relationship between discussion paces, learners' cognitive presence (one of the three main dimensions of the CoI framework to represent cognition), and learning achievements.

In this study, the main purpose was to investigate MOOC discussion data to reveal the relationship between discussion paces, learners' cognitive presence (one of the three main dimensions of the CoI framework to represent cognition), and learning achievements. To achieve this goal, we first collected discussion messages of the *Astronomy Talk* course with 2603 participants obtained from the Chinese MOOCs platform. Second, a total of 24,018 discussion messages were coded according to the cognitive presence of the CoI framework which was constructed by Garrison et al. (2001). Third, the cognitive presence, learning achievements, and their relationship with discussion paces were uncovered by jointly using lag sequential analysis, logistic regression, and grouped regression approaches. The study was guided by the following research questions:

- (1) What are the varying trends and distribution of cognitive presence on different discussion paces?
- (2) What are the differences of cognitive presence and their sequential patterns between different discussion paces of interactions?
- (3) What are the differences of cognitive presence and their sequential patterns between different achievement groups?
- (4) What is the influence of discussion paces on the relationship between cognitive presence and learning achievements?

The rest of the paper is structured as follows: Section 2 provides an overview of the theoretical foundation and related work. The methods in this study are detailed in Section 3, and the results are shown in Section 4. The discussion of the results is included in Section 5. Finally, the conclusion, limitations, and future work were described in Section 6.

2 Literature review

2.1 Community of Inquiry framework

Several models and frameworks representing student's cognition-related behavior in online discussion and physics learning environments have been proposed.

Among these methods, the Community of Inquiry (CoI) framework is one of the most studied structures when the objective is to describe the essential facets of social interactions and cognitive construction in online and blended education (Garrison et al., 2010). The CoI framework describes the educational experience in the online interactive environment with three key elements, which are cognitive presence, social presence, and teaching presence (Garrison et al., 2010; Kanuka & Garrison, 2004).

- (1) *Cognitive presence* is a primary component of the CoI framework, which reflects the process of acquiring and applying higher-order knowledge and it is also related to critical thinking and cognitive level.
- (2) *Social presence* represents the ability of learners to project themselves socially and emotionally in a community of inquiry, the main function of this dimension is to support cognitive and affective objectives of learning (Rourke et al., 1999).
- (3) *Teaching presence* is a significant influencing factor of learning achievements, student satisfaction, perceived learning, and sense of community (Garrison, 2007).

MOOC forums provide a platform for the exchange of ideas, asking questions, seeking answers, reporting errors and access course materials (Ramesh et al., 2014), students' cognitive and social interactive processes are hidden in their discussion data (Wang et al., 2015). Since the cognitive presence of the CoI framework provides a suitable theoretical framework for the classification of cognitive level in the MOOC discussion forum, this study uses it as a coding scheme. Garrison et al. (1999) define cognitive presence as “*the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication*”, which is built on Dewey's (1910) social constructivist view of learning and is implemented through a model of practical inquiry (Garrison et al., 2001) that defines four stages of the inquiry learning cycle:

- (1) *Triggering events*: In this phase, the learning cycle is triggered by an issue dilemma or divergence, which the instructor usually introduces in a formal educational setting.
- (2) *Exploration*: The main purpose of this phase is the exploration of ideas, where learners are asked to comprehend or grasp the nature of the issue, and then to collect a comprehensive range of relevant information. This is a divergent phase characterized by brainstorming, questioning, exchange of information.
- (3) *Integration*: In this phase, learners need to filter and synthesize the information generated in the phase of *exploration*, and later construct a meaning or possible solution based on continuous reflection and communication within the community of inquiry.
- (4) *Resolution*: In this phase the learner critically evaluates the solutions or hypotheses presented in the *integration* phase, leading to the resolution of the original problem. Often, however, the resolution of the original problem initiates a new learning cycle with a new *triggering events*.

In the CoI model, the above four phases represent the different levels of cognitive presence. The lowest level is the triggering events, while the highest level of cognitive presence is the resolution phase (Akyol & Garrison, 2011). In general, cognitive presence provides a description of the progressive phases of practical inquiry leading to the resolution of a problem or dilemma (Akyol & Garrison, 2011). On the other hand, the cognitive presence also provides a reliable scheme for qualitative analysis of interactive discussion data in MOOC forums. In this study, we used the content analysis to explore learners' cognitive presence hidden in MOOC discussion data, where we also used a coding scheme to assess the quality of cognitive presence developed by Garrison et al. (2001).

2.2 Cognitive presence and its sequential patterns analysis in the discussion forum

In recent years, researchers have conducted considerable studies on cognitive presence analysis. For instance, Kovanović et al. (2015) investigated the influence of technology use profiles (task-focused users, no-users, and highly intensive users) in asynchronous forums from the perspective of cognitive presence, and the result suggested that there was an important connection between technology use and students' cognitive presence. Gašević et al. (2015) proposed two different teaching presence strategies (externally-facilitated-regulation scaffolding and role assignment) and verified them through a quasi-experimental study. The results showed the effectiveness of role assignment to facilitate a high level of cognitive processing and presented a significant effect of the interaction between externally-facilitated regulation scaffolding and role assignment on cognitive presence. Olesova et al. (2016) found that the difficult questions were more likely to stimulate learners' high level of cognitive presence such as *integration* and *resolution*.

In addition, exploring sequential patterns of cognition-related behaviors is critical to uncover the temporal transition of learner cognitive presence, and lag sequential analysis (LSA) is usually used to uncover sequential relationships between each two adjacent learning behaviors in chronological order based on a statistical model (Bakeman & Gottman, 1997). To explore cognitive behavioral sequences in the learning process, Wu and Hou (2015) used LSA to explore the sequential discussion behavior patterns of knowledge construction behaviors and their difference among learners with different cognitive styles. Yang et al., (2015) proposed a two-tier test strategy and explored its effects on learners' behavior patterns. Huang et al., (2019) explored students' interactive behavior patterns and their learning sentiments by using both quantitative and lag sequential analysis in an online discussion forum. In this study, LSA, as a suitable method, was also used to explore the influence of discussion paces on the sequential pattern of cognitive presence.

Cognitive presence analysis is particularly necessary for tracking learners' cognitive processes (Coffield et al., 2004). The above studies indicated that cognitive presence analysis can help instructors explore the intrinsic and extrinsic factors that may influence learners' thinking, knowing, remembering, judging and problem-solving. However, these studies tend to analyze learners' cognitive presence from

a static perspective in the MOOC forum. Instead, there is a lack of research on how their cognition presence shifts and changes, that is, the sequential patterns of learners' cognitive presence are not yet clear. Thus, this study places more emphasis on the analysis of the sequential patterns of cognitive presence in different discussion paces and learning achievement groups from a dynamic perspective.

2.3 Influence of discussion paces on the cognitive presence and learning achievements

Despite numerous advantages, MOOC learning leads to a temporal and spatial separation between learners and instructors, possibly creating a barrier of interactive discourse information transmission, thus hindering the development of learners' cognitive presence (Wong et al., 2015). Since the discussion forums are the main way of interacting with peers in MOOC learning, the discussion paces (instructor/learner-paced) may be a key factor influencing the cognitive processing, knowledge constructing, and learning achievements. In this regard, instructors-paced discussion design and facilitation have been found to have a critical effect on learners' cognitive presence (Cohen et al., 2019). Xu et al. (2020) conducted a quasi-experiment to verify that instructor-paced discussion in WeChat-supported online learning had a positive influence on learners' cognitive level and learning achievements, but this study revealed learner-paced discussion did not influence it. Tsiotakis & Jimoyiannis (2016) also indicated that the instructor's active participation could promote members' engagement. Safitri et al. (2019) found that instructors-facilitation with specific guidance and cognitive strategies can help to improve students' creative thinking and academic performance.

In contrast, many empirical studies have presented different results. According to Tomkin and Charlevoix (2014), they used an A/B test to randomly assign MOOC participants a control group (learner-paced) or an intervention (instructor-paced) to identify the difference in learning achievements, participant rates, and students' satisfaction. Their result showed that the active interaction of professors had no significant effect on overall completion rates, student participation rates, and learning achievements. Similar results were presented in another study conducted by Zhao and Sullivan (2017), who concluded that instructor-paced interaction did not contribute directly to learners' cognitive presence and academic performance. Moore (2019) used LIWC as an analysis tool to code and analyze 2422 discussion posts, but neither instructor-paced nor learner-paced discussion had significantly influenced the students' cognitive processing and learning performance.

To sum up, the existing research generally focused on the effects of discussion paces on learners' cognitive and learning achievements, but the internal association mechanisms involved among the three factors need to be further explored. In addition, little research looked into the temporal sequence transitions and their distribution of learners' cognitive presence, thus it is necessary to conduct further research to better assist teachers to investigate learner cognitive processing dynamics and regulate the interactive teaching strategies in MOOC environment.

3 Methods

3.1 Research context and participants

The research context was a completely online course, namely *Astronomy Talk*, offered at the Chinese MOOCs platform (www.icourse163.org). This course focuses on the popularization science of astronomy, and its purpose is to enable learners to master the basic theories, fundamental knowledge and methods of astronomy, and to give them a sense of the charm of nature and science. Up to now, *Astronomy Talk* had opened 13 times and our research data was selected from the 11th course according to the quality of discussion posts, which was offered on a 15-week schedule ranging from December 2019 to March 2020. In this platform, learners can watch the instructional video and participate in forums under instructor-paced or learner-paced thematic discussion. The final learning achievements for this course consisted of weighted scores from the weekly test (50%), final exam (45%), and discussion performance (5%). In this course, only learners with a score equal to or higher than 60 points are eligible for the certificate (on a scale of 0–100).

The participants for this research delivered discussion forum posts under either an instructor-paced or learner-paced thematic discussion threads of the course *Astronomy Talk*. For this course, the number of registrations reached 36,402 from the beginning of the course to the present. In addition, 2603 participants (2602 learners and 1 instructor) enrolled in the 11th session of *Astronomy Talk*, and all participants have agreed to take part in the study. All participants were mostly from universities in China, and their common characteristic was a strong interest in astronomy. Only one instructor was involved in this MOOC course, namely Professor Xu. During the whole semester, Professor Xu needed to design various discussion topics related to the course content and encouraged learners to actively engage in the forum discussions. At the same time, in the instructor-paced group, Professor Xu also needed to respond to learners' questions at the end of each week's session and before the next session begins.

In addition, this study classified learners who got certificates into the high-achieving group and those who did not gain certificates into the low-achieving group (Peng & Xu, 2020). Thus, in this course, 122 learners are in the high-achieving group and the rest 2480 are in the low-achieving group.

3.2 Data collection

The research data in this study included all topic-related discussion posts of *Astronomy Talk* course collected from discussion forums of the Chinese MOOCs platform. In this study, we first performed a de-duplication operation on all posts and finally obtained a total of 24,018 posts (750 posts from the instructor and 23,268 posts from learners) after removing the meaningless posts containing only numbers and symbols. In addition, we also obtained some information corresponding to each post (e.g., time, poster, post type), as well as learners' information (e.g., user id, learning

duration, certificate record). To protect the privacy of participants, personal descriptors such as names, emails, phone numbers are removed before formal analysis.

3.3 Coding scheme

To investigate the relationship between discussion paces, cognitive presence, and learning achievements, and to provide a basis for the design of future MOOC discussion and facilitated strategies, a total of 24,018 messages were coded according to the cognitive presence of the CoI framework which constructed by Garrison et al. (2001). First of all, the coding scheme was discussed by two experienced coders. To improve the accuracy of coding and avoid content ambiguity, the sentence was used as the unit of analysis because it was more likely to contain a single concept, expression, or statement (Strijbos et al., 2006, Galikyan & Admiraal, 2019). Secondly, 2000 posts were randomly selected from the whole dataset (24,018) to be labeled by two coders independently, and the coded result reached an agreement statistic value of 0.849 (Cohen's Kappa), indicating a high level of consistency (Galikyan & Admiraal, 2019). All posts related to cognitive presence were coded into these four levels: 1) *triggering events*, 2) *exploration*, 3) *integration*, 4) *resolution*. The rest of the posts without cognitive units are coded as other "other". Thirdly, the two coders discussed and negotiated some inconsistencies in the coding process. After reaching a higher level of consensus, one of the coders coded all the remaining discussion posts (22,018). Table 1 shows the details of the coding rules.

Notably, after coding the posts based on the coding framework, it was found that some posts contained cognitive units (i.e., cognition-related posts), and also include non-cognitive units such as greeting and appreciation. For the follow-up analysis, we conducted a sentence-level cognition statistics for all posts and finally obtained 17,671 cognitive posts and 5597 non-cognitive posts.

3.4 Data analysis

Based on the four research questions posed in the study, the data analysis process was performed as follows:

To answer research question 1, we first divided the learners' cognition-related posts into instructor- and learner-paced groups and calculated the frequency and percentage of each dimension of cognitive presence. We also counted and visualized the temporal distribution of cognitive presence throughout the semester (15 weeks) under different discussion paces.

To address research questions 2 and 3, the Chi-square test and LSA were adopted to uncover differences in learners' cognitive presence and their sequential patterns under different discussion paces and learning achievement groups. In particularly, LSA was conducted using Generalized Sequential Querier 5.1 (GSEQ) software (Hou et al., 2015; Huang et al., 2019), and the following steps need to be followed to analyze the cognition-related posts in this study: 1) The coded data was feeded into the GSEQ in the prescribed format. 2) The GSEQ program was run to generate the .mds file. 3) Behavior sequence analysis was performed to obtain the behavior

Table 1 Coding scheme cognitive presence in MOOC discussions

Category	Code	Indicators	Example
Triggering events	S1	Recognizing the problem Sense of puzzlement	猎户座和天蝎座是处在北半球星空吗? / Are Orion and Scorpius in the northern hemisphere?
	S2	Divergence of ideas Exchanging ideas	我完全不同意他刚才的说法。 / I totally disagree with what he just said. 我想把我知道的计算方法分享在这里。 / I would like to share what I know about the calculation here.
		Suggestions for consideration Brainstorming	我觉得你最好使用单反相机拍星轨。 / I think you'd be better off using a DSLR to take star trails. 科技的发展让我们对地球有了更多的认识,也相信我们未来有机会去其他星球旅游或定居,甚至会发现外星人。 / The development of technology has given us a better understanding of the Earth and the belief that we will have the opportunity to travel or settle on other planets in the future, and even discover aliens.
Integration		Leaps to conclusions	黑洞也会向外释放能量,这些能量被释放完了黑洞可能会消失。 / The black hole will also release energy to the outside, the energy is released after the black hole may disappear.
	S3	Convergence—among group members	对这个问题我和前面那几位同学看法是一样的。 / I have the same opinion as the previous students on this issue.
		Convergence—within a single message	宇宙之外应该还会有更高级的文明和其他的宇宙,或者是说包含我们这个宇宙的更大的宇宙系统。 / There should be higher civilizations and other universes beyond the universe, or a larger cosmic system that contains our universe.
Resolution		Connecting ideas, synthesis	查了好多资料,类星体至今没有准确的说法。 / After checking a lot of information, there is no accurate term for quasars so far.
		Creating solutions	月球是离地球最近的天体,因此可以用地面上的三角测量法测量它的距离。 / The Moon is the closest celestial body to the Earth, so its distance can be measured by triangulation on the ground.
	S4	Application and solutions	尽量用手动对焦。晚上太暗,自动对焦经常不起作用,照出的月亮是虚的。如果照出来的月亮非常亮,掩盖了上面的细节,适当减少曝光时间。 / Try to use manual focus. At night is too dark, autofocus often does not work, the moon is illuminated is false. If the moon is very bright, covering up the details above, reduces the exposure time appropriately.

transition matrix and adjusted residuals matrix (Z values tables) for each group. 4) Based on the adjusted residual matrix of each group, we examined whether the temporal connections between cognitive sequences were statistically significant. The criterion is that, in a one-tailed test, the cognitive sequence with Z value greater than 1.96 is considered to be statistically significant at the level of 0.05. (Bakeman & Gottman, 1997; Sun et al., 2021), which can be used to identify if the tendency of transition from one behavior to another one significantly exceeds the expected value. 4) The sequential transitions of cognitive presence were visualized by Visio software, and then we compared the differences in cognitive sequential transitions between the different groups.

To answer research question 4, the logistic regression and grouped regression models were adopted to examine the moderating effect of discussion paces on the relationship between cognitive presence and learning achievements.

4 Findings

4.1 RQ1: What are the varying trends and distribution of cognitive presence on different discussion paces?

To address research question 1, we first calculated the frequency and percentage of each category of cognitive presence within coded discussion data. It is worth noting that cognition-related posts (17,671) were selected from the original posts (23,268). Table 2 presents the frequency and percentage of each level of cognitive presence under different discussion pacing types (instructor-paced and learner-paced discussion). In total, the most frequently observed cognitive presence was *exploration* (S2 = 8828) with an occurrence of 49.96%, followed by *integration* (S3 = 6808) which accounted for 38.53%. By contrast, the frequencies of *triggering events* (S1 = 907) and *resolution* (S4 = 1128) were lower (accounting for less than 12%). In addition, we also observed that learners in different discussion pacing groups differed not only in the frequency of postings but also in the distribution of the dimensions of cognitive presence. For instance, the number of cognition-related posts under instructor-paced discussions reached 17,180, much higher than the 491 posts under learner-paced discussions. As for instructor-paced discussion group, *exploration* (S2 = 8666) and *integration* (S3 = 6772) occurred most frequently, but *triggering events* (S1 = 657) and *resolution*

Table 2 The distribution of the levels of cognitive presence under instructor- and learner-paced discussion

	S1	S2	S3	S4	Total
Instructor-paced	657 (3.82%)	8666 (50.44%)	6772 (39.42%)	1085 (6.32%)	17,180
Learner-paced	250 (50.92%)	162 (32.99%)	36 (7.33%)	43 (8.76%)	491
Total	907 (5.13%)	8828 (49.96%)	6808(38.53%)	1128(6.38%)	17,671

(S4 = 1085) only accounted for about 10%. In contrast, the learner-paced discussion group tended to be more active in terms of *triggering events* (S1 = 250) that occupied the largest proportion and were followed by *exploration* (S2 = 162).

To further explore the temporal distribution of cognitive presence under different discussion paces throughout the whole semester, the weekly distribution of cognitive presence was aggregated and visualized in both instructor- and learner-paced discussion groups. As illustrated in Fig. 1, we can see that the four categories of cognitive presence change over time under different discussion pacing groups. Specifically, in the instructor-paced group, the *exploration* (S2) and *integration* (S3) of learners showed a pattern of increasing from week 3–9 and then fluctuating in the next few weeks. But *triggering events* (S1) and *resolution* (S4) remained at the low level of change with a small change throughout the semester. However, the learner-paced group seemed to be more active in the *triggering events* (S1) and *exploration* (S2), especially from week 3 to 15. The relatively high cognitive levels such as *integration* (S3) and *resolution* (S4) only peaked in weeks 11 and 10, respectively.

Fig. 1 Temporal distribution of cognitive presence with instructor-paced (a) and learner-paced (b) discussion in MOOC forum

4.2 RQ2: What are the differences of cognitive presence and their sequential patterns between different discussion paces of interactions?

To examine the difference in cognitive presence in terms of discussion pacing, the learners were divided into instructor- and learner-paced discussion groups based on the discussion guidance in the MOOC discussion forum. Then a Chi-square was conducted to compare the cognitive presence of learners in the two groups. By conducting a Chi-square test, we found that there was a significant difference in cognitive presence between the instructor- and learner-paced group ($\chi^2 = 2223.63$, $df = 3$, $p < 0.05$).

To probe the cognitive sequential patterns of students with different discussion paces, GSEQ software was used to perform the LSA with Z value as the indicator. GSEQ as a flexible application can process the encoded behavior data, and eventually generated adjusted residuals table of cognitive presence with different discussion pacing groups (shown in Table 3). In Table 3, the columns and rows represent the starting behavior and following behaviors respectively (Wu & Hou, 2015). Furthermore, the sequential transition diagrams of the cognitive presence were drawn to better display and compare the differences between sequential patterns of different paces of discussion groups (shown in Fig. 2 and Fig. 3). Notably, the arrowheads in these diagrams indicate the direction of behavior transfer, and numerical values represent the Z value.

As depicted in Fig. 2 and Fig. 3, there were several significant sequences of cognitive presence in both the instructor- and learner-paced groups. More specifically, in the instructor-paced discussion group, these significant sequences are “*triggering events* → *triggering events*” ($S1 \rightarrow S1$), “*exploration* → *exploration*” ($S2 \rightarrow S2$), “*integration* → *integration*” ($S3 \rightarrow S3$), “*resolution* → *resolution*” ($S4 \rightarrow S4$). As shown in Fig. 3, there were also four significant sequences of cognitive presence in the learner-paced group such as “*triggering events* → *exploration*” ($S1 \rightarrow S2$), “*exploration* → *triggering events*” ($S2 \rightarrow S1$), “*exploration* → *exploration*” ($S2 \rightarrow S2$), “*resolution* → *resolution*” ($S4 \rightarrow S4$). Furthermore, some differences in sequential patterns were also found in different

Table 3 Sequential analysis with different discussion pacing types (Adjusted residuals table)

Group		S1	S2	S3	S4
Instructor-paced	S1	17.48*	0.19	0.01	−3.23
	S2	0.94	61.22*	−21.97	−8.04
	S3	−1.46	−22.94	67.48*	−16.06
	S4	−2.29	−6.42	−17.9	54.36*
		S1	S2	S3	S4
Learner-paced	S1	−1.68	1.99*	1.57	−1.47
	S2	3.36*	3.73*	0.45	0.71
	S3	0.58	0.53	0.21	0.7
	S4	0.16	−0.48	−0.21	4.99*

Note: The z-values greater than 1.96 were considered significant, the boldface and one * was applied to mark the significance for each entry

* $p < 0.05$, ($Z > 1.96$)

Fig. 2 Transition diagram of cognitive presence for the instructor-paced discussion group

Fig. 3 Transition diagram of cognitive presence for the learner-paced discussion group

discussion pacing types by comparing the two transition diagrams of cognitive presence. For instance, the instructor-paced group showed repetitive action with respect to each of “*triggering events*”, “*exploration*”, “*integration*” and “*resolution*”, but there is a lack of transition between high levels of cognitive presence in the instructor-paced group. As shown in Fig. 3, in addition to the repetitive backtracking of “*resolution*” and “*exploration*”, we can find a cyclic trend between “*triggering events*” and “*exploration*” ($S1 \rightarrow S2 \rightarrow S1$).

Table 4 The frequency of the levels of cognitive presence under high- and low-achieving groups

	S1	S2	S3	S4	Total
Low-achieving group	808(5.00%)	8200(50.69%)	6220(38.45%)	948(5.86%)	16,176
High-achieving group	99(6.62%)	628(42.01%)	588(39.33%)	180(12.04%)	1495
All	907(5.13%)	8828(49.96%)	6808(38.53%)	1128(6.38%)	17,671

Table 5 Sequential analysis of learning achievement groups (adjusted residuals table)

		S1	S2	S3	S4
High-achieving group	S1	4.95*	0.58	0.44	−1.56
	S2	1.28	10.17*	2.52*	−2.83
	S3	−1.01	3.04	16.19	−4.66
	S4	−0.88	−1.75	−5.83	16.42*
		S1	S2	S3	S4
Low-achieving group	S1	15.2*	1.19	−1.48	−2.96
	S2	0.4	61.73*	−23.01	−8.65
	S3	−0.78	−24.31	66.06*	−14.93
	S4	−1.87	−7.43	−16.75	55.38*

Note: The z-values greater than 1.96 were considered significant, the boldface and one * was applied to mark the significance for each entry

* $p < 0.05$, ($Z > 1.96$)

4.3 RQ3: What are the differences of cognitive presence and their sequential patterns between different achievement groups?

To investigate the differences in learners' cognitive presence in terms of achievement groups, the cognition-related posts were divided into high- and low-achieving groups (shown in Table 4). Similar to question 2, the Chi-square test revealed a statistically significant relationship between the levels of cognitive presence and achievement groups ($\chi^2 = 109.869$, $df = 3$, $p < 0.05$). To further examine the sequential patterns of cognitive presence in high- and low-achieving groups, LSA was conducted to calculate the adjusted residuals Table (Z values) of coded cognition-related posts, and the results are shown in Table 5. Finally, the two transition diagrams (shown in Fig. 4 and Fig. 5) were visualized according to the significant sequence shown in Table 5.

Figure 4 depicts the significant sequences of students' cognitive presence in the high-achieving group. These significant sequences are “*triggering events* → *triggering events*” ($S1 \rightarrow S1$), “*exploration* → *exploration*” ($S2 \rightarrow S2$), “*resolution* → *resolution*” ($S4 \rightarrow S4$), and the sequential patterns demonstrated a significant transition from “*exploration*” to “*integration*” ($S2 \rightarrow S3$). As shown in Fig. 5, for the low-achieving group, these significant sequences seemed to be four repetitive

Fig. 4 Transition diagram of cognitive presence for high-achieving group

Fig. 5 Transition diagram of cognitive presence for low-achieving group

backtrackings such as “*triggering events* → *triggering events*” ($S1 \rightarrow S1$), “*exploration* → *exploration*” ($S2 \rightarrow S2$), “*integration* → *integration*” ($S3 \rightarrow S3$), “*resolution* → *resolution*” ($S4 \rightarrow S4$). Instead, the high-achieving group showed a meaningful transition from “*exploration*” to “*integration*” indicating an increase in cognitive presence significantly.

Table 6 Regression analysis between cognitive presence variables and learning achievements

Predictor Variables	B	S.E.	Wals	Sig.	Exp (B)
Triggering events	−2.262***	0.172	172.267	0.000	0.104
Exploration	0.751***	0.175	18.534	0.000	2.120
Integration	2.009***	0.228	77.456	0.000	7.455
Resolution	3.228***	0.155	431.015	0.000	25.233

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

4.4 RQ4: What is the influence of discussion pacings on the relationship between cognitive presence and learning achievements?

To answer research question 4, we first explored the relationship between cognitive presence and academic performance using binary logistic regression and suggested potential predictors of learning achievements (as shown in Table 6). By interpreting the result, we found that three dimensions of cognitive presence such as *exploration* ($B = 0.751$, $p < 0.001$), *integration* ($B = 2.009$, $p < 0.001$) and *resolution* ($B = 3.228$, $p < 0.001$) which all have a significantly positive effect on learning achievements. By contrast. The *triggering events* ($B = -2.262$, $p < 0.001$) demonstrated a significantly negative predictive effect on learning achievements.

To investigate whether the discussion pacings have a moderating effect on the relationship between cognitive presence and learning achievements, the grouped regression analysis was conducted in this section. A summary of our data analysis results is presented in Table 7. As seen in Table 7, we found that both instructor-paced ($\Delta F = 39.775$, $p < 0.001$) and learner-paced ($\Delta F = 39.775$, $p < 0.001$) moderated the relationship between learners' cognitive presence and learning achievements. By comparing the non-standardized coefficients ($B_{\text{learner-paced}} = 0.291$, $B_{\text{instructor-paced}} = 0.039$) shown in Table 7, it can be observed that the effect of cognitive presence on learning achievements was positive in both learner- and instructor-paced discussions. Furthermore, a simple slope plot was drawn to describe the moderating effect of different discussion pacing types. As shown in Fig. 6, the learning achievements of learners in both instructor-paced ($\beta = 0.048$, $p < 0.001$) and learner-paced ($\beta = 0.320$, $p < 0.001$) groups was closely related to their levels of

Table 7 Moderating effect on the relationship between cognitive presence and learning achievements

Model	Variables	B	SE	β	t	ΔR^2	ΔF	95%CI
learner-paced	constant	−0.041	0.038		−0.538	0.102	55.838***	[−0.096,0.055]
	cognitive	0.291	0.019	0.320	7.472***			[0.107,0.184]
instructor-paced	constant	0.064	0.008		4.060***	0.002	39.775***	[0.017,0.048]
	cognitive	0.039	0.003	0.048	6.307***			[0.013,0.025]

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Fig. 6 Simple slope plot of moderating effects

cognitive presence. However, the cognitive presence had more influence on learners' learning achievements under learner-paced discussion compared to the instructor-paced discussion.

5 Discussion

The present study explores the relationship between discussion pascings, learners' cognitive presence, and learning achievements in the MOOC discussion forum. Our special interest in the study mainly stems from the following reasons. First, the global epidemic COVID-19 has attracted more and more instructors and learners who regard MOOC as an important means of maintaining instruction and acquiring knowledge (Hodges et al., 2020). Meanwhile, the discussion pascings in MOOC learning has been considered as a potential impact factor on the cognitive presence of learners. Secondly, understanding the internal mechanisms of how discussion pascings affect learners' cognitive presence and learning achievements can help instructors adjust their online teaching strategies. Third, a preliminary descriptive statistic shows that there are obvious differences in the distribution of learners' cognitive presence and their sequential patterns. Besides, the two discussion pacing types, instructor-paced and learner-paced, had different levels of moderating effects on the relationship between cognitive presence and learning achievements. The following section further discusses the findings.

Concerning the distribution of cognitive presence, the result showed that the learners in the instructor-paced discussion group had a high level of cognitive presence in comparison with the learner-paced discussion group. The primary

reason may be that learners obtained an in-depth understanding of the related issues under the guidance of the instructor, thus reached a high level of cognition presence. This is consistent with previous findings (Joksimovi et al., 2015; Xu et al., 2020). According to Xu et al. (2020), the open questions (e.g., *If we were in a globular cluster, what would the night sky look like? Would there still be humans?*) proposed by the instructor could engage learners to participate in discussions, which may contribute to increase the quantity and quality of interactions and improve students' levels of cognitive presence. In terms of the weekly distribution, the difference in topic content may result in the fluctuation of learners' cognitive presence. To some extent, learners prefer to discuss topics related to course content and real life, which may trigger their deeper thinking. Thus, in order to help learners build cognitive presence, it is necessary for the instructor to design more inquiry and open domain discussion activities in the MOOC forum that are integrated with the learning materials.

Unlike the distribution of cognitive presence on teaching weeks, the sequential patterns indicate the dynamic and continuous process of learners' cognitive presence. By interpreting the findings, some interesting phenomena were discussed as follows:

- (1) Above all, the instructor-paced discussion tends to generate some cyclic back-tracking sequences such as “*triggering events* → *triggering events*” (S1 → S1), “*exploration* → *exploration*” (S2 → S2), “*integration* → *integration*” (S3 → S3), “*resolution* → *resolution*” (S4 → S4). These sequential patterns indicate that a certain degree of self-stability in learners' cognitive presence, which was consistent with the finding of previous research (Lin et al., 2014; Hou et al., 2015). Interestingly, we also found that there were significant sequential patterns of the instructor's “*triggering events*” to learners' “*triggering events*” (S1) and “*exploration*” (S2) under the instructor-paced group (as shown in Fig. 7). The

Fig. 7 Transition diagram of cognitive presence between the instructor and learners

results showed that learners are likely to be inclined to follow the instructor's *triggering events* (e.g., questioning behavior, knowledge guidance). Thus, the instructors' guidance style may be an important factor in causing a transition in learners' cognitive sequential patterns.

- (2) Second, the learner-paced group mostly focused on “*triggering events* → *exploration*” ($S1 \rightarrow S2$), “*exploration* → *triggering events*” ($S2 \rightarrow S1$). These patterns indicated that learners often turn to peers for help with problems they encounter in the learning process, then they would also seek to negotiate and construct a shared meaning to solve the given problem. After constructing a new meaning of the original problem, the new problems may be generated by learners. The result seemed to align well with the previous study (Wang et al., 2017), which indicates that learners achieve meaning construction of knowledge content through peer interaction. Although the learner-paced group lacked a high level of cognitive sequential transitions, it still included two significant sequential patterns between “*triggering events*” and “*exploration*” in comparison with the instructor-paced group. The reason may be two-fold: the one was that the discussion with peers may generate more *triggering events* and *exploration* in the learner-paced group, which proves that the cognitive presence in online learner-learner discussions tends to stay at its lower levels (Galikyan & Admiraal; Shea et al., 2010). While the other was that the instructor guidance is more conducive to the development of learner *integration* and *resolution* abilities. Similarly, this situation has been revealed in a previous study, which implied that the leadership role of the instructor tended to be powerful in facilitating high levels of cognition and thinking (Aviv et al., 2019; Garrison & Cleveland-Innes, 2005; Liu et al., 2019).
- (3) Third, the sequence of “*exploration* → *integration*” ($S2 \rightarrow S3$) was statistically significant in the high-achieving group. The result indicates that learners in the high-achieving group tend to share their insights and ideas with peers in response to the problem, then construct meanings about ideas generated in the *exploration* phase. On the other hand, the “*exploration* → *integration*” sequence also reveals a growth characteristic of the cognitive presence of learners in the high-achieving group. However, the sequential patterns were less involved in the cognitive presence of *resolution* in both the low- and high-achieving groups. The reason for the lack of *resolution* may be that the asynchronous text-based communication environment (e.g., MOOC forum) had some deficiencies in guiding discussions toward higher-order cognitive activities such as *resolution* (Garrison et al., 2001).

Regarding the relationship between cognitive presence and learning achievements, the result of logistic regression indicates that three cognitive presence categories such as exploration, integration and resolution showed significantly positive predictive effect on learning achievements, whereas the triggering events was found to be negatively associated with learning achievements. Correspondingly, Galikyan and Admiraal (2019) have also suggested that a certain level of cognitive presence,

specially, *integration* and *resolution* were considered as two significant predictors of final grades. In addition, unlike some previous studies that only emphasized the moderating effect of instructor-paced discussion (Lee et al., 2018; Ma et al., 2015), we found that the learner-paced discussion had a larger moderating effect on the relationship between cognitive presence and learning achievements. That is, in the learner-paced discussion group, learners' cognitive presence had a more positive impact on their learning achievements. The primary reason may be that learners in the learner-paced discussions tend to conduct self-regulated learning. MOOCs require learners to be able to self-regulate their learning paces including collaborative discussions, determining how and with what content they engage (Milligan & Littlejohn, 2014). Self-regulated learning not only contributes to learners' cognitive presence in MOOCs but also helps to improve learners' learning achievements (Azevedo & Cromley, 2004; Barnark-Brak et al., 2010; Hood et al., 2015). From this perspective, learners in the learner-paced discussion group may be more adept at using self-regulated learning strategies.

6 Conclusions, limitations, and future work

In this study, we collected and coded the discussion data of the *Astronomy Talk course* from the Chinese MOOCs platform. The cognitive presence, learning achievements, and their relationship with discussion paces were investigated throughout descriptive statistical approaches such as LSA, logistic regression, and grouped regression. The main findings were summarized as follows: First, learners might reach high levels of cognitive presence when involving the instructor-paced discussion. Meanwhile, the instructor's cognitive participation may result in the transition of learners' cognitive presence. Second, learners in the instructor-paced group tended to generate some cyclic backtracking sequences. However, the learner-paced discussion mostly focused on "*triggering events* → *exploration* → *triggering events*" ($S1 \rightarrow S2 \rightarrow S1$). Third, by comparing the cognitive sequential patterns between low- and high-achieving groups, we found that the sequence of "*exploration* → *integration*" ($S2 \rightarrow S3$) was statistically significant in the high-achieving group. Fourth, the low-achieving learners might stay in the *triggering events* phase, while the high-achieving learners might be inclined to engage in the discussion tasks related to *exploration*, *integration*, and *resolution*. Moreover, we also found that learner-paced discussion had a positive moderating effect on the relationship between cognitive presence and learning achievements.

Based on these findings, some targeted suggestions could be provided. First, the instructor should engage in learners' discussions in a MOOC forum according to their cognitive performance, and attach more importance to the learner-to-learner interactions in the learner-paced group. Second, the instructor needs to design more attractive topics with different guidance types (e.g., *Are Orion and Scorpius in the northern hemisphere?*) based on course content. Third, the MOOC forum should be organized with more interactive elements to facilitate online communications between students and instructors. The MOOC developer also can design personal self-learning modules (e.g., learning performance dashboard, social interactive

dashboard) supporting learners to conduct self-regulated learning and collaborative discussion.

The current study also has some limitations and further work is needed. First, only single-course discussion data from a MOOC was obtained for this research, which may limit the generalizability of the findings to other MOOCs and different styles of learners. Second, due to the unbalance sample size of discussion data between instructor- and learner-paced discussions, the analysis may lead to biased results by the Chi-square test. Third, manual coding of text data is time-consuming, and coding reliability can sometimes be affected by the coder's mood. Therefore, future work should attempt to obtain diversified discussion data from multiple courses or platforms to make the findings more generalizable. In addition, only discussion posts were used in this study, and we will also consider combining online learning behavioral log to analyze the learners' cognitive presence. Meanwhile, more effective data processing methods should be adopted to identify the cognitive presence in the MOOC environment, and robust deep learning algorithms should be developed to automate the classification for learners' cognitive presence in future research.

Funding This work was supported by the National Natural Science Foundation of China (Grant Nos. 62077017, 61977030, 61937001) and the Fundamental Research Funds of the Central Universities (Grant Nos. CCNU20TS032, 30106200548).

Declarations

Conflicts of interest/competing interests No conflict of interest.

References

- Alraimi, K. M., Zo, H., & Ciganek, A. P. (2015). Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education*, 80, 28–38 <https://doi.org/10.1016/j.compedu.2014.08.006>
- Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, 42(2), 233–250 <https://doi.org/10.1111/j.1467-8535.2009.01029.x>
- Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2019). Network analysis of knowledge construction in asynchronous learning networks. *Online Learning*, 7(3) <https://doi.org/10.24059/olj.v7i3.1842>
- Azevedo, R., & Cromley, J. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523–535 <https://doi.org/10.1037/0022-0663.96.3.523>
- Bakeman, R., & Gottman, J. M. (1997). Observing interaction: An introduction to sequential analysis. Cambridge university press.
- Barnark-Brak, L., Lan, W., & Paton, V. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distance Learning*, 11(1), 62–80 <https://doi.org/10.19173/irrodl.v11i1.769>
- Clow, D. (2013). MOOCs and the funnel of participation. In proceedings of the third international conference on learning analytics and knowledge (pp. 185–189). <https://doi.org/10.1145/2460296.2460332>.
- Coffield, F. J., Moseley, D. V., Hall, E., & Ecclestone, K. (2004). Learning styles and pedagogy in Post-16 learning: A systematic and critical review. Learning & Skills Research Centre.

- 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 <https://doi.org/10.1111/bjjet.12670>
- Dewey, J. (1910). How we think. *Rev.chil.Ortop.Traumatol*, 42(11), 11–12.
- Ferrari-Lagos, E., Martínez-Abad, F., & Ruiz, C. (2020, October). The importance of motivation and communication in MOOCs as elements to increase completion rates: A study at MOOCs on climate change. In eighth international conference on technological ecosystems for enhancing Multicultural-ity (pp. 1042–1047). <https://doi.org/10.1145/3434780.3436633>.
- Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Escaño, J. E. (2014). Improving the MOOC learning outcomes throughout informal learning activities. In proceedings of the second international conference on technological ecosystems for enhancing Multiculturality (pp. 611–617). <https://doi.org/10.1145/2669711.2669963>.
- Flanders, N. A. (1970). *Analyzing teaching behavior*. Addison-Wesley Pub. Co.
- Galikyan, I., & Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. *The Internet and Higher Education*, 43, 100692 <https://doi.org/10.1016/j.iheduc.2019.100692>
- Garrison, D. R. (2007). Online community of inquiry review: Social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks*, 11(1), 61–72 <https://doi.org/10.24059/olj.v11i1.1737>
- Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education*, 19(3), 133–148 https://doi.org/10.1207/s15389286ajde1903_2
- 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 [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- 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 <https://doi.org/10.1080/08923640109527071>
- Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, 13(1–2), 5–9 <https://doi.org/10.1016/j.iheduc.2009.10.003>
- 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 <https://doi.org/10.1016/j.iheduc.2014.09.006>
- Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014). Attrition in MOOC: Lessons learned from drop-out students. In international workshop on learning Technology for Education in cloud (pp. 37–48). Springer, .
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*, 27, 1–12.
- Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education*, 91, 83–91 <https://doi.org/10.1016/j.compedu.2015.10.019>
- Hou, H., Wang, S., Lin, P., & Chang, K. (2015). Exploring the learner's knowledge construction and cognitive patterns of different asynchronous platforms: Comparison of an online discussion forum and Facebook. *Innovations in Education and Teaching International*, 52(6), 610–620 <https://doi.org/10.1080/14703297.2013.847381>
- Hu, Y., Donald, C., Giacaman, N., & Zhu, Z. (2020, March). Towards automated analysis of cognitive presence in MOOC discussions: A manual classification study. In proceedings of the tenth international conference on Learning Analytics & Knowledge (pp. 135–140). <https://doi.org/10.1145/3375462.3375473>.
- Huang, C. Q., Han, Z. M., Li, M. X., Jong, M. S. Y., & Tsai, C. C. (2019). Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers & Education*, 140, 103589 <https://doi.org/10.1016/j.compedu.2019.05.015>
- Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, 87, 204–217 <https://doi.org/10.1016/j.compedu.2015.07.002>
- Kanuka, H., & Garrison, D. R. (2004). Cognitive presence in online learning. *Journal of Computing in Higher Education*, 15(2), 21–39 <https://doi.org/10.1007/BF02940928>

- Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention-a literature review. In EdMedia+ innovate learning (pp. 1305-1313). Association for the Advancement of computing in education (AACE).
- Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, 27, 74–89 <https://doi.org/10.1016/j.iheduc.2015.06.002>
- Lee, J., & Recker, M. (2021). The effects of instructors' use of online discussions strategies on student participation and performance in university online introductory mathematics courses. *Computers & Education*, 162, 104084. <https://doi.org/10.1016/j.compedu.2020.104084>
- Lee, S. J., Lee, H., & Kim, T. T. (2018). A study on the instructor role in dealing with mixed contents: How it affects learner satisfaction and retention in e-learning. *Sustainability*, 10(3), 850 <https://doi.org/10.3390/su10030850>
- Lin, P., Hou, H., Wu, S., & Chang, K. (2014). Exploring college students' cognitive processing patterns during a collaborative problem-solving teaching activity integrating Facebook discussion and simulation tools. *The Internet and Higher Education*, 22, 51–56 <https://doi.org/10.1016/j.iheduc.2014.05.001>
- Liu, Z., Yang, C., Rüdian, S., Liu, S., Zhao, L., & Wang, T. (2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interactive Learning Environments*, 27(5–6), 598–627 <https://doi.org/10.1080/10494820.2019.1610449>
- Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, 24, 26–34 <https://doi.org/10.1016/j.iheduc.2014.09.005>
- Milligan, C., & Littlejohn, A. (2014). Supporting professional learning in a massive open online course. *The International Review of Research in Open and Distance Learning*, 15(5), 197–213.
- Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments*, 27(5–6), 655–669 <https://doi.org/10.1080/10494820.2019.1610453>
- Moore, R. L., Yen, C. J., & Powers, F. E. (2021). Exploring the relationship between clout and cognitive processing in MOOC discussion forums. *British Journal of Educational Technology*, 52(1), 482–497 <https://doi.org/10.1111/bjet.13033>
- Olesova, L., Slavin, M., Lim, J. (2016). Exploring the effect of scripted roles on cognitive presence in asynchronous online discussions. Online learning (Newburyport, Mass.), 20(4), 34. <https://doi.org/10.24059/olj.v20i4.1058>.
- Peng, X., & Xu, Q. (2020). Investigating learners' behaviors and discourse content in MOOC course reviews. *Computers & Education*, 143(Jan.), 103671–103673 <https://doi.org/10.1016/j.compedu.2019.103673>
- Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., & Getoor, L. (2014). Understanding MOOC discussion forums using seeded LDA. Paper presented at the Proceedings of the ninth workshop on innovative use of NLP for building educational applications.
- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *The Journal of Distance Education/Revue de l'éducation Distance*, 14(2), 50–71 <http://hdl.handle.net/2149/732>
- Safitri, D., Setiawan, A., Suhandi, A., Malik, A., & Lisdiani, S. A. S. (2019, April). The effects of higher order thinking (hot) laboratory design in hooke law on student's creative thinking skills. In *journal of physics: Conference series* (Vol. 1204, no. 1, p. 012037). IOP publishing.
- Shea, P., Hayes, S., Vickers, J., Gozza-Cohen, M., Uzuner, S., Mehta, R., & Rangan, P. (2010). A re-examination of the community of inquiry framework: Social network and content analysis. *The Internet and Higher Education*, 13(1–2), 10–21 <https://doi.org/10.1016/j.iheduc.2009.11.002>
- Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29–48 <https://doi.org/10.1016/j.compedu.2005.04.002>
- Sun, Z., Lin, C. H., Lv, K., & Song, J. (2021). Knowledge-construction behaviors in a mobile learning environment: A lag-sequential analysis of group differences. *Educational Technology Research and Development*, 69(2), 533–551 <https://doi.org/10.1007/s11423-021-09938-x>
- Tomkin J H, Charlevoix D. Do professors matter? Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes [C]//Proceedings of the first ACM conference on Learning@ scale conference. 2014: 71–78. <https://doi.org/10.1145/2556325.2566245>.

- Tsiotakis, P., & Jimoyiannis, A. (2016). Critical factors towards analysing teachers' presence in on-line learning communities. *The Internet and Higher Education*, 28, 45–58 <https://doi.org/10.1016/j.iheduc.2015.09.002>
- Wang, S., Hou, H., & Wu, S. (2017). Analyzing the knowledge construction and cognitive patterns of blog-based instructional activities using four frequent interactive strategies (problem solving, peer assessment, role playing and peer tutoring): A preliminary study. *Educational Technology Research and Development*, 65(2), 301–323 <https://doi.org/10.1007/s11423-016-9471-4>
- 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 (pp.226-233).
- Wong, J. S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015, March). An analysis of MOOC discussion forum interactions from the most active users. In international conference on social computing, behavioral-cultural modeling, and prediction (pp. 452-457). Springer, Cham.
- Wu, S., & Hou, H. (2015). How cognitive styles affect the learning behaviors of online problem-solving based discussion activity: A lag sequential analysis. *Journal of Educational Computing Research*, 52(2), 277–298 <https://doi.org/10.1177/0735633115571307>
- Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, 2(3), 23–33.
- Xu, B., Chen, N. S., & Chen, G. (2020). Effects of teacher role on student engagement in WeChat-based online discussion learning. *Computers & Education*, 157, 103956 <https://doi.org/10.1016/j.compedu.2020.103956>
- 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 (Vol. 11, p. 14).
- Yang, T., Chen, S. Y., & Hwang, G. (2015). The influences of a two-tier test strategy on student learning: A lag sequential analysis approach. *Computers & Education*, 82, 366–377 <https://doi.org/10.1016/j.compedu.2014.11.021>
- Zhao, H., & Sullivan, K. P. (2017). Teaching presence in computer conferencing learning environments: Effects on interaction, cognition and learning uptake. *British Journal of Educational Technology*, 48(2), 538–551 <https://doi.org/10.1111/bjet.12383>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Zhi Liu^{1,2} · Xi Kong¹ · Sannyuya Liu^{1,2} · Zongkai Yang^{1,2} · Cuishuang Zhang²

Zhi Liu
zhiliu@mail.ccnu.edu.cn

Xi Kong
kongxicc@mails.ccnu.edu.cn

Zongkai Yang
13659885363@163.com

Cuishuang Zhang
zcs@mails.ccnu.edu.cn

¹ National Engineering Laboratory for Educational Big Data (NELEBD), Central China Normal University, Wuhan, Hubei, China

² National Engineering Research Center for E-Learning (NERCEL), Central China Normal University, 152 Luoyu Road, Wuhan, Hubei 430079, People's Republic of China