

Analyzing interactivity in asynchronous video discussions

Hannes Rothe¹, Janina Sundermeier¹, and Martin Gersch¹

¹Department Business Information Systems, Freie Universität Berlin, Germany
(hannes.rothe@fu-berlin.de, janina.sundermeier@fu-berlin.de,
martin.gersch@fu-berlin.de)

Abstract. Evaluating online discussions is a complex task for educators. Information systems may support instructors and course designers to assess the quality of an asynchronous online discussion tool. Interactivity on a human-to-human, human-to-computer or human-to-content level are focal elements of such quality assessment. Nevertheless existing indicators used to measure interactivity oftentimes rely on manual data collection. One major contribution of this paper is an updated overview about indicators which are ready for automatic data collection and processing. Following a design science research approach we introduce measures for a consumer side of interactivity and contrast them with a producer's perspective. For this purpose we contrast two ratio measures 'viewed posts prior to a statement' and 'viewed posts after a statement' created by a student. In order to evaluate these indicators, we apply them to Pinio, an innovative asynchronous video discussion tool, used in a virtual seminar.

Keywords: Online discussion, asynchronous video discussion, educational data mining, interactivity, higher education

1. Introduction

The development and spread of information and communication technologies as well as the increasing number of participants per course in higher education are only two reasons why the use of and the demand for technology-enhanced education tools are rising steadily [1]. Besides, the demand on peer-to-peer learning concepts and the integration of virtual communities into online learning environments is growing [2–4]. Especially the implementation and encouragement of asynchronous online discussions within technology-enhanced education are of wide interest in higher education [5]. Asynchronous online discussions favor the realization of Massive Open Online Courses (MOOCs) and many other types of online courses that allow students to study online at their own pace at different times (e.g. in different time zones) [6, 7].

Although the utilization of online discussions is of high interest for research on and practice of technology enhanced education, a research gap has been identified. Existing research on the evaluation of these applications only partially reflects the actual interactivity of students within asynchronous online discussions. Especially with the rise of learning analytics in general and educational data mining (EDM) in particular

quantitative indicators for the interactivity of students for educational purposes are of increasing interest. Most approaches in research on EDM related to asynchronous online discussions make use of ‘active’ indicators such as the ‘number of posts per students’ or ‘number of reviews per student’ in order to determine the interactivity and to evaluate the expected learning outcome [7, 8]. The role of ‘passive’ indicators, such as reading the posts and learning from it without actively taking part in a debate [9, 6], is only partially considered yet. Prior experiments and qualitative research showed that acquiring and assimilating information from peers by reading or watching contributions from other students has positive outcomes on the learning outcome [10, 6].

Hence, the intensity with which a student interacts with information from their peers should be included into proper interactivity evaluation. Following the approach of [11] we discern between two roles a student impersonates in online discussion - consumers and producers. Based on this differentiation, the central question of this paper is: *How can we use data mining to evaluate online discussions against the background of a multifaceted view on interactivity?* Thus, the purposes of this paper are to contribute to the ongoing debate on indicators for online discussions in two ways. First of all, we update the suggestions from Dringus and Ellis [12] about indicators used to evaluate online discussions which can be applied to EDM. Moreover, we introduce measures for the consumer’s side and contrast them with the producer’s perspective.

Following a design science research approach [13], we use an asynchronous video discussion tool to introduce extended indicators for the analysis of a consumer- and a producer dimension alike. Our research design starts by clarifying the problem field. We introduce current sets of indicators for online discussions within research on learning analytics and EDM. Afterwards we demonstrate the use of our indicators for assessing Pinio, an asynchronous (video) discussion application. The pertaining application was applied to the course Net Economy. Net Economy is an online course with 140 students from six institutes in higher education who engaged in several online discussions in 2013. The demonstration is supplemented by an explorative correlation study to assess the additional information gain from both indicators.

2. Objectives of a solution

2.1. Online Discussion

Discussions in higher education, whether online or offline, should encourage debates among students as well as between students and course instructors [14]. Typically we distinguish between synchronous and asynchronous discussions that mainly differ in terms of presence of the participants. The focus of this paper lies on asynchronous online discussions. Within these debates participants do not necessarily need to be present in a discussion forum at the same time. New entries within the online discussion are recorded and can be accessed at any time without a direct need to respond immediately [5, 15]. Thus, participants can decide when and where they take part in the online discussion. Hence they can flexibly integrate the online discussion into their schedules [16]. Moreover, online discussions are an important element to trigger

extensive and helpful dialogues between students [2, 17]. Several authors find that sufficient time to develop own thoughts and to prepare an appropriate statement are further benefits of online discussions [14, 18, 8, 15]. It has been shown that active participation in online discussions oftentimes lead to effective learning outcomes [7]. Especially students who feel less comfortable in face-to-face discussions profit from an anonymous discussion environment [14, 8]. Furthermore, students are able to prepare their answers and responses according to their own pace; they do not necessarily follow the pace of a learning group [19]. Apart from the before mentioned advantages, several authors identified also a couple of disadvantages related to online discussions. One of which is the unpredictable amount of time participants have to wait for replies to their former posts [20]. The authors Dringus and Ellis [12] found that participants might even be discouraged when facing an over proportional long waiting time. Moreover, the structure of the online discussion forum is an important factor influencing whether some post remains unread and unanswered because participants lose track of what has been written.

The role of the educator, namely instructor or course designer, is to moderate online discussions in a manner that irrelevant or even wrong contributions are clarified. Additionally instructors keep track of a discussion to ensure relevance of postings for the course topics, reduce off-topic and support a positive debate culture. In order to do so, the course instructors need to evaluate the quality of a discussion. Following the brief introduction on factors influencing online discussion, this is a complex task. In the following we will discuss, whether this process of quality assessment can be supported by information technology and measures from the field of educational data mining.

2.2. Educational Data Mining in Online Discussions

Considering the fact, that educators in e- or blended learning settings mostly interact through (web-based) technology with students, potential transparency of student's actions mostly decreases. Lecturers who give their presentation via video streams can only guess if students are watching them at home or switch to another browser tab. Also online discussions may quickly become either very complex or stay hardly initiated; as a result educators can either hardly review the amount of data or stay unclear why a discussion does not thrive [8]. Educators giving traditional lectures may be able to recognize a distracted audience and intervene promptly, for instance by cold-calling. Nevertheless web-based technologies impede this strategy at a first glance. A few years ago Learning Analytics gained common recognition as a field of research that may resolve these issues. The general attention on Learning Analytics within research on educational technology grew tremendously since the publication of the Horizon Report 2020 by the European Union [21]. Following the literature on Learning Analytics its main task lies within the collection, processing and evaluation of usage data in the educational sector [22–24]. A recently developing branch within the field Learning Analytics is called Educational Data Mining. EDM uses techniques from Business Intelligence and Data Mining to transform data into information using algorithms to support decision making processes [25]. For example text mining, clus-

tering or rule setting methods can be used to help educators interpret large amounts of data which are mostly collected by web services, like web-based trainings, course management systems and also online discussion tools. For this purpose key indicators need to be created to find answers for hypotheses regarding educational technologies in general and online discussion tools in particular. Following this initial step these indicators may be directly collected or processed using raw data. Sources of data are diverse. To test hypotheses regarding educational technologies we need to advance the data mining process. Subsequently manual data collection is an obstacle. We need to use data that can be collected machine-aided to enable automatic data analysis. Within educational technology server-sided data (like log files) as well as client-sided data (like client-triggered events via javascript) are common [25]. Therefore our main task to evaluate the quality of an online discussion tool is finding relevant and quantifiable indicators that could be collected automatically.

To assess the quality of online discussion within a course, we ask for the interactivity of students with other students, the educator and the learning content [26–28]. Dringus & Ellis [12] were not able to identify ‘the’ key indicators for participation in online discussions, even though they presented a list of indicators based on a thorough literature review. They subsume that relevant constructs and ‘the associated indicators are found in the literature, but only in piecemeal’. Their study nevertheless contains many indicators that are collected manually. Additionally they propose indicators that could be of use for data mining. Prominent indicators which measure interactivity between humans (student-to-student as well as student-to-educator) within online discussion boards are the amount of replies or posts per user. Other indicators, like the amount of accesses to a discussion board, are used to assess the level of human-computer or human-to-content interaction [7]. These indicators nevertheless only partially reflect the interaction of students and the online discussion. ‘Lurking’ or content ‘consumption’ are not directly assessed.

Interactivity in online discussions is embodied by active participation - or knowledge ‘production’ - as well as passive ‘consumption’. Active participants post answers to existing contributions or initiate new discussion streams. Passive consumption of online discussions involves reading of central contributions and finding relevant information without contributing or answering to an existing discussion [7]. In literature, the term “lurker” is often adopted in this context to describe students who are only observing the ongoing discussion and who remain themselves silent [7, 8]. The passive consumption in contrast reflects to a type of interactivity that may be linked to a prior or following content production.

Research mostly argues about indicators for active types of interactivity in the context of online discussions. Several authors measure the access to online discussions boards as well as the number of posts or reviews per students in order to determine the effectiveness of online discussions in terms of learning outcomes [29–31, 7, 9]. [6] examined in experiments the number of views per user and posting. The results of this approach show that a combination of active posting within discussion forums and passive consumption – like reading of posts – contribute to the learning process. Even though the authors Jyothi et al. [8] acknowledge that lurking has some kind of contribution to the learning process they focus on indicators for active participation, such as

access to the discussion forum. In order to gain further insights related to the passive consumption of asynchronous online discussions, Wise et al. [9] intend to operationalize the passive consumption of online discussions in terms of reading. For this purpose, the authors measure time difference between two activities and assume that the students dedicate a certain amount of the mean time to read prior posts. In a similar manner, [31] observes the number of website hits per student in order to determine the passive consumption of the discussions. These indicators are based on the assumption, that students visit the web page to read thoroughly through postings, which may lead to a biased interpretation. Students may dedicate their time spent on a site with activities unrelated to the actual discussion, e.g. switching the browser tab without closing the discussion. Another approach, adopted by several authors, is to count the number of posts that are marked as ‘read’ or ‘important’ [32, 29, 33]. Instead of measuring passive consumption they observe manual or automatic ‘ticking’ of a checkbox. These indicators may be biased by students who do not mark every single post.

In contrast to [12] we present a list of indicators which were used for machinable data collection and data processing (see table 1). The displayed results follow a literature review of peer-reviewed articles regarding data-mining and online discussion for educational purposes.¹ Our overview shows that most conducted research makes use of indicators for ‘active’ participation such as the access to the discussion forum and the number of posts per user in order to enhance existing research. Indicators ‘passive’ interactivity are less often determined. In the following we develop and demonstrate indicators to determine passive indicators of asynchronous online discussions ready for data mining techniques.

Table 1. A list of indicators that can be collected and processed automatically

	Indicator	Operationalization	Reference, cf.
	amount of posts created	-	[7, 9, 12, 29, 30, 34, 35]
	Time difference between two posts	-	[12, 36]
	amount of words / sentences	amount of text elements (regular expressions)	[33]
active	amount of reviews / edits	-	[9, 29]
	change in subject	change in title/subtitle of post	[8]
	response	reply to a post with and without mentioning a prior author within this post	[8, 12]
	detect key words, phrases related to topic	text mining for keywords and phrases	[12]

¹ Databases: ScienceDirect, EBSCO, ERIC, IEEE Xplore

	detect genre of posting using text mining	cascade classification model for detecting a "discussion, comment, reflection, information sharing and scaffolding"	[37]
	centrality measures	network analytical approach on a micro level: Degree, Betweenness, Closeness, Auth/Hub Centrality	[8, 32, 33]
	density of the discussion	network analytical approach on a macro level	[8]
	amount of reads	action 'read' estimated by time spent on a page	[9, 35]
passive	amount of website hits	amount of page hits	[31, 38]
	amount of 'scans'	action 'scan' estimated by time below 6.5 seconds per word	[9]
	marked as "read"	amount of posts that are marked as "read" or "important" by students	[29, 32, 33]

3. Assessing the ,consumer‘ perspective

3.1. Institutional Context

We build upon our findings from the literature review to evaluate an asynchronous video discussion tool which was used in the course ‘Net Economy’. ‘Net Economy’ is a virtual seminar that is offered in collaboration across several locations. The setting targets participants with heterogeneous educational backgrounds in the fields of business and economics, business information systems, and educational sciences. Besides teaching about Entrepreneurship in the Net Economy the virtual collaborative learning setting focusses on spanning across different cultural backgrounds from Germany (Bochum, Berlin, Clausthal, Soest), Ukraine (Simferipol), and Indonesia (Jakarta). 140 students participated in the course in 2013.

Every Net Economy class is divided into three phases: ‘preparation phase, knowledge development phase’ and ‘case study phase’. Throughout the course, project work is conducted in small teams of six students and across locations, both in terms of team composition as well as presentation and discussion of findings. By separating these phases, learning and working processes are structured as a project with the use of predefined milestones. Preparation phase consists of a team building process, where students get in touch with each other. For this purpose we integrated a virtual social network into the seminar [3]. The following knowledge development phase consists of pre-recorded video lectures and team assignments regarding the course topics. During the final case study phase every team applies what they learned before. Every finding is documented and discussed in steering committees and team websites following phase-specific assignments. Steering committees and final presen-

tations are held at each location and are merged together through synchronous video conferencing.

Besides the virtual social network the course is complemented by an innovative, asynchronous video discussion tool called Pinio. Pinio² is a web tool that contains video statements of lecturers and students. Each statement takes up to thirty seconds and usually contains a key argument of a discussant. In contrast to traditional text-based asynchronous discussion tools the video tool reduces anonymity. As a result every argument is complemented by facial expression and gestures. Every participant of the class was allowed to start and reply to discussions. During Knowledge Development Phase we had two initial discussion threads in the first week. Participation in at least one of these discussions was mandatory. In the following weeks starting a discussion or replying to a post was completely voluntary. During Case Study Phase each of the 23 teams had to start a discussion containing their first idea for a new business model. Additionally each team had to reply to at least two discussions from other teams to give them feedback for their ideas.

3.2. Indicators for an asynchronous video discussion

To assess the interactivity of students in online discussion we need to incorporate passive consumption and active participation. While active participation could be easily measured by the amount of video comments of a student, the influence of passive consumption can be manifold. First of all we need to gather data on consumption. While merely capturing the duration spent on a web site has been used to assess 'reading', this measure may be difficult. The duration is indirectly measured by the time difference between two hits or events on a website [9]. Depending on the online discussion software it is difficult to distinguish between reading one or multiple posts.

In Pinio every video post has a maximum duration of 30 seconds. A video is measured as being played when a student clicks on the pertaining button and keeps playing it until the end unless the browser tab is changed or closed. Once a video is finished, an event is triggered that sends data for a complete video presentation to the database. Therefore viewing a discussion post is more reliably logged, then merely guessing the amount of time spent on a web site. Based on the received data we aggregated indicators for the consumption of a video statement.

We distinguish between two sets of indicators that refer to the process perspective of learning. On the one hand we processed the amount of viewed comments (v) in a discussion (D) prior to a statement (p_x) at time (t). We call it *viewPrior*. While watching comments beforehand we might get insights into the intensity with which a student prepares his or her comment. Additionally watching prior comments is necessary for linking an argument to the overall discussion. On the other hand we measured *viewAfter* as the amount of comments that have been watched after a student contributed to a discussion. We assume that following a discussion after posting a statement signals engagement and an involvement in the learning content that reaches beyond an explicitly or implicitly created need within a course. The absolute number of views

² see also www.pinio.me

depends on the moment a statement is made by a student. If he or she contributes lately to a debate, there are many videos posted beforehand and less videos will be stated afterwards. Hence, to compare results of viewAfter or viewPrior we need to standardize both indicators (see eq. 1 and 2 for both ratio variables).

$$\text{Ratio of viewPrior: } rvp_x = \frac{\sum_{t_0}^{t_x-1} v_{D,t}}{\sum_{t_0}^{t_x-1} p_{D,t}} \quad (1)$$

$$\text{Ratio of viewAfter: } rva_x = \frac{\sum_{t_x+1}^T v_{D,t}}{\sum_{t_x+1}^T p_{D,t}} \quad (2)$$

For this purpose we designed the ratio of consumed or watched videos posted before or after a statement in relation to the amount of existing and following statements at that time ($p_{D,t}$). Therefore we are able to assess the percentage of viewed videos prior and after contributing to a discussion.

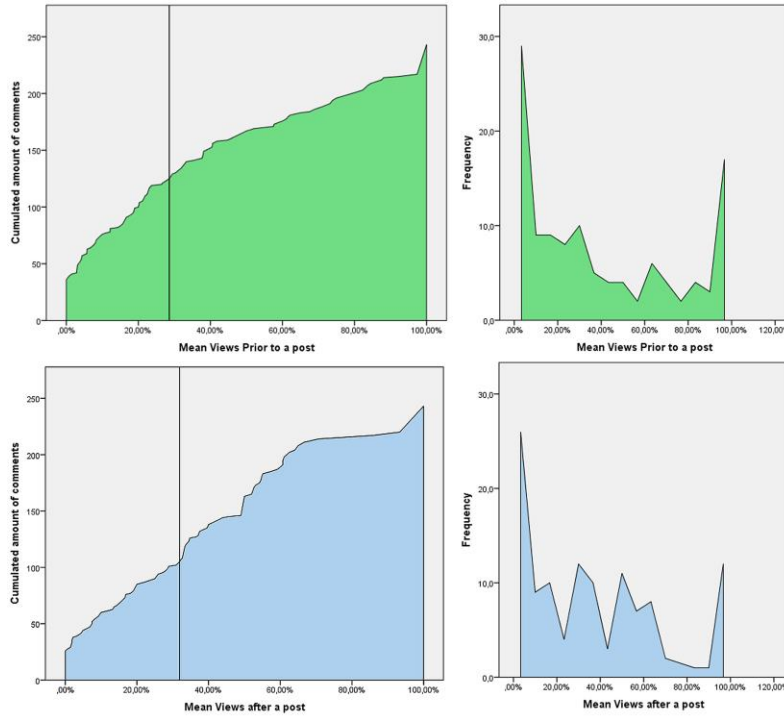


Fig. 1. Histograms and distributions, vertical lines represent median views

During the course 270 video statements were created. In sum, 6,483 videos have been watched. Hence students watched by far more videos than they recorded, which

is in line to the findings of [38] for text-based online discussions. Figure 1 presents the distribution of mean ratios for viewPrior and viewAfter of every student within the class Net Economy. Additionally we present the distribution of the pertaining variables in contrast to the cumulated amount of comments. As we can see for both variables, the distributions of viewPrior and viewAfter are multimodal. Apparently both indicators are skewed towards each end of the scale (0% and 100%). As we relate these distributions to the amount of comments, we gain insights into the relationship between the active and passive type of interactivity. A large amount of comments is made without watching prior posts at all. There are also many students who only state their argument and do not track the following dispute. Nevertheless, we see a higher median as well as comparably greater marginal values for the upper 50-percentile for viewAfter in comparison to viewPrior. Subsequently there are some students who show involvement above the level of what they are asked for within the given task.

To assess the added value of both indicators in terms of information gain, we measure correlation between indicators for passive and active interactivity. We assume an information gain, if indicators quantify different dimensions of interactivity. Hence, we expect low correlation between each indicator. Since viewPrior and viewAfter show multimodal distribution we choose Kendall's- τ correlation. Kendall's- τ is a rank correlation coefficient that does not require a certain distribution of the variables. Results of the correlation analysis are shown in table 2.

Table 2. Kendall's tau correlation of indicators for active/passive types of interactivity

			Number of comments	Mean Views Prior to a post	Mean Views after a post
Kendall's tau correlation	Amount of Comments	Correlation Coefficient	1.000		
		Sig. (2-tailed)			
		N	140		
	Mean Views Prior to a post	Correlation Coefficient	-0.057	1.000	
		Sig. (2-tailed)	0.430		
		N	116	116	
	Mean Views after a post	Correlation Coefficient	0.144*	-0.041	1.000
		Sig. (2-tailed)	0.045	0.525	
		N	116	116	116

*. Correlation is significant at the 0.05 level (2-tailed).

As we can see, the only significantly positive relationship exists between viewAfter and the amount of comments. We assume that higher involvement and motivation in the pertaining debate or the learning content are positively associated to the amount of viewed comments after adding a statement. Hence, both indicators may be related to the same construct. A following qualitative study or an experiment may research the causation behind this occasion. Nevertheless, we see that viewPrior and viewAfter

as well as viewPrior and amount of comments are not correlated. Subsequently we gain more information about the passive type of interactivity within an online discussion.

4. Conclusion

Our literature review showed that some studies built upon the work of Dringus and Ellis [12] meanwhile. The strengthened focus on educational data mining within technology enhanced learning had a positive effect on research about quality assessment of online discussions in higher education. Nonetheless, we see that assessing how effective an application for online discussion actually is, remains a complex task. Many indicators are still only partially validated and there is still a lot of room for enhancing these indicators. In contrast to many indicators for the active ‘producer perspective’, we found that the passive ‘consumer perspective’ is weakly considered in EDM. Nevertheless this perspective promises valuable insights into the interactivity between students and the discussion technology. For this purpose we demonstrated the use of two variables, ‘mean views prior to a student’s statement’ and ‘mean views after a student’s statement’. As a matter of machine-aided routine we may now differentiate between a student who posted two videos and watched none of the former comments, and a student who posted only one video, but followed the whole discussion.

Our research calls for a multitude of follow up questions. First of all, behavioristic research could validate the relationship between the active and passive dimensions of interactivity within online discussion. Furthermore the presented indicators were weakly correlated, but contain valuable information about the student’s interactivity with the online discussion system. Hence, applying these indicators in a cluster analysis for discussant types may lead to more insights. In his research [39] argued for a typical distribution in the dichotomy between very active and rather passive students, who use web 2.0 tools in educational settings. Further research may ask whether we’ll also find a normal distribution between those types of students in online discussion. Furthermore we might examine the influence of deviations from a typical distribution for learning success and learner’s satisfaction.

The results of this paper are limited to some extent. The study focused solely on EDM. Qualitative research on online discussion was widely neglected. Combining quantitative and qualitative approaches would be beneficial to gain further insights [31, 9]. Additionally qualitative methods or experiments are fundamental to test assumptions on causality. In contrast to former research we applied our indicators to a new type of online discussion: asynchronous videos. Further research should investigate the difference between the progress of online discussions in traditional discussion boards compared to asynchronous video discussions. We suggest applying an experiment to test the causality of the associated hypotheses.

References

1. Bures EM, Abrami PC, Amundsen C (2000) Student motivation to learn via computer conferencing. *Research in higher Education* 41(5): 593–621
2. Swan K, Shea P (2005) The development of virtual learning communities. *Learning together online: Research on asynchronous learning networks*: 239–260
3. Weber P, Rothe H (2013) Social networking services in e-learning. In: Bastiaens T, Marks G (eds) *Education and Information Technology 2013: A Selection of AACE Award Papers*, AACE, vol 1. Chesapeake, pp. 89–99
4. Kear K (2004) Peer learning using asynchronous discussion systems in distance education. *Open Learning: The Journal of Open, Distance and e-Learning* 19(2): 151–164
5. Hammond M (2005) A review of recent papers on online discussion in teaching and learning in higher education. *Journal of Asynchronous Learning Networks* 9(3): 9–23
6. Cheng CK, Paré DE, Collimore L et al. (2011) Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance. *Computers & Education* 56(1): 253–261
7. Webb E, Jones A, Barker P et al. (2004) Using e-learning dialogues in higher education. *Innovations in Education and Teaching International* 41(1): 93–103
8. Jyothi S, McAvinia C, Keating J (2012) A visualisation tool to aid exploration of students' interactions in asynchronous online communication. *Computers & Education* 58(1): 30–42
9. Wise AF, Perera N, Hsiao Y et al. (2012) Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *The Internet and Higher Education* 15(2): 108–117
10. Beaudoin MF (2002) Learning or lurking?: Tracking the “invisible” online student. *The Internet and Higher Education* 5(2): 147–155
11. Tobarra L, Robles-Gómez A, Ros S et al. (2014) Analyzing the students' behavior and relevant topics in virtual learning communities. *Computers in Human Behavior* 31: 659–669
12. Dringus LP, Ellis T (2005) Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education* 45(1): 141–160
13. Peffers K, Tuunanen T, Rothenberger MA et al. (2007) A design science research methodology for information systems research. *Journal of management information systems* 24(3): 45–77
14. Harasim L (2000) Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education* 3(1): 41–61
15. Kaye A (1989) Computer-mediated communication and distance education. In: Mason R, Kaye A (eds) *Mindweave: Communication, computers, and distance education*. Pergamon, New York
16. Gibbs W, Simpson LD, Bernas RS (2008) An analysis of temporal norms in online discussions. *International Journal of Instructional Media* 35(1): 63
17. Woo Y, Reeves TC (2007) Meaningful interaction in web-based learning: A social constructivist interpretation. *The Internet and Higher Education* 10(1): 15–25
18. Jonassen DH, Kwon II H (2001) Communication patterns in computer mediated versus face-to-face group problem solving. *Educational technology research and development* 49(1): 35–51
19. Prester GE, Moller LA (2001) Exploiting opportunities for knowledge-building in asynchronous distance learning environments. *Quarterly Review of Distance Education* 2(2): 93–104
20. Peters VL, Hewitt J (2010) An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education* 54(4): 951–961

21. Johnson L, Adams Becker S, Cummins M, Estrada V, Freeman A, Ludgate H (2013) NMC Horizon Report: 2013 Higher Education Edition. <http://www.nmc.org/publications/2013-horizon-report-higher-ed>. Accessed 15 Feb 2013
22. Campbell JP, DeBlois PB, Oblinger DG (2007) ACADEMIC ANALYTICS. *Educause Review* 42(4): 40–57
23. Elias T (2011) Learning Analytics: Definitions, Processes and Potential. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
24. Siemens G, Long P (2011) Penetrating the fog: Analytics in learning and education. *Educause Review* 46(5): 30–32
25. Romero C, Ventura S (2007) Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33(1): 135–146
26. Romero-Zaldivar V, Pardo A, Burgos D et al. (2012) Monitoring student progress using virtual appliances: A case study. *Computers & Education* 58(4): 1058–1067
27. Bernard RM, Abrami PC, Borokhovski E et al. (2009) A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research* 79(3): 1243–1289
28. Moore MG (1989) Editorial: Three types of interaction. *American Journal of Distance Education* 3(2): 86–89
29. Hamuy E, Galaz M (2010) Information versus communication in course management system participation. *Computers & Education* 54(1): 169–177
30. Mazzolini M, Maddison S (2007) When to jump in: The role of the instructor in online discussion forums. *Computers & Education* 49(2): 193–213
31. Thomas MJW (2002) Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning* 18(3): 351–366
32. Bayer J, Bydzovská H, Géryk J, Obšivac T, & Popelnský L (2012) Predicting drop-out from social behaviour of students. *Proceedings of the 5th International Conference on Educational Data Mining*
33. Romero C, López M, Luna J et al. (2013) Predicting students' final performance from participation in on-line discussion forums. *Computers & Education* 68(0): 458–472. doi: 10.1016/j.compedu.2013.06.009
34. Thomas MJW (2002) Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning* 18(3): 351–366
35. Hung J, Zhang K (2008) Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*
36. Kumar V, Chadha A (2011) An Empirical Study of the Applications of Data Mining Techniques in Higher Education. *International Journal of Advanced Computer Science and Applications* 2(3): 80–84
37. Lin F, Hsieh L, Chuang F (2009) Discovering genres of online discussion threads via text mining. *Computers & Education* 52(2): 481–495
38. Ebner M, Holzinger A, Catarci T (2005) Lurking: An underestimated human-computer phenomenon. *Multimedia, IEEE* 12(4): 70–75
39. Lehr C (2011) Web 2.0 in der universitären Lehre. http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000035056, received at 2nd February 2014