



Sharing is Learning: Using Topic Modelling to Understand Online Comments Shared by Learners

Kok Khiang Lim^(✉) and Chei Sian Lee

Wee Kim Wee School of Communication and Information, Nanyang Technological University,
Singapore, Singapore

w200004@e.ntu.edu.sg, leeecs@ntu.edu.sg

Abstract. Learners have utilised coding video tutorials to learn new programming languages or enhance their existing skillset. Past studies have focused on content creators' perspective (e.g. motivation to produce video tutorials) or understanding the learner perspective focusing on the outcome of the learning. However, the research on the learning process focusing on learners' sharing behaviour when and after watching the video tutorial was limited. This study aims to address this gap by analysing learners' online comments shared on the video hosting platform to infer learning behaviour from the self-regulated learning perspective. Learners' comments from 24 video tutorials were collected from a popular YouTube coding channel and analysed using the probabilistic topic modelling method. Ten latent topics were uncovered. The findings indicated the presence of three self-regulated learning behaviours. Interestingly, the learners' comments comprised a high proportion of comments related to sharing of coding-related questions, suggesting that learners not only use the commenting platform to provide feedback but also as a means to seek clarification. In addition, this finding also informed the content creators on the areas to engage the learners or refine course content.

Keywords: Online comment · Online learning · YouTube · Self-regulated learning · Video tutorial

1 Introduction

Coding video tutorial is typically produced using the digital video recording from the computer screen as the coding was typed on the screen along with audio narration to guide learners [1]. Coding video tutorial takes advantage of the rich media, visual and audio, afforded by the video to break down the complex procedural task of coding to support learning [2]. Learners would learn by observing the coding demonstration such as scripting, debugging and compiling the codes, and practising the coding as guided by the video. This form of learning has positive contributions to task relevance, self-efficacy and improve students' engagement [3, 4].

Not surprisingly, YouTube has become one of the popular video content sites to host video tutorials. It allows learners the flexibility and autonomy to search and access videos through its many dedicated learning channels [5]. In addition, as with most social network sites, the social features embedded in YouTube enable learners to publicly share their comments about their learning experiences or feedback on the video content [6, 7]. Consequently, these users-generated comments became a ready pool of large volume of data, enabling researchers to conduct studies such as users' behaviour and sentiments. Extant studies have examined the sentiments and toxicity expressed in the YouTube comments [8, 9], classified YouTube comments according to the thematic topics of interest [1, 10], and analysed the perceptions and behaviour of viewers [11].

However, the study on learners' learning behaviour after consuming the YouTube video tutorials is limited and little is known about what drives the learning behaviour. The characteristics of learning from an online environment, such as a video tutorial, require learners to adopt self-regulated learning (SRL) in order to achieve their learning outcome [12, 13]. Thus, this study will address this gap by analysing learners' comments shared on the video hosting platform. Specifically, YouTube comments will be collected and analysed using the probabilistic topic modelling method.

2 Related Work

2.1 Online Video Tutorial and Self-regulated Learning

Online video tutorials have become a prevalent teaching tool to deliver instructional content across various subjects such as languages [14] and health sciences [15]. In particular for coding, its popularity resides in the video tutorial's ability to demonstrate coding concepts [1, 5, 16, 17], convey the outcome of compiling and running a code [5, 17], and its interactivity that allows learners to search and navigate the video content [18, 19]. This form of learning has recently incorporated into online learning alongside mainstream online learning platforms, such as massive online open courses [20].

A main characteristic of online learning is that it requires learners' ability to self-regulate their learning to achieve a positive learning outcome. The influence of SRL in the online learning environment has been studied extensively, and the studies have shown positive results [13, 21, 22]. SRL strategies in an online learning environment can be broadly categorized into *goal setting*, *environment structuring*, *task strategies*, *help-seeking* and *self-evaluation* [23]. Goal setting requires the learner to set a learning goal and develop a plan to achieve that goal [24]. Environment structuring involves the selection or creation of effective and conducive condition for learning [24]. Task strategies refer to understanding tasks and identifying the appropriate approaches or methods to learn [24]. Help-seeking is defined as choosing an appropriate method to seek assistance in guiding oneself to learn [24]. Self-evaluation refers to the comparison between the attained performance against a standard and provides reasons for that success and failure [24].

Araka et al. [25] reviewed 30 past works of literature on the methods used to measure SRL in an online learning environment. The study found that the common methods used were self-reporting questionnaire and survey, and data mining and analytical methods on learners' activities [25]. However, the use of online comments with computational analysis to measure SRL in an online learning environment is few. Thus, this study proposed using online comments as a viable source of data that could be quickly harvested and analysed to understand learners' sharing behaviours from the lens of SRL.

2.2 Analysing YouTube Comments

YouTube comment is a form of publicly shared user-generated content on an online platform. It holds information or opinion that the user voluntarily contributes in a non-intrusive and unrestrictive manner [26, 27]. The online comments served as an important communications channel for the users to feedback and share their thoughts and opinions on their video consumption experiences [27, 28]. Various approaches have been taken to analyse, identify and categorise the YouTube comments. Obadimu et al. [9] used the Support Vector Machines (SVM), a supervised classification method, to automatically classify the comments to facilitate the filtering of unacceptable YouTube comments. Similarly, [1] used the SVM to effectively detect useful YouTube comments to help understand their viewers' concerns and needs. On the other hand, [10] utilised an off-the-shelf content analysis package, Text Miner (SAS Institute Inc.), to analyse YouTube comments to uncover healthy eating habits.

This study will adopt an increasingly popular unsupervised probabilistic topic modelling, Structural Topic Modelling (STM) [29]. STM extends from Latent Dirichlet Allocation [30] to incorporate documents metadata into its model in the form of covariates, for example, date and ratings of reviews, to enhance the allocation of words to the latent topics in the documents [29]. Reich et al. [31] have evaluated STM's utility and reliability to analyse a large volume of online learning data from discussion forums, surveys and course evaluations to "find syntactic patterns with semantic meaning in [the] structured text" (p. 156). In sum, topic modelling has the potential to identify latent topics related to learning behaviours and experiences from learners' comments.

3 Methodology

3.1 Sampling and Data Collection

Learners' comments were collected from one of the popular YouTube channels, the freeCodeCamp, which curates coding related video tutorials contributed by the coding community to help learners code. As of end-2019, freeCodeCamp has more than 3 million subscribers with over 1,200 videos in its channel. The videos were sorted according to their popularity. Those with more than a million views and published before 2019 were selected. The selection criteria imposed were to ensure that the data collected were representative of its popularity over one year. Due to the downloading limitation set by YouTube, the data collection via the Python v3.8 script was performed over three days, between 27 December 2020 to 29 December 2020. In total, 52,431 comments from 24 videos ($M = 2,185$, $SD = 3,437$) were collected (See Table 1). The collected data is pre-processed and analysed using R statistical software v4.0.

Table 1. Descriptive statistics for the 24 YouTube videos.

	Total	Mean	SD	Maximum	Minimum
Views	73,132,823	3,047,200	4,013,501	20,795,380	1,000,518
Comments	52,431	2,185	3,437	17,517	419

3.2 Data Pre-processing

Data pre-processing is performed before any computation analysis as it will help improve the overall data quality and the relevancy of the raw dataset. As YouTube comments is a form of user-generated content that is unstructured and free-text, this step is particularly important. Data pre-processing is used to (1) eliminate any data noises that were not meaningful (e.g. punctuation, numbers, non-alphanumeric characters and stop-words), unusual (e.g. ‘cifoymysye’) or highly re-occurrence words (e.g. ‘course’) that might otherwise skew the analysis; (2) correct the misspelt words and typographic mistakes that were common in the user-generated content; (3) filter-off data that were not relevant to the study such as non-English language (R ‘cld2’ package), duplicates, empty or comments that were posted earlier than the year 2019; and (4) resolve words in the comments to its lemma (R ‘textstem’ package), that is, to convert the word to its based or dictionary form (e.g. from ‘learning’ to ‘learn’). In addition, a total of 9,482 misspelt words were identified by R ‘hunspell’ package, of which 435 misspelt words (e.g. ‘ecssssssssssstatic’) were unable to be resolved and thus, removed. The remaining misspelt words were auto-corrected. After pre-processing, 35,334 valid comments were used for topic modelling.

3.3 Topic Modelling

Before applying the STM to the dataset, the number of topics (k) needs to be defined. A range of k between 5 to 15, in intervals of 1, was applied to the model to estimate the best k for the dataset. The estimated k was evaluated based on three metrics: held-out likelihood estimation, semantic coherence and residuals [29, 32]. The held-out likelihood is a measure of the model’s predictive power [29]. Semantic coherence is the words with the highest probability that occur together in a given topic [32]. Residuals measure whether there is an overdispersion of the multinomial variance in the STM method [29]. Thus, the best k would have a high held-out likelihood, high semantic coherence and low residuals measures [29, 32].

The best k was then further fine-tuned and evaluated to build the best fitting STM. The STM was then assessed based on exclusivity and semantic coherence [29]. Exclusivity in this instance measures the difference between topics based on the similarities of word distribution in the various topics, that is, a topic is considered exclusive if the top words do not appear in other topics. The fine-tuned k STM was applied to the dataset for the topic modelling analysis. The results were manually assigned with suitable labels to describe the topics based on their associated keywords and comments.

4 Results

The STM identified ten topics and computed its expected topic proportion (γ) within the dataset (See Appendix). The assignment of the labels was first performed by one researcher and then confirmed by the second researcher. The labelling was based on the logical association between the top keywords and their associated comments.

Four topics comprise Topic 4, 7, 9 and 10 represented coding-related questions and constituted about 33% of the overall expected topic proportion. The four topics were labelled as ‘Sharing doubts and questions’ ($\gamma = .077$), ‘Sharing technical questions on coding’ ($\gamma = .084$), ‘Sharing programming questions’ ($\gamma = .069$) and ‘Sharing programming error messages’ ($\gamma = .101$). Although the four topics were relatively similar, their associated comments revealed that the topics addressed different aspects of the coding video tutorial. For example, Topic 4 mainly shared how to execute the codes and set up the coding environment. Topic 7 was related to a database query. Topic 9 was associated with Python programming, and Topic 10 was related to general code debugging. Nonetheless, these were indicators that the learners were proactive in learning, reaching out to clarify their doubts and queries in response to their learning. Thus, Topics 4, 7, 9, 10 could be inferred as a form of help-seeking strategy from the SRL perspective [24].

Two topics (Topic 2 and 6) revealed that learners had exhibited positive learning behaviours. Topic 2, labelled as ‘Sharing compliments’, has the highest proportion ($\gamma = .186$) of comments with keywords such as ‘thank’, ‘tutorial’, and ‘great’. Its associated comments “...I love your tutorials ... and all the knowledge it offers...”, further implied that the topic is related to complimenting the instructors or video tutorials. Topic 6, labelled as ‘Sharing learning practices’ ($\gamma = .118$), had keywords on “good”, ‘understand’ and ‘help’ and comment such as “...followed this course completely on Monday...bought exam and studied... passed the exam...”, thus reflected how and what learners had achieved from the video tutorial. Topic 2 and 6, therefore, suggested that learning behaviours were reflective of SRL in both the self-evaluation and goal setting strategies [23]. By sharing compliments, learners attributed and explained their learning success as part of self-evaluation. The sharing of learning practices presented how learners learned and achieved their desired learning outcome in goal setting.

Next, three topics (Topic 1, 3 and 5) were observed to be feedback towards the video tutorial. Topic 1 and Topic 5 had a similar expected topic proportion ($\gamma = .075$) after rounded to three decimal places and were labelled as ‘Sharing general feedback’ and ‘Sharing information on video segment’ respectively. Topic 3 was labelled as ‘Sharing specific feedback on video content’ ($\gamma = .146$). In Topic 1, a learner commented that the used of multiple examples in the video was useful as it helped to reinforce the learning, while in Topic 5, learners shared the timing of the video segment much like a table of contents to ease in search of video content. Meanwhile, in Topic 3, learners were more targeted in their comments that included “in some parts of the video his voice sounds like Thor’s voice, Walter Thor to the rescue!!!”. Lastly, Topic 8 was labelled as ‘Sharing of fun content’ ($\gamma = .068$) and had the lowest expected topic proportion. This topic represented the sharing of functional codes by the learners in response to the video tutorials.

5 Discussion and Conclusion

This study is one of the earliest studies to examine learners’ sharing behaviours from the lens of SRL. The topic modelling approach uncovered the latent topics and their associated comments that indicate how learners learn and how they share and interact with other learners. Out of the ten topics identified, six sharing behaviours are related to learning behaviours. Based on the SRL perspective, these behaviours are further categorised into three SRL strategies and inferred as goal setting, help-seeking and self-evaluation (See Table 2). However, it is important to note that the sharing behaviours are expected to change over time, according to evolving learners’ needs. The other four topics identified were learners’ feedback on the video tutorial.

Table 2. Relationship between SRL strategies and identified topics.

SRL strategies	Topics
Goal setting	6
Help-seeking	4, 7, 9, 10
Self-evaluation	2

Of interest is the category on sharing of coding-related questions comprises Topic 4, 7, 9 and 10, which were representative of a help-seeking SRL strategy. This result was unexpected as it was contrary to a past study showing that help-seeking was not effective in an online learning environment [33]. According to [23], help-seeking implies an in-person meeting or through email or online consultation with peers or instructors, while [24] defined help-seeking as “choosing specific models, teachers or books to help oneself to learn” (p. 79). In this finding, help-seeking appeared to be an effective dominating self-regulating coping behaviour. It presented a potential alternative definition of help-seeking in a social learning environment. Learners would seek help publicly through social features within the learning platform. Given that coding is a technical skill that

often requires guidance and clarifications, it is not too surprising that learners would use social features as a means to “communicate” with instructors, peers or the public to seek help and guidance.

The results also found that learners do utilise the social features within the video hosting platforms to reflect on their learning experiences (Topic 2) and attribute their learning success (Topic 6). These sharing provided insights into learners’ learning behaviours synonymous with self-evaluation and goal setting in the SRL strategies. On the other hand, Topics 1, 3, 5 and 8, while not typical of SRL, were feedback towards the coding video tutorial and content that could inform the content creators on the areas to engage the learners, address learners’ concerns or refine the course content. Subtly, these topics represented an indirect form of environment structuring in SRL strategies. The topics address the softer aspect of the environment: video tutorial and content, rather than the physical environment.

In conclusion, this research provides a new perspective to the study of how sharing behaviours are associated with online learning behaviour. Specifically, this research demonstrated a feasible approach to use topic modelling such as STM to infer learning behaviour from user-generated content. Here, keywords were extracted, and together with their associated comments, were further evaluated for labelling. Furthermore, this research also gives an insight that learners do exhibit their learning activities in online comments. Though the number of learning behaviours extracted is small, it represents how the learners had reacted to consuming the video tutorials. Thus, it can hypothesize that different video tutorials could potentially vary learning behaviour.

Three limitations are identified in this study. First, the dataset was restricted to coding video tutorials and thus could not be generalized to other video tutorials as learners may exhibit different learning behaviours. Second, as this is a preliminary study on sharing behaviour from video tutorials, replies associated with learners’ comments were not collected to reduce any potential differences that could arise during the computational analysis. Third, the labelling of the topic is manual and subjective to the researchers’ interpretation. However, it should be noted that the two researchers participated in this exercise to reduce potential personal biases of the researchers.

Future work could expand the dataset to include other types of video tutorials (e.g. teaching language) that are less technical. Having video tutorials on different subjects would provide a contrasting comparison of the learning behaviour. Another limitation relates to the targeted audience in the sampled videos. The learners consuming the learning content in this study are general knowledge seekers who are interested in computer programming related content. Since YouTube is also a learning platform for other types of learners (e.g. children), it will be worthwhile to replicate this study with types of learners. With the findings from this study and the feasibility of using topic modelling on user-generated content, future work could further expand to include replies from its associated comments.

Appendix

See Table A-1.

Table A-1. Topic modelling results from YouTube comments.

Topic: Label*	Expected topic proportion (γ)	Top ten keywords	Extract of comments [#]
Topic 1: Sharing general feedback	0.075	follow, helpful, office, comment, example, step, note, reference, information, key	“Love this course! Learning coding and this is so simple and helpful, quick examples one right after the other gives quick imprinting and easy to remember information. 40 mins in and I feel really comfortable already”
Topic 2: Sharing compliments	0.186	thank, tutorial, great, love, awesome, beginner, free, amaze, video, basic	“I love you. I love your tutorials. I love free code camp and all the knowledge it offers. But I simply ask 2 things from you. atleast a banner ad so you guys make some money. and please please use dark theme...”
Topic 3: Sharing specific feedback on video content	0.146	video, like, watch, time, start, explain, finish, long, speed, subtitle	“In some parts of the video his voice sounds like Thor's voice, Walter Thor to the rescue!!!”
Topic 4: Sharing doubts and questions	0.077	know, app, project, read, website, lesson, idea, nothing, talk, wow	“I get this error when trying to run the "Hello World" "C++ Test Project - Debug": The compiler's setup (GNU GCC Compiler) is invalid, ...” Does anyone know how to fix this???”
Topic 5: Sharing information on video segment	0.075	new, open, react, password, script, software, log, var, check, access	“00:04:22 comment your Javascript code 00:05:58 data types and variables 00:09:14 storing values with assignment ... 00:11:33 initializing variables ... 00:12:00 uninitialized variables ...”
Topic 6: Sharing learning practices	0.118	good, understand, help, easy, job, exam, pass, useful, practice, excellent	“Booked exam on Sunday for Thursday. Followed this course completely on Monday, half of Tuesday. Bought practice exams and studied them through for Tue,Wed, half of Thurs. Studied for 5 hrs each dday. Passed the exam today ie Thursday. This course is excellent as beginner to understand AWS terminologies, but you shall need some practice exams as well to be confident when giving the exam.”
Topic 7: Sharing technical questions on coding	0.084	work, text, world, database, late, channel, hard, hope, age, advance	“from tkinter import * root = Tk() myLabel = Label(root, text="Hello world") root.mainloop() I did this according to the video. But the "Hello world" did not appear on the window. Why is that?”

(continued)

Table A-1. (continued)

Topic 8: Sharing fun content	0.068	guess, word, console, game, count, friend, lose, leave, simple, en- ter	<p>“Here’s my animal guessing game. I had a little fun adding some features to it after doing a little research. This will tell you how many guesses you have left and also how many letters in your guess were correct. it also lets the user give up by typing “exit”.</p> <pre>import sys ...”</pre> <p>“can someone help with the multi quiz game</p>
Topic 9: Sharing programming questions	0.069	question, change, link, test, life, answer, color, confuse, image, bad	<p>from coding</p> <pre>import Question question_prompts = ["What colour are apples?\n(a) Red/Green\n(b) Orange\n(c) Purple\n\n", "What colour are Bananas?\n(a) Teal\n(b) Magenta\n(c) Yellow\n\n", "What colour are strawberries?\n(a) Yel- low\n(b) Red\n(c) Blue\n\n"]...</pre>
Topic 10: Sharing programming error messages	0.101	error, help, create, user, problem, set, result, save, click, enter	<pre>“...elseif (\$op == ""){ echo \$num1 * \$num2; } else { echo "Invalid operator"; } What is wrong in this program? Can any- one please help?”</pre>

*Labels are manually assigned.

#Typos are inherent errors in the comments.

References

1. Poché, E., Jha, N., Williams, G., Staten, J., Vesper, M., Mahmoud, A.: Analyzing user comments on YouTube coding tutorial videos. In: IEEE/ACM 25th International Conference on Program Comprehension, Buenos Aires, pp. 196–206 (2017). <https://doi.org/10.1109/ICPC.2017.26>
2. Storey, M.-A., Singer, L., Cleary, B., Figueira Filho, F., Zagalsky, A.: The revolution of social media in software engineering. In: FOSE 2014: Future of Software Engineering Proceedings, New York, USA, pp. 100–116 (2014). <https://doi.org/10.1145/2593882.2593887>
3. van der Meij, J., van der Meij, H.: A test of the design of a video tutorial for software training. *J. Comput. Assist Learn.* **31**(2) (2014). <https://doi.org/10.1111/jcal.12082>
4. Carlisle, M.: Using YouTube to enhance student class preparation in an introductory Java course. In: SIGCSE 2010: Proceedings of the 41st ACM Technical Symposium on Computer Science Education, New York, USA, pp. 470–474 (2010). <https://doi.org/10.1145/1734263.1734419>
5. MacLeod, L., Bergen, A., Storey, M.-A.: Documenting and sharing software knowledge using screencasts. *Empir. Softw. Eng.* **22**(3), 1478–1507 (2017). <https://doi.org/10.1007/s10664-017-9501-9>

6. Dubovi, I., Tabak, I.: An empirical analysis of knowledge co-construction in YouTube comments. *Comput. Educ.* **156**, 103939 (2020). <https://doi.org/10.1016/j.compedu.2020.103939>
7. Zhou, Q., Lee, C.S., Sin, S.C.J., Lin, S., Hu, H., Ismail, M.F.: Understanding the use of YouTube as a learning resource: A social cognitive perspective. *Aslib J. Inf. Manag.* **72**(3), 339–359 (2020). <https://doi.org/10.1108/AJIM-10-2019-0290>
8. Lee, C.S., Osop, H., Goh, D., Kelni, G.: Making sense of comments on YouTube educational videos: a self-directed learning perspective. *Online Inf. Rev.* **41**(5), 611–625 (2017). <https://doi.org/10.1108/OIR-09-2016-0274>
9. Obadimu, A., Mead, E., Hussain, M.N., Agarwal, N.: Identifying toxicity within YouTube video comment. In: Thomson, R., Bisgin, H., Dancy, C., Hyder, A. (eds.) *SBP-BRiMS 2019*. LNCS, vol. 11549, pp. 214–223. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21741-9_22
10. Siersdorfer, S., Nejdil, W., Pedro, J.S.: How useful are your comments? Analyzing and predicting YouTube comments and comment rating. In: *WWW 2010: Proceedings of the 19th International Conference on World Wide Web*, Raleigh, North Carolina, USA (2010). <https://doi.org/10.1145/1772690.1772781>
11. Teng, S., Khong, K.W., Sharif, S.P., Ahmed, A.: YouTube video comments on healthy eating: descriptive and predictive analysis. *JMIR Public Health Surveill.* **6**(4) (2020). <https://doi.org/10.2196/19618>
12. Johnson, G., Davies, S.: Self-regulated learning in digital environments: theory, research, praxis. *Br. J. Res.* **1**(2), 1–14 (2014). <http://hdl.handle.net/20.500.11937/45935>
13. Zhou, Q., Lee, C.S., Sin, S.C.J.: Using social media in formal learning: investigating learning strategies and satisfaction. *Proc. Assoc. Inf. Sci. Technol.* **54**(1), 472–482 (2017). <https://doi.org/10.1002/pr2.2017.14505401051>
14. Brook, J.: The affordances of YouTube for language learning and teaching. *Hawaii Pacific University TESOL Working Paper Series* **9**(2), 37–56 (2011)
15. Burke, S., Snyder, S.: YouTube: an innovative learning resource for college health education courses. *Int. Electron. J. Health Educ.* **11**, 39–46 (2008)
16. Ellmann, M., Oeser, A., Fucci, D., Maalej, W.: Find, understand, and extend development screencasts on YouTube. In: *SWAN 2017: Proceedings of the 3rd ACM SIGSOFT International Workshop on Software Analytics*, New York, USA, pp. 1–7 (2017). <https://doi.org/10.1145/3121257.3121260>
17. MacLeod, L., Storey, M.A., Bergen, A.: Code, camera, action: how software developers document and share program knowledge using YouTube. In: *Proceedings of the 23rd IEEE International Conference on Program Comprehension*, Florence, Italy, pp. 104–114 (2015). <https://doi.org/10.1109/ICPC.2015.19>
18. Kim, J., Guo, P.J., Cai, C.J., Li, S.W., Gajos, K.Z., Miller, R.C.: Data-driven interaction techniques for improving navigation of educational videos. In: *UIST 2014: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, New York, USA, pp. 563–572 (2014). <https://doi.org/10.1145/2642918.2647389>
19. Pavel, A., Reed, C., Hartmann, B., Agrawala, M.: Video digests: a browsable, skimmable format for informational lecture videos. In: *UIST 2014: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, New York, USA, pp. 573–582 (2014). <https://doi.org/10.1145/2642918.2647400>
20. Swan, K.: Research on online learning. *J. Asynchronous Learn. Netw.* **11**(1), 55–59 (2007). <https://doi.org/10.24059/olj.v11i1.1736>
21. Jansen, R.S., van Leeuwen, A., Janssen, J., Conijn, R., Kester, L.: Supporting learners' self-regulated learning in massive open online courses. *Comput. Educ.* **146** (2020). <https://doi.org/10.1016/j.compedu.2019.103771>

22. Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G., Paas, F.: Supporting self-regulated learning in online learning environments and MOOCs: a systematic review. *Int. J. Hum.-Comput. Interact.* **35**(4–5), 356–373 (2019). <https://doi.org/10.1080/10447318.2018.1543084>
23. Barnard, L., Lan, W.Y., To, Y.M., Paton, V.O., Lai, S.L.: Measuring self-regulation in online and blended learning environments. *Internet High. Educ.* **12**, 1–6 (2009). <https://doi.org/10.1016/j.iheduc.2008.10.005>
24. Zimmerman, B.J.: Academic study and the development of personal skill: a self-regulatory perspective. *Educ. Psychol.* **33**(2), 73–86 (1998). <https://doi.org/10.1080/00461520.1998.9653292>
25. Araka, E., Maina, E., Gitonga, R., Oboko, R.: Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018). *Res. Pract. Technol. Enhanc. Learn.* **15**(1), 1–21 (2020). <https://doi.org/10.1186/s41039-020-00129-5>
26. Naab, T.K., Sehl, A.: Studies of user-generated content: a systematic review. *Journalism* **18**(10), 1256–1273 (2016). <https://doi.org/10.1177/1464884916673557>
27. Yoo, K.H., Gretzel, U.: What motivates consumers to write online travel reviews? *Inf. Technol. Tour.* **10**(4), 283–295 (2008). <https://doi.org/10.3727/109830508788403114>
28. Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. *J. Comput.-Mediat. Commun.* **13**(1) (2007). <https://doi.org/10.1109/EMR.2010.5559139>
29. Roberts, M.E., Stewart, B.M., Tingley, D., Airoidi, E.M.: The structural topic model and applied social science. Neural Information Processing Society (2013). <http://scholar.harvard.edu/dtingley/node/132666>
30. Blei, D.M., Ng, A., Jordan, M.: Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003). <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
31. Reich, J., Tingley, D., Leder-Luis, J., Roberts, M.E., Stewart, B.: Computer-assisted reading and discovery for student generated text in massive open online courses. *J. Learn. Anal.* **2**(1), 156–184 (2014). <https://doi.org/10.18608/jla.2015.21.8>
32. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: *EMNLP 2011: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 262–272. Association for Computational Linguistics (2011)
33. Vilkova, K., Shcheglova, I.: Deconstructing self-regulated learning in MOOCs: in search of help-seeking mechanisms. *Educ. Inf. Technol.* **26**(1), 17–33 (2020). <https://doi.org/10.1007/s10639-020-10244-x>