Computers & Education 140 (2019) 103594

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Detecting cognitive engagement using word embeddings within an online teacher professional development community

Thushari Atapattu*, Menasha Thilakaratne, Rebecca Vivian, Katrina Falkner

School of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia

ARTICLE INFO

Keywords:
Teacher professional development
MOOC
Word embeddings
ICAP framework
Online learning

ABSTRACT

Research states that effective teacher professional development (PD) engages teachers as active learners and co-creators of content. However, it is yet to be known whether such pedagogy impacts on cognitive engagement. We adopt the ICAP Framework to measure cognitive engagement in a teacher PD Massively Open Online Course (MOOC). We use word embeddings to automate the identification of teachers' community contributions as representing 'active' engagement by manipulating course materials, or 'constructive' engagement through the generation of new knowledge. We explored individual variation in engagement across units. Our findings demonstrated that the participants' cognitive engagement is influenced by the nature of MOOC tasks. We adopted a manual content analysis approach to explore constructive contributions. From 67 cases considered, all but one case was identified as containing 'constructive knowledge', providing a solid basis for replicating our proposed methodology to analyse cognitive engagement within the community-centric MOOC models.

1. Introduction

Teacher professional development (PD) programs are increasingly adopting Massive Open Online Courses (MOOCs) as a model of delivery (Falkner, Vivian, Falkner, & Williams, 2017; Laurillard, 2016). MOOC pedagogy that blends instruction and community learning aligns well with teacher PD needs as they allow teachers to build subject confidence and connect with like-minded colleagues, as well as enable teachers to participate in PD at their convenience, from anywhere, anytime with an Internet connection (Tseng & Kuo, 2014). In a review of epistemological and social dimensions of teaching in MOOCs (Toven-Lindsey, Rhoads, & Lozano, 2015), the majority of pedagogical practices currently used in MOOCs primarily adopt objectivist-individual approaches, with some adopting constructivist-group or individual-constructivist approaches combined with objectivist-individual. They identified that although discussion boards were available in the majority of courses, they were primarily used as a supporting discussion between participants and instructors rather than for innovative and meaningful knowledge-sharing and collaboration.

Varying MOOC models have emerged, such as traditional, teacher-centred courses (known as xMOOCs), as well as community courses (Joksimović et al., 2018), referred to as cMOOCs, adopting connectivist principles that place a greater emphasis on community (Siemens, 2012). The creation of professional community networks, in which teachers can share resources, pedagogy, and research, are valuable sources of support for the generation of new knowledge and teacher professional growth (Brennan, 2012). Based on this growing evidence, a number of emerging PD courses have adopted models that engage teachers as agents of their professional learning experiences through Communities of Practice (CoP) (Wenger, 1998) and co-creation (Brennan, 2012; Kleiman,

* Corresponding author.
E-mail address: thushari.atapattu@adelaide.edu.au (T. Atapattu).

https://doi.org/10.1016/j.compedu.2019.05.020
Received 28 January 2019; Received in revised form 26 May 2019; Accepted 28 May 2019
Available online 31 May 2019
0360-1315/

Wolf, & Frye, 2013; Laurillard, 2016; Tseng & Kuo, 2014) (Falkner et al., 2017). These models adopt constructionist and connectivist ideals by placing teachers at the center of resource creation, as co-designers of their professional learning experience and by harnessing a community platform to facilitate this engagement. Through a connectivist approach, instructors engage learners as problem solvers for their own contexts (Huang, 2002) in which they construct knowledge through active thinking and reflection (Ruey, 2010). However, creating MOOC communities that provide participants with an opportunity to deepen their understanding and to learn from one another is a challenge for instructional designers and best practices in this space is yet to be fully understood (Toven-Lindsey et al., 2015).

Such community-centric models pose a challenge in finding scalable and effective ways to measure teacher professional growth as they rely less on conventional graded online learning activities (such as quizzes), and instead on active participation and community contributions. Connectivist approaches are time-consuming and challenging to measure as the focus is not necessarily on the output, but the quality of the learning process and the way that individuals connect new knowledge to pre-existing knowledge (Huang, 2002). Self-evaluations through surveys (Kennedy, Coffrin, de Barba, & Corrin, 2015; Y.; Wang, 2014) and quantification of community contributions, such as post frequency, post size, the use of sophisticated and concrete words (Crossley et al., 2015; Xing & Gao, 2018; Yang, Sinha, Adamson, & Rose, 2013) have been previously utilised to measure participant success in MOOC forums. However, deeper measures of cognitive engagement that support a movement toward understanding participants' cognitive processes and knowledge integration are required.

Chi and Wylie's (2014) Interactive, Constructive, Active, and Passive (ICAP) framework presents a guide for evaluating the cognitive engagement of contributions within online communities. Although initially developed for face-to-face learning, this framework has been utilised within MOOC studies to measure the association between course materials and discussion contributions (X. Wang, Yang, Wen, Koedinger, & Rose, 2015). More specifically, the framework provides a measure for new knowledge construction that involves the generation of new ideas beyond the course materials ('constructive' engagement) and the identification of 'active engagement' involving the manipulation or rephrasing of course materials. However, prior work that has utilised this framework has relied on significant manual work with qualitative content analysis (X. Wang et al., 2015).

In this paper, we present a novel automated approach using a language modeling technique called neural word embeddings (Doc2Vec) (Le & Mikolov, 2014) to measure the cognitive engagement of a teacher PD MOOC for K-6 Digital Technologies education. We analyse teachers' community contributions across tasks to understand whether they demonstrate constructive behaviour as they progress through the PD MOOC. We couple this technique with qualitative manual content analysis to validate the machine-extracted constructive contributions.

Informed by a theoretical framework, our work establishes a novel linguistic-related model by bridging the gap between technology and methodology to efficiently measure cognitive engagement in community-centric MOOCs without manual analysis.

2. Related work

Limited studies have focused on the association between discussion contributions and course materials as an approach to measure learners' cognitive behaviour (Wise, Cui, Jin, & Vytasek, 2017). Wang et al. (2015) suggest that on-topic (or content-related) discussions are correlated with more learning than off-topic discussions. Dowell, Brooks, Kovanovic, Joksimovic, and Gasevic (2017) explored whether the learners are on- or off-topic in their discussions of subsequent offerings of courses and found that they are mostly on-topic over time. However, the majority of this research neglects the fact that off-topic (but content-related) contributions that go beyond the course material may potentially result in the construction of new knowledge and hence, be worthwhile to explore further.

Wang, Wen, and Rose (2016) developed a qualitative coding scheme to explore the cognitive behaviour within online discussions and their correlation with learning gain, based on the ICAP framework (Chi & Wylie, 2014). The authors suggest that higher-order thinking behaviours, such as being socially interactive and constructively adding new knowledge, demonstrate learning gains. Vellukunnel et al. (2017) also utilised the ICAP framework within a Computer Science education community and found that participants' posts that reflected some level of constructive problem-solving are positively correlated with course grades.

The Community of Inquiry Model (COI) (Garrison, Anderson, & Archer, 2000) presents three constructs as being important to learning: cognitive, social, and teaching. The 'cognitive' construct associates learners' knowledge construction within a learning community and involves four phases, including a triggering event, exploration, integration, and resolution. The COI model has been adopted as a theoretical framework within online learning communities (Kovanović et al., 2018). The COI model supports both constructivist and connectivist learning approaches. Constructivism, suggesting that humans construct knowledge and meaning through their experiences, and connectivism which captures the way in which social learning technologies facilitate learning and the sharing of knowledge through online communities (Siemens, 2012). This study is based on a course that is designed to engage learners in constructivist learning by engaging their prior knowledge and experiences and inviting them to share professional knowledge, experiences and ideas within a professional learning community that encapsulates connectivist principles.

The analysis of cognitive behaviour within online communities involves qualitative content analysis, which typically requires significantly large manual labour. Building on the work of Wang et al. (2016), our work focuses on bridging the gap between technology, theory and methodology by proposing an approach based on neural embedding techniques. We summarise the underlying theoretical framework in the next section.

3. Theoretical framework

We frame our problem in understanding the cognitive behaviour of a MOOC community as a problem on the association between community contributions and corresponding course materials. We present the theoretical background that supports the measurement of this association within an online context.

The ICAP framework (Chi & Wylie, 2014) has been utilised in several studies to measure cognitive engagement in face-to-face and online environments (Lutz et al., 2018; Vellukunnel et al., 2017; X.; Wang et al., 2016). We select the ICAP framework to ground our research problem as it is also previously utilised to solve a similar kind of problem (X. Wang et al., 2016) and because the framework is grounded in the learning theory of 'constructivism', making it well suited within a PD context that engages teachers as active learners and co-creators (Falkner et al., 2017; Brennan, 2012).

Although Chi and Wylie's (2014) original framework was developed to understand classroom conversational data, the ICAP framework has been proven to be effectively adopted within the online context to understand learners' cognitive engagement (X. Wang et al., 2016). ICAP categorises learners' cognitive behaviour in discussions into four modes:

- Interactive: interactive behaviours should meet two criteria: 1) both partners' utterances must be primarily *constructive,* and 2) sufficient degree of turn-taking must occur.
- Constructive: learners generate or produce additional externalised outputs or products beyond what was provided in the learning materials.
- Active: learners' engagement with instructional materials can be operationalised as active if some form of overt motoric action or physical manipulation is undertaken.
- Passive: learners being oriented toward and receiving information from the instructional materials without overtly doing anything else related to learning.

Our work adopts only two modes as the focus of this study: 'constructive' behaviour that demonstrates the contributions beyond materials and 'active' behaviour that is aligned well with the materials (see Table 2 for examples in each mode). We omit 'passive' behaviour as our PD MOOC expects active behaviour by requesting participants to contribute to the online community. We also exclude 'interactive' behaviour as interactivity is not a key focus of our community platform (Google plus) or MOOC design. Instead, our teacher community is used as a knowledge sharing database where interactivity through commenting, voting, sharing is less obvious in our dataset.

Table 1
Community task options in each unit.

Unit Name & Topics	Community Task Options
Unit 2: Data – patterns Computational Thinking and Data; Types of Data; Representations of Data; Pattern Recognition; Visualising Data.	1) Collect and visualise data from your classroom. 2) Find & share a useful data source for use in a lesson. 3) Think of an activity for PD that involves teachers collecting, organising and presenting data.
Unit 3: Representation Data Encoding and Decoding; Digital Data; An Introduction to Binary.	1) Create a lesson/PD session that explores how some form of digital data is represented and transferred. 2) Describe an activity that involves children encoding and decoding data. 3) Create or share a resource for explaining binary or digital data.
Unit 4: Digital Systems Hardware and Software; Input and Output; Introduction to Networks.	1) Develop an inquiry question about computers or networks. 2) Create a classroom resource to explore computer in comparison to humans. 3) Share a lesson for exploring past and present technologies.
Unit 5: Information Systems Introduction to Information Services; Social Protocols; Encryption; Purposes and Uses of Technology.	1) Create a lesson for sharing information online safely or behaving appropriately. 2) Find an example of an information service that collects and presents information to users. 3) Share how students can present work online safely/engage in teamwork.
Unit 6: Algorithms & Programming Introduction to Algorithms; Computational Thinking Skills; Unplugged & Plugged Algorithms; Introduction to Programming.	1) Create a resource that explains a Computational Thinking skill. 2) Design an activity that explores sequencing of instructions. 3) Create an algorithm flowchart for an everyday activity and explain how you could use this in the classroom.
Unit 7: Visual Programming Introduction to Visual Programming Environments; Programming in Blockly; Programming in Scratch.	1) Design an activity that builds on an existing Scratch project. 2) Design an activity that incorporates a visual programming environment. 3) Use a visual programming environment to create your own project, which is a basis for an activity.

Table 2
Mapping of ICAP framework to the language model.

Mode	Revised definition (X. Wang et al., 2016)	Language model	Example
Active	student is displaying engagement with the course materials in the post by paraphrasing, repeating, or mapping resources	Semantic similarity between n-grams of course material and posts are in close proximity (document vectors in the same direction, i.e. cosine of angle is near to 1)	*"I used both **Blockly** and **Scratch** to introduce the concept of **Visual Programming** to one of my classes. **Blockly** gave students clear goals and **instructions** whereas **Scratch** allowed students to be more open […]" (Unit 7)*
Constructive	proposing a new idea that is related to the course but go beyond what is covered in the course	Semantic similarity between n-grams of course material and posts deviate from each other (document vectors in the opposite direction, i.e. cosine of angle is near to −1)	*"In my Year 1 class I had a set of **Bee Bots**. In our **literacy rotations** we did two activities that focused of sequencing steps, 1 - We used the **Bee Bots** and the **alphabet mat** to work together to **spell sight words**. The children were given a colour group and when they finished spelling the word they would say it. reinforcing the sight word in an engaging way." (Unit 6)*

4. Language modeling

Our work introduces an approach to minimise the labour-intensive qualitative content analysis of MOOC communities and instead automatically measure the cognitive engagement of discussions. To achieve this, we adopt a language modeling technique called neural word embeddings (Le & Mikolov, 2014; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to understand the community contributions and their association with course materials. We utilise a state-of-the-art neutral embedding technique called Doc2Vec (Le & Mikolov, 2014). A Doc2vec model is a shallow neural network with three layers; input, hidden and output. In comparison to Wang's, Wen, & Rose (2016) content-related approach where coders manually analysed participant posts and code according to the categories of ICAP framework, our work focuses on a linguistic-related approach where latent similarity is measured between course materials and classify posts into one of two chosen ICAP categories (i.e. constructive or active).

We utilise unsupervised neural embeddings over the widely used bag-of-words model (i.e. the model creates a vocabulary of known words and measures the presence of each word in a given document) as the latter ignores the semantic aspects and word order. Our technique not only embeds semantics but also similarity between course materials and community posts using *distributional hypothesis,* i.e. it assumes that two words in a similar context are semantically related.

Prior to building the Doc2Vec model, we extract n-grams from the text. N-grams represent the terms that are constructed from *n* number of tokens (e.g. 'visual programming' as a bi-gram without considering it as two separate tokens as 'visual' and 'programming'). We followed a data-driven approach to detect the n-grams (Thilakaratne, Falkner, & Atapattu, 2018) in the course materials and community posts by using a formula based on collocation patterns described by Mikolov et al. (2013).

4.1. Doc2Vec

Doc2vec (Le & Mikolov, 2014), an extension to the Word2vec model (Mikolov et al., 2013), is an unsupervised technique to generate dense vector representations (i.e. most of the values in the vector are non-zero) for documents. The created vector representations can be used to facilitate content similarity of the documents. This recently evolved embedding technique has been successfully applied to many recent Natural Language Processing (NLP) related tasks and is at the forefront of the research trend (Hashimoto, Alvarez-Melis, & Jaakkola, 2015). Therefore, we selected *Doc2vec* as the main technique to model the documents over traditional topic discovery techniques such as Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998) and Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003).

We constructed Doc2vec model for each course unit to measure the similarity between 'participants' posts' and 'lesson' vector (Fig. 1). For example, consider a lesson about *'Algorithms and Programming'* that mainly discusses *visual programming* (Lesson 1 in Fig. 1). We extracted all the user posts related to this lesson. If a user has posted more than one post, we concatenated all of his/her posts to represent that user. Afterwards, we created vector space for each lesson and its corresponding user posts using Doc2vec model (represented by the circle in Fig. 1). *L1* in Fig. 1 represents the lesson vector representation, and u*N* represents the vector representation of the user posts of *N*th user. A vector is a multidimensional continuous floating point numbers (vector size is 300 in our experiments) where each point captures a dimension of meaning. As a result, documents that have similar vector representations resides in close proximity in the vector space. For example, in Fig. 1, the vector *u8* is closer to the lesson vector *L1* because the *user 8* also discusses *visual programming* in their posts (i.e. *u8* is the nearest neighbour of lesson vector). In other words, the *lesson* and *user* 8 has similar vector representations. In contrast, the vector *u10* of *user 10,* who discusses *Beebot* in their post is located far away from *L1* representing the semantic dissimilarity of the two documents. In other words, *user 10* may integrate new knowledge in constructing their posts different from what was taught in the lesson. We analysed this semantic change to differentiate *active* and *constructive* users.

We used *Gensim* API (Rehurek & Sojka, 2010) to construct the Doc2vec models. The vector size, window size (i.e. the size of the context window that slides within the document to analyse the proximity of words), minimum word count and iterations used for Doc2vec training are 300, 10, 5, and 20 respectively. *Distributed memory* was selected as the training algorithm over the *distributed bag*

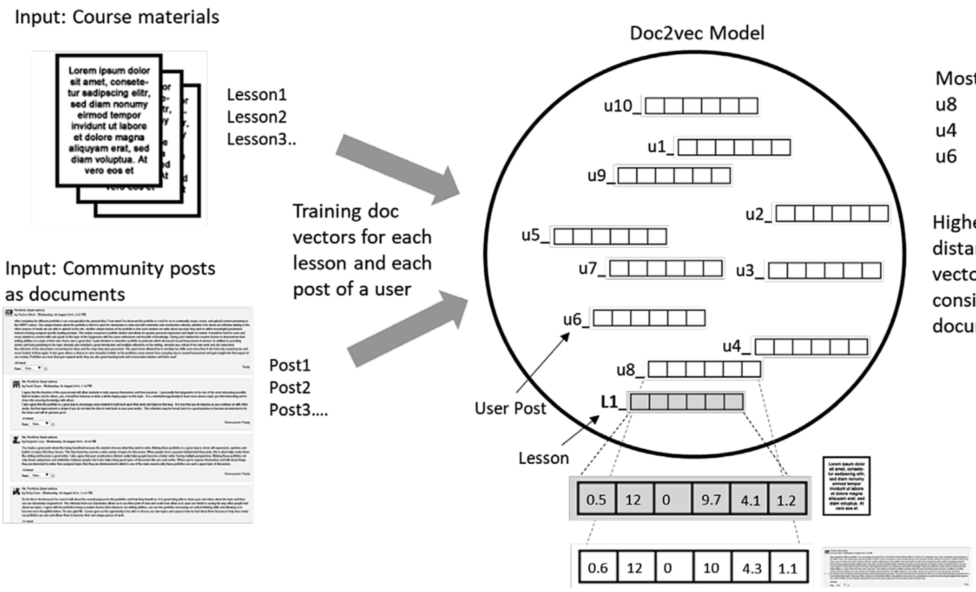


Fig. 1. Overview of the language model.

of words since the former tends to perform well (Le & Mikolov, 2014) and considered as the default model in Gensim API. This variant of doc2vec model act as a memory that remembers what is missing from the current context or the topic of the document (Le & Mikolov, 2014). The constructed Doc2vec model was analysed by performing nearest neighbour search to gain insights into how participant's posts vary from the lesson content as described below.

4.2. Cosine similarity

The nearest neighbour search was performed using cosine as the similarity metric. More specifically, we analysed how participant '*posts*' are positioned in the vector space related to the '*lesson*' vector. If a post is positioned closely to the lesson vector (i.e. a post with high cosine similarity score), we assumed that the participant had not deviated much from what is being learnt. In contrast, a post with low cosine similarity represents a participant that has integrated new knowledge to the post.

5. Research study design

In this study, we aim to understand the types of cognitive contributions teachers make to an online PD MOOC and the extent to which teachers are demonstrating constructive and active behaviour through their contributions. To facilitate this, we formulate our first research question to understand the nature of course materials, community posts, and participants.

RQ1. *How can we cluster participants based on their contributions and what types of contributions emerge across units?*
The findings lead us to explore participants' new knowledge integration through qualitative analysis and answering our second research questions.

RQ2. *What contents of the community contributions are mostly constructive?*
By answering this question, we gain understanding of new knowledge integration within the K-6 Digital Technologies curriculum in Australia.

5.1. Context

We explore these questions by adopting a case study approach of a self-sustained, self-paced MOOC developed to support Kindergarten to Year 6 (K-6) level teachers' PD in implementing the new [Country] Digital Technologies curriculum. The program has had over 19,000 enrolments, indicating the great demand for online PD across Australia '(Falkner et al., 2017).

The MOOC content is delivered through a traditional MOOC structure via the Coursebuilder platform, consisting of individual teaching units around key CS concepts (data patterns & representation, digital & information systems, algorithms & programming, and visual programming), exemplars of practice, featuring short videos and web-based content. This MOOC adopts a blend of individual-objectivist and individual-constructivist pedagogy (Arbaugh & Benbunan-Fich, 2006), with knowledge being transmitted within a contextualised manner, and through engaging individual participants in authentic professional tasks that are dependent on engagement with the subject matter.

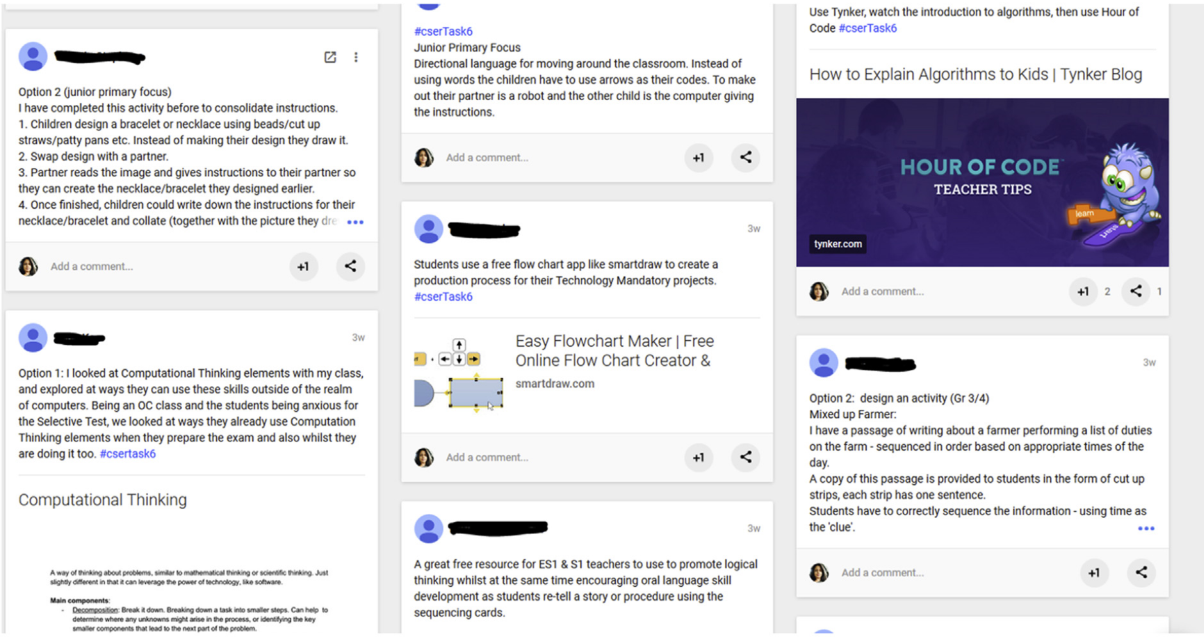


Fig. 2. Screenshot of the community.

The MOOC community is provided through a separate medium (Google plus) to assist in sustained community engagement beyond individual cohorts (Fig. 2). It forms an accessible database of learning and teaching ideas and resources that participants can continue to return to, search for and reuse. Engagement in the MOOC community is driven by MOOC activities and a community-centric design, which has been presented in detail previously '(Falkner et al., 2017).

The course presents 7 units, featuring sub-lessons within each that present a combination of web content, videos and practical classroom examples for various topics within the Digital Technologies curriculum. Participants are required to complete a task at the end of each unit which could involve sharing an idea for a classroom activity, experience or resource as part of the requirement for completing the MOOC for a certificate. This includes contributing a minimum of seven posts that correspond to seven units. Table 1 shows community task options that the participant can choose from each unit.

5.2. Dataset

Our dataset is a combination of participants' community posts and course materials.

5.2.1. Course documents

We generated text documents from the course web content and videos for six units (unit 2 to 7). Videos from YouTube were transcribed automatically and verified manually by the first author of this paper. The first introductory unit is eliminated from the analysis due to having noisy and irrelevant data such as participant introductions and networking.

5.2.2. Community posts

We extract 'task-related' community posts (task 2 to 7) from 1045 participants who have completed the course exit survey between 1st of August 2016 to 31st of August 2018. Our platform dedicates separate 'tabs' for each task (see Section 5.1 - context). The completion of the exit survey implies that the dataset contains a complete picture of community contributions of the selected cohort.

Our final dataset contained 7454 posts extracted from 1045 users (Mean = 6.71, SD = 2.09).

5.3. Methods

5.3.1. Pre-processing

MOOC tasks (see Table 1 - 'community task options') that are repeated in community posts were eliminated automatically using a String matching algorithm. Accordingly, task questions were not included in 'course documents'. The sentences of each unit and posts were extracted using NLTK's *Punkt Sentence Tokenizer* (Loper & Bird, 2002). Prior to building language models, each sentence was pre-processed by removing stop words, URLs, hashtags, individual numbers (except numbers inside terms such as 3D), punctuations, terms constituent of single letters (such as e.g., *i.e.*), and performed lemmatisation.

Other than these general pre-processing steps, our corpus is less noisy since course materials are well-written and formal as they were developed by instructors. Also, community posts are linguistically advanced compared to typical MOOC discussion posts as our MOOC has been designed for a particular cohort – Australian teachers, who have completed a Bachelor of Education and who are required to have a standard level of English in order to teach K-6.

As discussed in the 'language model' above, we obtain a cosine similarity score between course materials and posts to understand the discussion behaviour (i.e. constructive or active). We use the term 'similarity score' throughout the manuscript to refer the association between discussion posts and materials within the range of −1 ('constructive' end) to +1 ('active' end). Table 2 lists the mapping of constructive and active modes of ICAP to our language model. The examples in the last column demonstrate posts from the 'active' (e.g. *Scratch*, *Blockly*, *visual programming* are key concepts covered in Unit 7 - see Table 1) and 'constructive' (e.g. *Beebots, literacy rotations, alphabet mat* are new knowledge that is not being discussed in Unit 6) modes.

5.3.1.1. RQ1. Since RQ1 explores participants' discussion behaviour throughout PD across tasks, we obtained the similarity score of each participant per unit. We create document vectors per participant in each unit by aggregating multiple posts per unit task. Document vectors are also constructed from the course content of each unit. We selected the Two-step clustering method (Norusis, 2008) to cluster participants' similarity scores in each unit. Two-step clustering method is efficient in clustering a large number of cases (approximately a thousand in our problem). 'Euclidean distance' was set as the distance measure. Even though Two-step clustering supports automatic selection of number of clusters, it is sensitive to the order of cases. Therefore, we have used the Silhouette method to obtain optimal number of clusters in each unit (Kaufman & Rousseeuw, 1990) by sorting the cases in different random orders. This optimal number of clusters was used to perform Two-step clustering.

5.3.1.2. RQ2. To answer RQ2, we manually analysed the contributions of mostly constructive participants. A threshold of −0.5 was identified to extract participants whose similarity scores are less than −0.5. Manual analysis was performed on these posts by an author of this paper who did not contribute to algorithm development. The analyst is also an instructor of the MOOC and had expertise with the content.

6. Results & discussions

6.1. Answering RQ1

In our first research question, we focus on understanding the patterns of participants' community contributions in relation to materials and their change across tasks. Using Two-step clustering algorithm (Norusis, 2008), we clustered each participants' 'similarity score' between posts and course materials. In order to obtain the optimal number of clusters for our dataset, we used the Silhouette method. As shown in Fig. 3, three clusters were found as optimal (highest Silhouette index) in Unit 3. A similar analysis has been conducted for all the other units, and optimal number of clusters found was 3. The two-step clustering algorithm is applied to our dataset by setting three fixed number of clusters.

Table 3 shows the cluster sizes, means and standard deviations of the centroids. Cluster 1 contained participants with higher similarity scores (i.e. cosine of angle is near to 1). According to our interpretation in Table 2, this cluster contains 'active' participants. Cluster 2 contained participants with similarity scores closer to '0', implying that these participants are neither strongly active nor strongly constructive, however, contribute to content that is a combination of both ('overlap' cluster). Conversely, cluster 3 included participants with the lowest similarity scores (i.e. cosine of angle is near to −1). This cluster suggests that the participants' contributions have largely deviated from course materials and they are likely contributing to constructing new knowledge ('constructive').

Fig. 4 depicts three clusters across units. We have used the Silhouette coefficient for cohesion and separation to measure the validity of our clusters (Rousseeuw, 1987). Cluster 'cohesion' measures how closely related are objects within a cluster while 'separation' measures how distinct or well-separated a cluster from other clusters. Silhouette coefficient ranges from −1 to 1, and if it is

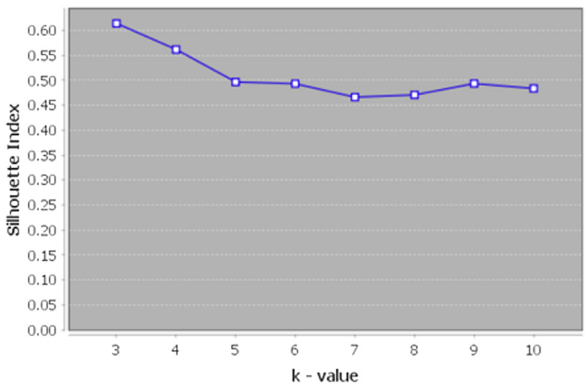


Fig. 3. Silhouette analysis for k ranging from 3 to 10 (Unit 3).

Table 3
Descriptive statistics of clusters.

Unit	Average Silhouette coefficient	Cluster 1			Cluster 2			Cluster 3		
		Mean	SD	n	Mean	SD	n	Mean	SD	n
L2	0.9	0.347	0.048	510	0.002	0.079	304	−0.419	0.110	139
L3	0.8	0.260	0.074	512	0.008	0.089	273	−0.488	0.152	154
L4	0.7	0.501	0.150	292	0.093	0.129	256	−0.499	0.235	388
L5	0.6	0.352	0.151	347	0.002	0.095	240	−0.426	0.180	337
L6	0.6	0.461	0.160	234	0.114	0.120	432	−0.491	0.227	220
L7	0.7	0.535	0.068	287	−0.018	0.140	256	−0.418	0.059	297

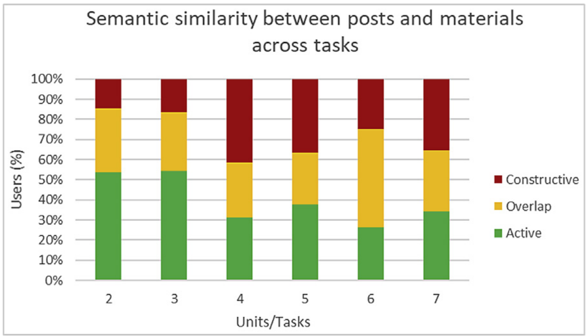


Fig. 4. Participants' community contributions related to materials across tasks.

closer to 1, it is interpreted as a good clustering solution. As an example, Fig. 5 shows the cluster quality of Unit 7. The second column of Table 3 lists the average silhouette coefficient in each cluster solution.

Fig. 4 presents some visible differences in terms of constructive, active and blended ('overlap') behaviour between the MOOC units. In general, it is apparent that the semantic similarity between posts and materials are somewhat decreasing across units (visible with 'green' bars), however, that some units have a similar profile. In particular, we see units 2 and 3 as well as units 4, 5 and 7 as having similar patterns of constructive, active and overlap engagement. The slight change in 'active' engagement may suggest that participants are more likely to be actively paraphrasing, repeating the concepts taught in the course at the beginning of the PD (as seen in units 2 and 3). The variations in constructive and blended ('overlap') knowledge-sharing across the MOOC units may suggest that behaviour is influenced by the types of activities proposed in the MOOC (see Table 1) and how teachers are invited to draw on knowledge.

As an example, units 2 and 3 were about 'data', with Unit 2 being most aligned with familiar Mathematics and Science classroom activities and as containing the least new Digital Technologies discipline knowledge to teachers. In Unit 2, participant language is more closely aligned with course material, and interesting observation of our community is that there appears to be a high number of cases in the community that involve re-sharing one of our course exemplars – an activity of collecting student data and representing it as a house for whole-class discussion. Below we present an example of a course exemplar (Fig. 6) and how a teacher has re-shared their experience in implementing the exemplar activity in the classroom. We can see in this example how closely the teacher's post is to the course exemplar, demonstrating 'active' behaviour.

We see in Fig. 4 that in both Unit 2 and 3 participants stay more closely to course content, suggesting that teachers are less confident to propose ideas and use language that deviate from those found in the course. Conversely, throughout the course, we see variations across the units where participants are likely to produce content that is different from course material (visible with 'red' bars). This may suggest that participants integrate course knowledge with their professional and pedagogical knowledge to construct new knowledge, thus moving further away from relying on course materials.

To elaborate further, in Unit 4 and 5, we observe a larger portion of 'constructive' contributions. As illustrated in Table 1, both of these units invite participants to engage in more personalised activities, such as reflecting on information systems they use in their

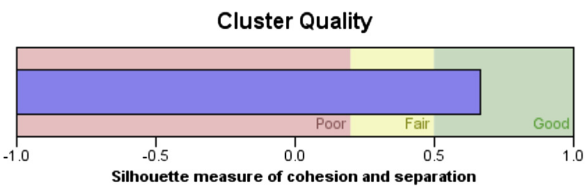


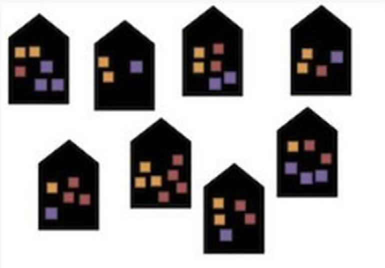
Fig. 5. Cluster quality of Unit 7 (Average Silhouette coefficient = 0.7).

In this activity, we target the following objectives:

- Collect, explore and sort data, and use digital systems to present the data creatively
- Recognise and explore patterns in data and represent data as pictures, symbols and diagrams
- Collect, access and present different types of data using simple software to create information and solve problems
- In this activity we record data about the family members of each student - recording how many parents they have, how many brothers and sisters, how many pets, etc. This can be varied and altered according to the student cohort, and the information that you already know about each student.

Each family member is recorded or represented, as a different colour block on each student's house. Each student could prepare their own data by making their own house picture:

And then these can be collected to show all of the families in the class:



This information can then be analysed and ordered according to different questions:

- Who has the largest number of family members in total?
- Who has the most brothers and sisters?
- Who has the most pets?

Fig. 6. Examples of active behaviour in Unit 2 - a sample from course material (above), a sample post (below).

everyday lives, their first experiences with computers and inquiry questions they would pose to students. These types of activities resulted in a large amount of ‘new knowledge’ generated (constructive behaviour) within the community and reflected the nature of the activity, which is about participants' personal experiences, the generation of new questions and use of technology which is unique to them.

Although we can see that there are variations in behaviour across the MOOC units and that some units have similar patterns of engagement, it is challenging to determine the extent to which types of MOOC activities impact on engagement as teachers in our MOOC were able to choose from a set of activity options within each unit. Furthermore, it may be that patterns of engagement were influenced by other factors, such as quality of content delivery, topic difficulty or teacher confidence with a topic. Future research in this space could explore factors that may impact on the types of knowledge contributions shared.

Identifying variations in behaviour motivated us to further analyse the ‘mostly constructive’ cluster (‘red’ bars) to understand whether these teachers are actually contributing new knowledge that is relevant to the community, and if so, what types of new content.

6.2. Answering RQ2

We explore our second research question by conducting manual content analysis. As discussed in Section 5.3, one annotator

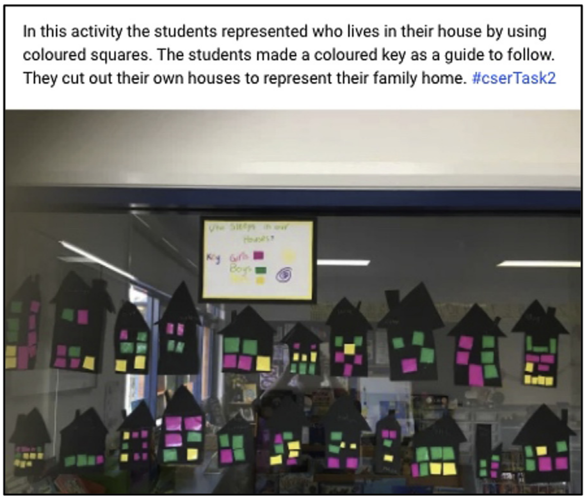


Fig. 6. (*continued*)

examined 39 and 28 participants' posts (similarity score < -0.5) from Unit 5 and 6, respectively. A threshold of -0.5 was selected to determine whether the posts that are largely deviated from course materials contain off topics instead of constructive knowledge. We selected two later units for analysis because the participants demonstrated constructive behaviour (Fig. 4) and the two topics were different from one another. From the 67 cases considered, all but one case from Unit 5 was identified as containing 'constructive contributions', with the one post having only 'active contributions' (i.e. not new)'. This suggests that the performance of our proposed language model is recommended for understanding the constructive behaviour of discussions. Table 4 presents the results from the qualitative analysis and the active and constructive knowledge that teachers brought to the community for units 5 and 6.

In all constructive posts reviewed, it was evident that in Unit 5, participants re-emphasising existing course language in their posts, such as 'creating and sharing information and online environments safely', and lesson ideas that were proposed in the course, such as sharing student work. Constructive knowledge teachers bring to this unit relate to new digital tools and platforms being suggested, with new ideas on how they can be used in the classroom. In particular, many of the 'constructive' posts introduced 'creating and sharing' work in online environments (e.g. SeeSaw) relating to opportunities for students' to share work with families as an authentic context, as well as the use of collaborative tools (e.g. GSuite) for students to engage in teamwork. For example, a teacher shared both a tool and the context of sharing work:

> *"'**Seesaw is an app that I use in my classroom** to allow the students to create work together and **then share it with their parents** and other students in the class"* (Unit 5)

This constructive behaviour aligns with the goals of the MOOC activities which invited teachers to share ways they engage students sharing work and creating work online together and what tools and environments they might use.

In terms of Unit 6, a number of new core Digital Technologies language and concepts are introduced from Unit 6, including algorithms and key programming constructs (Table 1). This topic is quite new to Australian teachers, with the majority having never encountered such content. The course materials introduced key language associated with algorithms (e.g. patterns, logical sequence,

Table 4
Summary of a qualitative concept analysis of 'mostly constructive' participants.

Unit & Topic	Active contributions	Constructive contributions
Unit 5 (Information systems)	Language: Creating and sharing information in safe online environments. Context: Sharing student work, sharing online safely, collaborating online, communicating online.	Language: eSafety/cybersafety, ownership. Resources: GSuite tools, Seesaw, Blogs, LMS, Applications. Contexts: Sharing work with families/parents, collaboration within tools (e.g. GSuite such as Google Docs, OneNote), holding classroom discussion using tools, acquiring a 'digital license', creating reflective journals, exploring commenting behaviour, seeking and receiving feedback online. Pedagogy: Modeling behaviour, teamwork, differentiation.
Unit 6 (Algorithms)	Language: Patterns, direction, logical sequence, arrangement, instructions and steps. Contexts: Use of storybooks for sequencing, board game creation, the use of drawing, writing or telling instructions for exploring algorithms.	Contexts: Painting, sight words, narratives (e.g. frames of a cartoon strip, historical events). Resources: Sphero, BeeBots, YouTube, new Book title suggestions (e.g. Princess and the Pea). Pedagogy: Use of cards (for sequencing), ability grouping, using grid paper for direction, reinforcing learning, student engagement.

order) and provided examples of contexts where algorithms are suitable (e.g. game creation, storybook narratives). New knowledge contributed by participants involved new examples of contexts in which algorithms can be introduced (e.g. literacy rotations), the introduction of new resources (e.g. storybook titles and technology such as robotics), as well as pedagogical considerations (e.g. ability grouping, reinforcement, teaching aides such as cards). An example being,

> "*I read a story to the children, **The Princess and the Pea**. I'd **printed the story in pictures from a free resource** I'd found online. I then had the **students put the cards** in the correct order of the story we'd just read.*" (Unit 6)

The introduction of digital technology to support programming activities (e.g. robotics and visual programming tools) were not introduced until the following unit (7); however, we can see participants making this connection earlier to the topic of algorithms by introducing technology such as Bee-Bots and Sphero within algorithm activities, resulting in constructive behaviour.

The activities invite participants to share lesson ideas and resources that connect with these key concepts and language, but also with their classroom practice. As a result, we see a greater number of these posts are either overlapping or different to course materials. This type of behaviour, toward the end of the MOOC, is what we would expect to see, as participants are moving toward generating new ideas and connecting new discipline knowledge with other learning areas and classroom practice. Unit 7 takes this topic further, introducing programming for the first time, with participants producing an even mix of similar, overlapping and different posts to course material.

It is evident that some participants introduce new contextualised examples, resources or pedagogy that is not explicitly taught in the course materials or that is taught later in the course. Furthermore, we see participants presenting new contexts for introducing Digital Technologies, such as within English, Mathematics and new activity ideas for teaching the content. Additionally, some teachers have presented new pedagogical strategies for the topics, such as using teamwork, modeling behaviour, grouping by ability and differentiated tasks, which introduces new teacher professional expertise into the MOOC community.

7. General discussion

This study sought to explore how participants in an online teacher PD MOOC for Digital Technologies integrate or build on professional knowledge presented in the course. We found that participants within the MOOC engaged in various cognitive engagement behaviours, such as constructive (new knowledge), active (remixing course materials) and a combination of both (integrating constructive knowledge and active knowledge). The results also identify that the cognitive engagement of participants changes throughout the course and that the more constructive contributions may increase when course activities invite participants to draw on their personal experiences and perspectives to explicitly propose new classroom ideas.

As per previous studies (X. Wang et al., 2016; X. Wang et al., 2015), we have demonstrated the application of the ICAP framework to identify cognitive engagement regarding active and constructive participant behaviours within the MOOC community. Furthermore, we identify a third cognitive mode emerge that involves the application of both constructive and active knowledge. We presented a new linguistic-related approach to automate the measurement of cognitive engagement within an online community that can complement traditional approaches of manual content analysis (Kovanović et al., 2019; X.; Wang et al., 2016).

According to the authors' knowledge, this is the first study that utilises neural word embeddings (Doc2vec model) to bridge the gap between technology and theory to propose a novel methodology for MOOC content analysis without relying on manual coding data due to the unsupervised nature of our language model. Word embeddings integrate semantic aspects and word order and measure the semantic similarity between course materials and community contributions using the distributional hypothesis as a basis. When building the Doc2Vec model, we utilise a data-driven approach to extract n-grams from the text (i.e. course materials and posts) using an approach contributed to the computational linguistic community through authors' previous work (Thilakaratne, et al., 2018).

Our evaluation study in RQ2 demonstrates that from the 67 cases considered, all but one case was identified as containing 'constructive knowledge', providing a solid basis for replicating our proposed methodology to analyse cognitive engagement within the community-centric MOOC models. In future works, we intend to utilise newness/givenness measure (Cai et al., 2005; McCarthy et al., 2012) as a baseline to further validate our proposed methodology without relying on manual validation.

Our findings provide insight into teachers' cognitive engagement within a PD MOOC that can inform PD MOOC pedagogy. Prior literature suggests that effective teacher online PD engages teachers as active learners and co-creators of content (Brennan, 2012; Kleiman et al., 2013; Laurillard, 2016; Falkner et al., 2017). Results indicate the value of online teacher PD that is designed in a way to engage teachers as active learners, and that our PD MOOC model has encouraged and facilitated constructive and active contributions related to professional practice.

Through cluster analysis across tasks, we discovered that the MOOC community is comprised of posts similar to course content (active), overlaps course content (active and constructive) and that is different to course content (constructive). Closer analysis of 'mostly constructive' contributions identified that within these posts participants are contributing new ideas in the form of language, contexts, pedagogy, and resources, however, they were also building on the language and key concepts used in the course materials. This behaviour is favourable, as it suggests that teachers are integrating course knowledge and building on it with their own professional knowledge and experience. This constructive behaviour aligns with the intentions of designing the MOOC with teachers as active co-creators of course knowledge. The desire is that participants bring their professional expertise and contextualise course content to suit their own contexts; we can see this behaviour reflected through the constructive examples reviewed in this study. In the context of community-centric MOOCs, this behaviour is of value to the MOOC community, as other participants can benefit from new ideas, knowledge and experiences shared by others within the shared community of practice.

Our analysis of types of knowledge contributions (active and constructive) across various MOOC units reveals that there are also variations in the types of knowledge-sharing behaviours between units, with some having a larger portion of constructive behaviour. Our analysis of the posts, within our understanding of the context of activities proposed, suggests that participants' contributions and the extent that they contribute new knowledge may be influenced by the types of activities proposed in the PD. Such findings have implications and opportunities for PD MOOC design and future research in terms of exploring how activities can more actively promote constructive or active behaviours. Activities that invite participants to share lesson ideas, experiences and resources that connect with key course concepts and language, but also with their classroom practice, resulted in contribution results that depict a greater number of these posts as either overlapping (active) or different (constructive) to course materials.

We summarise our main contributions below;

- Technical

The use of neural word embeddings over typical machine learning models using bag-of-words as features or relying on other linguistic features (e.g. LIWC (Pennebaker, Francis, & Booth, 2001), Coh-metrix (McNamara, Graesser, McCarthy, & Cai, 2014)) is promising as our model is fully automated and unsupervised in nature. Therefore, our model does not require manually annotated data, relieving the labour-intensive task of the qualitative content coding.

- Evaluation approach for community-centric MOOCs

Informed by a theoretical framework, our work proposes a novel approach to evaluate cognitive engagement within a community-centric teacher PD MOOC by measuring the association between discussion contributions and course materials.

- MOOC pedagogy

We identify that constructive behaviours in a community-centric MOOC model involve the application of discipline language from the course material as integrated with new ideas, such as contextualised examples, pedagogy and resources. The inclusion of questions and activities that foster constructive behaviours can engage teachers in PD and as active participants in the co-creation of professional knowledge and practice.

We summarise several implications for researchers and practitioners below;

- Replication

Our language model is available upon request for replication.

- Implications for practice

1. Analysing MOOC community contributions in terms of active and constructive engagement has value in terms of identifying what course content participants are engaging with as well as types of new knowledge participants are bringing to the community.
2. MOOC activities that encourage participants to reflect on their experiences, implementation of classroom activities and to generate new questions of inquiry, resource suggestions and lesson ideas can support constructive behaviour.

This paper measures teacher cognitive engagement within a single case, a Digital Technologies MOOC. Future research could seek to determine if similar results are discovered across other teacher PD MOOCs, typical community-centric MOOC models, within Digital Technologies and for other subject areas.

Similar to other empirical studies, there are several limitations in our study. When obtaining the similarity score between posts and materials, we calculate an aggregate score per participant. It is likely that a participant's behaviour varies between posts even within the same lesson. Moreover, participants could demonstrate multiple cognitive behaviours within the same posts. Since the modes in the ICAP framework is not rigid, our aggregate score to categorise participants as constructive or active is suitable only to understand patterns about professional growth. However, these measures might mislead when categorising individual's discussion behaviour. Alternatively, a weighting scheme can be implemented by allocating more weights to higher order thinking behaviours (i.e. constructive) over others to categorise ‘individual’ teachers.

Due to our incorporation of the participants who completed the exit survey, our sample ignores the teachers who are progressing in the course. This may lead to some information loss during the analysis.

Another limitation of our model is that we do not take social learning or demographics into account. In other words, teachers' discussion behaviour can be influenced by peers or their individualities. It is likely that teachers learn from the community, adapt to community norms, which is an intrinsic aspect of the community-centric model.

Our research identifies that constructive contributions need to be explored within the context of the question being asked. For example, across our MOOC, teachers can engage in self-directed learning through the tasks they choose to complete (e.g. to share a personal experience, a lesson idea or to create a resource). Future research could explore how different types of MOOC activities that are proposed influence how participants contribute to a MOOC community and which types of activities active and constructive behaviours.

8. Conclusion

Our work presents a novel automated approach to identify cognitive engagement of discussions within community-centric models and the value of analysing types of cognitive engagement for understanding how participants engage in professional learning. Using 7454 posts extracted from 1045 participants, we explore whether participants contribute new knowledge to a PD MOOC by moving further away from the course materials and shown examples in their postings and the extent to which teachers bring their own professional experiences, knowledge and ideas to the MOOC community. Our findings suggest that, in general, teachers construct new knowledge by integrating and contributing ideas about discipline, resources, and pedagogical knowledge and that constructive behaviour may be influenced by the types of MOOC tasks participants are invited to complete. Using manual content analysis, we demonstrate that from the 67 cases considered, all but one case was identified as containing ‘constructive knowledge’, providing a solid basis for replicating our proposed methodology to analyse cognitive engagement within the community-centric MOOC models.

Statement on open data, ethics and conflicts of interest

This research was approved by the University of Adelaide Human Research Ethics Committee. Authors declare no conflicts of interest. Teacher online community is openly accessible.

Declaration of interest

None.

Acknowledgement

This work was supported by Google Australia & New Zealand.

References

Arbaugh, J. B., & Benbunan-Fich, R. (2006). An investigation of epistemological and social dimensions of teaching in online learning environments. *The Academy of Management Learning and Education, 5*(4), 435–447.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Brennan, K. (2012). *ScratchEd: Developing support for educators as designers.* Retrieved from http://web.media.mit.edu/~kbrennan/files/Brennan_ScratchEd_Meetups.pdf.

Cai, Z., Dufty, D., Graesser, A. C., Hempelmann, C. F., McCarthy, P. M., & McNamara, D. S. (2005). Using LSA to automatically identify givenness and newness of noun phrases in written discourse. *Paper presented at the annual meeting of the cognitive science society*.

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

Crossley, S., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., et al. (2015). Language to completion: Success in an educational data mining massive open online class. *Paper presented at the proceedings of the 7th international conference on educational data mining*.

Dowell, N., Brooks, C., Kovanovic, V., Joksimovic, S., & Gasevic, D. (2017). The changing patterns of MOOC discourse. *Paper presented at the proceedings of the fourth ACM conference on learning @ scale*.

Falkner, K., Vivian, R., Falkner, K., & Williams, S. (2017). Reflecting on three offerings of a community-centric MOOC for K-6 computer science teachers. *Paper presented at the Proceedings of the SIGCSE Technical Symposium on Computer Science Education*.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education model. *The Internet and Higher Education, 2*(2–3).

Hashimoto, T. B., Alvarez-Melis, D., & Jaakkola, T. S. (2015). Word, graph and manifold embedding from Markov processes. *Computing Research Repository* arXiv preprint, arXiv:1509.05808.

Huang, H. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Education Technology, 33*(1).

Joksimović, S., Dowell, N., Poquet, O., Kovanović, V., Gašević, D., Dawson, S., et al. (2018). Exploring development of social capital in a CMOOC through language and discourse. *The Internet and Higher Education, 36*, 54–64. https://doi.org/10.1016/j.iheduc.2017.09.004.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data.* New york, USA: John Wiley & Sons.

Kennedy, G., Coffrin, C., de Barba, P., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. *Paper presented at the proceedings of the fifth international conference on learning analytics and knowledge*.

Kleiman, G. M., Wolf, M. A., & Frye, D. (2013). *The digital learning transition MOOC for educators: Exploring a scalable approach to professional development.* (Retrieved from).

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., ... Gašević, D. (2018). Exploring communities of inquiry in massive open online courses. *Computers & Education, 119*, 44–58. https://doi.org/10.1016/j.compedu.2017.11.010.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., ... Gašević, D. (2019). Examining communities of inquiry in Massive Open Online Courses: The role of study strategies. *The Internet and Higher Education, 40*, 20–43. https://doi.org/10.1016/j.iheduc.2018.09.001.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028.

Laurillard, D. (2016). The educational problem that MOOCs could solve: Professional development for teachers of disadvantaged students. *Research in Learning Technology, 24*(1).

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Paper presented at the proceedings of the 31st international conference on international conference on machine learning*.

Loper, E., & Bird, S. (2002). NLTK: The Natural Language toolkit. *Paper presented at the Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics Philadelphia* (Pennsylvania).

Lutz, B., Ironside, A., Hunsu, N., Groen, C., S.,B., Adesope, O., et al. (2018). Measuring engineering students' in-class cognitive engagement: Survey development informed by contemporary educational theories. *Paper presented at the ASEE annual conference & exposition*.

McCarthy, P. M., Dufty, D., Hempelmann, C. F., Cai, Z., McNamara, D. S., & Graesser, A. C. (2012). Newness and givenness of information: Automated identification in written discourse. In M. M. Philip, & B.-D. Chutima (Eds.). *Applied Natural Language processing: Identification, investigation and resolution* (pp. 457–478). Hershey, PA, USA: IGI Global.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge, M.A: Cambridge University

Press.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Paper presented at the proceedings of the 26th international conference on neural information processing systems*.
Norusis, M. J. (2008). *SPSS 16.0 guide to data analysis* (2nd ed.). Prentice Hall.
Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001.* Mahway: Lawrence Erlbaum Associates.
Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Paper presented at the LREC 2010 Workshop on new challenges for NLP frameworks.*.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Ruey, S. (2010). A case study of constructivist instructional strategies for adult online learning. *British Journal of Education Technology, 41*(5).
Siemens, G. (2012). *MOOCs are really a platform.* ElearnSpace. Retrieved from http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/.
Thilakaratne, Menasha, Falkner, Katrina, & Atapattu, Thushari (2018). Automatic detection of cross-disciplinary knowledge associations. *Proceedings of {ACL} 2018, student research workshop* (pp. 45–61). Association for Computational Linguistics. https://www.aclweb.org/anthology/P18-3007.
Toven-Lindsey, B., Rhoads, R. A., & Lozano, J. B. (2015). Virtually unlimited classrooms: Pedagogical practices in massive open online courses. *The Internet and Higher Education, 24*, 1–12. https://doi.org/10.1016/j.iheduc.2014.07.001.
Tseng, F., & Kuo, F. (2014). A study of social participation and knowledge sharing in the teachers' online professional community of practice. *Computers & Education, 72*, 37–47. https://doi.org/10.1016/j.compedu.2013.10.005.
Vellukunnel, M., Buffum, P., Boyer, K. E., Forbes, J., Heckman, S., & Mayer-Patel, K. (2017). Deconstructing the discussion forum: Student questions and computer science learning. *Paper presented at the proceedings of the SIGCSE technical symposium on computer science education*.
Wang, Y. (2014). MOOC learner motivation and learning pattern discovery. *Paper presented at the proceedings of the 7th international conference on educational data mining*.
Wang, X., Wen, M., & Rose, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Paper presented at the proceedings of the sixth international conference on learning analytics & knowledge*.
Wang, X., Yang, D., Wen, M., Koedinger, K., & Rose, C. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. *Paper presented at the proceedings of the 8th international conference on educational data mining*.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge: Cambridge University Press.
Wise, A. F., Cui, Y., Jin, W., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education, 32*, 11–28. https://doi.org/10.1016/j.iheduc.2016.08.001.
Xing, W., & Gao, F. (2018). Exploring the relationship between online discourse and commitment in Twitter professional learning communities. *Computers & Education, 126*, 388–398. https://doi.org/10.1016/j.compedu.2018.08.010.
Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. *Paper presented at the proceedings of the NIPS data-driven education workshop*.