Contemporary Educational Technology, 2023, **15**(3), ep438
ISSN: 1309-517X (Online)

OPEN ACCESS

Research Article

Using natural language processing to analyze elementary teachers' mathematical pedagogical content knowledge in online community of practice

Jiseung Yoo [1]
0000-0002-2613-6446

Min Kyeong Kim [1*]
0000-0002-6788-9890

[1] Ewha Womans University, Seoul, SOUTH KOREA
[*] Corresponding author: mkkim@ewha.ac.kr

Citation: Yoo, J., & Kim, M. K. (2023). Using natural language processing to analyze elementary teachers' mathematical pedagogical content knowledge in online community of practice. *Contemporary Educational Technology, 15*(3), ep438. https://doi.org/10.30935/cedtech/13278

ARTICLE INFO

Received: 3 Mar 2023

Accepted: 1 May 2023

ABSTRACT

This study focuses on how teachers' pedagogical content knowledge (PCK) of mathematics may differ depending on teacher interactions in an online teacher community of practice (CoP). The study utilizes data from 26,857 posts collected from the South Korean self-generated online teacher CoP, 'Indischool'. This data was then analyzed using natural language processing techniques; specifically, text classification with word2vec, BERT, and machine learning classifiers was used. The results indicate that the texts of posts can predict the level of teacher interactions in the online CoP. BERT embedding and classifier exhibited the best performance, ultimately achieving an F1 score of .756. Moreover, topic modeling utilizing BERT embedding is used to uncover the specific PCK of teachers through high- and low-interaction posts. The results reveal that high-interaction posts with numerous likes and replies demonstrate more in-depth reflections on teaching mathematics and refined PCK. This study makes two significant contributions. First, it applies a data science framework that allows for the analysis of real data from an actual online teacher community. Secondly, it sheds light on the intricacies of knowledge management in an online teacher CoP, an area that has to this point received limited empirical attention.

Keywords: online community of practice, natural language processing, mathematics pedagogical content knowledge, knowledge management

INTRODUCTION

With the technological advancements that have been made in the last several years, online teacher communities have rapidly proliferated as a form of web-based networking for purposes ranging from teacher professional development to bonding and trust (Greenhow et al., 2020; Lantz-Andersson et al., 2018; Macià & Garcia, 2016). Through the active participation of teachers and the development of platforms such as social media, these online communities, which provide a way for teachers to connect beyond the constraints of traditional face-to-face communities, have come to develop their own distinct roles and functions. As a result, there has been a wide range of research aiming to understand these developments by examining aspects of teacher learning in communities (Booth, 2012; Rehm & Notten, 2016; Tseng & Kuo, 2014), their relationship to teaching and learning practices (Goodyear et al., 2014; Hu et al., 2018), emotional ties and support (Davis, 2013; Trust, 2015), and networks in online teacher communities (Gao et al., 2017; Xie & Luo, 2019).

Online teacher communities share a unique characteristic in that most of their users are working teachers. Online experiences naturally influence teachers' professional practices in their classrooms, and their in-class activities also affect online activities. Informal learning experiences allow participants in online communities

Copyright © 2023 by authors; licensee CEDTECH by Bastas, CY. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/).

to hone their skills as they expand their tacit knowledge with accumulated practices and discussion (Faraj et al., 2011). However, previous research has also pointed out some limitations in teacher learning within online teacher communities, such as unequal contributions in knowledge sharing, superficial exchanges without reflection, low quality of information, and the potential for "quick exchanges" of materials that lack professional evaluation and reflection (Bergviken-Rensfeldt et al., 2018; Lantz-Andersson et al., 2018; Macià & Garcia, 2016).

Online teacher learning communities, like offline teacher learning communities, should not confine participant members' expertise within the framework of technical rationality; instead, they should aim to turn teachers into reflective practitioners as opposed to technical experts (Schön, 1990). However, as shown in the previous studies mentioned above (Bergviken-Rensfeldt et al., 2018; Lantz-Andersson et al., 2018; Macià & Garcia, 2016), the participation of some teachers in online teacher communities seems to be limited to cherry-picking materials and resources on a daily basis. If so, does the online teacher community contribute to in-depth teacher learning and professional development?

Online Community of Practice

A number of studies have explored the impact of online teacher communities on teaching and learning, with most of these studies analyzing these communities through the lenses of professional learning networks (PLNs) (Krutka & Carpenter, 2016; Trust et al., 2016) and online communities of practice (CoPs) (Booth, 2012; Leung, 2022; Tseng & Kuo, 2014). While PLNs emphasize informal learning without a specific goal, online CoPs–as described by Wenger et al. (2011) have a shared understanding of the goal of organizational learning, and members develop their expertise through the sharing of both tangible and intangible practices and experiences (Luo et al., 2020). In online CoPs, teachers can be seen to constitute a group that collectively creates, shares, and applies knowledge through ongoing interaction (Barab et al., 2003; Staudt Willet & Carpenter, 2021).

The unique aspect of online CoPs is that they enable deep organizational learning through daily and voluntary participation. According to Davenport (2001), the messages, posts, comments, and files shared between members in online CoPs are digital representations of learning processes and outcomes. These stored materials themselves are a source of organizational learning. Teachers can actively participate in knowledge creation and utilization while preserving their autonomy, and they can also collaborate with other teachers to develop teaching and learning ideas and share teaching materials (Booth, 2012; Brown & Munger, 2010). Further, the inclusive and horizontal communication structure of online CoPs facilitates more active reflection and practice sharing, which leads to higher acceptance of new knowledge and change (Hu et al., 2018). Studies have focused on the type of teacher learning taking place in online CoPs, with Ryu and Lee (2017) organizing an online teacher community with the aim of improving technological pedagogical and content knowledge (TPACK) of elementary preservice teachers, and Liljekvist et al. (2021) found that teachers' collaboration in a math-related online CoP was focused on improving PCK.

On the other hand, several studies have also highlighted the challenges involved in ensuring the quality of knowledge management within online teacher communities (Brown & Munger, 2010; Lantz-Andersson et al., 2018; Rehm & Notten, 2016; Seo, 2011); the literature suggests that there are two main challenges: participation and learning patterns. The imbalance between the active engagement of a few and the passive involvement of many can limit the sustainability of online teacher communities (Brown & Munger, 2010). Moreover, the materials that are shared, although they are useful and impactful, are often consumed as finished products rather than being further improved. This "take and go" approach to consuming knowledge (Lantz-Andersson et al., 2018, p. 311) can lead to shallow organizational learning and hinder the development of the teacher community.

Knowledge Sharing in Online Communities of Practice

Online CoPs can serve as a space for social interaction and knowledge sharing within the community, thus supporting the learning and professional development of teachers within such communities. Organizational science research has indicated that knowledge management within online communities relies on the reciprocal sharing of information and practices among individuals who have strong interrelationships within networks (Nonaka & Takeuchi, 2007; Hansen, 1999). This perspective aligns with Wenger et al.'s (2011) CoP, which suggests that learning occurs through collaborative activities and shared practices within a community of individuals who share a common interest.

The dynamics and outcomes of teacher learning can be affected by cumulative differences in interactions within online CoPs, which can lead to variations in the availability and utilization of knowledge (Burkink, 2002; Fesenmaier & Contractor, 2009). The patterns and volumes of interaction among members shape the structure of the network and create hubs that have a high degree of connectivity and are central to the flow of information. The knowledge that is located at both the hubs and the periphery of the network can change over time, and it is influenced by the degree of connectivity among members as well as their knowledge-sharing behaviors. Jin and Wang (2022) stated that hubs containing high-energy level nodes can be found in the upper layers of the knowledge network whereas those with low-energy level nodes are situated in the lower layers.

Moreover, active engagement and interaction among members can lead to more sophisticated and evolved knowledge within a network by allowing an actor to enter the value-creating processes of other parties, support them, and benefit from them (Barker, 2015). Therefore, networks with more active and frequent interactions among members are expected to generate deeper and higher levels of knowledge due to the increased opportunities for knowledge sharing and exchange.

Previous studies have used various indicators to measure the interactions that occur during teacher learning in online CoPs. For example, Hu et al. (2018) focused on the process by which teachers select and curate materials required for curriculum or teaching and learning operations as a way of acquiring knowledge. Dibie and Sumner (2016) attempted to capture the interactions in the network by classifying knowledge utilization behaviors into groups such as conversation frequency, selection, click, and saving.

Teachers' Pedagogical Content Knowledge

The formation, evolution, and transfer of knowledge in online CoPs are facilitated by participant engagement, and this study uses the concept of pedagogical content knowledge (PCK) to describe the outcome of teacher learning in online CoPs. PCK, as introduced by Shulman (1987), is a type of teacher knowledge that extends beyond the subject matter itself by including representations and practical strategies. It is considered a central aspect of teacher knowledge and has been further refined and defined by various scholars since it was first introduced (Ball & Bass, 2003; Ball et al., 2008; Grossman, 1989). One of the most influential redefinitions of PCK in the context of mathematics education is the concept of mathematical knowledge for teaching (MKT) or content knowledge for teaching mathematics (CKTM) (Ball et al., 2008; Hill et al., 2007), which encompasses both content knowledge (CK) and PCK and refers to the mathematical knowledge that teachers need to effectively teach the subject.

Previous models of PCK have identified key elements of effective teaching and integrated these components into the practice of mathematics instruction. PCK encompasses both subject matter knowledge and the practical strategies used to convey that knowledge to students. It represents the intentional development of a teacher's understanding of what and how to teach mathematical concepts in a specific context. Hill et al. (2007) proposed a comprehensive framework for analyzing mathematics and PCK that reorganizes the PCK components that have been previously identified into five categories: common content knowledge, specialized content knowledge, knowledge of content and students, knowledge of content and teaching, and knowledge of assessment.

Previous studies examining online teacher communities have focused on how teachers learn and develop professionally, with a particular focus on their PCK through sharing within these communities. Lee et al. (2020) conducted a review of the literature on online teacher communities and discovered that members frequently utilized these communities to gather and distribute information related to teaching, with a particular emphasis on curricular and technological resources. Similarly, Liljekvist et al. (2015) and van Bommel et al. (2020) noted that more than half of the contents shared in teachers' Facebook groups was related to professional knowledge and PCK. Leung (2022) used natural language processing (NLP) to analyze the specific contents of teachers' practical knowledge in online CoPs and found that teachers shared and co-created teaching ideas, activities, and tips specific to the subject area, which are findings that align with Shulman's (1987) concept of PCK.

Table 1. Dataset

Variable	Data type	Variable	Data type
Date	Integer	Comment	Object (text)
Title of post	Object (text)	# Likes	Integer
Content of post	Object (text)	# Replies	Integer

Context of the Case Study

The case study of 'Indischool' selected for this research is a web-based community of Korean elementary school teachers that operates as a non-profit and is supported by the voluntary participation of its members. Since 2001, 'Indischool' has been recognized as a model of collaborative learning and knowledge sharing that breaks away from traditional teaching cultures of individualism and non-interference with the aim of achieving improved curriculum organization and teaching practices (Hur & Hara., 2007; Kim, 2008; Seo, 2011). With a membership of 149,071 elementary school teachers as of January 2022, the community is widely used by the vast majority of all such teachers in Korea, as the total number of elementary school teachers in the country is 195,037. The majority of elementary school teachers with MOE certification are members, and they have shared an average of 80,000 educational materials, 50,000 posts, and 67,000 comments each year since 2015 (Indischool, 2022).

Previous studies have evaluated the case of 'Indischool' as a model of collaborative learning and shed light on its characteristics and limitations as a CoP (Hur & Hara, 2007; Kim, 2008; Seo, 2011). Those studies have found that teacher learning in 'Indischool' contributes to professional growth in several dimensions, including the creation and sharing of instructional materials and knowledge, reflection on practice, the formation of horizontal networks and emotional support, and connections with offline teacher communities (Hur & Hara, 2007). Moreover, Kim (2008) identified four types of shared knowledge in 'Indischool', including teaching and learning materials, tacit knowledge, life as a teacher, and the creation of a new school culture. However, 'Indischool' has also been noted to have certain limitations in promoting teacher professionalism. For example, Seo (2011) found that teachers download and consume shared materials on the morning of the school day rather than using them for preparation, while Hur and Hara (2007) reported that teachers primarily used 'Indischool' for the ease of access to and conversion of materials, rather than for sharing technical skills and tips or deeply considering learner learning.

This study aims to provide a looking glass into the implications of knowledge management in online teacher CoP. To comprehensively understand the characteristics of teachers' PCK in online CoP, it is necessary to first consider the effect of teachers' reactions, feedback, and interactions as forms of knowledge utilization of knowledge management. Using advanced machine learning techniques, this paper presents a classification model based on features from text data predicting teacher interactions in online CoP. We have also conducted a content-based analysis using BERT topic modeling and presented the difference in teacher PCK with high and low interactions. We seek to answer three research questions in this study:

RQ1. What textual features can be found in the posts of 'MathLibrary' in 'Indischool'?

RQ2. Can level of teacher interactions in online CoP be predicted by the features of the text in posts?

RQ3. How does elementary teachers' mathematical PCK differ in posts of online CoP with high and low interactions?

METHODS

Data Collection

This research extracted a significant volume of data from an online CoP, 'Indischool', which is the largest self-generated teacher community in South Korea (**Table 1**). 'Indischool' is a platform that offers several categories by subjects, such as literacy, mathematics, and science, as well as topics such as classroom management, policy & politics, and Q&A. The subject-specific spaces in 'Indischool'–referred to as the 'library'–allow it to serve as a platform on which teachers can share their tacit knowledge and teaching materials related to their curriculum plans, rather than a forum for free communication on general topics. We collected data from the 'MathLibrary' since it is considered a valuable source of information on elementary school

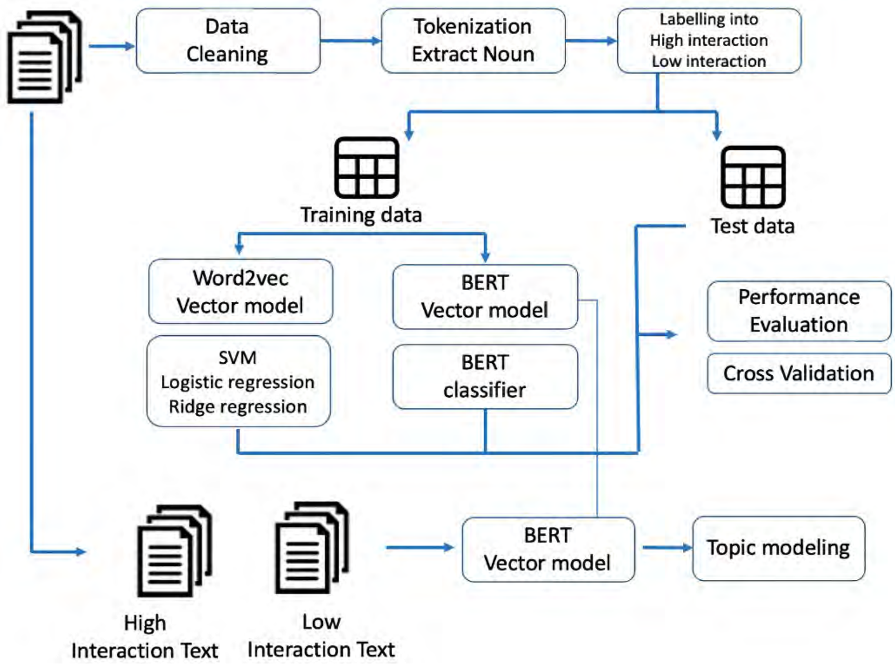

Figure 1. Data analysis process (Source: Authors)

teachers' mathematical PCK. As of May 2022, 26,857 posts and 50,035 replies had been uploaded in 'MathLibrary' on 'Indischool', and the posts were collected by Python. The dataset contains the titles of the posts, complete texts of the posts and comments, number of replies, number of likes, number of attached files, and post time and date. Each post's title and full texts are combined to create a corpus dataset.

Figure 1 describes the research framework. The proposed framework is composed of three parts and eight steps. To answer research question 1, we used term frequency-inverse document frequency (TF-IDF) analysis to highlight significant words. TF-IDF values were calculated to determine the most crucial words in the posts of 'MathLibrary' by examining the words with the highest TF-IDF values for each document. The significant words were then analyzed to determine the commonly found textual elements in the posts of 'MathLibrary'. This analysis is expected to provide insights into the vocabulary utilized and the information shared in these posts.

To answer RQ2, we pre-trained a supervised model using text data and a label as target data. The binary label was derived from the interaction score, which indicates how much teachers liked and reacted to the posts of 'MathLibrary'. Before training, we used Word2Vec and bidirectional encoder representations from transformers (BERT) to prepare a pre-trained embedding for each document. The embedded documents and labels were split into training, validation, and test sets. Word2Vec embedded data was passed to support vector machines (SVM) and ridge regression, and BERT embedding was combined with a BERT classifier to train the classification model.

To answer RQ3, this research used topic modeling utilizing BERT to identify different topics as teacher knowledge through high and low interactions. As a classical topic modeling framework, latent Dirichlet allocation (LDA) model was unable to train the weight and contexts of text because of the limitations of bag of words (BoW) model. Moreover, BERT embedding, and classifier attained the highest performance in this study, so topic modeling using BERT was utilized to determine what latent topics existed in each dataset with high and low interaction. The difference in extracted topics from each set demonstrated how teacher knowledge in online CoP could vary depending on the extent of their interaction and engagement

Labeling Techniques

In this study, the values of 'number of likes' and 'number of replies' were used to generate binary labels. Based on the descriptive statistics of the number of replies and number of likes, the interaction score was derived from the following equation scales. The binary label for each case in 'Indischool' dataset was calculated using the third quartile of the interaction score as shown below. The choice of the third quartile as the threshold was based on the results of multiple simulation tests, which showed that the precision and accuracy of the classification model were highest when using the third quartile as a reference point. For each post in the dataset, the binary label was allocated to its text.

Interaction score=($\frac{X_{num\ of\ repies} \cdot X_{num\ of\ likes}}{X_{num\ of\ repies} + X_{num\ of\ likes}}$), where label is 1 (positive), high interaction, if interaction score ≥3rd quartile and it is 0 (negative), low interaction, if interaction score <3rd quartile.

Pre-Processing

Before the model training, we performed pre-processing to clean the raw text data. Using pyKoSpacing and pyHanspell, which are packages to pre-process Korean, punctuations and errors were removed. Moreover, the package 'customized KoNLPy' converted text into root forms and extracted nouns. To capture educational terms, such as 'teachers' books' and 'mathematical practices', the user dictionary was set before extracting nouns. All the common Korean stop words were removed. To obtain more accurate results, general terms such as 'mathematics', 'teacher', 'grade', and 'class' were also removed. This study trialed with two types of models: machine learning classification with word embedding and transformer-based classification.

Word2Vec

Word embedding is a language representation model used for mapping words to vectors in vector spaces. It represents words in vector space with several dimensions (Mikolov et al., 2013). Word2Vec is one of the most popular methods for constructing word embedding. It can capture the meaning of a word in a document, semantic similarity, and the relations of words with other words based on embedded vectors. We used Word2Vec from the Gensim package with a vector size of 300 and a skip-gram model that predicts the surrounding words. This study averaged word-embedded vectors in each post to obtain a single fixed-size vector that represents the whole text. By computing all word vectors into document-level vectors, dimensionally reduced vectors presented the semantic features of the posts, and these can be used as input to the text classification model.

BERT

Built on the revolutionary transformer architecture, BERT has caused a stir in NLP by achieving state-of-the-art results across various tasks. BERT is a multi-layer transformer-based encoder that has been pre-trained in both directions for NLP applications. BERT is intended to learn the context of a sentence as it is being trained through self-supervised learning with the aid of a masked language model. KoBERT (the Korean BERT pre-trained case) (SKT Brain, 2019) was released after the BERT-based-multilingual-case (BERT-M-case) proved insufficient for the Korean NLP task. KoBERT was pre-trained exclusively on the Korean corpus. To improve KoBERT, 'Berttokenization' and 'BertForSequenceClassification' were used in this study. The fine-tuning process involved using the proposed model on a text dataset with a batch size of 16, along with training for three epochs with a learning rate of 0.00001. The classification accuracy of the BERT-based model was further improved with the use of the proposed loss function, which is based on AdamW optimization.

Support Vector Machine and Ridge Regression

This study performed SVM and ridge regression using Sckit-learn in python to fit text classification models and predict teacher interactions. SVM classification is an algorithm that can be used to solve two-class problems. This machine learning algorithm uses support vectors that display the training points of the decision function in n-dimensional space. The vectors are designed in space such that they can be separated linearly with the maximum margin. A hyperplane must be chosen such that it is as far away from the margin of both classes as possible. In text classification, documents with labels are used to train the SVM via Scikit-learn module in Python. This algorithm performs well on large data sets.

Ridge regression classifier (RRC) is an adaptation of the ridge regression technique, which is widely used in regression problems. This classification algorithm is as simple as converting the response variables of regression models with ±1 and categorical classes to represent class label assignments. However, it adds a penalty term to the cost function, which reduces complexity, thus preventing overfitting and elevating generalization. RRC also benefits from the closed-form solution of the ridge regression method. Classification models from Scikit-learn using ridge regression rely on iterative optimization, so they typically compute faster.

Table 2. Top-30 TF-IDF words

No	Keyword	TF-IDF	No	Keyword	TF-IDF	No	Keyword	TF-IDF
1	Resource	760.199	11	Decimal	411.371	21	File	257.591
2	Problem	752.039	12	Revision	358.923	22	Area	234.120
3	Video	685.343	13	Game	352.647	23	Review	217.343
4	Worksheet	624.121	14	Assessment	350.045	24	Graph	200.642
5	Fraction	582.773	15	Subtraction	315.470	25	Calculation	194.790
6	Online	486.121	16	Addition	309.450	26	YouTube	193.402
7	Division	468.723	17	Workbook	306.544	27	Activities	192.309
8	Share	460.423	18	Demand	301.676	28	Quiz	184.247
9	Textbooks	426.914	19	Wrap-up	295.039	29	Puzzle	184.013
10	Multiplication	417.160	20	Storytelling	284.891	30	Ratio	178.106

Table 3. Classification model score

Model	Accuracy	F1 (micro)
TF-IDF + SVM	.720	.674
Average Word2vec + ridge regression	.689	.716
Average Word2vec + SVM	.706	.752
BERT classification (KoBERT)	.772	.756

Topic Modeling With BERT

To uncover latent topics within texts, this study employed topic modelling with BERT using Umap, k-means, and c-TF-IDF. The contextual meaning of the texts was calculated, which was then used to generate clusters representing different topics. Based on BERTopic algorithm (Grootendorst, 2022), pre-trained sentence transformer model was used to determine semantic document similarity through dimensionality reduction with Umap technique and clustering of sentence embeddings through K-Means. The optimal number of topics was determined by using the elbow method and evaluating the Silhouette score. After clustering, c-TF-IDF was used to calculate the relevance of each word based on its frequency across the documents and classes.

RESULTS AND FINDINGS

This section validates the performance of the text classification model using embedding and the ML classifier mentioned in the preceding method section. BERT classifier achieved a high accuracy score of 77.2% when categorizing postings in 'Indischool' based on the level of teacher interactions. Content-based analysis using topic modelling was conducted to discover how PCK differed between the two sets of texts from posts.

RQ1: Textual Features in Posts on MathLibrary in Indischool

Table 2 shows the top 30 keyword features extracted from 'Indischool' data set. We can see that sharing educational resources is one of the characteristics of online CoPs. Many words rank high in the keyword features list, such as "resource", "quiz", "video", "worksheet", and "file". Further, texts in online CoPs often include relatively mathematical concepts, such as "fraction", "division", and "graph", as characters to appear in the data. Those teaching strategies-related words make up a proportion of the top 30 keyword features, and some of them are repeated more than once. Additionally, the repertoire of teaching ("wrap-up", "review") and some specific words related to assessment are utilized to create a sense of teacher PCK.

RQ2: Text Classification Model via Interaction Level in Mathlibrary

The performance of the classification model that predicted the degree of teacher interactions on 'Indischool"s mathematics community is presented in **Table 3**. The results were derived by applying several document embedding techniques and classification methods. The baseline model provided a starting point for comparing and evaluating future progress. Given the proportion of labels, the prediction accuracy baseline was 72%. The performance of the models improved in the order of average Word2vec embedding with Ridge linear regression and the SVM model. The model using BERT embedding and classifier showed the best performance, and the accuracy implies that the model could predict about 77% of teacher interactions based on the text of posts in the math community of 'Indischool'. AUC (area under ROC curve) score of 0.79 on **Figure 2** means that BERT classifier model can distinguish between high and low interaction text reasonably well.

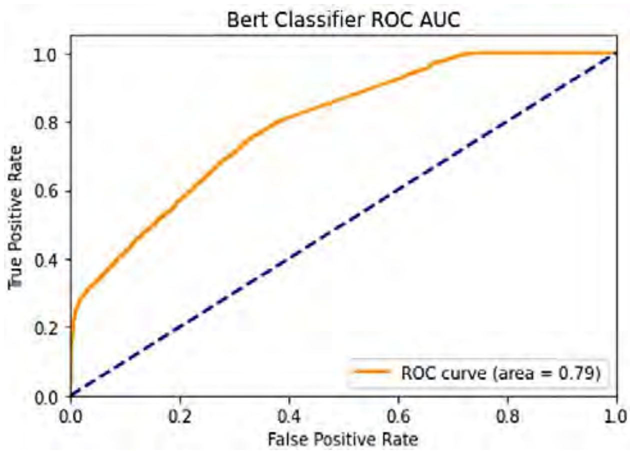

Figure 2. BERT classification model ROC curve (Source: Authors)

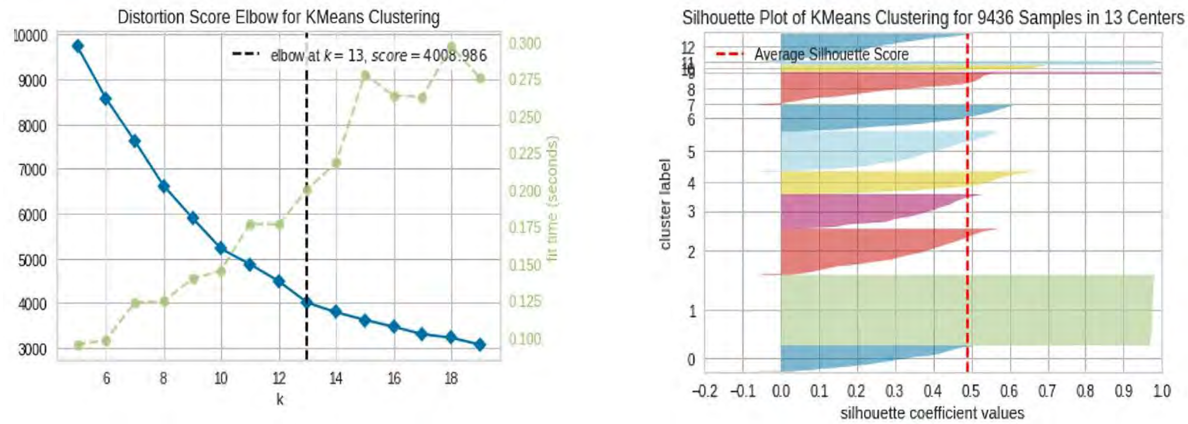

Figure 3. Elbow method (left) & Silhouette score (right) of high interaction dataset (Source: Authors)

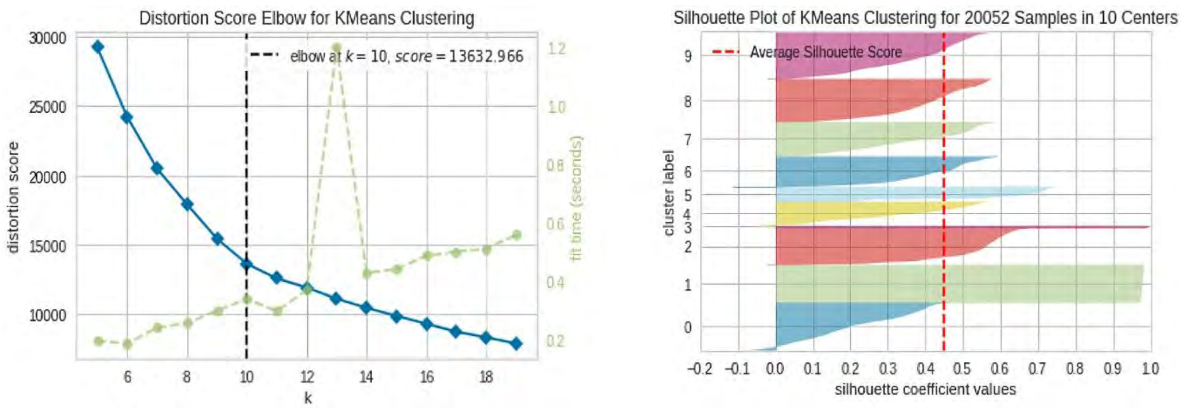

Figure 4. Elbow method (left) & Silhouette score (right) of low interaction dataset (Source: Authors)

RQ3: Elementary Teachers' Mathematical PCK in MathLibrary by Interaction Level

The text-interaction classification model of BERT classified the whole data set into two categories, posts with high interaction and posts with low interaction. Topic modelling with BERT determined which teacher knowledge was formed as a topic in each category. This study used elbow and Silhouette methods to accurately determine the optimal number of topics (**Figure 3** and **Figure 4**), and the optimal number of topics based on those methods was 13 and 10 for each dataset. Considering both the Silhouette score and the number of posts for each topic cluster, we finalized 11 topics for high-interaction posts and nine topics for low-interaction posts.

The topics with high teacher interaction scores in online CoP are, as follows (**Table 4**): contextual open-ended tasks with a focus on teaching strategies aimed at developing students' problem-solving skills by

Table 4. Topics from high interaction posts

No	Topics	Top c-TF-IDF keywords	%
0	Re-customize other teachers' curriculum	Standards, revision, & version	8.19
1	Alternative concrete & manipulative materials	Materials, kit, & textbook	11.49
2	Misconceptions about a concept	Misconception, ratio, & difficulties	8.57
3	Contextual open-ended tasks	Creativity, puzzle, & activities	11.79
4	Alternative-assessment strategies: Project & writing	Writing, debate, & letter	6.64
6	How to use ICT: Geometry	VR/AR, cube, & width	9.91
7	History of mathematics & iconic numbers	π, 0, rule, & induction	7.81
8	Data analysis: From design to analysis	Graph, research, class, & trend	7.85
9	Idea for STEM lessons	Interdisciplinary, science, & life,	8.31
10	Building & sharing repertoire by units	Introduce, practice, & routine	7.45
11	Mathematical learning difficulties	Operation, basics, & slow learner	10.76

Note. Exclude too small clusters topic no.5 (0.94%) & topic no.12 (0.29%) & Total 100%

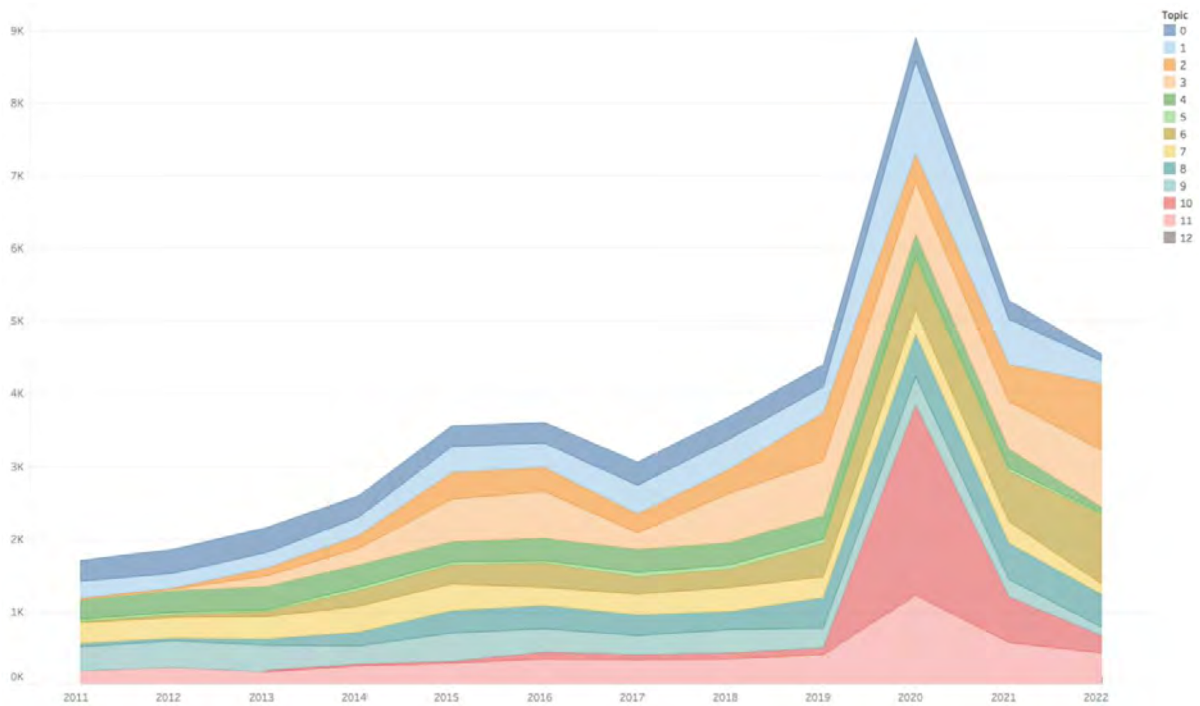

Figure 5. High interaction text topic dominance per year (Source: Authors)

sharing projects and class materials account for the largest single topic with 11.79%. Although the conversation around alternative assessment strategies as a tool for evaluating student achievement has been active since 2013 due to the Korean government's policy of process-oriented assessment, it accounts for the smallest proportion among the topics, and there is no mention of formative or diagnostic assessment in the topic. There are four clusters dedicated to the sharing and development of teacher strategies and experiences for supporting the learning of common content knowledge and specialized content knowledge: How to use ICT, history of mathematics and iconic numbers, data analysis, and ideas for STEM lessons.

Moreover, as depicted in **Figure 5**, the topic of "building and sharing repertoire by units" exhibited a significant increase during coronavirus pandemic, with a large number of posts focused on teachers creating their own types of lesson plans and routines for both in-school and remote instruction in blended-learning systems. Discussions of basic learning and math difficulties continue to be active and have increased in post-COVID-19 era.

A breakdown of topics with low teacher interaction scores is presented in **Table 5**, and a time series analysis of topic shares is shown in **Figure 6**. Materials for drawing and art tasks, with ready-to-use resources, account for the highest percentage (14.34%) of posts with low teacher interaction. Topics such as arithmetic operation practice, storytelling math resources, and measuring: kinesthetic activities are characterized by teacher collaboration focused on file sharing, rather than extensive conversation. The same can be said for

Table 5. Topics from low interaction posts

No	Topics	Top c-TF-IDF keywords	%
0	Concrete & manipulative materials	Stickers, files, & cards	9.56
1	Arithmetic operation practices	Operation, drills, & practice	11.35
2	Re-customize other teachers' curriculum	Worksheet, version, & template	12.09
3	All about EXAM	Exam, quiz, & grading	12.72
4	Measuring: Kinesthetic activities	Length, game, & activities	12.62
6	Multimedia resources: Videos, pictures, etc.	Platform, ppt, & YouTube	9.10
7	Storytelling math	Harry Potter, online, & movie	11.18
8	Drawing & art tasks	Game, drawing, & interests	14.34
9	Puzzle and quiz: Warm up & wrap up	Puzzle, link, & game	6.13

Note. Exclude too small clusters cluster no.5 (0.92%) & Total 100%

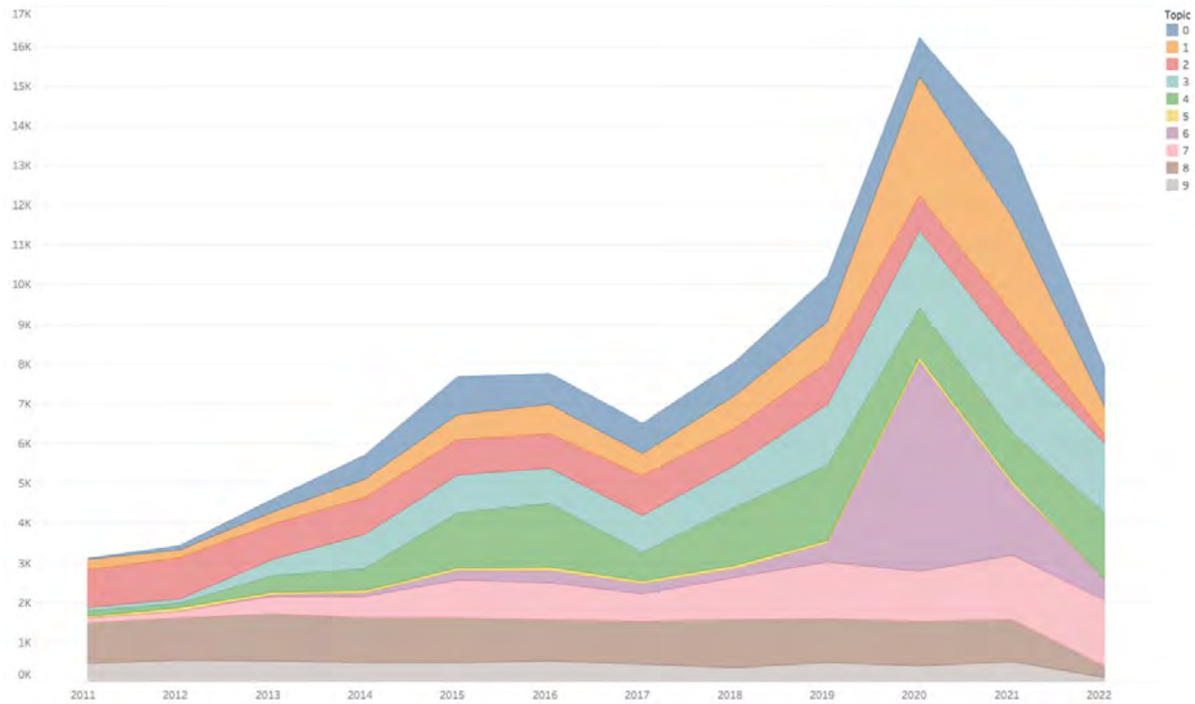

Figure 6. Low interaction text topic dominance per year (Source: Authors)

formalized paper assessments and exams topic, where low levels of interaction are observed among teachers as they share resources, and where teachers do not engage in meaningful discussion. Multimedia resources, including videos and pictures, are not as commonly used in teaching and learning math compared to other subjects such as social studies and science.

As depicted in **Figure 6**, topic related to multimedia resources has been low. Despite this, teachers have been actively sharing and developing their own multimedia resources, particularly in response to requirements for remote learning during COVID-19 pandemic.

DISCUSSION

This study analyzed the large-scale data of 'MathLibrary' section of the online teacher CoP, 'Indischool', using NLP techniques. In this section of 'Indischool', teachers actively engaged with each other by posting likes or comments as well as downloading and using attached resources, as they discussed their ideas and experiences teaching mathematics. The results showed that text classification models, particularly those using BERT embeddings and classifiers, were able to accurately predict the level of teacher interaction with 77% accuracy. This implies that the semantic characteristics of the text in each post are related to the level of interaction the posts are receiving. The second analysis applied topic modeling using BERT to posts classified by interaction level and found that posts with high and low interaction had different topics.

Teachers discussed a range of topics related to mathematics and curriculum implementation in the online teacher CoPs. Through topic modeling, it was revealed that teachers shared their knowledge and developed expertise in areas such as assessment, teaching strategies, and the unique characteristics of math subjects and techniques. This is consistent with the findings of Faraj et al. (2015), who stated that both explicit and tacit knowledge can be shared and developed among participants in online communities, thus providing numerous opportunities for expertise.

The eleven topics from posts with high interaction revealed a new, innovative approach as well as a deeper connection to real-life contexts. Teachers are concerned with designing inquiry-focused curriculums rather than simply focusing on mathematical knowledge. As demonstrated by Liljekvist et al. (2021), the participants of online CoPs can share and enrich various facets of PCK through discussion. The discussions in 'Indischool' encompass a broad range of subjects related to teaching mathematics that extend beyond practical methods and effective materials alone. They delve into more scholarly topics, such as misconceptions involving teaching and alternative assessment methods. Moreover, teachers discussed new ideas to address external factors and resolve those unexpected factors in teaching math while considering their students' circumstances.

These results suggest that teacher interactions play a crucial role in co-creating knowledge. While it may be challenging to identify the precise moment of PCK evolution in the conversation, large-scale datasets and their features do show differences in teachers' knowledge about what and how to teach. This implies that feedback and reaction can enhance CoP and make teacher collaboration more visible, thus extending the reach and sustainability of organizational learning (Davenport, 2001), ultimately creating a repository of interactions that foster the reflective practice and in-depth personal learning (Hara & Kling, 2002).

The dataset of the low interaction group consists of posts that did not receive many comments or likes from teachers. Most topics in this dataset are either rich teaching resources or information required for daily classroom activities. PCK in these topics focuses on classroom practices–such as one-off lesson ideas, materials, and tips–rather than in-depth knowledge about systemic strategies for math curricula or reflections on improving teaching mathematics (van Bommel et al., 2020). This finding is similar to those of recent empirical studies (Peng, 2019; Zöller et al., 2020) examining open-source software (OSS) projects, which discovered that weak ties in networks are well-suited for exchanging new practical tactics and promoting the continuous importation of external resources. While the stimulation of external heterogeneous knowledge may be less effective than knowledge sharing and co-creation in a small group of leaders, it is still a source of emerging knowledge and innovation. In the context of online teacher CoPs, the influx of resources from outside the network, along with experience from various backgrounds and contexts, increases the likelihood that teachers will generate knowledge to solve problems and improve professional development in the long run (Greenhow et al., 2020).

Here, it is necessary to consider that the quality of networked knowledge in online CoPs is dependent on participants' knowledge sharing activities and learning experiences. In other words, small differences in these activities can lead to differences in accumulated interactions and the knowledge structure of online CoPs. Initially marginalized knowledge can form a knowledge stack through mass tagging, linking, and downloading behavior, thus causing knowledge to rise from lower levels to the core level (Bozkurt et al., 2020; Hillman et al., 2021). As such, it is important to explore sophisticated methods for capturing diverse forms of engagement behaviors and incentivizing quality learning activities from them. It is possible for managers' strategies to facilitate information systems (Kim et al., 2022; Shahbaznezhad et al., 2021) for teachers' subgroups in online communities, which can elevate stronger ties (Matranga & Silverman, 2020), thereby aiding in the transformation of individuals from passive lurkers to active contributors. Further research can explore more specific internal and external factors that enhance teacher interaction and collaborative knowledge growth.

Our study has limitations that must be considered, which provide directions for further research. As a case study of a single online community, the results and insights obtained may not be representative of other knowledge-sharing online teacher CoPs. There is a need for future studies to determine the generalizability of our findings to other similar communities. Moreover, we cannot determine the optimal percentage of teacher likes and replies to posts that would classify participation and contribution within a community (the percentage in our study is the 3rd quartile). Further, the findings herein, which focused on the texts and interactions among community members, do not provide a comprehensive understanding of how this knowledge is applied in teachers' teaching practice or the level of PCK possessed by teachers. More research is needed to explore how teachers' PCK, know-how, and tips, which are discussed in the textual content of online communities, are implemented in the classroom using qualitative research methods. Integrating NLP techniques with latent profile analysis to examine the correlation between teachers' engagement behavior and their level of PCK could also represent a promising approach.

Author contributions: JY: conceived & designed analysis, collected data, contributed data & analysis tools, performed analysis, & wrote paper & **MKM:** conceived & designed analysis, contributed data & analysis tools, & wrote paper. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: The authors declared that since the study used pre-existing data that is openly accessible and did not necessitate approval from an ethics committee, ethical review and approval were waived.

Declaration of interest: Authors declare no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request. The code supporting this research is available from the following source: https://github.com/JiseungYoo/COP-PCK

REFERENCES

Ball, D. L., & Bass, H. (2003). Toward a practice-based theory of mathematical knowledge for teaching. In B. Davis, & E. Simmt (Eds.), *Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group* (pp. 3-14).

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407. https://doi.org/10.1177/0022487108324554

Barab, S. A., MaKinster, J. G., & Scheckler, R. (2003). Designing system dualities: Characterizing a web-supported professional development community. *The Information Society, 19*(3), 237-256. https://doi.org/10.1080/01972240309466

Barker, R. (2015). Management of knowledge creation and sharing to create virtual knowledge-sharing communities: A tracking study. *Journal of Knowledge Management, 19*, 334-350. https://doi.org/10.1108/JKM-06-2014-0229

Bergviken-Rensfeldt, A., Hillman, T., & Selwyn, N. (2018). Teachers 'liking' their work? Exploring the realities of teacher Facebook groups. *British Educational Research Journal, 44*(2), 230-250. https://doi.org/10.1002/berj.3325

Booth, S. E. (2012). Cultivating knowledge sharing and trust in online communities for educators. *Journal of Educational Computing Research, 47*(1), 1-31. https://doi.org/10.2190/EC.47.1.a

Bozkurt, A., Koutropoulos, A., Singh, L., & Honeychurch, S. (2020). On lurking: Multiple perspectives on lurking within an educational community. *The Internet and Higher Education, 44*, 100709. https://doi.org/10.1016/j.iheduc.2019.100709

Brown, R., & Munger, K. (2010). Learning together in cyberspace: Collaborative dialogue in a virtual network of educators. *Journal of Technology and Teacher Education, 18*(4), 541-571.

Burkink, T. (2002). Cooperative and voluntary wholesale groups: Channel coordination and interfirm knowledge transfer. *Supply Chain Management: An International Journal, 7*(2), 60-70. https://doi.org/10.1108/13598540210425812

Davenport, E. (2001). Knowledge management issues for online organizations: 'Communities of practice' as an exploratory framework. *Journal of Documentation, 57*(1), 61-75. https://doi.org/10.1108/EUM0000000007077

Davis, T. (2013). Building and using a personal/professional learning network with social media. *The Journal of Research in Business Education, 55*(1), 1-13.

Dibie, O., & Sumner, T. (2016). Using weak ties to understand the resource usage and sharing patterns of a professional learning community. *Social Network Analysis and Mining, 6*(1), 6-27. https://doi.org/10.1007/s13278-016-0335-z

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science, 22*(5), 1224-1239. https://doi.org/10.1287/orsc.1100.0614

Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading collaboration in online communities. *MIS Quarterly, 39*(2), 393-412. https://doi.org/10.25300/MISQ/2015/39.2.06

Fesenmaier, J., & Contractor, N. (2001). The evolution of knowledge networks: An example for rural development. *Community Development, 32*(1), 160-175. https://doi.org/10.1080/15575330109489697

Gao, F., Luo, T., & Zhang, K. (2012). Tweeting for learning: A critical analysis of research on microblogging in education published in 2008-2011. *British Journal of Educational Technology, 43*(5), 783-801. https://doi.org/10.1111/j.1467-8535.2012.01357.x

Goodyear, V. A., Casey, A., & Kirk, D. (2014). Tweet me, message me, like me: Using social media to facilitate pedagogical change within an emerging community of practice. *Sport, Education and Society, 19*(7), 927-943. https://doi.org/10.1080/13573322.2013.858624

Greenhow, C., Galvin, S. M., Brandon, D. L., & Askari, E. (2020). A decade of research on K–12 teaching and teacher learning with social media: Insights on the state of the field. *Teachers College Record, 122*(6), 1-72. https://doi.org/10.1177/016146812012200602

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv, 05794*. https://doi.org/10.48550/arXiv.2203.05794

Grossman, P. L. (1989). A study in contrast: Sources of pedagogical content knowledge for secondary English. *Journal of Teacher Education, 40*(5), 24-31. https://doi.org/10.1177/002248718904000504

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly, 44*(1), 82-111. https://doi.org/10.2307/2667032

Hara, N., & Kling, R. (2002). Communities of practice with and without information technology. In E. G. Toms (Ed.), *Proceedings of the 65th American Society for Information Science and Technology Annual Meeting* (pp. 338-349). Information Today. https://doi.org/10.1002/meet.1450390137

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111-156). Information Age Publishing.

Hillman, T., Lundin, M., Rensfeldt, A. B., Lantz-Andersson, A., & Peterson, L. (2021). Moderating professional learning on social media: A balance between monitoring, facilitation and expert membership. *Computers & Education, 168*, 104191. https://doi.org/10.1016/j.compedu.2021.104191

Hu, S., Torphy, K. T., Opperman, A., Jansen, K., & Lo, Y. J. (2018). What do teachers share within socialized knowledge communities: A case of Pinterest. *Journal of Professional Capital and Community, 3*(2), 97-122. https://doi.org/10.1108/JPCC-11-2017-0025

Hur, J. W., & Hara, N. (2007). Factors cultivating sustainable online communities for K-12 teacher professional development. *Journal of Educational Computing Research, 36*(3), 245-268. https://doi.org/10.2190/37H8-7GU7-5704-K470

Indischool. (2022). Indischool statistical data. *Brunch*. https://brunch.co.kr/@indischool/70

Jin, F., & Wang, Y. (2022). The emergence of community knowledge in a complex network-based folksonomy self-organization mode. In *Proceedings* of the *11th International Conference on Software and Information Engineering* (pp. 62-65). https://doi.org/10.1145/3571513.3571524

Kim, D. H. (2008). The development process of network-based community of practice for teachers' knowledge sharing and expertise development: A case study of Indischool. *Journal of Educational Technology, 24*(2), 1-30. https://doi.org/10.17232/KSET.24.2.1

Kim, H., Hwang, S., Kwak, Y., & Choi, J. (2022). Can online community managers enhance user engagement? Evidence from anonymous social media postings. *Knowledge Management Research, 23*(2), 211-228. https://doi.org/10.15813/kmr.2022.23.2.011

Krutka, D. G., & Carpenter, J. P. (2016). Participatory learning through social media: How and why social studies educators use Twitter. *Contemporary Issues in Technology and Teacher Education, 16*(1), 38-59.

Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and Teacher Education, 75*, 302-315. https://doi.org/10.1016/j.tate.2018.07.008

Lee, D., Jung, J., Shin, S., Otternbreit-Leftwich, A., & Glazewski, K. (2020). A sociological view on designing a sustainable online community for K–12 teachers: A systematic review. *Sustainability, 12*(22), 9742. https://doi.org/10.3390/su12229742

Leung, J. (2022). An NLP approach for extracting practical knowledge from a CMS-based community of practice in e-learning. *Knowledge, 2*(2), 310-336. https://doi.org/10.3390/knowledge2020018

Liljekvist, Y. E., Randahl, A. C., van Bommel, J., & Olin-Scheller, C. (2021). Facebook for professional development: Pedagogical content knowledge in the center of teachers' online communities. *Scandinavian Journal of Educational Research, 65*(5), 723-735. https://doi.org/10.1080/00313831.2020.1754900

Liu, J., Wei, J., Liu, Y., & Jin, D. (2022). How to channel knowledge coproduction behavior in an online community: Combining machine learning and narrative analysis. *Technological Forecasting and Social Change, 183*, 121887. https://doi.org/10.1016/j.techfore.2022.121887

Luo, T., Freeman, C., & Stefaniak, J. (2020). "Like, comment, and share"–professional development through social media in higher education: A systematic review. *Educational Technology Research and Development, 68*(4), 1659-1683. https://doi.org/10.1007/s11423-020-09790-5

Macià, M., & García, I. (2016). Informal online communities and networks as a source of teacher professional development: A review. *Teaching and Teacher Education, 55*, 291-307. https://doi.org/10.1016/j.tate.2016.01.021

Matranga, A., & Silverman, J. (2020). An emerging community in online mathematics teacher professional development: An interactional perspective. *Journal of Mathematics Teacher Education, 25*, 63-89. https://doi.org/10.1007/s10857-020-09480-2

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv, 3781*. https://doi.org/10.48550/arXiv.1301.3781

Nonaka, I., & Takeuchi, H. (2007). The knowledge-creating company. *Harvard Business Review, 85*(7-8), 162.

Peng, G. (2019). Co-membership, networks ties, and knowledge flow: An empirical investigation controlling for alternative mechanisms. *Decision Support Systems, 118*, 83-90. https://doi.org/10.1016/j.dss.2019.01.005

Rehm, M., & Notten, A. (2016). Twitter as an informal learning space for teachers!? The role of social capital in Twitter conversations among teachers. *Teaching and Teacher Education*, *60,* 215-223. https://doi.org/10.1016/j.tate.2016.08.015

Ryu, K., & Lee, Y. (2017). Effects of online teacher learning community activities linked with internship course for the improvement of elementary pre-service teacher's TPACK. *The Journal of Korean Teacher Education, 34*(2), 417-437. https://doi.org/10.24211/tjkte.2017.34.2.417

Schön, D. A. (1990). *Educating the reflective practitioner*. Jossey-Bass.

Seo, K. (2011). Collaborative professional development of online teacher community. *The Journal of Korean Teacher Education, 28*(1), 133-161. https://doi.org/10.24211/tjkte.2011.28.1.133

Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The role of social media content format and platform in users' engagement behavior. *Journal of Interactive Marketing, 53*(1), 47-65. https://doi.org/10.1016/j.intmar.2020.05.001

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

SKT Brain. (2019). *Korean BERT pre-trained cased (KoBERT)*. https://github.com/SKTBrain/KoBERT

Staudt Willet, K. B., & Carpenter, J. P. (2021). A tale of two subreddits: Change and continuity in teaching-related online spaces. *British Journal of Educational Technology, 52*(2), 714-733. https://doi.org/10.1111/bjet.13051

Trust, T. (2015). Deconstructing an online community of practice: Teachers' actions in the Edmodo math subject community. *Journal of Digital Learning in Teacher Education, 31*(2), 73-81. https://doi.org/10.1080/21532974.2015.1011293

Trust, T., Krutka, D. G., & Carpenter, J. P. (2016), "Together we are better": Professional learning networks for teachers. *Computer & Education, 102*, 15-34. https://doi.org/10.1080/21532974.2016.1208506

Tseng, F. C., & Kuo, F. Y. (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

van Bommel, J., Randahl, A. C., Liljekvist, Y., & Ruthven, K. (2020). Tracing teachers' transformation of knowledge in social media. *Teaching and Teacher Education, 87*, 102958. https://doi.org/10.1016/j.tate.2019.102958

Wenger, E., Trayner, B., & De Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework*. Cambridge University Press.

Xie, Q., & Luo, T. (2019). Examining user participation and network structure via an analysis of a Twitter-supported conference backchannel. *Journal of Educational Computing Research, 57*(5), 1160-1185. https://doi.org/10.1177/0735633118791262

Zöller, N., Morgan, J. H., & Schröder, T. (2020). A topology of groups: What Github can tell us about online collaboration. *Technological Forecasting and Social Change, 161*, 120291. https://doi.org/10.1177/0735633118791262

◆❖◆