



# Investigation of Pre-service Teachers' Conceptions of the Nature of Science Based on the LDA Model

Manman Wang<sup>1</sup> · Shoubao Gao<sup>1</sup> · Weiling Gui<sup>1</sup>  · Jianqiang Ye<sup>2</sup> · Shuaishuai Mi<sup>1</sup>

Accepted: 18 February 2022 / Published online: 6 April 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

This study used the Latent Dirichlet Allocation (LDA) topic model to analyze pre-service teachers' views on the nature of science (NOS). This approach can be used to automate the classification of documents, and at the same time, the researcher does not need to deduce with a NOS framework prior to evaluation. Participants were 155 pre-service teachers studying at the Shandong Normal University in China. To gather our data, we used an open questionnaire, namely, the Views of Nature of Science Questionnaire—Form C (VNOS-C). LDA topic modeling was used to classify the document, which was divided into 12 topics. By comparing the LDA topic modeling results with the theoretical framework behind the VNOS-C questionnaire, we categorized these 12 topics into eight descriptive aspects of the NOS: The Empirical Nature of Scientific Knowledge, Observation, Inference, and Theoretical Entities in Science, Scientific Theories and Laws, The Theory-Laden Nature of Scientific Knowledge, The Social and Cultural Embeddedness of Scientific Knowledge, The Myth of The Scientific Method, The Tentative Nature of Scientific Knowledge, and The Nature of Scientific Theory. The results show that pre-service teachers usually hold naive or mixed views of the NOS. In addition, each aspect of NOS is not independent of each other but interrelated and influencing each other. In the future, more consideration can be given to the relationship between each aspect of NOS.

**Keywords** Nature of science · Latent Dirichlet allocation (LDA) topic modeling · Pre-service teachers

---

✉ Weiling Gui  
guiweiling@sdu.edu.cn

✉ Shuaishuai Mi  
621001@sdu.edu.cn

<sup>1</sup> School of Physics and Electronics, Shandong Normal University, Jinan 250358, Shandong, China

<sup>2</sup> College of Chemistry & Materials Engineering, Wenzhou University, Wenzhou 325035, Zhejiang, China

## 1 Introduction

There is a general consensus in the current education reform to develop students' view of the nature of science (NOS) (Bell et al., 2011). Teaching and learning related to the NOS can contribute to epistemic insight, that is, NOS-related pedagogies can help us recognize how and why we should understand science and the power and limitations of science (Erduran & Kaya, 2018). Research has recognized that teacher NOS views play a central role in student understanding of the NOS and that improving pre-service teachers' understanding of the NOS is an important first step in improving student understanding of the NOS (Abd-El-Khalick & Lederman, 2000; Bell et al., 2011; Osborne et al., 2003). To improve pre-service teacher views of the NOS, the first step was typically to assess pre-service teacher views of the NOS (Lederman, 1992). Such investigations can introduce teacher-trainers to the view of the NOS held by pre-service teachers and help teacher-trainers prepare their own teaching.

In previous studies (e.g., Abd-El-Khalick., 2013; Akerson et al., 2017; Ozgelen et al., 2013; Zion et al., 2020), the data processing methods used by researchers to assess pre-service teachers' view of the NOS can be divided into two kinds: using hand-coded documents and helping researchers process the data with the help of instruments. In both methods, researchers analyze terms, keywords, and phrases that contain the view of NOS, but participants' responses may involve sentences that do not contain terms, keywords, and phrases related to their view of NOS, even though these sentences may still express their view of NOS (Ozgelen et al., 2013; Zion et al., 2020). In addition, previous researchers assessed the NOS of pre-service teachers based on a NOS framework. The basic idea is that researchers need to deduce a NOS framework through research literature before assessing the pre-service teachers' view of the NOS and then assess the view of the NOS of pre-service teachers based on the NOS framework (e.g., Abd-El-Khalick., 2013; Akerson et al., 2017; Demirdöğen et al., 2016; Zion et al., 2020). In this way, there may be different results due to the different NOS frames deduced by researchers.

LDA topic modeling is an important text analysis technology (Ramage et al., 2009). LDA topic modeling method can connect words with similar meanings, distinguish the specific usage of multi-meaning words, and then automatically classify a large number of documents with high quality according to the meaning of the sentence and the related situation described (Alghamdi & Alfalqi, 2015). In other words, LDA topic modeling not only analyzes sentences containing NOS-related keywords or phrases, but also includes those that can express participants' view of the NOS but do not contain NOS-related keywords or phrases. In addition, LDA topic modeling is an unsupervised classification method (Roberts et al., 2014). This method can be used to categorize a large number of unmarked documents to get at the underlying topics in the documents (Momtazi, 2018). This means that this method does not require the researcher to deduce a NOS framework. Researchers can determine the meaning and label of each topic based on the LDA classification results (Bastani et al., 2019), which is the NOS view held by the participants. This reduces the risk of affecting the research results due to the different NOS frameworks deduced by the researcher.

## 2 Literature Review

There is widespread consensus that effective teachers are a critical factor for student learning (Hanuscin et al., 2011). Some studies (e.g., Gess-Newsome & Lederman, 1995) show that the relationship between teachers' ideas and their practice in the classroom is not a

direct or a simple linear relationship; that is, teacher views of the NOS cannot be directly transformed into classroom practice to improve student views of the NOS, and it is clear that teachers cannot effectively design and teach courses on concepts they do not understand (Abd-El-Khalick et al., 1998). In addition, Abd-El-Khalick (2013) suggests that teachers with a well-informed understanding of NOS are in a better position to build robust inquiry learning environments where they can openly draw students' attention to relevant NOS ideas. Therefore, it is vital to improve teacher understandings of the NOS. However, as highlighted above, doing so requires determining the current views of the NOS held by pre-service teachers and providing information to teacher trainers.

To date, scholars have conducted numerous investigations on pre-service teachers. For example, Lederman (2007) summarizes the various instruments and trends in the assessment of NOS. Scholars have doubted the validity of one type of such instruments—standardized and convergent paper and pencil instruments—for the evaluation of pre-service teachers' NOS views (Lederman et al., 2002) for the following reasons: (1) respondents perceive and interpret an instrument's project in a similar way to the instrument's developer, and (2) such standardized instruments usually reflect the experts/academics who developed the instruments views on the NOS, which, because subjective, are necessarily biased (Abd-El-Khalick et al., 1998). Compared with such a standardized instrument, an open questionnaire allows participants to explain their understanding of the NOS (Lederman et al., 2002). Therefore, the use of open questionnaires can more well-informed assess pre-service teacher views of the NOS. The data processing methods of researchers on open questionnaires can be divided into two kinds: one is completely by hand-coding processed, and the other is that researchers process the data with the help of instruments.

The first type involves the researcher processes the data by hand-coding. Mesci et al. (2020) assessed pre-service teachers' views of the NOS with the Views of Nature of Science Questionnaire [VNOS-270] (Lederman et al., 2002). The method used in their study is based on an existing classification proposed by Khishfe and Abd-El-Khalick (2002) to identify participants' NOS views. The researchers based their analysis on a participant's ability to express the meaning of NOS aspects using words, expressions, and phrases in response to a question and then to categorize participants into specific aspects of the NOS (e.g., tentative) on the basis of their reactions to each item and decide on their views on specific aspects of the NOS (informed, naïve, or not categorized). Other scholars (e.g., Akerson et al., 2017; Demirdöğen et al., 2016; Zion et al., 2020) who employed open questionnaires also used a similar method for data analysis.

The second type involves researchers process data with the help of instruments. Ozgen et al. (2013) administered the Views of Nature of Science Questionnaire Version B (VNOS-B) to 37 pre-service science teachers (PSTs) to assess their conceptions of the NOS. The PSTs' responses to the VNOS-B were word-processed and entered into the NVivo 8 qualitative data analysis software; subsequently, they designed a three-stage data analysis technology. First, the author assigned a phrase or word relevant to the aspect of the NOS addressed in the PSTs' statements (e.g., subject, empirical, subjective). Second, sentences containing phrases or words that reflected a view of the NOS (e.g., subjectivity) were reviewed in more detail; next, the participants' views on specific aspects of the NOS (e.g., subjectivity) were considered. Third, the opinions expressed in all the sentences were classified and then divided into naïve, mixed, and informed according to their degree of consistency with the contemporary consensus on the NOS (Lederman, 1992). Sorensen et al. (2012) used a similar method of data analysis involving written tasks (discussing questions about the NOS by writing articles) and interviews as data sources. This style of data analysis used software to qualitatively analyze the meanings of the keywords and phrases used by participants, locate the sentences corresponding to the keywords and

phrases, classify sentences according to the meaning of the sentences, and determine the participants' views of the NOS.

Scholars generally consider an open questionnaire more difficult to analyze than standardized and convergent paper and pencil instruments (Roberts et al., 2019) because it uses "manual coding." "Manual coding" means that the researcher (or the researcher uses software) specifies phrases or words related to the NOS mentioned in the responses based on the respondents' responses to the questionnaire and then classifies the views of the respondents as naive, mixed, or informed according to their agreement with the opinions expressing consensus views. "Manual coding" presents added difficulty because it typically requires researchers to divide the questionnaire into several dimensions according to their theoretical expectations or exemplary past studies and, moreover, also requires several human coders, whose work is compared (Artstein & Poesio, 2008; Robertset al., 2019). The biggest limitation of using the "manual coding" method of data processing is that the basic units of analysis are terms, keywords, and phrases that contain views of the NOS, and then this method will ignore some information (Akerson et al., 2017; Demirdöğen et al., 2016; Mesci et al., 2020; Ozgelen et al., 2013; Zion et al., 2020). More specifically, participant responses may involve sentences that do not contain terms, keywords, and phrases related to their views of the NOS even though such sentences may still express their views of the NOS. Therefore, using the above two kinds methods cannot well-informed determine participant views of the NOS. In addition, whether researchers analyze data manually or by means of instruments, researchers need to process and analyze data based on a NOS framework. For example, Mesci et al. (2020) conduct data analysis according to the classification method proposed by Khishfe and Abd-El-Khalick (2002), while Ozgelen et al. (2013) code using the method of Lederman et al. (2002). We do not deny the rationality of the classification or coding methods proposed by the researchers (such as Abd-El-Khalick, 2012; N. G. Lederman et al., 2002). However, the results may be affected by the fact that the researchers used different NOS frameworks for data processing. In order to make up for these limitations, we may need a new data processing method.

Topic modeling is a prominent document analysis technique that has been widely accepted in many communities, such as machine learning and the social sciences (Ramage et al., 2009). The approach is divided into several more specific modeling techniques, such as latent semantic indexing (LSI) (Deerwester et al., 1990), probabilistic latent semantic analysis (PLSA) (Hofmann, 2001), latent Dirichlet allocation (LDA) (Blei et al., 2003), and correlated topic models (Lafferty & Blei, 2006). Of these models, LDA topic modeling is most widely applied (Morstatter & Liu, 2018). The LDA model is a hierarchical Bayesian model that can connect words with similar meanings and distinguish the particular usage of a word with multiple meanings (Alghamdi & Alfalqi, 2015). However, the purpose of topic modeling is not to understand the meaning of words in a document (Buenaño-Fernandez et al., 2020); instead, topic modeling seeks to determine the topics under study in a collection of documents (Momtazi, 2018). In other words, LDA topic modeling is different from the software analysis techniques used by Ozgelen et al. (2013). LDA topic modeling does not only consider sentences containing terms, keywords, and phrases that contain a view of the NOS but analyzes all viewpoints of the participants. In addition, LDA topic modeling is an unsupervised classification method (Roberts et al., 2014). This method can be used to categorize a large number of unmarked documents to get at the underlying topics in the documents (Momtazi, 2018). To better understand the process of topic modeling, it is helpful to imagine it as the opposite of creating a document from an existing topic (Foster & Inglis, 2019). For example: "Topic 1" as "creativity." The paragraph describing "creativity" may contain words such as creation, reasoning,

imagination, and cleverness, but each of these words has a different probability of representing Topic 1, and words with high probability are better at expressing Topic 1 than words with low probability. We may analyze a completed document for three topics and find that the document comprises 50% of the words in Topic 1, 30% in Topic 2, and 20% in Topic 3. Topic modeling is the reverse of this process: in topic modeling, we start with documents and calculate their most suitable topic composition. This means that researchers can infer participants' views of the nature of science directly from the classification results of the LDA topic modeling, without requiring them to deduce a NOS framework from the literature or previous research results, which may avoid the risk of influencing results due to differences in NOS frameworks.

To sum up, the research questions of this paper are as follows:

1. What views of NOS do pre-service teachers hold based on using LDA topic modeling?
2. Compare the results of the LDA topic modeling method with the NOS framework behind the questionnaire used in this study (Views of the Nature of Science Questionnaire—Form C (VNOS-C)).

### 3 Method

#### 3.1 Participants

The participants of this study consisted of 155 pre-service middle school science teachers (64 males and 91 females, with an average age of 20) studying physics at the Shandong Normal University in China. All participants had similar social and educational backgrounds: they studied physics, chemistry, and biology in high school and had completed courses in mechanics, electromagnetics, thermology, and the history of physics, to name but a few. All participants are from the course physics Experimental Design (and Research). This is their degree course, and everyone in their major is going to take this course. The training objectives of this course are as follows: (1) learn to use theoretical knowledge of physics to guide experiments; (2) deepen the training of analyzing problems and solving practical problems and improve the skills and level of experimental measurement; (3) strengthen the cultivation of the whole process of completing scientific experiments; and (4) try the research of scientific research subjects.

The questionnaire survey was conducted in a supervised classroom environment with a class that functioned as a unit, and all students participated voluntarily. Considering the openness of the VNOS-C questionnaire, we did not set a time limit for its completion. The participants generally took 90–120 min to complete the questionnaire. In addition, we informed them in advance that we just wanted to know their views on NOS-related issues, and their responses to the questionnaire would not have any impact on their course. When answering each item, we encouraged the participants to express their own views as much as possible and to use specific examples to support their views.

#### 3.2 Data Collection and Instruments

Although philosophers, historians, sociologists, and educators have often disagreed about the specific meaning of the NOS, they agree on its most important aspects, for example, currently it would be difficult to reject the theory-laden nature of scientific observations or

defend a deterministic/absolutist or empiricist conception of NOS (Lederman et al., 2002). We agree with this general consensus because existing instruments to assess teacher and student views of the NOS based on these understandings have proved reliable and valid (Pavezet al., 2016), and moreover, the NOS aspects that inform these instruments are also affirmed by many curriculum documents across nations (Abd-El-Khalick, 2014; Olson, 2018).

This study used the Views of the Nature of Science Questionnaire—Form C (VNOS-C) (Lederman et al., 2002) as its instrument for data collection. The VNOS-B, VNOS-C, and VNOS-D questionnaires are intended to assess the same aspects of the NOS, with the differences being either the additional context-specific questions in forms B and C or the developmental appropriateness and language of VNOS-D. In addition, the two questionnaires, VNOS-B and VNOS-C, take a long time. VNOS-C usually takes 1.5 h, while VNOS-D is generally less than 1 h (Lederman, 2007). Since this study is a textual analysis of pre-service teachers' responses to the NOS, pre-service teachers are required to express their views in maximum detail. Therefore, the VNOS-C questionnaire was selected in this study.

### 3.3 The VNOS-C Questionnaire

The VNOS-C questionnaire consists of ten items that can be used to assess pre-service teacher views on the eight aspects of the NOS, namely, (1) The Empirical Nature of Scientific Knowledge: Science relies on the observation of nature and makes judgments according to the observed results. However, scientists do not always observe the natural phenomena directly but need to interpret them from the theoretical framework through human perception or with the help of specific instruments. (2) Observation, Inference, and Theoretical Entities in Science: observation is not the same as inference, and students need to make a clear distinction between the two concepts; observation is a descriptive statement of natural phenomena, while inference is a statement of phenomena that are not directly accessible to the senses. (3) Scientific Theories and Laws: scientific theory is a highly developed and proven internal interpretative system, which is usually based on assumptions and axioms, and cannot be directly measured, only supported by indirect evidence and verified by theory; a law is a descriptive statement between observable phenomena. Theory and law are two different types of knowledge. They do not transform each other, nor are they subordinate. (4) The creative and imaginative nature of scientific knowledge: although science is empirical, the development of scientific knowledge involves the creativity and imagination of both non-scientists and scientists. (5) The Theory-Laden Nature of Scientific Knowledge: scientists' prior knowledge, beliefs, life circumstances, and expectations influence their work, and these factors may influence the research methods they use and the interpretation of their observations. (6) The Social and Cultural Embeddedness of Scientific Knowledge: science influences, and is influenced by, a variety of cultural elements and domains, including social structures, worldviews, power structures, and philosophical, religious, political, and economic factors. (7) The Myth of the Scientific Method: one of the misunderstandings of the scientific method is for students to think that scientists at work always follow certain procedures, such as comparison, test, speculation, and hypothesis. Indeed, scientists sometimes follow certain procedures; however, in the process of the development of scientific knowledge, there is no single scientific method. (8) The Tentative Nature of Scientific Knowledge: scientific knowledge (e.g., facts, theories, and laws), while reliable and enduring, is by no means absolute or certain (Abd-El-Khalick, 2012; N.

G. Lederman et al., 2002; N. Lederman, 2007). In addition, in order to further explore the views of participants, N. G. Lederman et al. (2002) formulated an interview protocol for NOS when using the questionnaire and selected some participants to explain their answers to an item in VNOS-C. However, an increasing number of qualitative researchers have realized that interview is not a neutral way to obtain data but the result of interaction and mutual influence between two people (or more) (McDonald, 2010). In other words, the data obtained from the interviews are not only from the participants, and cannot fully represent their views. Therefore, in this study, only open-ended questionnaires were used to evaluate participants' views on the NOS, instead of questionnaires combined with interviews. However, in order to obtain participants' opinions more clearly on each item, we required participants to explain their opinions with specific examples when answering each item in the VNOS-C questionnaire—for example, the second item in the VNOS-C questionnaire: “What is an experiment? Please give specific examples to support your opinion.”

In this study, participants responded to the 10 items in the VNOS-C questionnaire, and each participant's responses was made into a separate document, with each document containing a participant's view of the NOS according to the questionnaire. One hundred fifty-five questionnaires were collected after excluding questionnaires with incomplete responses.

### 3.4 Pre-processing and Statistical Analysis

This section describes in detail how to use the method of LDA topic modeling for data analysis. The LDA algorithm is a probabilistic generative model of a topic. Its basic idea is that a document is composed of a random mixture of potential topics, and each document is a function of potential variables called topics (Buenaño-Fernandez et al., 2020). In other words, each document is modeled as a mixture of bag-of-words topics, and each topic is a discrete probability distribution, which defines the probability of each word appearing on a given topic. Figure 1 shows the detailed operation of the LDA algorithm. LDA assumes that each document ( $M$ ) composed of  $N$  words can be expressed as the probability distribution of Dirichlet on the potential topics. The parameters  $\beta$  and  $\alpha$  are the Dirichlet priors of  $\Phi_k$  and  $\theta_d$ , respectively. Values of these parameters larger than 1 will result in a smooth distribution of topics or words, while values below 1 will result in a sparse distribution of fewer topics or words (Bastani et al., 2019). The words in the document are observable variables ( $W_{d,n}$ ), the topic distribution for each document ( $\theta_d$ ) and the topic assignment

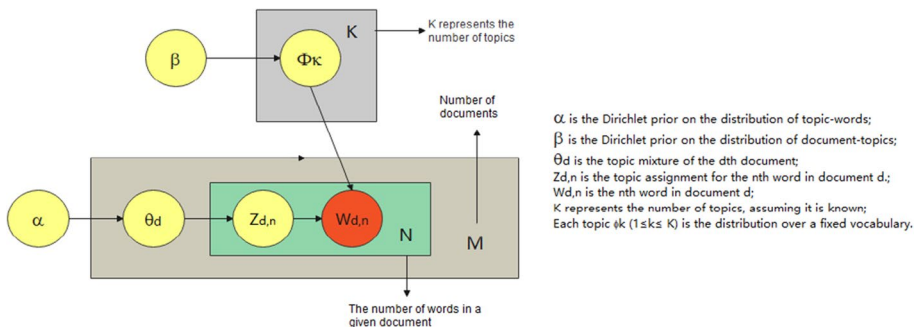


Fig. 1 Latent Dirichlet allocation (LDA) for topic modeling (Momtazi, 2018)

for each word ( $Z_d$ ) are unknown (hidden) variables, and we need to infer the unknown variables from the observable variables. LDA has the significant advantage of using Bayesian learning to infer unobserved structures by computing their posterior distributions from joint distributions (Bastani et al., 2019). In other words, LDA topic modeling can provide two main outputs: topics  $\Phi_k$  and their proportions (importance) in each document  $\theta_d$ . In this study, LDA topic modeling is used to discover the topic of the documents.

In this study, our data comes from the responses of pre-service teachers to the VNOS-C questionnaire. Each participant’s answer to the questionnaire is made into an electronic document, and all documents constitute a text database. Figure 2 provides an overview of the flow through which we processed the data using LDA in this study. Before applying the LDA to model to the text body, we pre-processed the text body, which involved deleting special symbols that were not materially significant for classification, such as numbers, English letters (such as m, g), punctuation marks, special symbols, and full-width character stops (Foster & Inglis, 2019). We obtained all the significant words in each document by pre-processing our text database (the red circle in Fig. 1:  $W_{d,n}$ ). In addition, when using

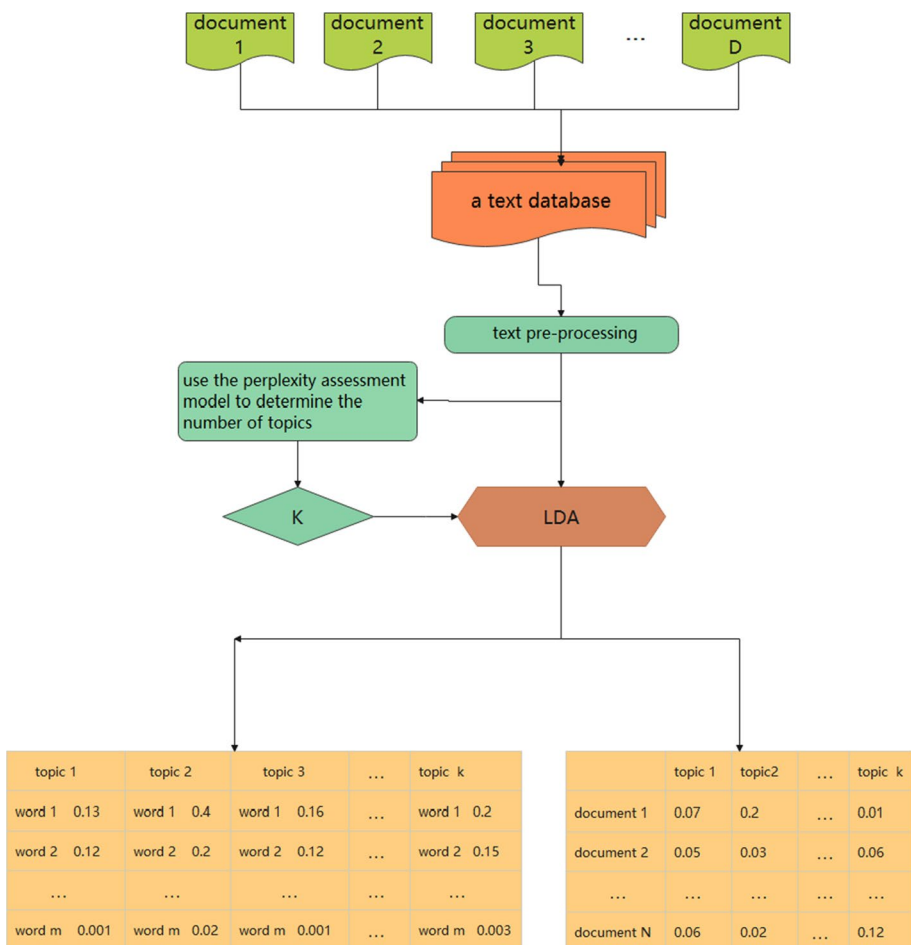


Fig. 2 An overview of LDA



topic modeling for analysis, it is necessary to predetermine the number of topics the algorithm should find (Blei et al., 2003). Perplexity is a commonly used indicator in LDA topic modeling (Jacobi et al., 2015). In addition, Jacobi et al. (2015) stress that perplexity should be only used to initially determine the number of topics—in social sciences, the interpretability of the subject is the more important goal. Therefore, according to perplexity (we tested from 2 to 50 topics in steps of 1) and the interpretability of the LDA topic modeling result, we determined that 12 topics should be included in our text database. Finally, LDA topic modeling was performed. Through the operational LDA algorithm, we obtained topics  $\Phi_k$  and their proportions (importance) in each document  $\theta_d$ . Each topic is represented by a set of words (Momtazi, 2018). The proportion of each word in a topic is different. The higher the proportion value is, the more representative it is of the topic. The total proportion value of all words is equal to 1.

The results of LDA topic modeling are the document topic probabilistic model and discrete probability distribution of different words with probability vectors for each topic. Researchers can determine the labels and meanings of topics based on the results of LDA classification (Bastani et al., 2019). In other words, researchers do not need to reason about a NOS framework but can determine the NOS views held by participants based on the labels and meanings of each topic. In the results of LDA topic modeling, each topic is composed of a set of words, and the proportion (importance) of each word in the topic is different. The greater the proportion, the better it can explain the topic. Through experiments and checks, Morstatter and Liu (2018) found that the proportion of all words contained in a certain topic is ranked from the largest to the smallest. Then you can use the first 20 words to identify the hashtag and meaning of the topic. Sugimoto et al. (2011) used the same approach and also analyzed the document with the proportion associated with each topic to better unpack their topic results. The statistical analysis method of the current study uses the same method as Sugimoto et al. (2011), focusing on the top 20 words and the document with the highest correlation with each topic to determine the label and meaning of each topic.

## 4 Research results

### 4.1 The Pre-service Teachers' Views of the NOS

This section presents the LDA results. Table 1 lists the top 20 words associated with each topic and the proportion value of each word, arranged in descending order of proportion values. In each topic, the higher the proportion of the word, the more closely related the word is to the topic, or the better the word explains the meaning of the topic. In addition, we can observe that the same word appears in more than one topic, for example, “science” (appearing in topics 2, 3, 7, 9, 10, 11, and 12, respectively) and “theory” (appearing in topics 1, 4, 5, 6, 11, and 12, respectively). The same terms, phrases, or keywords express different meanings in different contexts (Jaeger et al., 2019). If we consider the word “science” separately from the context, the meaning is the same, but the meaning of the word “science” in different contexts is different. When the same word appears in different topics, it provides a context for the word, and the meaning expressed is also different.

Sugimoto et al. (2011) determined the label and meaning of each topic according to the top 20 words most related to the topic and the documents most representative of the topic. This study adopts the same method as Sugimoto et al. (2011). Since each document is long

**Table 1** Twenty most probable words of each topic (proportion value retains three effective numbers)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
Theory 0.194	Experiment 0.0961	Science 0.176	Theory 0.0970	Experiment 0.121	Humanity 0.0808	Science 0.142	Data 0.0542	science 0.186	Science 0.172	Theory 0.0699	Science 0.159
Phenomenon 0.117	Explore 0.0940	the study 0.0401	Law 0.0721	Science 0.0898	Experiment 0.0764	nucleus 0.0380	Analysis 0.0428	Experiment 0.136	Law 0.0940	Study 0.0551	Development 0.121
Hypothesis 0.0575	Science 0.0732	Dinosaur 0.0341	Newton 0.0512	Model 0.0824	Theory 0.0496	Atom 0.0327	Conclusion 0.0390	knowledge 0.108	Objective 0.0638	Method 0.0262	Society 0.0794
Experiment 0.0490	Data 0.0481	Extinct 0.0297	Physical 0.0500	Atom 0.0725	Verification 0.0396	Particle 0.0300	Object 0.0390	Find 0.0511	Thing 0.0372	Standard 0.0245	Culture 0.0540
Explanation 0.0484	Subject 0.0425	Earth 0.0265	put forward 0.0405	put forward 0.0627	Find 0.0389	Verification 0.0281	Design 0.0369	Error 0.0204	Development 0.0304	Logic 0.0207	species 0.0364
Species 0.0310	Conclusion 0.0364	resulting in 0.0265	Mechanics 0.0282	Structure 0.0308	World 0.0371	prove 0.0264	Learn 0.0356	create 0.0202	Study 0.0250	Science 0.0197	value 0.0281
Creation 0.0250	Truth 0.0310	Species 0.0236	Quantum 0.0230	Species 0.0308	Prove 0.0201	Country 0.0251	law 0.0339	Certain 0.0200	Religion 0.0239	System 0.0186	Experiment 0.0246
Observation 0.0234	Rule 0.0214	Volcanic 0.0234	Classic 0.0207	Electron 0.0231	Idea 0.0194	Experiment 0.0239	Collect 0.0331	Imagination 0.0196	influences 0.0213	Experience 0.0173	Basis 0.0191
Data 0.0221	Creature 0.0214	Meteorite 0.0216	System 0.0198	Theory 0.0216	View 0.0194	create 0.0209	change 0.0259	Feature 0.0165	Imagination 0.0204	Form 0.0150	Theory 0.0186
Imagination 0.0207	Condition 0.0210	Culture 0.0212	Field 0.0196	Truth 0.0203	Guess 0.0187	Process 0.0199	Livelihood 0.0219	Study 0.0158	Fact 0.0200	Philosophy 0.0148	Test 0.0181
Conformity 0.0186	Structure 0.0196	break out 0.0210	Movement 0.0193	Subject 0.0177	Think 0.0170	Philosophy 0.0194	Verification 0.0205	Instrument 0.0145	Movement 0.0161	Object 0.0133	Characteristic 0.0160
Feature 0.01337	Process 0.0160	Subject 0.0205	Find 0.0184	Stage 0.0156	Universe 0.0138	observed 0.0165	Fall 0.0205	Method 0.0140	Explanation 0.0159	phenomenon 0.0131	Era 0.0151
Definition 0.0101	Nature 0.0160	Knowl- edge 0.0205	quantum mechanics 0.0170	set up 0.0148	Society 0.0128	Conclusion 0.0163	Knowledge 0.0200	Explore 0.0137	Particle 0.0142	view 0.0121	process 0.0150
Change 0.00970	Law 0.0142	Substance 0.0199	Experiment 0.0149	Scattering 0.0143	Angle 0.0126	Electron 0.0162	Biological 0.0196	Use 0.0131	Thinking 0.0140	Test 0.0121	Practice 0.0131

Table 1 (continued)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
Perfect 0.00873	Evidence 0.0140	Atom 0.0183	History 0.0117	Development 0.0132	Development 0.00981	Livelihood 0.0158	Experiment 0.0188	Factor 0.0121	Law 0.0134	concept 0.0116	Technology 0.0130
Verification 0.00795	Purpose 0.0123	the reason 0.0170	Energy 0.0107	Overthrow 0.0111	Life 0.00981	Practice 0.0152	Thing 0.0183	law 0.00977	Cognition 0.0120	Principle 0.0116	create 0.0125
Support 0.00757	Animals 0.0117	Hit 0.0165	Einstein 0.0107	Create 0.0111	Method 0.00957	Political 0.0144	Manner 0.0168	Rule 0.00941	Subjective 0.00932	difference 0.0114	Fusion 0.0123
Reality 0.00718	Inference 0.0113	Occur 0.0152	Happening 0.0105	Formula 0.0110	Produce 0.00957	Bombard- ment 0.0139	means 0.0154	Include 0.00906	set up 0.00816	Mathematics 0.0108	Technology 0.0120
Evolution 0.00718	Facts 0.0108	Create 0.0121	Religion 0.00909	Provide 0.00978	Thinking 0.00934	Time 0.0135	process 0.0137	design 0.00887	Gene 0.00758	language 0.0106	Livelihood 0.0118
Verification 0.00679	Guess 0.009172	Planning 0.0118	the study 0.00886	Law 0.00884	Learn 0.00911	Judgment 0.0133	Variety 0.0135	Already 0.00853	Nature 0.00719	Classification 0.0102	Occur 0.0106

and contains multiple topics, we chose only the excerpts that best highly relevant to each topic. The excerpts that best represent Topic 1 are as follows:

- ... Experiments are used to test theories or verify certain phenomena. At this stage, most of the experiments that students are exposed to concern learning a certain theory in theoretical classes, and then consolidating that theoretical knowledge through experiments and checking for deficiencies. Students are familiarized with the equipment and processes of experiments so that they can keep in touch with higher-level experiments after re-creation. An experiment enables researchers to develop something that they think can be established based on the theory. The experiment at this time is the process of trying to find an ideal state for a specific goal...
- ... Changes will occur and scientific theories will change; however, due to this uncertainty, we possess the need to learn and continue to develop. If a scientific theory remains the same, we only need to learn its application; the theory does require studying. With the development and progress of society and the advancement of experimental equipment, we have secured better conditions to study, verify, improve, and apply the scientific and theoretical knowledge we have gained, maintain a questioning attitude, and discover problems. It is precisely because scientific theories will change that they can create more powerful possibilities, which is why we spend our energy learning about them...
- ... A scientist's determination of the characteristics of a species is limited to their theory, which can explain the characteristics of existing phenomena. By observing and experimenting, they come up with a theory that can describe these characteristics while distinguishing them from the characteristics of other species. However, these characteristics will change, and their accuracy is difficult to verify. Scientists can derive theories from phenomena, but without a representative representation of the present, it is difficult to predict. For example, there is no good way to completely screen for genetic changes in a recessive gene, including as it relates to the current pandemic (new coronavirus disease outbreak), and it is difficult to fully grasp its characteristics and control it in a short time...

Because of the language differences, it may be confusing to read, and in order to make it easier for the reader to obtain the excerpted information, we only select some sentences from the excerpt. Topic 1 selects the following sentences:

- ... Experiments are used to test theories or verify certain phenomena... an experiment enables researchers to develop something that they think can be established based on the theory...
- ... scientific theories will change... with the development and progress of society and the advancement of experimental equipment, we have secured better conditions to study, verify, improve, and apply the scientific and theoretical knowledge we have gained, maintain a questioning attitude, and discover problems...
- ... a scientist's determination of the characteristics of a species is limited to his theory... these characteristics will change, and their accuracy is difficult to verify...

The sentences that best represent the other topics are shown in Table 2.

From Tables 1 and 2, we can determine the label and meaning of each topic. Due to space limitations, we take Topic 1 as an example to explain in detail how to understand the focus depicted in Topic 1 based on the two types of information mentioned above.

For Topic 1, among the 20 words with the highest proportion value listed above, “theory” has the highest proportion value for the topic and is the most representative word. Phenomenon, hypothesis, experiment, explanation, observation, data, verification, perfection, and confirmation construct an experimental context; meanwhile, species, conformity, characteristics, definition, and evolution construct a biologically related context. The two contexts seem irrelevant if viewed separately: one describes experiments, while the other describes organisms; however, both groups contain high-contribution words for the

**Table 2** Document fragments highly relevant to each topic

Topic	Document fragments highly relevant to each topic
Topic 1	<p>...Experiments are used to test theories or verify certain phenomena...an experiment enables researchers to develop something that they think can be established based on the theory...</p> <p>...scientific theories will change...with the development and progress of society and the advancement of experimental equipment, we have secured better conditions to study, verify, improve, and apply the scientific and theoretical knowledge we have gained, maintain a questioning attitude, and discover problems..</p> <p>...a scientist's determination of the characteristics of a species is limited to his theory...these characteristics will change, and their accuracy is difficult to verify..</p>
Topic 2	<p>...an experiment is an inquiry conducted to verify the correctness of a truth...the process of verifying other truths through known truths is also an experiment..</p> <p>...science is about organizing knowledge and obtaining regular and truthful experiences...scientific subjects comprise scientific truths, they are truths obtained by scientists after continuous trials and experiments...knowledge of scientific disciplines can be used to solve problems..</p>
Topic 3	<p>...different scientists get different reasons for the extinction of dinosaurs based on the same data, because different scientists think differently...scientists could not use existing data to experiment on how the dinosaurs went extinct..</p>
Topic 4	<p>...when scientists develop a theory, that scientific theory is different from the original scientific theory. The key is that the applicable conditions of the theory are no longer the same. For example, when Newton's law of motion was proposed..</p> <p>... Scientific theories are different from scientific laws. Scientific theory is a set of research methods and science for a specific field, and should contain scientific laws..</p>
Topic 5	<p>... In the beginning, Thomson proposed... Rutherford used...and found that... Prove the nuclear structure..</p>
Topic 6	<p>... Science uses real and natural methods to help people understand the world. It does not involve too many human emotions..</p> <p>... The development of scientific knowledge certainly requires experiments...although some experiments were created in a certain way..</p>
Topic 7	<p>...scientists infer the position of the nucleus from the motion of the atom... Later, various microscopes with high magnification were used..</p>
Topic 8	<p>... I believe experimentation refers to a series of actions in which learners or researchers design and perform specific operations when exploring an unknown problem or verifying a known problem. For example, in physics, to verify Newton's second law, we..</p>
Topic 9	<p>... Experiments are based on existing knowledge, using instruments and methods of..... to conduct a scientific exploration process of a certain understanding or knowledge...</p> <p>... Imagination and creativity are used to plan and design the collection of data...</p>
Topic 10	<p>... Science is the exploration of the objective laws of the world, and its goal should be the general laws that cannot be changed by people and their observation and thinking..</p>
Topic 11	<p>... Science is a theory that accurately reflects the nature and laws of the world..</p> <p>... Scientific theories have rigorous logic and systematic characteristics, and scientific theories are the tools based on the processing of the consensus of known sensory information.</p>
Topic 12	<p>... I believe that science is the product of advanced technology in an era based on social development, which can reflect the level of scientific and technological development...</p> <p>... I believe that science is integrated with social and cultural values. First...than..</p>

topic—therefore, there must be a certain connection between them. In addition, certain leftovers are present, such as creativity, imagination, and change. It seems that these two contexts involve change and creation. Therefore, we assume that the focus of Topic 1 is the understanding of scientific theories.

To further determine whether our judgment on the topic is correct, we need to combine articles with the highest probability value to the topic, as shown in the results section. The first part is the answer to the second question, “What is an experiment?” of the VNOS-C questionnaire. Experiments are used to test theories, verify phenomena, consolidate the theoretical knowledge learned, and create and verify some ideas or theories. The second part is the answer to the fourth question of the questionnaire: “After scientists develop a theory, will the theory change?” The theory will change. The reason for the change is social progress and equipment. The third part is the seventh question of the questionnaire, “How certain are scientists about their characterization of what a species is? What species evidence do you think scientists used to determine what a species is?” Since scientists used their own theories to determine the characteristics of the species, the characteristics will change, and, subsequently, the theory will also change. Through the analysis of the documents with the highest probability value, we observed that these three parts are also the pre-service teachers’ understandings of scientific theories, which are consistent with our earlier conjectures. The determination of the other topics adopts the same method and process as Topic 1. The results of each topic are shown as follows:

- Topic 1: Based on the experimental view of scientific theory, improvements in technology and methods lead to changes in theory.
- Topic 2: Pre-service teachers believe that science is the truth obtained by scientists after multiple experiments and is used to solve practical problems.
- Topic 3: Scientists came to different conclusions with the same set of data because they could not reproduce the experimental process and prove the accuracy of the data.
- Topic 4: Scientific theories are different from scientific laws. Theories and laws are in a hierarchical relationship. Theories contain several laws, and multiple laws constitute a theory.
- Topic 5: The pre-service teacher simply stated the scientist’s process of obtaining the atomic structure.
- Topic 6: Science uses an objective method to show the world that already exists and uses experimental methods to prove its ideas.
- Topic 7: Scientists combine experimental phenomena to prove their speculations about the atomic structure.
- Topic 8: The overall structure of the experiment or the specific process and purpose of the experiment.
- Topic 9: Experiments are a means of discovering scientific knowledge and can also be used to test or verify the accuracy of certain scientific knowledge.
- Topic 10: Science is an objective method to explore the law of development of things in nature.
- Topic 11: Theory is a way to understand scientific knowledge, or it can be used to predict certain scientific knowledge.
- Topic 12: The development of science is integrated with social and cultural values. Science can promote cultural development. Simultaneously, the nature of society and cultural values can, in turn, influence science.

## 4.2 Comparison of the LDA Topic Modeling Results and the NOS Framework Behind the VNOS-C Questionnaire

In this section, we compare the results of LDA topic modeling with the theoretical framework behind the VNOS-C questionnaire.

The consensus among philosophers, historians, sociologists, and educators about the specific definition and meaning of NOS at a given point in time or historical period, as well as some aspects (such as the distinction between observation and inference, the scientific method, and scientific theories and laws) of NOS mentioned in scientific education documents, form the theoretical framework behind the VNOS-C questionnaire (Lederman et al., 2002). The VNOS-C questionnaire contains eight specific aspects of NOS, namely: The Empirical Nature of Scientific Knowledge, Observation, Inference, and Theoretical Entities in Science, Scientific Theories and Laws, The Theory-Laden Nature of Scientific Knowledge, The Social and Cultural Embeddedness of Scientific Knowledge, The Myth of The Scientific Method, The Tentative Nature of Scientific Knowledge, and The Creative and Imaginative Nature of Scientific Knowledge (see the first column of Table 3).

According to the classification criteria of Lederman et al. (2002), he divided the NOS into "More Naive Views" and "More Informed Views." However, in dividing the LDA modeling results, we found different views from "More Naive Views" and "More Informed Views." According to the classification criteria of Lederman et al. (2002), some topics express more than "More Naive Views," but do not reach the level of "More Informed Views." We classify these topics as "Mixed Views." Therefore, we categorize our results into "More Naive Views," "More Informed Views," and "Mixed Views." The research results of Mesci (2020) are similar to ours. When processing data, he also found that there may be different understandings of a certain aspect of the view of the NOS. Therefore, in his research, the situation that there are two understandings of the same aspect of the NOS are classified as "Mixed Views."

Take "The Empirical Nature of Scientific Knowledge" as an example: If a topic expresses similar to: (1) Science is concerned with facts. We use observed facts to prove that theories are true (Bell et al., 2011; Lederman et al., 2002). (2) Science is concerned only with experiments and mathematical numbers in the laboratory (Mesci, 2020). The topics are considered to have "More Informed Views." If participants hold opinions similar to: (1) much of the development of scientific knowledge depends on observation.... [But] I think what we observe is a function of convention. I don't believe that the goal of science is (or should be) the accumulation of Observable Facts (Lederman et al., 2002). (2) Rather than...science involves abstraction, one step of abstraction after another (Lederman et al., 2002). (3) Science is not limited to studying only visible events. Although scientists do not directly see many cases with their eyes, they make observations and explanations based on empirical data as a result of these observations. For example, scientists recently found a new planet for a long years later based on observations (Bell et al., 2011). The topics are considered having "More Naive Views." If it is between two views, it is a "Mixed Views" (see Table 3). Topic 2: "Pre-service teachers believe that science is the truth obtained by scientists after multiple experiments and is used to solve practical problems" and Topic 10: "Science is an aim method to explore the law of development of things in nature" are closest to "More Naive Views"; compared with Topic 2 and Topic 10, Topic 6 (Science uses an aim method to show the world that already exists and uses experimental methods to prove its ideas) and Topic 9 (Experiments are a means of discovering scientific knowledge and can also test or verify the accuracy of certain scientific knowledge) not only express

**Table 3** Comparison of the LDA topic modeling results and the NOS framework behind the VNOS-C questionnaire

NOS framework	Previous researchers used VNOS to obtain the NOS view		LDA topic modeling results	
	More naive views	More informed views	More naive views	More informed views
The Empirical Nature of Scientific Knowledge	<ul style="list-style-type: none"> <li>Science is concerned with facts. We use observed facts to prove that theories are true (Bell et al., 2011; Lederman et al., 2002)</li> <li>Science is concerned only with experiments and mathematical numbers in the laboratory (Mesci, 2020)</li> </ul>	<ul style="list-style-type: none"> <li>Much of the development of scientific knowledge depends on observation...[But] I think what we observe is a function of convention. I don't believe that the goal of science is (or should be) the accumulation of observable facts (Lederman et al., 2002)</li> <li>Rather...science involves abstraction, one step of abstraction after another (Lederman et al., 2002)</li> <li>Science is not limited to studying only visible events. Although scientists do not directly see many cases with their eyes, they make observations and explanations based on empirical data as a result of these observations. For example, scientists recently found a new planet for a long years later based on observations (Bell et al., 2011)</li> </ul>	<p>Topic 2</p> <p>Topic 10</p>	<p>Topic 6</p> <p>Topic 9</p>



**Table 3** (continued)

NOS framework	Previous researchers used VNOS to obtain the NOS view		LDA topic modeling results	
	More naive views	More informed views	More naive views	More informed views
Observation, Inference, and Theoretical Entities in Science	<ul style="list-style-type: none"> <li>• Scientists are quite certain about the structure of the atom. First scientists only had theories of how the atom was constructed until they were able to use very powerful microscopes to see how they were constructed (Lederman et al., 2002)</li> <li>• Scientists are strongly sure of the structure of an atom due to chemical structures observed among the elements. The electron microscope presented strong evidence (visually) of the structure (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>• Science is based on observations by sensory organs and technological instruments, and inferences are made by associating these observations with existing knowledge. (Mesci, 2020)</li> </ul>	Topic 5	Topic 7
Scientific Theories and Laws	<ul style="list-style-type: none"> <li>• Newton's Law #1 is proven and through various testing and experiments it has come to be known as a proven law. Theories, however, have not been proved enough to be changed into laws (Bell et al., 2011)</li> <li>• A scientific law is somewhat set in stone, proven to be true...A scientific theory is apt to change and be proven false at any time. (Mesci, 2020)</li> <li>• Scientific theories are not accepted by everyone, while scientific laws are proven and accepted by everyone (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>• A scientific law describes quantitative relationships between phenomena such as universal attraction between objects. Scientific theories are made of concepts that are in accordance with common observation or go beyond and propose new explanatory models for the world (Lederman et al., 2002)</li> </ul>	Topic 4	

**Table 3** (continued)

NOS framework	Previous researchers used VNOS to obtain the NOS view		LDA topic modeling results	
	More naive views	More informed views	More naive views	More informed views
The Theory-Laden Nature of Scientific Knowledge	<ul style="list-style-type: none"> <li>• [Scientists reach different conclusions] because the scientists were not around when the dinosaurs became extinct, so no one witnessed what happened....I think the only way to give a satisfactory answer to the extinction of the dinosaurs is to go back in time to witness what happened (Lederman et al., 2002)</li> <li>• Science is about the facts and could not be influenced by cultures and society (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>• Both conclusions are possible because there may be different interpretations of the same data. Different scientists may come up with different explanations based on their own education and background or what they feel are inconsistencies in others ideas (Lederman et al., 2002)</li> </ul>	Topic 3	Mixed views
The Social and Cultural Embeddedness of Scientific Knowledge	<ul style="list-style-type: none"> <li>• Science has a particular method of going about things, the scientific method (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>• Of course culture influence the ideas in science. It was more than a 100 years after Copernicus that his ideas were considered because religious beliefs of the church sort of favored the geocentric model (Lederman et al., 2002)</li> </ul>		Topic 12
The Myth of The Scientific Method	<ul style="list-style-type: none"> <li>• Science has a particular method of going about things, the scientific method (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>• When you are in sixth grade you learn that here is the scientific method and the first thing you do this, and the second thing you do that and so on...That's how we may say we do science, but [it is different from]...the way that we actually do science (Lederman et al., 2002)</li> </ul>	Topic 6 Topic 8 Topic 9	

**Table 3** (continued)

NOS framework	Previous researchers used VNOS to obtain the NOS view		LDA topic modeling results	
	More naive views	More informed views	More naive views	More informed views
The Tentative Nature of Scientific Knowledge	<ul style="list-style-type: none"> <li>Scientific knowledge is certain, and does not change in any way, because it is the same everywhere (Mesci, 2020)</li> </ul>	<ul style="list-style-type: none"> <li>Everything in science is subject to change with new evidence and interpretation of that evidence. We are never 100% sure about anything because...negative evidence will call a theory or law into question, and possibly cause a modification (Lederman et al., 2002)</li> </ul>	More naive views	More informed views
The Creative and Imaginative Nature of Scientific Knowledge	<ul style="list-style-type: none"> <li>Scientific studies develop in the direction of logic and proof, so there is no room for imagination and creativity in science. Science is expressed in definite mathematical numbers. (Mesci, 2020)</li> </ul>	<ul style="list-style-type: none"> <li>Logic plays a large role in the scientific process, but imagination and creativity are essential for the formulation of novel ideas ... to explain why the results were observed (Lederman et al., 2002)</li> <li>Scientists use imagination and creativity in their work. Creativity and imagination play an important role at every moment of a study. Research questions, the procedures, results, and conclusions are shaped by creativity and imagination. (Mesci, 2020)</li> </ul>	More naive views	More informed views

**Table 3** (continued)

NOS framework	Previous researchers used VNOS to obtain the NOS view		LDA topic modeling results	
	More naive views	More informed views	More naive views	More informed views
The Nature of Scientific Theory	<ul style="list-style-type: none"> <li>A theory is an untested idea, or an idea that is undergoing additional tests, generally it hasn't been proved to the satisfaction of the scientific community (Lederman et al., 2002)</li> </ul>	<ul style="list-style-type: none"> <li>In the vocabulary of a scientist the word theory is used differently than in the general population. It does not mean someone's idea that can't be proven. It is a concept that has considerable evidence behind it and has endured the attempts to disprove it (Lederman et al., 2002)</li> </ul>	<p>More naive views</p> <p>Topic 1</p> <p>Topic 11</p>	Mixed views

that experiments can discover scientific knowledge but also verify scientific knowledge, but did not show “Science is concerned with facts” or “Science is concerned only with experiments and mathematical numbers in the laboratory” (“More Informed Views”), and so on; therefore, Topics 6 and 9 are classified as “Mixed Views.”

According to the label and meaning of each topic, we can see that through the LDA topic modeling method, the NOS view of pre-service teachers is divided into 12 independent topics, which means that pre-service teachers hold 12 mainstream views. The 12 views obtained by LDA topic modeling were compared with the views of participants obtained by previous researchers. The 12 views generated by the LDA modeling may use different terms and keywords than the previous participants describe their understanding of the topic, but they express the same meaning, and LDA topics are divided into this view. To sum up, the comparison between the LDA topic modeling results and the theoretical framework behind the VNOS-C questionnaire is shown in Table 3.

## 5 Discussion

In the results section, we present 12 topics that were formed using LDA topic modeling. In this section, further explanation is provided based on the results of LDA topic modeling and the view of the NOS held by pre-service teachers. We compared the 12 views obtained from the LDA topic modeling with those obtained from the previous evaluation of pre-service teachers using the VNOS-C questionnaire to determine the views of pre-service teachers. This paper uses The Empirical Nature of Scientific Knowledge and Observation, Inference, and Theoretical Entities in Science to explain the level of view of NOS represented by each topic.

**The Empirical Nature of Scientific Knowledge** Topics 2 and 9 emphasize the objectivity of science, and pre-service teachers believe that science is a method used to explore the truth of nature. In their studies, Lederman et al. (2002), Bell et al. (2011), and Mesci (2020) believed that pre-service teachers with similar views held naive views. Therefore, this study believed that Topic 2 and Topic 10 indicated that pre-service teachers held naive views. Topics 6 and 9 propose that experiments are a way of discovering scientific knowledge. People use experiments to test and prove their guesses about scientific knowledge, theorems, and phenomena. That is to say, pre-service teachers believe that scientific knowledge needs to be supported by empirical evidence. In this process, people’s ideas about scientific knowledge are involved. They regard scientific claims as a mixture of objective and personal. Compared with the research results of Lederman et al. (2002) and Bell et al. (2011), the view of the NOS on Topics 6 and 9 falls between informed and naive, so we take Topic 6 and 9 as a mixed view.

**Observation, Inference, and Theoretical Entities in Science** Question 6 of the VNOS-C questionnaire asked participants to answer the question, “How do scientists determine the structure of atoms? What concrete evidence do you think scientists use to show what atoms look like?” In Topic 5, although the pre-service teachers described in detail the process of atomic discovery, such as what methods scientists used to discover what ultimately determined the structure of the atomic nucleus, we found that their answers were the same as the statements in high school textbooks. In high school physics textbooks, the discovery process of atoms appears as the background of atomic teaching, and all pre-service teachers

have learned this content in high school. In this study, they just stated the content of high school physics textbooks, but did not reflect their thoughts, such as whether scientists were innovative in the process of discovering atomic mechanisms. Therefore, we speculate that pre-service teachers only state high school knowledge without really understanding the relationship between observation, reasoning, and theoretical entity. Topic 7: the pre-service teacher mentioned reasoning from observation and the use of the microscope for direct proof of atomic structure. We believe that Topic 5 and Topic 7 show that pre-service teachers have naive views.

This research uses the above two aspects as examples to explain. Table 3 shows the level of NOS view represented by each topic. Compared with the previous research results, if a topic is consistent with a naive view, the topic represents a naive view. If the informed view is consistent, it is an informed view, and if it is between the two views, it is a mixed view.

According to the label and meaning of each topic, we can see that through the LDA topic modeling method, the NOS view of pre-service teachers is divided into 12 independent topics, which means that pre-service teachers hold 12 mainstream views, does not mean that each point of view is independent and not connected in any way.

Lederman et al. (2002) emphasized that there is no one-to-one correspondence between participant responses to the VNOS project and the NOS aspect. Knowledge of a certain project is more biased toward a certain aspect. In other words, a participant's understanding of one aspect of the NOS can be reflected in multiple items on the questionnaire. Similarly, each topic does not necessarily represent only one aspect of the NOS. For example, one topic can represent two aspects of the NOS, or two topics can describe the same aspect of the NOS; however, they are just two of the many views regarding this aspect. For example, Topic 1 is about scientific theories and is more inclined toward the tentative aspect of scientific knowledge; in addition, the top 20 words that contribute to Topic 1 also contain words, such as creativity and imagination. This shows that pre-service teachers believe that the process of changing scientific theories requires human imagination and creativity. Obviously, Topics 5 and 7 are answers to how scientists discovered the atomic structure. Topic 5 is only a statement of the work done by scientists in the process of discovering the atomic model, which is mainly based on facts. Topic 7 involves the process of discovering atomic models. During this process, scientists make inferences based on experimental phenomena and finally determine the atomic structure model. According to understandings of each topic, we divided the pre-service teachers' understandings of the NOS into the following aspects: The Tentative Nature of Scientific Knowledge (Topic 1); The Empirical Nature of Scientific Knowledge (Topics 2, 6, 9, and 10); The Subjectivity of Scientific Knowledge/Theory-Laden (Topic 3); The Difference and Connection Between Scientific Theory and Scientific Laws (Topic 4); Reasoning and Theoretical Entities (Topics 5 and 7); Scientific Method (Topics 6, 8, and 9); The Essence of Scientific Theory (Topics 1 and 11); and The Integration of Science, Social Life, and Culture (Topic 12).

There are two differences between the LDA topic modeling results and the NOS framework behind the VNOS-C questionnaire. First, the NOS framework of the VNOS-C questionnaire takes creativity and imagination as an aspect of the view of the NOS; however, our LDA topic modeling analysis method did not include creativity and imagination within one topic but distributed them across multiple topics. Second, compared with the NOS framework of the VNOS-C questionnaire, the LDA topic modeling results take The Nature of Scientific Theory as an aspect of the NOS.

Investigating the pre-service teachers' views on the NOS in China, the USA, and Turkey, Liang et al. (2009) found that the percentage of Chinese pre-service teachers holding "informed" views on creativity and imagination is zero. Liang et al. (2009) reasons that this does not

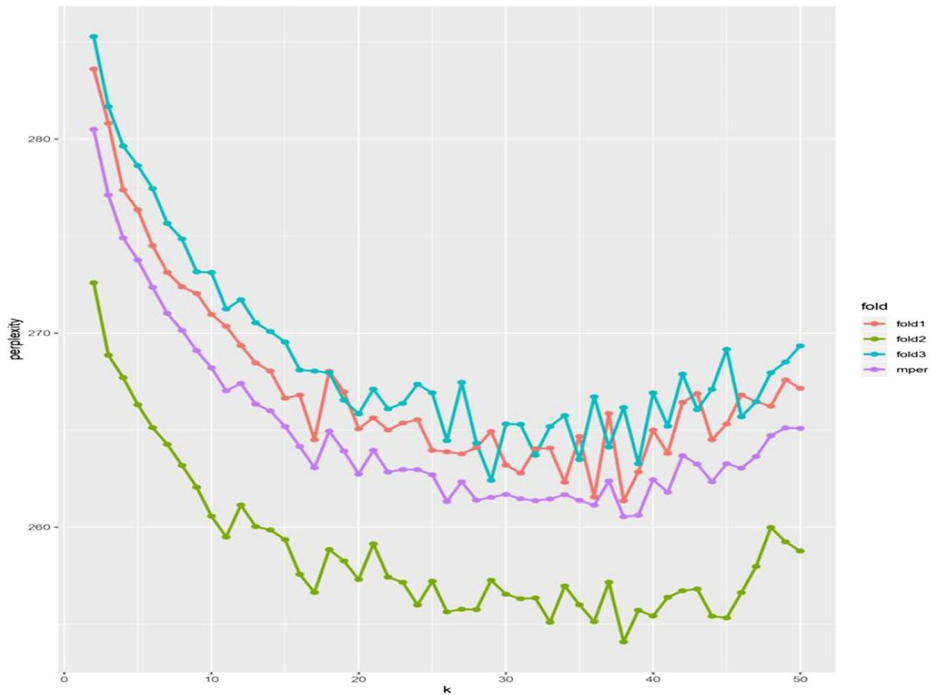
necessarily mean that Chinese pre-service teachers do not have an “informed” view of the NOS but may instead signal that their scoring standards are too strict and the written answers of pre-service teachers are too short or incomplete to be considered “informed.” In this research, topics of creativity and imagination are not formed because they do not exist alone; instead, they are reflected in the process of experimentation and theoretical inquiry. Therefore, creativity, imagination, conjecture, and other words related to imagination and creativity are assigned to the appropriate corresponding topics, such as Topics 1, 3, and 6 in our study. This may mean that participants' understandings of creativity and imagination are not isolated but rather expressed in other aspects of the NOS, such as the tentative nature of scientific knowledge, scientific knowledge is not immutable, and the process of scientific knowledge change. Validating this hypothesis requires more empirical evidence. Notably, these results of our LDA topic modeling might reflect the shortcomings of the approach of Lederman and his colleagues (Abd-El-Khalick, 2006; Bell, 2006; Cobern & Loving, 2001; Hanuscinet al., 2006; Khishfe & Lederman, 2006) for determining the NOS by using a consensus view. Irzik and Nola (2011) point out that each aspect of the view of the NOS proposed by Lederman and his colleagues lacks sufficient systematic unity; science is so rich and so dynamic that characterizing the necessary and sufficient conditions for being a “science” in a way that does justice to this richness and complexity will always be quixotic. Therefore, in future studies, researchers may pay more attention to the relationship between some aspects (such as tentative and subjective) involved in NOS, so as to propose more appropriate strategies for improving pre-service teachers' view of the NOS.

Zion et al. (2020) divided the assessment of pre-service teachers' understanding of scientific theory into two aspects: the first is the definition of scientific theory, which involves the relationship between scientific theory and scientific laws; second, whether the scientific theory will change or not is related to the transient NOS. In other words, the scientific theory relates to two aspects of the NOS view, and it can be placed as one of the two. This also shows that each aspect of the view of the NOS may not be separated, but they are integrated and interact with each other. When cultivating the previous teachers' view of the NOS, we should not only focus on some aspects but should have more understanding of the relationship between each aspect of the view of the NOS.

In addition, taking Topic 1 as an example, we found that pre-service teachers believe that scientific theories will change according to the results. This does not mean that all pre-service teachers believe that scientific theories will change, just that the number of participants that believe that scientific theories remain unchanged is so small that it does not appear as a separate topic. When this ratio is large, it will appear as a separate topic in the results, just like Topics 5 and 7. These two topics represent one aspect of the NOS, but they appear as two topics. This is explained by the fact that all participants hold a large proportion of these two views.

## Appendix

When using topic modeling for analysis, it is necessary to predetermine the number of topics the algorithm should find (Blei et al., 2003). Perplexity is a commonly used indicator in LDA topic modeling (Jacobi et al., 2015). Notably, in language modeling, perplexity reduces the probability of test data monotonically and is algebraically equivalent to the inverse of the average geometric probability per word—the lower the perplexity value, the



**Fig. 3** Calculated complexity value from 2 to 50 topics

better the topic modeling effect (Blei et al., 2003). Therefore, we used the perplexity of the retained test set to evaluate the model and determine the optimal number of topics. To calculate the value of perplexity in this study, we created three folds. Each fold was tested from 2 to 50 topics in steps of 1. Figure 3 shows the values of our three folds and their averages. The figure shows that the lowest perplexity value occurred between 35 and 40 topics. However, Jacobi et al. (2015) stress that perplexity should be only used to initially determine the number of topics—in social sciences, the interpretability of the subject is the more important goal. In addition, we also observed that our curve moved up and down at 11 topics and that fluctuations became more obvious at 15 topics; therefore, we defined our topic as between 11 and 15 topics and thus conducted LDA topic modeling for 11 topics to 15 topics. According to the interpretability of the LDA topic modeling result, we determined that 12 topics should be included in our text database.

**Acknowledgements** The study was funded by two projects:

(1) This work was supported by the first postgraduate OMO course group establishment project (physics teaching theory course group) of the Shandong Normal University of China.

Grant Recipient: Shoubao Gao

(2) This work was supported by the National Natural Science Foundation of China (NSFC) under grant number [72004166].

Grant Recipient: Jianqiang Ye

**Author Contribution** Manman Wang: Methodology, formal analysis, investigation, and writing—original draft.

Shoubao Gao: Funding acquisition.

Weiling Gui: Supervision, project administration, and funding acquisition.



Jianqiang Ye: Advice on the use of analytical methods and funding acquisition.  
 Shuaishuai Mi: Conceptualization and writing—review and editing.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

## References

- Abd-El-Khalick, F. (2012). Examining the sources for our understandings about science: Enduring confluences and critical issues in research on nature of science education. *International Journal of Science Education*, *34*(3), 353–374. <https://doi.org/10.1080/0950>
- Abd-El-Khalick, F. (2013). Teaching with and about nature of science, and science teacher knowledge domains. *Science & Education*, *22*(9), 2087–2107. <https://doi.org/10.1007/s11191-012-9520-2>
- Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education*, *22*(7), 665–701. <https://doi.org/10.1080/09500690050044044>
- Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, *82*(4), 417–436.
- Abd-El-Khalick, F. (2006). Over and over and over again: College students' views of nature of science. In L. B. Flick & N. G. Lederman (Eds.), *scientific inquiry and nature of science: Implications for teaching, learning, and teacher education* (pp. 389–425). Springer Netherlands. <https://doi.org/10.1007/978-1-4020-5814-118>
- Abd-El-Khalick, F. (2014). The evolving landscape related to assessment of nature of science. *Handbook of research on science education* (pp. 621–650). Taylor; Francis.
- Akerson, V. L., Pongsanon, K., Park Rogers, M. A., Carter, I., & Galindo, E. (2017). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science. *International Journal of Science and Mathematics Education*, *15*(2), 293–312. <https://doi.org/10.1007/s10763-015-9690-x>
- Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications*, *6*(1). <https://doi.org/10.14569/IJACSA.2015.060121>
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications*, *127*(1), 256–271. <https://doi.org/10.1016/j.eswa.2019.03.001>
- Bell, R. L., Matkins, J. J., & Gansneder, B. M. (2011). Impacts of contextual and explicit instruction on pre-service elementary teachers' understandings of the nature of science. *Journal of Research in Science Teaching*, *48*(4), 414–436. <https://doi.org/10.1002/tea.20402>
- Bell, R. L. (2006). Perusing padora's box: Exploring the what, when, and how of nature of science instruction. In L. B. Flick & N. G. Lederman (Eds.), *scientific inquiry and nature of science: Implications for teaching, learning, and teacher education* (pp. 427–446). Springer Netherlands. <https://doi.org/10.1007/978-1-4020-5814-119>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.
- Buenaño-Fernandez, D., Gonzalez, M., Gil, D., & Lujan-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*, *8*, 35318–35330. <https://doi.org/10.1109/ACCESS.2020.2974983>
- Coburn, W. W., & Loving, C. C. (2001). Defining "Science" in a multicultural world: Implications for science education. *Science Education*, *85*(1), 50–67. [https://doi.org/10.1002/1098-237X\(200101\)85:1%3c50::AID-SCE5%3e3.0.CO;2-G](https://doi.org/10.1002/1098-237X(200101)85:1%3c50::AID-SCE5%3e3.0.CO;2-G)
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.
- Demirdöğen, B., Hanuscin, D. L., Uzuntiryaki-Kondakci, E., & Köseoğlu, F. (2016). Development and nature of preservice chemistry teachers' pedagogical content knowledge for nature of science. *Research in Science Education*, *46*(4), 575–612. <https://doi.org/10.1007/s11165-015-9472-z>

- Erduran, S., & Kaya, E. (2018). Drawing nature of science in pre-service science teacher education: Epistemic insight through visual representations. *Research in Science Education*, 48(6), 1133–1149. <https://doi.org/10.1007/s11165-018-9773-0>
- Foster, C., & Inglis, M. (2019). Mathematics teacher professional journals: What topics appear and how has this changed over time? *International Journal of Science and Mathematics Education*, 17(8), 1627–1648. <https://doi.org/10.1007/s10763-018-9937-4>
- Gess-Newsome, J., & Lederman, N. G. (1995). Biology teachers' perceptions of subject matter structure and its relationship to classroom practice. *Journal of Research in Science Teaching*, 32(3), 301–325. <https://doi.org/10.1002/tea.3660320309>
- Hanuscin, D. L., Akerson, V. L., & Phillipson-Mower, T. (2006). Integrating nature of science instruction into a physical science content course for preservice elementary teachers: Nos views of teaching assistants. *Science Education*, 90(5), 912–935. <https://doi.org/10.1002/sce.20149>
- Hanuscin, D. L., Lee, M. H., & Akerson, V. L. (2011). Elementary teachers' pedagogical content knowledge for teaching the nature of science. *Science Education*, 95(1), 145–167. <https://doi.org/10.1002/sce.20404>
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1/2), 177–196. <https://doi.org/10.1023/A:1007617005950>
- Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education*, 20(7), 591–607. <https://doi.org/10.1007/s11191-010-9293-4>
- Jacobi, C., van Atteveldt, W., & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106. <https://doi.org/10.1080/21670811.2015.1093271>
- Jaeger, S. R., Roigard, C. M., Jin, D., Vidal, L., & Ares, G. (2019). Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research. *Food research international* (Ottawa, Ont.), 119, 895–907. <https://doi.org/10.1016/j.foodres.2018.10.074>
- Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, 39(7), 551–578. <https://doi.org/10.1002/tea.10036>
- Khishfe, R., & Lederman, N. (2006). Teaching nature of science within a controversial topic: Integrated versus nonintegrated. *Journal of Research in Science Teaching*, 43(4), 395–418. <https://doi.org/10.1002/tea.20137>
- Lafferty, J., & Blei, D. (2006). Correlated topic models. In Y. Weiss, B. Schölkopf et al. (Eds.). *Platt* (Eds.), *Advances in neural information processing systems*. MIT Press. <https://proceedings.neurips.cc/paper/2005/file/9e82757e9a1c12cb710ad680db11f6f1-Paper.pdf>
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331–359. <https://doi.org/10.1002/tea.3660290404>
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497–521. <https://doi.org/10.1002/tea.10034>
- Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–879). Mahwah, NJ: Lawrence Erlbaum.
- Liang, L. L., Chen, S., Chen, X., Kaya, O. N., Adams, A. D., Macklin, M., & Ebenezer, J. (2009). Pre-service teachers' views on nature of scientific knowledge development. *Int J of Sci and Math Educ*, 7, 987–1012. <https://doi.org/10.1007/s10763-008-9140-0>
- McDonald, C. V. (2010). The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *Journal of Research in Science Teaching*, 47(9), 1137–1164. <https://doi.org/10.1002/tea.20377>
- Mesci, G. (2020). The influence of PCK-based NOS teaching on pre-service science teachers' NOS views. *Science & Education*, 29, 743–769. <https://doi.org/10.1007/s11191-020-00117-7>
- Mesci, G., Schwartz, R. S., & Pleasants, B.A.-S. (2020). Enabling factors of preservice science teachers' pedagogical content knowledge for nature of science and nature of scientific inquiry. *Science & Education*, 29(2), 263–297. <https://doi.org/10.1007/s11191-019-00090-w>
- Momtazi, S. (2018). Unsupervised latent Dirichlet allocation for supervised question classification. *Information Processing & Management*, 54(3), 380–393. <https://doi.org/10.1016/j.ipm.2018.01.001>
- Morstatter, F., & Liu, H. (2018). In search of coherence and consensus: Measuring the interpret ability of statistical topics. *Journal of Machine Learning Research*, 18(169), 1–32. <http://jmlr.org/papers/v18/17-069.html>

- Olson, J. K. (2018). The inclusion of the nature of science in nine recent international science education standards documents. *Science & Education*, 27(7–8), 637–660. <https://doi.org/10.1007/s11191-018-9993-8>
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What “ideas-about-science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40(7), 692–720. <https://doi.org/10.1002/tea.10105>
- Ozgelen, S., Yilmaz-Tuzun, O., & Hanuscin, D. L. (2013). Exploring the development of preservice science teachers' views on the nature of science in inquiry-based laboratory instruction. *Research in Science Education*, 43(4), 1551–1570. <https://doi.org/10.1007/s11165-012-9321-2>
- Pavez, J. M., Vergara, C. A., Santibañez, D., & Cofré, H. (2016). Using a professional development program for enhancing Chilean biology teachers' understanding of nature of science (NOS) and their perceptions about using history of science to teach NOS. *Science & Education*, 25(3–4), 383–405. <https://doi.org/10.1007/s11191-016-9817-7>
- Publications, S., Sugimoto, C. R., Li, D., Russell, T. G., Finlay, S. C., & Ding, Y. (2011). The shifting sands of disciplinary development: Analyzing North American Library and Information Science dissertations using latent Dirichlet allocation. *Journal of the American Society for Information Science*, 62(1), 185–204. <https://doi.org/10.1002/asi.21435>
- Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 248–256. <https://www.aclweb.org/anthology/D09-1026>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, 91(2). <https://doi.org/10.18637/jss.v091.i02>
- Sorensen, P., Newton, L., & McCarthy, S. (2012). Developing a science teacher education course that supports student teachers' thinking and teaching about the nature of science. *Research in Science & Technological Education*, 30(1), 29–47. <https://doi.org/10.1080/02635143.2012.671767>
- Zion, M., Schwartz, R. S., Rimerman-Shmueli, E., & Adler, I. (2020). Supporting teachers' understanding of nature of science and inquiry through personal experience and perception of inquiry as a dynamic process. *Research in Science Education*, 50(4), 1281–1304. <https://doi.org/10.1007/s11165-018-9732-9>

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