

# 잠재의미분석방법을 통한 학교보건 연구동향 분석

신선희<sup>1</sup> · 박윤주<sup>2</sup>

한국교원대학교 교육정책전문대학원 겸임교수<sup>1</sup>, 한국교원대학교 교육정책전문대학원 부교수<sup>2</sup>

## Trend Analysis of School Health Research using Latent Semantic Analysis

Seon-Hi Shin<sup>1</sup> · Youn-Ju Park<sup>2</sup>

<sup>1</sup>Adjunct Professor, The Graduate School of Educational Policy and Administration, Korean National University of Education

<sup>2</sup>Associate Professor, The Graduate School of Educational Policy and Administration, Korean National University of Education

### ABSTRACT

**Purpose:** This study was designed to investigate the trends in school health research in Korea using probabilistic latent semantic analysis. The study longitudinally analyzed the abstracts of the papers published in 「The Journal of the Korean Society of School Health」 over the recent 17 years, which is between 2004 and August 2020. By classifying all the papers according to the topics identified through the analysis, it was possible to see how the distribution of the topics has changed over years. Based on the results, implications for school health research and educational uses of latent semantic analysis were suggested. **Methods:** This study investigated the research trends by longitudinally analyzing journal abstracts using latent dirichlet allocation (LDA), a type of LSA. The abstracts in 「The Journal of the Korean Society of School Health」 published from 2004 to August 2020 were used for the analysis. **Results:** A total of 34 latent topics were identified by LDA. Six topics, which were 「Adolescent depression and suicide prevention」, 「Students' knowledge, attitudes, & behaviors」, 「Effective self-esteem program through depression interventions」, 「Factors of students' stress」, 「Intervention program to prevent adolescent risky behaviors」, and 「Sex education curriculum, and teacher」 were most frequently covered by the journal. Each of them was dealt with in at least 20 papers. The topics related to 「Intervention program to prevent adolescent risky behaviors」, 「Effective self-esteem program through depression interventions」, and 「Preventive vaccination and factors of effective vaccination」 appeared repeatedly over the most recent 5 years. **Conclusion:** This study introduced an AI-powered analysis method that enables data-centered objective text analysis without human intervention. Based on the results, implications for school health research were presented, and various uses of latent semantic analysis (LSA) in educational research were suggested.

**Key Words:** Trend analysis; Latent semantic analysis; School health research; Latent dirichlet allocation (LDA)

## 서론

### 1. 연구의 필요성

최근 다양한 학문 분야에서 그동안 수행되어온 연구 성과를

체계적으로 분석하는 연구동향에 대한 연구들이 활발히 이루어지고 있다. 이러한 연구동향에 대한 분석 연구는 크게 두 가지 목적을 제시하고 있다. 첫째, 특정 분야의 연구 주제와 방법이 현재에 이르기까지 어떻게 변화되었는지에 대한 분석을 통해 해당 분야의 이론 및 연구방법 등 지식기반을 확충하도록

### Corresponding author: Youn Ju Park

The Graduate School of Educational Policy and Administration, Korean National University of Education, 250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju 28173, Korea.

Tel: +82-43-230-3406, Fax: +82-43-230-3401, E-mail: 20green@knue.ac.kr

Received: Dec 11, 2020 / Revised: Dec 13, 2020 / Accepted: Dec 14, 2020

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

한다[1]. 둘째, 연구동향의 분석을 통해 변화된 환경의 변화 및 요구에 대응하는 연구 방향을 제시함으로써 미래 연구의 발전적 방향을 모색하도록 한다. 이러한 관점에서 연구 동향 분석은 ‘연구에 대한 연구(research on the research)’로서 특정 학문 분야의 종전 연구 성과를 분석하고, 변화의 특징을 도출함으로써 향후 연구 방향을 제시하고, 보다 미래지향적인 연구를 촉진하는데 목적을 두고 있다 할 것이다[2].

이와 같은 연구동향분석의 필요성에도 불구하고 학교보건 분야에 대한 연구동향분석은 타 학문영역에 비하여 현저히 미흡한 실정이라 할 것이다. 국내에서 학교보건 관련 연구동향을 종합적으로 분석한 연구는 10여 년 전에 이루어진 ‘학교보건 관련 국내 연구동향 분석’[3], ‘최근 학교보건의 연구동향 분석’[4]에 불과하고, 2010년 이후에는 학교보건 분야만을 중점적으로 분석한 관련 연구가 전혀 없는 것으로 조사되었다. 또한, 종전 연구동향분석은 연구주제나 학술지 논문을 기초로 일차원적 데이터에 대한 단순한 패턴분석과 시계열 데이터에 대한 추세분석이 주로 이루어져 연구자의 주관적·자의적 판단이 개입될 개연성으로 인해 본질적인 연구의 한계가 있었다[5].

최근 인공지능 등의 발달로 연구동향 분석에 있어서도 보다 과학적·객관적인 접근이 가능하게 되었다. 대표적인 연구방법으로 최근 텍스트의 심층적 의미 분석을 활용하는 잠재의미 분석(Latent Semantic Analysis, LSA)을 들 수 있다. LSA에는 다양한 모형들이 있지만 대표적으로 많이 사용되는 모델은 잠재디리클레할당(Latent Dirichlet Allocation, LDA)로서 보편적으로 사용되고 있다. LDA는 AI 알고리즘으로 텍스트에 잠재되어 있는 주제들을 조건확률모형을 통해 분석하는 기법으로 단어, 구, 문장 등에 대한 텍스트마이닝을 통해 심도있는 분석을 가능하게 한다. 최근 우리나라 학계에서는 LDA의 이러한 장점으로 정책 동향이나 연구 동향을 분석하려는 시도가 활발히 이루어지고 있다. Hong [6]는 SW교육 정책을 다룬 409건의 지역 뉴스 기사들을 LDA를 사용해 분석하여 기사에서 출현한 잠재주제들을 제시하고 정책 발전을 위한 시사점을 모색하였다. Kim과 Pack [7]은 교육부의 대학교조개혁 평가 관련 보도 자료 25개와 국내 종합일간지의 기사 625개를 수집하여 LDA를 사용하여 분석한 후 관련 정책 쟁점들을 분석하였다.

본 연구는 LDA를 학교보건연구 동향 분석연구에 적용하여 학교보건연구의 주제들이 어떻게 변화되어 왔는지에 대한 논의를 통해 학교보건연구의 정체성 및 미래 연구방향에 대하여 고찰하였다. 나아가 연구대상으로는 우리나라 학교보건 관련

학회지 중 대표적인 학회지인 학교보건학회지에 수록된 논문으로 제한하여, 학교보건학회지의 정체성과 향후 학술지 발전 방향에 시사점을 제시하고자 하였다.

## 연구 방법

### 1. 잠재의미분석 모형의 이론과 활용

최근 인공지능 분야 중 기계학습과 자연어 처리 분야에서 적용이 확산되고 있는 확률적 잠재의미분석의 대표적 모형으로 LDA (Latent Dirichlet Allocation)와 CTM (Correlated Topic Model)를 들 수 있다[8-10]. 이들은 수량화된 데이터가 아닌 문서들로 이루어진 말뭉치 데이터의 심층적 내용 구조를 분석하여 내재적인 주제들을 찾아내는 텍스트마이닝 기법을 사용한다. LDA는 특정주제를 다룬 텍스트 문서는 그 안에 단어들의 분포가 존재하고 어떤 확률분포에 따라 혼합되어 존재한다는 전제에서 출발한다. 따라서 적합한 확률모형을 사용하여 데이터 분석을 시도하면 궁극적으로 문서를 대표하는 잠재주제를 찾아낼 수 있고 말뭉치에 존재하는 주제들의 세트를 찾아낼 수 있다고 설명한다[9,11].

LDA 관점에서 보면 영문 초록은 말뭉치로 구성되어 있는 바, 말뭉치에 내재되어 있는 각 잠재주제들의 확률분포로부터 나온 무작위 혼합표본으로서 위계적 구조를 가진 한 확률모형으로 볼 수 있다[9,11]. LDA는 분석단위가 단어, 구, 문장 등 여러 형태로써 단어나 구의 경우 문장에서의 순서 정보는 사용되지 않기 때문에 ‘a bag of words’ 모형이라고도 불리기도 한다. 모형의 모수치를 추정할 때 잠재주제들의 수는 알려져 있다고 가정하는 바, 가령 잠재주제들의 수가  $K$ 일 때 LDA 모형은 다음과 같이 세 단계로 구성된다:

[1 단계] 각 잠재주제 내에서 용어들의 확률분포인  $\beta$ 는  $\delta$ 를 모수치로 하는 디리클레 확률분포를 따른다고 가정한다. 이때 잠재주제들은 상호 독립적이라고 가정한다.

$$\beta \sim \text{Dirichlet}(\delta)$$

[2 단계] 각 문서의 잠재주제의 확률분포인  $\theta$ 는  $\alpha$ 를 모수치로 하는 디리클레 확률분포를 따른다고 가정한다.

$$\theta \sim \text{Dirichlet}(\alpha)$$

[3 단계] 한 문서의 단어의 수를  $N$ 이라고 하고 어떤 잠재주제,  $z_i$ 가 선택되었을 때  $i$ 번째 단어,  $w_i$ 는 다항확률분포를 따른다고 가정한다.

(i)  $w_i$ 에 대하여  $z_i \sim \text{Multinomial}(\theta)$ .

(ii)  $z_i$ 와  $\beta$ 가 정해졌을 때 한 문서에서 어떤 단어  $w_i$ 는 다항

확률분포를 따른다. 이 조건확률을 기호로 표현하면 다음과 같다:  $P(w_i | z_i, \beta)$ .

위와 같이 구성된 LDA 모형의 모수치를 추정하기 위하여 한 개의 초록을 기준으로 데이터의 로그우도를 표현하면 아래 수식(1)과 같다.  $M$ 개의 초록으로 이루어진 전체 말뭉치에 대한 로그우드는 각 초록의 주변 확률분포를 모두 곱하여 로그를 취하면 얻을 수 있다. 아래 수식(2)는 전체 말뭉치(Corpus;  $C$ )에 대한 로그우드를 보여 준다.

$$\text{Log} \left( p(w|\alpha, \beta) = \text{Log} \int p(\theta|\alpha) \left\{ \prod_{z_i} \left[ \prod_{i=1}^N p(w_i|z_i, \beta) p(z_i|\theta) \right] \right\} d\theta \right) \quad (1)$$

$$\text{Log} \left( p(C|\alpha, \beta) = \text{Log} \prod_{j=1}^M \int p(\theta|\alpha) \left\{ \prod_{z_i} \left[ \prod_{i=1}^N p(w_i|z_i, \beta) p(z_i|\theta) \right] \right\} d\theta \right) \quad (2)$$

LDA 모형의 모수치 추정방식에는 라플라스근점 방식, 변분근점(VEM) 방식, MCMC 방식 등이 소개된 바 있다[11]. 본 연구에서는 topicmodels 이라는 R 패키지를 사용하였으며 이 패키지가 제공하는 LDA 모수 추정방식으로 VEM과 Gibbs 샘플링을 사용하였다. 추후로 이 두 방식을 각각 LDA-VEM과 LDA-Gibbs 샘플링으로 언급하였다.

한편, 앞서 언급했듯이 LDA를 비롯한 확률적 잠재의미분석 모형들을 적용할 때 잠재주제들의 수는 알려져 있다고 가정한다. 하지만 실제 데이터 분석 상황에서 연구자는 잠재주제들의 수를 미리 알 수 없다. 따라서 잠재주제들의 수를 2~K 개로 다양하게 가정하고 모형을 여러 번 돌리게 된다. 즉 잠재주제 수에 따라 다양한 LDA 모형들을 적용한다. 그리고 분석 결과를 비교한 후 현행 데이터에 가장 적합한 수를 가진 LDA 모형을 최종적으로 선택하여 결과를 해석한다. 잠재주제들의 적합한 수를 판단하는데 사용되는 대표적인 지표들은 Perplexity, Empirical likelihood, Harmonic mean, AIC, BIC 등과 같은 우도-기반 수량화 지표들이 활용되어진다[12-15]. 하지만 우도-기반 지표들의 경우 적용 시 신뢰하기 어려운 결과를 초래하기도 하여 보완의 필요성이 제기되고 있다[16,17]. 이를 보완하는 방법으로 문서의 의미구조가 유사한 것들로 실제로 묶이는지 연구자가 살펴보거나, 잠재의미분석 모형을 활용한 것과 연구자가 직접 분석한 것의 일치 정도에 따라 모형의 적합도를 가늠하기도 한다[16,18-20]. 복잡한 통계모형을 사용하는 분석에서 최상의 모형을 선택하는데 모형의 해석력(interpretability)이 판단의 기준 중 하나가 되는 경우는 드물지 않다. 잠재집단분석(latent class analysis)에서도 해석력은 모형선택의 중요 준거 중 하나이다[21].

본 연구에서는 R 패키지인 topicmodels에서 제공하는 Per-

plexity, AIC, BIC를 사용하였으며 학교보건연구 분야 전문가를 포함한 연구팀의 내용분석 의견과 가장 일치하는 모형을 최종으로 선택하여 결과를 해석하였다. 이 통계치들은 값이 작을수록 더 적합한 모형임을 가리킨다. Perplexity의 경우 Grun 등[11]의 논문에서 제시한 방식을 따라 데이터를 무선적으로 10 세트로 나누어 하나는 검사용 데이터로, 나머지는 훈련용 데이터로 사용하여 Perplexity를 계산하였다. 잠재주제 수는 2~50까지 하나씩 증가하며 모형을 구분하였다. 따라서 각 잠재주제 수마다 10개의 Perplexity가 산출되었으며 더 나아가 추정방식인 LDA-VEM과 LDA-Gibbs 샘플링에 따라 별도의 세트로 각각 산출되었다. AIC와 BIC는 모형의 예측 정도보다는 데이터에 대한 모형 적합도에 초점을 둔 통계치이므로 입력정보를 최대화할 수 있는 전체 말뭉치에 적용하여 산출하였다. 따라서 이 통계치들은 잠재주제 수마다 각 1개의 추정치를 얻었다.

가장 적합한 잠재주제 수를 확정한 후 이 수만큼의 잠재주제를 추출한 모형을 최종 모형으로 결정하고 이 모형을 통해 각 초록의 잠재주제 확률분포인  $\theta$  벡터를 얻는다. 한 초록의  $\theta$  벡터에서 가장 큰 확률값과 연계되어 있는 잠재주제에 그 초록을 배정함으로써 공통의 주제를 반영하는 초록들을 묶어 내었다. 또한 각 잠재주제와 가장 밀접한 연관성을 보이는 상위 용어들 즉 모형의  $\beta$  값이 가장 큰 순위에 해당하는 용어들을 추출하였다. LDA의 심층적 내용분석을 통해 출현한 잠재주제들에 대하여 연구팀은 동일 주제군으로 묶인 초록들의 내용을 검토한 후 잠재주제들에 대한 명칭을 생성하였다. 이 과정에서 LDA가 도출한 각 잠재주제별 상위 용어들을 참고하였다. 이러한 LDA의 결과를 바탕으로 본 연구는 해마다 학술지에 실리는 연구들이 다루는 대표적인 주제 분포가 어떻게 달라지는지 살펴보았다.

## 2. 데이터와 분석 도구

2004년부터 2020년 8월까지 17년 동안 학술지 「한국학교보건학회지」에 게재된 논문의 영문 초록 389개 중 초록이 없거나 기준에 맞지 않는 7개를 제외하고 총 382개를 분석에 사용하였다. 연구 기간에 속하는 대부분의 초록은 ‘Purpose’, ‘Methods’, ‘Results’, ‘Conclusion’이라는 헤딩과 함께 네 부분으로 구성되어 있었다. 분석 자료는 연구자가 다루려는 주제를 심층적으로 분석하기 위해 영문초록의 ‘Purpose’와 ‘Conclusion’에 속하는 텍스트를 대상으로 분석하였다.

LDA 분석에 앞서 초록의 각 영어 문장은 자연어 처리 분석

을 통해 단어의 기본형인 레마(lemma)로 파싱(parsing)하여 품사별로 UPOS 태그(Universal POS tags)를 붙였다. 품사 중 명사만을 추려서 분석에 사용하였다. 연구 주제의 심층적 분석을 위하여 'school health'와 같은 보편적이며 광범위하고 모호한 단어는 제외하였다. 이 외에도 'purpose', 'goal', 'aim', 'findings', 'results', 'data', 'methods' 등 연구의 주제와 상관 없는 보편적 단어들도 분석에서 제외하였다. 파싱과 LDA 분석을 위하여 R 버전 4.0.2와 RStudio 버전 1.3.1073을 사용하였다.

## 연구결과

### 1. 연구결과1

총 382개 영문초록이 사용되었으며 파싱 후 얻어진 품사별 빈도는 Figure 1과 같다. 본 LDA 분석에 포함된 품사는 명사로 한정하여, 앞서 설명한 기준에 의해 일부 용어들을 제외하고 1,123개의 용어와 382개 초록으로 이루어진 382×1123 크기의 빈도행렬을 입력 데이터로 사용하였다.

### 2. 연구결과2

잠재주제 수를 결정하기 위하여 AIC, BIC와 10개의 훈련용-검사용 무선분할 데이터 세트에서 계산된 Perplexity 통계량을 산출하여 비교하였다. Figure 2는 10개 데이터 세트에 적용된 두 모수치 추정방식에 따른 Perplexity의 패턴을 보여준다. LDA-Gibbs 샘플링 추정방식으로 산출된 Perplexity의 최솟값이 가리키는 잠재주제 수는 무선적으로 분할된 10개 데이터 세트 사이에 일정하지 않았다. 이에 각 잠재주제 수별로 10개

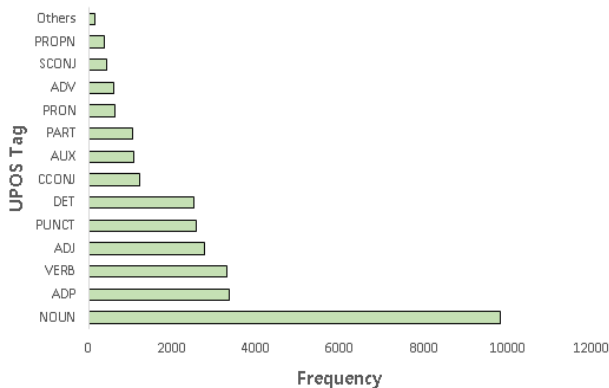


Figure 1. Frequency distribution of UPOS tags.

의 데이터 세트에서 나온 Perplexity의 평균값을 계산한 후 이 평균값이 가장 작은 잠재주제 수를 찾아 최종 모형을 확정하였다. 확정된 모형의 잠재주제 수는 34개였다.

### 3. 연구결과3

잠재주제 수가 34개인 LDA모형에 의해 도출된 주제의 명칭을 붙이기 위하여 공통의 잠재주제로 묶인 초록들의 내용을 살펴 보았으며 각 잠재주제와 가장 긴밀한 연관성을 보이는 상위 용어들, 즉 각 잠재주제 내에서 상위  $\beta$ 값을 보인 용어들을 추출하여 참고하였다. 아래 Table 1은 LDA-Gibbs 샘플링 방식이 추출한 34개 잠재주제들과 초록 수 및 발표년도, 각 주제별 연관성이 높은 상위 8개 용어를 보여준다. 그리고 각 잠재주제의 명칭은 연구팀이 공통 잠재주제로 묶인 초록들의 내용을 분석한 결과 붙인 명칭이다.

잠재주제별 초록의 발표 횟수를 살펴보면 학생들의 건강지식, 태도, 행동(주제 21번), 우울증 개입을 통한 효과적인 자아존중감 프로그램(주제 27번), 학생 스트레스 요인(주제 28번), 청소년기 위험 행동 예방 개입 증진 프로그램(주제 16번), 청소년 우울증과 자살 예방(주제 5번), 성교육 교육과정(주제 26번) 주제들이 이 학술지에서 가장 빈번하게 다루어진 것들을 알 수 있다. 각 주제는 최소 20개의 연구에서 다루어진 것으로 파악되었다.

아래 Figure 3은 빈번하게 다루어진 주제들의 연도별 발표 추이를 각각 보여준다. 막대그래프의 서로 다른 색깔은 각기 다른 주제들을 나타내고 있다. Figure 3의 주제들은 17년 동안에 대체로 꾸준히 발표되었음을 알 수 있다. 다만 「청소년기 우울증과 자살 예방」과 관련된 주제 5는 다소 늦게 다루어지기 시작해서 가장 최신평까지 발표되는 주제였다. 「청소년기 위험

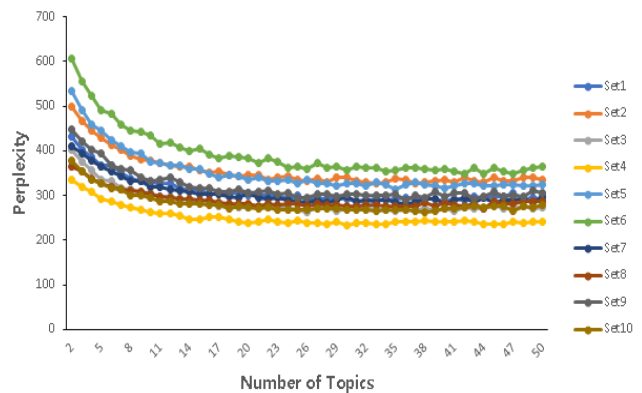


Figure 2. Pattern of perplexity by LDA-Gibbs sampling.

**Table 1.** 34 Latent Topics Identified by LDA

Latent topic	Publication year (Number of abstracts)	Number of abstracts	Top 8 terms	Name of latent topic
1	2020 (2) 2019 (1) 2016 (1) 2015 (1) 2013 (3) 2012 (1) 2010 (2) 2009 (1) 2007 (1) 2006 (1)	14	Addiction internet student relationship ability phone intervention correlation	Student addition to internet
2	2018 (1) 2017 (2) 2011 (1) 2010 (1) 2005 (1)	6	Group effect therapy intervention happiness aggression bath edema	Effect of group treatment
3	2017 (1) 2016 (1) 2012 (1) 2010 (2) 2007 (1) 2006 (1) 2005 (1) 2004 (2)	10	Factor status age life symptom characteristic fatigue perception	Socioeconomic factor and status
4	2018 (1) 2014 (1) 2011 (1) 2010 (1) 2009 (1) 2008 (1) 2007 (1) 2004 (1)	8	Type girl dysmenorrhea gender body consciousness boy equality	Female adolescent gender sensitivity and menstrual pain
5	2020 (1) 2018 (1) 2017 (3) 2016 (3) 2015 (2) 2014 (3) 2013 (1) 2012 (3) 2010 (1) 2009 (2) 2008 (1)	21	Depression adolescent ideation suicide sleep factor prevention support	Adolescent depression and suicide prevention
6	2019 (1) 2017 (1) 2013 (3) 2012 (1) 2010 (2) 2009 (2) 2008 (1) 2007 (1) 2006 (1) 2005 (2) 2004 (1)	16	Program promotion activity student university intervention relationship level	Health promotion activity
7	2016 (1) 2013 (1) 2009 (1) 2008 (2) 2007 (1) 2005 (1)	7	Obesity weight management factor diabete type control condition	Management factor for child obesity
8	2020 (1) 2018 (1) 2017 (1) 2016 (1) 2009 (1) 2005 (1)	6	Effect symptom student syndrome alexithymia dermatitis dust effort	Effects of air quality on students
9	2017 (1) 2016 (1) 2015 (3) 2012 (2) 2010 (1) 2008 (1) 2004 (1)	10	Area system community model problem disease response future	Regional system model for disease problems
10	2017 (2) 2010 (1) 2009 (1) 2008 (1) 2006 (1)	6	Center awareness pregnancy quality toothpaste amount daycare number	Pregnancy, mouth, meal, etc.
11	2019 (1) 2012 (1) 2011 (1)	3	Indicator statistic approval need population graduate persistence utilization	School Health Indicators
12	2020 (2) 2018 (1) 2016 (2) 2015 (1) 2013 (2) 2009 (3) 2008 (1) 2005 (2)	14	Program effectiveness prevention student empowerment management violence ability	Effective prevention program for students
13	2019 (1) 2018 (2) 2017 (1) 2016 (3) 2014 (1) 2013 (1) 2012 (1) 2006 (1) 2005 (1)	12	Student factor program parent support need peer addiction	Parental support factor for student addition
14	2016 (1) 2015 (1) 2010 (1) 2008 (1) 2006 (1)	5	Satisfaction appearance degree student difference importance gender recognition	Appearance satisfaction degree
15	2020 (2) 2019 (1) 2016 (2) 2015 (1) 2014 (2) 2010 (1) 2009 (1) 2008 (1) 2007 (1) 2006 (1)	13	Smoking prevention smoke student cigarette influence alcohol cessation	Smoking prevention and factors for no smoking
16	2020 (1) 2019 (4) 2018 (2) 2017 (3) 2016 (3) 2015 (1) 2013 (2) 2011 (1) 2009 (2) 2007 (2) 2004 (1)	22	Adolescent risk behavior factor behaviors program intervention promotion	Promotion program for intervening adolescent risky behavior
17	2017 (2) 2014 (1)	3	Job demand nurse burnout satisfaction engagement resource teacher	Job demands for health teacher

**Table 1.** 34 Latent Topics Identified by LDA (Continued)

Latent topic	Publication year (Number of abstracts)	Number of abstracts	Top 8 terms	Name of latent topic
18	2020 (1) 2019 (1) 2018 (2) 2016 (2) 2015 (1) 2011 (1) 2010 (2) 2005 (1)	11	Adolescent family relationship community drug function hope level	Support for the relationship between adolescent family and the local community
19	2017 (1) 2015 (1) 2011 (4) 2010 (1) 2009 (1) 2007 (1) 2006 (2)	11	Student education life training aid support skill Cpr	Student health education
20	2016 (1) 2015 (1) 2014 (1)	3	Student difference confidence bullying sense learning bioethic characteristic	Student bullying
21	2017 (3) 2016 (2) 2015 (4) 2014 (2) 2013 (1) 2012 (2) 2011 (3) 2010 (3) 2009 (2) 2008 (1) 2007 (1) 2006 (2) 2005 (1) 2004 (2)	29	Education student knowledge attitude effect program behavior media	Students' knowledge, attitudes, behavior
22	2009 (2)	2	Youth disparity inequality adolescent population region country evidence	Adolescence inequality
23	2018 (1) 2016 (1) 2014 (2) 2012 (1) 2011 (1) 2009 (2) 2008 (1) 2007 (1)	10	Safety education injury system accident review treatment compensation	Safety education and management system for accident and injury
24	2019 (1) 2017 (1) 2015 (1) 2013 (1) 2008 (1) 2004 (2)	7	Policy problem support placement process teacher protocol change	Policy matters on environment, pollution, etc.
25	2017 (1) 2016 (1) 2012 (3) 2011 (1) 2006 (1) 2004 (1)	8	Teacher job management competency education performance development need	Personnel management for teachers
26	2020 (1) 2019 (1) 2018 (1) 2017 (1) 2015 (1) 2014 (1) 2013 (1) 2012 (1) 2010 (2) 2009 (2) 2006 (2) 2005 (4) 2004 (2)	20	Education sex teacher curriculum material teaching contents educator	Sex education and curriculum, and teacher
27	2020 (1) 2019 (2) 2018 (2) 2017 (3) 2016 (1) 2015 (1) 2014 (3) 2013 (1) 2012 (1) 2011 (3) 2010 (2) 2009 (2) 2008 (2) 2007 (3) 2004 (1)	28	Self esteem program efficacy effect student image depression	Effective self-esteem program through intervention for depression
28	2019 (1) 2018 (3) 2017 (3) 2016 (1) 2015 (1) 2014 (3) 2012 (2) 2011 (2) 2010 (2) 2007 (2) 2006 (1)	21	Stress factor program student level grade job characteristic	Factors for students' stress
29	2020 (1) 2019 (2) 2018 (1) 2017 (3) 2016 (1)	8	Factor intention vaccination effect knowledge benefit practice disaster	Preventive vaccination and factors for effective vaccination
30	2018 (1) 2017 (1) 2016 (3) 2015 (1) 2010 (1) 2008 (1)	8	Drink student consumption energy education caffeine information city	Providing information on high caffeine drink
31	2018 (2) 2015 (3) 2014 (1) 2013 (1) 2012 (1) 2010 (1) 2009 (2) 2007 (1) 2005 (1)	13	College student breakfast habit behavior disorder time woman	Unversity students' health behavior
32	2016 (1) 2012 (1) 2010 (1) 2007 (1) 2006 (1) 2005 (1) 2004 (1)	7	Survey year death system development direction incidence trend	Information and trends of accident and death
33	2017 (1) 2012 (1) 2011 (1) 2010 (3) 2006 (1) 2005 (1) 2004 (3)	11	Child behavior order mother attention deficit influence level	Parental influence on child health behavior
34	2019 (1) 2013 (1) 2011 (1) 2010 (1) 2009 (1) 2007 (1) 2005 (1) 2004 (2)	9	Problem student family care rate drinking characteristic prevalence	Parental care for student problematic behaviors including drinking

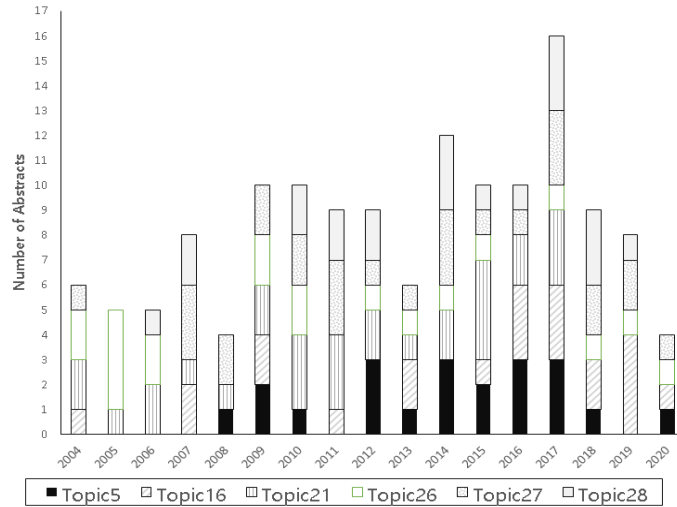


Figure 3. Yearly trends of frequent topics.

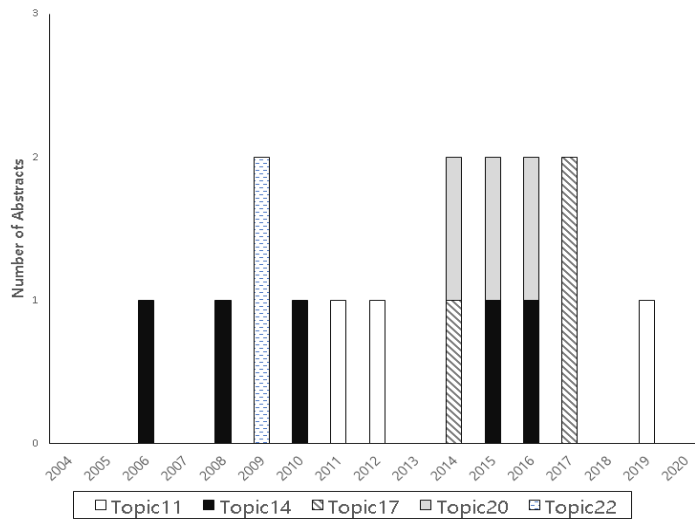


Figure 4. Yearly trends of infrequent topics.

행동 예방 개입 증진 프로그램」과 관련된 주제 16과 「성교육 교육과정」과 관련된 주제 26, 그리고 「우울증 개입을 통한 효과적인 자아존중감 프로그램」과 관련된 주제 27은 최근 3년 빠짐없이 등장한 주제들이었다.

이와 반대로 청소년 건강형평성(주제 22번), 학교보건지표(주제 11번), 보건교사 직무(주제 17번), 학생 따돌림(주제 20번), 외모 만족도(주제 14번) 주제들은 이 학술지에서 상대적으로 드물게 다루어진 것들이었다. 이 범주에 속하는 각 주제는 5개 이하의 연구에서만 주로 다루어진 내용이었다. 아래 Figure 4는 드물게 다루어진 주제들의 연도별 발표 추이를 보여 준다. Figure 4의 주제들은 비교적 단기간에 드물게 다루어

지는 경향성을 보이고 있다.

Figure 5는 각 잠재주제가 학술지에 등장하는 시기를 보여 준다. 2004년부터 2020년 8월 까지 기간 동안에 34개 잠재주제들이 학술지에 게재된 시기를 도표로 제시하고 있어 연구 주제의 추이로 살펴 본 학술지의 연구 동향을 쉽게 파악할 수 있도록 하였다.

## 논 의

본 연구는 2004년부터 2020년 8월호까지 학교보건학회지에 게재된 연구의 영문초록을 LDA방법을 적용한 학교보건연

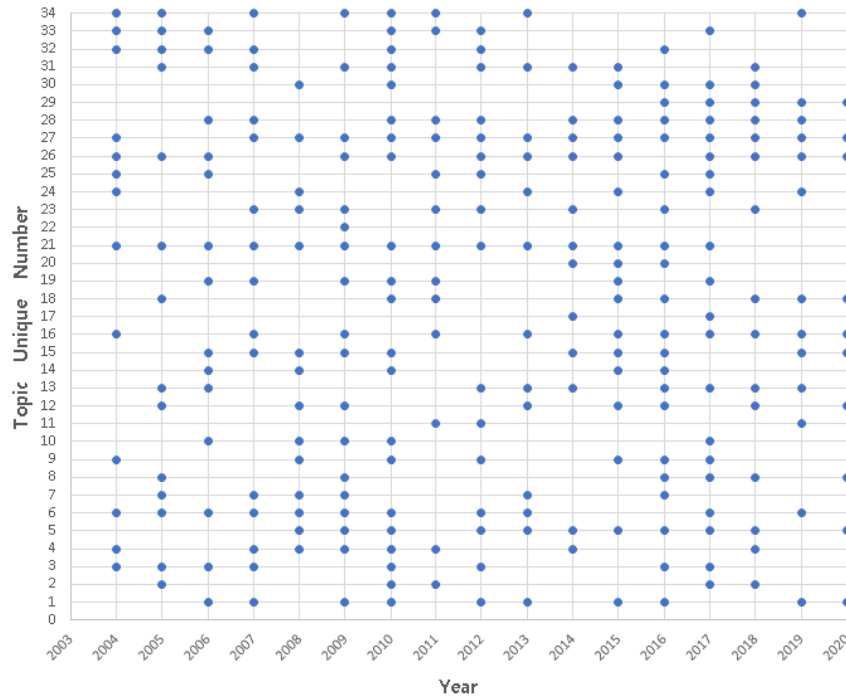


Figure 5. Yearly trends of latent topics.

구동향 분석을 통해 최근 중시되고 있는 연구주제를 다음과 같이 도출하였다: 1) 학생들의 건강지식, 태도, 행동, 2) 우울증 개입을 통한 효과적인 자아존중감 프로그램, 3) 학생 스트레스 요인, 4) 청소년기 위험 행동 예방 개입 증진 프로그램, 5) 청소년 우울증과 자살 예방, 6) 성교육 교육과정 등.

학술지에서 빈번하게 다루어지고 있는 우울증 개입, 학생 스트레스 요인, 청소년 우울증과 자살 예방 등은 정신건강 측면이 강조되고 있는 오늘날 학교건강환경의 변화를 학교보건 학회지 게재 연구들도 적절하게 반영하고 있음을 보여주고 있다. 이와 동시에 건강은 학생 스스로 증진할 수 있는 대상임을 전제하여 학생의 건강증진 역량에 초점을 둔 연구도 증가하는 추세로 건강증진의 보편적 개념이 학교보건연구에도 확산되고 있음을 시사하고 있다. 아울러 위기 청소년에 대한 사회적 이슈를 반영하여 위험행동 예방 또는 개입 프로그램에 대한 연구 또한 증가하는 추세이다. 이러한 연구가 증가되고 활성화된 배경에는 이와 같은 사회적 환경과 2005년 도입되어 매년 시행되고 있는 청소년건강행태온라인 조사결과가 축적됨에 따라 데이터 기반 연구의 토대가 마련된 것이 일부 영향을 주었을 것으로 판단된다. 또한 최근 미투 운동 등 사회적으로 성 문제가 핵심이슈로 등장하면서 학생 성교육 관련 연구들이 종전에 비해 확연하게 증가하는 동향임을 알 수 있었다. 학교보건 연구동향에 대한 본 연구의 결과는 학교보건학회의 게재 논

문들이 다루는 주제들이 보편적으로 연구 수행 당시의 학교보건 이슈를 적절히 반영하고 있음을 시사하고 있다.

2000년 이후 건강형평성에 대한 사회적 관심이 높아지고 있으며, 국민건강증진종합계획에서도 건강불평등을 중요한 정책 주제로 다루고 건강형평성을 제고하기 위하여 지표 개발 등의 다양한 노력이 이루어지고 있다. 그러나 학교보건 연구동향에 대한 본 연구의 결과 학생들의 건강형평성에 대한 연구는 거의 다루어지고 있지 않음을 보여준다. 학교보건 연구 측면에서 학교연구집단의 건강형평성을 분석하고 연구할 수 있는 종합적인 학교보건 관련 지표가 마련되고 자료가 제공되어야 할 정책적 필요성을 제시하고 있으며 아울러 학교보건 연구 측면에서 건강형평성에 대한 관심이 더욱 높아져야 함을 나타낸다.

이와 함께 학교보건의 특성을 반영하는 깊이 있는 처방적·실천적 성격의 연구가 이루어지지 못하는 등 미흡한 측면도 분석되어 이에 대한 논의가 필요한 것으로 사료된다. 주요한 분석결과는 다음과 같다. 첫째, 학교건강환경을 주제로 하는 연구는 극히 낮은 빈도로 수행된 것으로 나타났다. 2000년 이후 학생들의 단순한 질병 예방이나 위기 상황의 해결에서 보다 적극적으로 학교보건문제를 해결하고 학교구성원의 건강을 증진하기 위한 방향으로 접근하고 있다[22,23]. 학교보건의 통합적 특성으로 건강교육과 건강서비스뿐만 아니라 보다 쉽게 건강한 행동을 하는 학교환경을 조성하는 분야도 중요한 연구주



제이나 이에 대한 연구는 거의 다루어지지 않고 있다.

향후 학교건강증진 관련 국가정책, 협력적 건강증진 거버넌스, 학교건강증진 평가체계 등을 포함한 학교보건에 영향을 미치는 체계적·환경적 측면에서의 연구들이 요구되고 있다. 둘째, 건강 초점에서 웰빙으로 개념을 확장하고 있는 학교보건의 선진 동향을 반영하는 연구들이 미흡한 것으로 나타났다. 최근 선진각국의 학교교육정책의 목적을 개인과 사회의 웰빙으로 새로운 방향 설정을 하면서 학교보건연구분야도 이러한 변화를 반영하는 연구들이 활발히 수행되고 있으나, 우리나라의 경우 학교보건분야의 주된 연구주제로 보기에는 아직 미흡하여 초기 단계에 머무르고 있는 실정으로 분석되었다. 향후 학교보건연구는 신체적 건강과 정신적 건강이 통합되고, 나아가 교육과 건강이 통합되는 웰빙의 관점에서 접근함으로써 새로운 학교건강증진의 전략 및 과제를 모색할 필요가 있다.

이와 같이 본연구의 분석결과는 지금까지 우리나라 학교보건연구는 외부 환경의 급속한 변화에 따라 연구주제 또한 적절하게 변화하여 왔음을 보여주고 있으나, 미흡한 측면 또한 제시하고 있어 향후 이와 관련한 심도 있는 논의가 요구된다. 학교보건 연구동향 분석결과에 대한 논의와 함께 연구의 분석방법으로 활용한 LDA 적용에서 드러난 방법론적 한계점에 대해서도 토의가 필요하다. 토픽명 선정 시 전문가의 판단에 의한 라벨링은 타당성 검증에 대한 지속적 보완이 필요하며 빈도가 높은 상위 단어 중 연구자의 주관적 판단에 의한 단어 선정은 객관성 확보를 위한 판단의 여지가 있다. 이러한 LDA 연구의 본질적 한계는 향후 지속적으로 보완되어야 할 것이며, 연구의 객관성 확보를 위하여 타 연구방법을 병행 사용하는 것도 바람직한 방법이라 할 것이다.

## 결 론

연구동향 분석은 현재까지 수행되었던 연구 실태 및 성과를 점검하는 과정을 통해 미래 연구방향에 대한 시사를 얻는 목적에 초점을 두고 있다. 본 연구는 LDA를 활용하여 학술지 「한국학교보건학회지」에 발표된 논문들에 대한 심층적 내용분석을 통해 주제 분포가 세월이 흐를수록 변화하는지 여부와 변화 형태를 파악하였다. 이를 통해 우리나라 학교보건연구에 대한 동향과 추이를 가늠해 볼 수 있었으며 분석 결과를 바탕으로 교육적 시사점을 제시하였다. 최근 지구촌 환경의 급속한 변화로 인한 코로나 등 건강문제의 심각한 위협으로 학생의 정상적 일상을 방해하면서 정신 건강에 심각한 부정적 영향을 미치고 있다. 이러한 상황에서 학교보건은 그 어느 때 보다 중요시

되고 있는 바, 학교보건 연구는 외부 환경의 급격한 변화에 대응하기 위한 지식기반을 확충하고, 나아가야 할 방향 및 과제를 제시해야 하는 책무를 갖는다 할 것이다. 특히, 학교보건은 각 나라의 역사·정치·경제·사회·문화·교육체계의 복합적 특징을 반영하기 때문에 각국의 환경에 적합한 고유한 연구들이 수행되어야 한다. 한국학교보건 학회지에 게재된 연구들의 동향분석을 통해 드러난 본 연구의 분석결과는 학교보건연구의 주제들은 대체적으로 시대적, 환경적 요구를 반영하여 변화되어 온 것으로 나타났다. 향후 보완해야 할 연구과제로서 교육과 건강을 통합적으로 바라보는 웰빙 관점에서의 학교보건 접근과 함께 이를 지원하는 국가 정책 및 거버넌스 차원의 연구를 필요로 하고 있다.

한편, 본 연구에서 사용한 LDA는 객관적·과학적 텍스트 심층분석 도구로 최근 활용이 확산되고 있는 분석방법으로 정신건강 등 심층 분석을 요구하는 학교보건의 특성상 타 분야에 비하여 LDA의 활용도가 보다 크다 할 것이다. 또한 데이터에 바탕을 둔 연구 분석방법으로 연구자의 자의적·주관적인 해석이라는 비판에서 벗어나 신뢰도와 타당도가 높은 분석 결과를 도출할 수 있는 장점을 가진다. 향후 학교보건연구 분야에서 LDA를 적용한 연구가 증가할 것으로 사료되는 바, 본 연구는 학교보건연구에서 아직 보편화되지 않고 있는 LDA를 연구방법으로 채택한 선도적 연구라는 점에서 연구의 부가적인 의미를 가진다 할 것이다.

## CONFLICTS OF INTEREST

The authors declared no conflict of interest.

## REFERENCES

1. Robert Donmoyer, Michael Imber, James Joseph Scheurich. The knowledge base in educational administration: multiple perspectives. Albany: State University of New York Press; 1995. p. 1-31.
2. Bae SH, Jang HY, Oh SH, Jang CS, Lee TH. Vocational education research in Korea: a trend analysis of the Journal of Vocational Education Research. Journal of Vocational Education. 2013;32(2):45-71.
3. Jung JS, Kim JS. The analysis of the research trends related to school health in Korea. Journal of Korean Society of School Health. 2004;17(1):85-95.
4. Kwon SJ. Anaysis of research trends on school health. Journal of Korean Academy of Community Health Nursing. 2008;19(1):101-111.

5. Jo HI, Kim JW, Lee BG. A study on research trends of block-chain using LDA topic modeling: focusing on United States, China, and South Korea. *Journal of Digital Contents Society*. 2019;20(7):1453-1460.  
<https://doi.org/10.9728/dcs.2019.20.7.1453>
6. Hong YH. Topic analysis of software education policy: focused on Busan regional newspapers. *Journal of the Korean Official Statistics*. 2019;24(2):52-77.  
<https://doi.org/10.22886/JKOS.2019.24.2.52>
7. Kim JE, Pack SG. Analysis of issues on the college and university structural reform evaluation using text big data analytics. *Asia-Pacific Education Review*. 2016;17(3):409-436.  
<https://doi.org/10.15753/aje.2016.09.17.3.409>
8. Blei DM, Lafferty JD. A correlated topic model of science. *The Annals of Applied Statistics*. 2007;8(1):17-35.  
<https://doi.org/10.1214/07-aoas114>
9. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research*. 2003;3:993-1022.
10. Zhai C, Massung S. Text data management and analysis: a practical introduction to information retrieval and text mining. San Rafael, California: Association for Computing Machinery and Morgan & Claypool Publishers; 2016. p. 329-388.
11. Grun B, Hornik K. Topicmodels: an R package for fitting topic models. *Journal of Statistical Software*. 2011;40(13):1-30.  
<https://doi.org/10.18637/jss.v040.i13>
12. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest, Hungary: Akademiai Kiado; 1973. p. 267-281.
13. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978;6(2):461-464.  
<https://doi.org/10.1214/aos/1176344136>
14. Tein JY, Coxe S, Cham H. Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*. 2012;20(4):640-657.  
<https://doi.org/10.1080/10705511.2013.824781>
15. Shin S, Seo D, Park M. Effectiveness of Latent Class Analysis (LCA) techniques to handle count data with excessive zeros in topic analysis. Poster session presented at: The 32nd florida artificial intelligence research society; 2019 May 19-22; Florida, USA.
16. Ponweiser M. Latent dirichlet allocation in R [diploma thesis]. Iowa: Institute for Statistics and Mathematics; 2012. p. 1-138.
17. Wallach HM, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. Pittsburgh: In Proceedings of the 26th Annual International Conference on Machine Learning; 2009. p. 1105-1112.
18. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. *Neurocomputing*. 2009;72:1775-1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
19. Chang J, Blei DM. Relational topic models for document networks. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*. 2009;81-88.
20. Griffiths TL, Steyvers M. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(suppl 1):5228-5235.  
<https://doi.org/10.1073/pnas.0307752101>
21. Collins LM, Lanza ST. Latent class and latent transition analysis. Hoboken, NJ: John Wiley & Sons; 2010. p. 77-110.
22. Chang CG. Development of school health indicator system for the health promotion of school children. *Journal of the Korean Society of School Health*. 2012;25:204-213.
23. Park YJ. Looking for a new perspective on school health promotion. *Journal of the Korean Society of School Health*. 2018; 31:157-166. <https://doi.org/10.15434/kssh.2018.31.3.157>