



토픽모델링을 활용한 국내외 수학교육 평가 연구 동향 분석

손 태 권* · 황 성 환**†

*한국교원대학교 대학원 학생, **서울가주초등학교 교사

An Analysis of Domestic and International Research Trends of Assessment in Mathematics Education Using Topic Modeling

Son, Taekwon * · Hwang, Sunghwan **†

*Graduate Student, Korea National University of Education, South Korea, sontaekwon7@gmail.com

**Teacher, Seoul Gaju Elementary School, South Korea, ihwang413@gmail.com

초록. 수학교육 평가는 과거 성취도 측정을 목적으로 시행되던 결과 중심의 평가에서 학생의 사고와 교육과정, 교수·학습, 프로그램, 교사 지식과 질, 정의적 영역에 대한 평가 등으로 그 연구 영역을 넓혀왔다. 이러한 시점에서 그동안 국내 수학교육 평가가 어떠한 방향으로 수행되어왔는지 살펴보기 위해 2000년부터 2020년 8월까지 KCI 등재지에 게재된 237편의 논문과 SSCI에 게재된 857편의 논문을 LDA 토픽 모델링을 활용하여 연구 동향을 면밀히 분석하였다. 그 결과, 국내외 연구는 서로 다른 키워드 분포를 보였으며 국내 연구는 시기별로 집중되는 연구 주제가 달라지는 경향을 보였다. 또한 국내외의 수학교육 평가 연구는 총 19개의 토픽으로 분류되었으며, 그 중 5개의 유사한 토픽 9개의 상이한 토픽이 출현하였다. 최근 연구 관심이 증가한 토픽은 국외의 ‘교육과정 평가’와 국내의 ‘학생의 역량 평가’로 나타났다. 이상의 결과를 바탕으로 국내 수학교육 평가 연구의 향후 발전 방향과 시사점을 도출하였다.

핵심어: 평가, 연구 동향, 비교 연구, 토픽 모델링, LDA

ABSTRACT. Research into assessments in mathematics education has broadened from summative assessments focusing on students achievement to those focusing on students' cognition, curriculum, teaching and learning, teacher knowledge and quality, and affective domains. Considering these changes in assessment research, we examined the characteristics of domestic assessment research in mathematics education by comparing 237 articles published in KCI journals with 857 articles published in SSCI journals from 2000 to 2020 August. We used LDA topic modeling to examine research trends over time. The findings indicated that there were different keyword distributions by period between domestic and international mathematics education journals. There were nineteen research topics in both journals; five topics were similar while nine topics were different. In addition, the hot topics in international and domestic mathematics education journals were found to be curriculum assessment and student competency assessment, respectively. Based on these findings, we discussed practical implications for the development of assessment research in domestic mathematics education.

KEY WORDS: assessment, research trends, comparative study, topic modeling, LDA(Latent Dirichlet Allocation)

† corresponding author

1. 서론

교육은 사회에서 바람직하다고 규정한 목표를 달성하기 위해 이루어는 의도적인 행위이며, 평가는 교육이 의도한 목표를 달성했는지의 여부를 판단하는 수단이다(Chae, 2001). 이처럼 교육과 평가는 불가분의 관계로 얽혀있으며, 교육이 수천 년에 걸쳐 이어져온 것과 같이 평가의 역사도 그 맥을 함께 해왔다(Kilpatrick, 1993).

교육의 지향점과 방법이 시대에 따라 변화하듯, 평가의 흐름 또한 교육의 변혁과정과 함께 변화되어 왔다. 과거의 평가는 학습 후 학습자의 성취수준의 향상을 측정하거나 경쟁을 통한 순위 매김을 목적으로 시행된 총괄평가가 주를 이루었다. 그러나 학습 결과보다는 과정이 강조되고 학생의 성취도가 수업, 교육과정, 교사, 학습 동기, 학습 자료 등 다양한 요인에 영향을 받는다는 사실이 밝혀짐에 따라 평가는 과정중심평가, 교육과정과 교수·학습 평가, 프로그램 평가, 교사 지식과 질 평가, 정의적 영역에 대한 평가 등으로 그 영역을 넓혀왔다(Ministry of Education [MOE], 2017; NCTM, 1995; Wilson, 2007). 이러한 패러다임의 변화에 따라 평가는 다양한 분야에서 활용되어 왔으며, 그 결과는 교육정책과 교육과정 개정 방향을 정립하는 데 중요한 기초자료이자 학생, 교사, 학부모 등에게 환류되어 학생의 학습을 지원하고 개선하기 위한 토대로서의 역할을 수행해왔다(MOE, 2017).

한편, 평가의 패러다임 변화와 그 중요성에도 불구하고 수학교육에서 평가에 관한 연구는 다른 영역에 비해 주목을 받지 못한 것으로 보인다. 예컨대, 1963년부터 2019년까지 국내 수학교육 연구동향을 분석한 Pang, Sun, Cho, Lee, Kim, Kim et al.(2019)은 평가 연구가 전체 연구의 6.71%에 불과하므로 평가 연구의 확산이 필요하

다고 지적하였다. 2010년부터 2016년까지 초등 수학교육 연구 동향을 분석한 Kim & Pang(2017) 또한 평가 연구가 전체 연구의 극히 일부에 불과하므로 균형 잡힌 연구가 수행될 필요가 있음을 주장하였다. 이러한 선행연구들의 결과는 수학교육이 다양한 분야에서 양적·질적으로 성장을 이루었으나 이에 따르는 수학교육 평가 연구는 상대적으로 부족했음을 의미한다. 그러나 평가가 수학교육의 성장과 학생의 성취도 향상을 뒷받침하는 중요한 요소임을 감안할 때, 국내외 수학교육의 평가 연구 동향을 파악하여 어떤 연구가 수행되고 있는지 그리고 어떤 연구가 부족한지에 대한 현 주소를 되돌아보고 향후 수학교육 평가 연구가 나아가야 할 방향에 대해 고민해볼 필요가 있다.

그동안 수학교육 연구 동향을 메타적으로 분석한 연구는 지속적으로 이루어져 왔다. 예컨대, 수학교육 전반(Pang et al., 2019), 교실 담론과 상호작용(Cho, Kwon, Bae, & Lee, 2014), 초기 대수교육(Han & Kwon, 2018), 수학교사의 교육과정 자료 활용(Ku & Lee, 2020) 등이 있다. 이처럼 수학교육의 여러 영역에서 연구 동향을 파악하려는 시도가 이어져왔으나 평가 연구 동향에 초점을 두고 분석한 연구는 거의 찾아볼 수 없는 실정이다. 중등 수학교육과 관련된 32개의 평가 연구 문헌을 분석한 Kim & Jung(2012)의 연구가 있었으나 분석 대상이 중등교육에 국한되고 분석한 문헌의 수가 제한적이므로 수학교육 평가에 관한 전반적인 연구 동향을 파악하기에는 한계가 있다. 게다가 연구 동향을 분석한 대부분의 연구들이 수집한 문헌을 일일이 사람의 손으로 분류하는 방식을 취하므로, 대량의 문헌을 다루기에 많은 시간과 노력이 필요할 뿐만 아니라 연구자의 주관성에 따라 그 결과가 크게 달라질 수 있다는 제한점이 있다(Son & Lee,

2020). 이러한 한계를 벗어나 데이터 자체의 특징에 초점을 맞추어 감춰진 의미를 발굴하는 토픽 모델링(topic modeling) 기법이 주목을 받고 있다(Blei & Lafferty, 2009). 토픽 모델링을 통해 데이터 분류와 그 의미 파악에 드는 노력을 줄일 수 있으며 단어 패턴에 따라 효율적으로 데이터 이면에 숨겨진 의미를 파악할 수 있다(Blei, Ng, & Jordan, 2003)

이에 본 연구에서는 토픽모델링 기법을 사용하여 국내외 수학교육 평가에 관한 연구동향을 광범위하게 분석하였다. 국내외 수학교육 평가 연구의 분류와 흐름을 파악하고 서로 비교함으로써 향후 국내 수학교육 평가 연구의 발전을 위한 교육적 시사점을 제공하고자 한다.

II. 이론적 배경

1. 평가의 목적과 연구 동향 분석

평가는 교육에 대한 시대적 요구를 반영하여 변화하며, 역으로 평가 내용과 방식 그리고 결과가 교사의 교수법, 교육과정 개정, 교육정책 수립에 변화를 초래하기도 한다. 예컨대, OECD DeSeCo(Definition and Selection of Competencies) 프로젝트를 통해 개인의 성장과 사회의 발전을 위한 핵심역량이 강조되면서 수학적 과정과 역량에 대한 평가가 세계적으로 주목을 받았다(Kim, Park, & Lee, 2015). 이에 따라 학습 결과에 대한 평가보다는 과정중심평가가 강조되면서 학생의 변화와 성장에 초점을 둔 교수·학습으로의 변화가 촉구되어 왔다(Hwang & Kim, 2018). 이처럼 교육과 평가는 개별적인 관계가 아니라 서로에게 영향을 미치며, 평가가 교육에 어떠한 기여를 하느냐에 따라 그 목적은 달라진

다. 이러한 관점에서, 평가는 추구하는 목적에 따라 다양하게 분류될 수 있다. MOE(2017)는 평가가 수업에 통합된다는 점에 주목하여 평가를 1) 학생의 학습 과정과 그 결과로 학습된 지식에 대한 정보를 얻기 위한 학습 과정에 초점을 둔 평가 2) 학생이 교육과정에서 제시된 성취기준에 도달했는지에 대한 정보를 얻기 위한 것으로 학습 결과에 초점을 둔 평가 3) 교사의 수업에 대한 정보를 파악하기 위한 평가의 세 가지 목적으로 분류하였다. NCTM(1995)의 수학 교수를 위한 평가 기준에서는 평가의 목적을 학생의 성장과정 모니터링, 교사의 교수 결정을 위한 자료, 성취도 측정, 프로그램의 효과성 판단으로 구분하였다. 이와 유사하게 Williams(2007)는 평가를 학생의 학습을 돕기 위한 형성 평가, 학생의 성과와 성취도를 확인하기 위한 가이드로서의 총괄 평가, 교수와 교육 프로그램의 질을 측정하기 위한 평가로 구분하였다. 또한 Niss(1993)는 평가의 결과 정보가 어떤 주체에게 제공되는지에 따라 학생, 교사, 그리고 학부모와 교육기관 및 교육행정가로 평가를 분류하였다.

이처럼 바라보는 관점과 추구하는 목적에 따라 평가의 분류가 달라지듯이, 그동안 수학교육 평가 연구 동향을 분석한 연구들도 다양한 관점에서 평가 연구를 분류해왔다. 4,559편의 국내 수학교육 연구 동향을 분석한 Pang et al.(2019)은 평가 일반(관련 이론, 방향성, 관점 등), 평가 방법(수행 평가, 포트폴리오 등), 평가 문항 및 기준 개발, 평가 실태 및 인식을 평가 연구의 하위 영역으로 구분하였다. Park & Kim(2011)은 2005년부터 2009년까지 국내외 수학교육 논문 587편을 분석하였으며, 평가 연구를 평가 결과 분석, 평가 실태 및 평가 방법 개선, 문항 및 기준 개발로 구분하였다. Ha, Pang, & Ju(2010)은 2005년부터 2009년까지 초등수학교육 논문 383편을 분석하고 평가 연

구를 새로운 평가 방향, 평가 실태, 평가문항 및 평가기준 개발·분석, 평가 방법, 평가 결과 분석 및 활용으로 구분하였다. 최근 토픽 모델링을 활용하여 수학교육 연구 동향을 분석한 연구도 수행되고 있다. 국내외 수학교육 연구 동향을 토픽 모델링으로 비교하고 분석한 Shin(2020)은 평가와 관련된 키워드로 평가, 문항, 시험, 기준, 성취도, 학교, 학생, 질문, 반응을 제시하였다. 또한 1968년부터 2015년까지 ESM과 JRME 논문을 분석한 Inglis & Foster (2018)는 과제, 모델링, 지식, 인지, 역량, 결과, 수준, 해결, 질, 성과, 성취도, 정보를 평가와 관련된 키워드로 제시하였다.

이상의 선행연구를 종합해보면, 수학교육에서 평가는 학생의 학습과 성취도, 교사의 교수, 학교와 교육과정, 평가 일반과 방법 및 문항 등에 관한 일련의 활동을 모두 포함하며 평가 분류 방법에 대한 연구자들의 일치된 견해는 존재하지 않는다는 것을 알 수 있다. 또한 수학교육 평가 연구 동향에 대한 대부분의 연구가 평가 자체에 초점을 맞추기보다는 수학교육 연구 동향 전반을 살펴보고 그 하위 주제로서 평가에 대한 분석이 이루어지고 있었다. 따라서 본 연구에서는 수학교육 평가 연구 자체에 초점을 맞추고 자동화된 데이터 분석 기법인 토픽 모델링을 활용하여 국내외 수학교육 평가 연구의 동향을 면밀히 살펴보고자 한다.

2. 토픽 모델링

토픽 모델링은 텍스트 마이닝(text mining) 기법의 한 종류로서 주어진 텍스트 데이터의 집합에서 토픽이라는 문서들의 집합을 탐색하기 위한 통계적 모델이다(Son & Lee, 2020). 연구자의 손으로 처리하기에 너무 많은 노력과 시간이 필요한 비정형화된 텍스트 데이터로부터 이전에는 찾을 수 없었던 의미 있는 정보와 구조를 도출

해낼 수 있다(Blei & Lafferty, 2009). 토픽 모델링의 종류에는 STM(Structural Topic Modeling), pLSA(probabilistic Latent Semantic Analysis), LSI(Latent Semantic Indexing) 등이 있으나 그 중 가장 널리 활용되는 토픽 모델링 기법은 잠재디리클레 할당(Latent Dirichlet Allocation; 이하 LDA)이다. LDA는 관측할 수 있는 단어들의 패턴을 통해 특정 토픽의 특성을 추론하며, 잠재적인 토픽들이 무작위로 혼합되어 하나의 문서를 이루고 있다고 가정한다(Blei et al., 2003). 즉, 관측 가능한 단어를 업데이트하며 토픽에 할당되는 단어와 문헌의 잠재적인 변수에 대한 확률 추정치를 개선해나간다. Figure 1은 LDA의 문헌 생성과 관측된 단어의 토픽 할당 과정을 시각화한 모델이다.

D : 말뭉치 내 문헌들의 집합

K : 토픽의 개수

N : 문헌에 속한 단어들의 집합

θ_d : d 번째 문헌에 대한 토픽의 확률 분포($d = 1, 2, \dots, D$). α 를 매개변수(parameter)로 하는 디리클레 분포($\theta \sim Dir(\alpha)$)를 따름

$Z_{d,n}$: d 번째 문헌의 n 번째 단어에 대한 토픽 할당($n = 1, 2, \dots, N$). θ_d 를 매개변수로 하는 다항 분포를 따름($Z_{d,n} \sim Multinomial(\theta_d)$)

$w_{d,n}$: 관측이 가능한 d 번째 문헌의 n 번째 단어, 토픽 $Z_{d,n}$ 에 따른 다항 확률 분포를 따름

$$P(w_{d,n} | Z_{d,n}, \beta)$$

α : 문헌에 대한 토픽들의 사전 확률 분포. 디리클레 분포의 매개변수

β_k : k 개의 토픽에 대한 단어들의 확률 분포, η 을 모수로 하는 디리클레 분포를 따름

Figure 1의 하단 모델에서 원으로 표현된 노드(node)는 무작위 확률 변수, 음영은 관측 가능

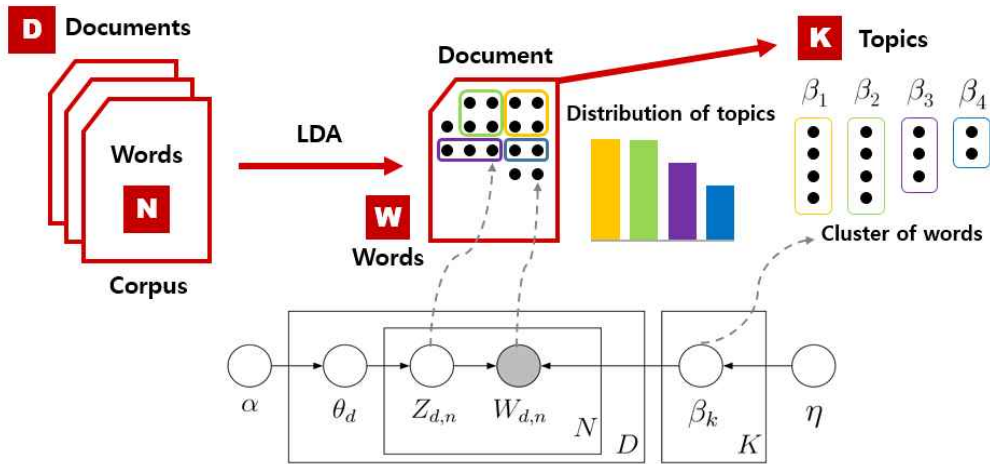


Figure 1. Graphical model representation of LDA(adapted from Blei, 2012; Nabli, Djemaa, & Amorn, 2018)

여부, 직사각형의 판(plate)은 과정의 반복을 의미한다. 여기서 관측 가능한 변수(observed variable)는 $w_{d,n}$ 이며 θ_d 와 $Z_{d,n}$ 는 관측되지 않는 잠재 변수(hidden variable)이다(Blei, 2012). 따라서 문헌에 어떤 토픽들이 할당되는지 파악하려면 관측 가능한 변수인 $w_{d,n}$ 로부터 잠재 변수들을 추론해나가는 과정이 필요하다. 그러나 정확한 확률 분포가 가정되지 않으므로 매개변수 α , η 와 관측한 증거 $w_{d,n}$ 을 통해 잠재 변수의 확률 분포를 갱신하며 결과를 추론해 나간다.

이 과정을 구체적으로 살펴보자. Figure 1은 D개의 문헌과 N개의 단어를 포함한 말뭉치에 대하여 K를 4로 정하고 LDA를 수행했을 때 처리과정을 표현한 시각적 모델이다. 먼저 연구자가 매개변수 α 를 정의하면 d번째 문헌에 대한 토픽의 확률 분포 θ_d 와 θ_d 의 분포에 주제를 무작위로 추출하는 과정인 $Z_{d,n}$ 이 결정된다. 또한 매개변수 η 를 정의하면 토픽에 대한 단어 분포 $\beta_{1:4}$ 가 결정된다. 이 후 단어들의 집합 $w_{d,n}$ 은 $Z_{d,n}$ 에 따라 4개의 토픽에 할당되며 $\beta_{1:4}$ 에 따라 토픽에 대한 단어들이 표집되어 문헌에 기술된

다. 이러한 과정이 N번 반복되면 4개의 토픽으로 구성된 단어들의 집합으로 문헌이 생성되며, D개의 문헌에 반복 수행되면 문헌의 집합인 말뭉치가 생성된다. LDA 기법은 위에서 기술한 과정들을 역으로 실시하며, 관측 가능한 단어 $w_{d,n}$ 으로부터 문헌의 토픽 확률 분포 θ_d 와 단어들의 확률 분포 β_k 를 추정하는 방법이다(Shin, 2020).

III. 연구 방법

1. 분석 대상 및 기간

국외 토픽 모델링 연구의 경우 일반적으로 영향력 지수(impact factor)를 기준으로 학술지를 선정한다(Son & Lee, 2020). 본 연구에서는 Williams & Leatham(2017)이 영향력 지수에 따라 선정한 20개의 국제 수학교육 학술지 중 SSCI (Social Science Citation Index)에 등재된 수학교육 학술지 5개를 분석 대상으로 선정하였다. 국내 수학교육 학술지의 경우 수학교육 전문 학술단체

인 <대한수학교육학회>, <한국수학교육학회>, <한국학교수학회>, <한국초등수학교육학회>에서 발간한 학술지 7종을 대상으로 선정하였다(Son & Lee, 2020). 분석 기간은 2000년부터 2020년 8월까지로 설정하였으며, 이는 대부분의 국내외 수학교육 전문 학술지가 2000년 이후 각각 KCI와 SSCI 등재지로 선정되었기 때문이다(Shin, 2020). 이에 본 연구에서는 KCI, Web of Science, ProQuest의 데이터베이스에서 2000년 1월 1일부터 2020년 8월 31일까지의 KCI 논문 3,459편과 SSCI 논문 4,679편을 수집하였다. 수집한 논문 중 수학교육 평가와 관련된 논문을 선별하기 위해 ‘수학(mathematics)’, ‘평가(assessment)’, ‘검사(test)’, ‘측정(measurement)’의 키워드를 사용해 데이터를 선별하고 최종적으로 KCI 논문 237편과 SSCI 논문 857편을 분석 대상으로 선정하였다. 논문 수집과 수집한 논문의 선별과정은 Figure 2와 같으며, 선정된 구체적인 학술지별 논문 편수는 Table 1에 정리하였다.

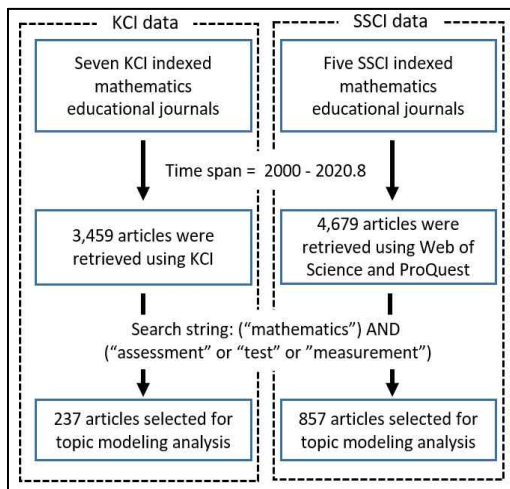


Figure 2. Data retrieval process

Table 1. The number of papers reviewed for each journal

| | Journal | Count |
|------|-------------|-------|
| KCI | 수학교육학연구 | 36 |
| | 학교수학 | 46 |
| | 수학교육 | 29 |
| | 수학교육논문집 | 27 |
| | 초등수학교육 | 10 |
| | 한국초등수학교육학회지 | 33 |
| | 한국학교수학회논문집 | 56 |
| | Total | 237 |
| SSCI | JRME | 113 |
| | ESM | 227 |
| | ZDM | 345 |
| | JMTE | 95 |
| | MTL | 77 |
| | Total | 857 |

토픽 모델링 연구에서는 수집한 서지정보들 중 제목, 키워드, 초록을 주로 활용한다. 이 중 제목과 키워드는 연구의 구체적인 내용을 반영하지 못하는데 반해, 초록은 연구에서 발견한 가치 있는 내용을 대표하며 연구의 목적과 문제, 발견한 사실을 포함한다(Chen, Zou, Cheng, & Xie, 2020; Griffiths & Steyvers, 2004). 따라서 본 연구에서는 키워드와 제목은 분석 대상에서 제외하고, 국내외 논문들에 대한 분석의 통일성을 꾀하기 위해 영문 초록만을 분석하였다.

2. 전처리(pre-processing)

전처리는 두 단계 과정을 거쳐 진행하였으며 R(version 4.0.2)을 분석 프로그램으로 활용하였다.

첫 번째 단계에서는 수집한 데이터에서 단어를 정제하는 과정을 수행하였다. 먼저 일상적으로 사용되는 the, is, a, an 등과 같은 불용어(stop-word)와 @, &, \$, % 등의 특수 기호를 제거하였으며, 단어의 어근(stem)과 의미를 기준으로 단어를 정제하였다. 예컨대, ‘Programs’와 같이

복수 형태의 대소문자 표기가 혼합된 단어들을 소문자 단수 형태의 ‘program’으로 변환하였으며, ‘prospective teacher’와 ‘pre-service teacher’를 ‘preservice teacher’로 통일하였다. 또한 ‘number’, ‘term’와 같이 수학에서 중요한 단어가 ‘the number of’, ‘in term of’와 같이 일상적인 용어로 사용되는 경우를 찾아 제거하였다.

두 번째 단계에서는 TF-IDF(Term Frequency - Inverse Document Frequency)를 적용하여 단어의 출현 빈도를 조절하였다. TF-IDF는 단어 빈도와 역문서 빈도의 곱으로 계산되며, 모든 문서에 흔하게 등장하는 단어를 걸러내는 역할을 수행한다. 예를 들어, 단어 빈도(TF)만을 기준으로 LDA를 수행한다면 ‘paper’와 ‘article’과 같이 대부분의 문서에 등장하는 단어가 상위 키워드로 추출되므로 역문서 빈도(IDF)를 곱하여 이를 조절하는 것이다. 따라서 TF-IDF를 너무 작은 값으로 설정할 경우 일상적인 단어의 출현 빈도가 높아지고, 지나치게 높은 값을 설정할 경우 수학 교육에서 중요한 단어들이 삭제될 수 있으므로 유의할 필요가 있다. 본 연구에서는 TF-IDF 값을 1사분위수(first quartile) 이상으로 설정하였으며, 국내외 논문에서 같은 기준을 적용하였다.

3. 최적의 토픽 수 및 토픽 명 선정

LDA분석은 R의 ‘topicmodels’ 패키지를 활용하였다. 매개변수 α 는 0.1로 설정하였으며, 깁스 샘플링(gibbs sampling) 방식으로 1,000회를 반복하였다. 최적의 토픽 개수를 찾는 방법에 대하여 연구자들 간 일치된 견해는 없으나, 혼잡도(perplexity; Inglis & Foster, 2018), 조화 평균(harmonic mean; Griffiths & Steyvers, 2004), 코사인 유사도(cosine similarity; Cao, Xia, Li, Zhang, & Tang, 2009). 일관성(coherence; Blair, Bi, &

Mulvenna, 2020) 등이 널리 활용된다. 본 연구에서는 K 값을 2에서 30까지로 설정하고 모델의 적합성을 평가하기 위한 내부 지표인 혼잡도 지수의 추이를 살펴보았다. 혼잡도 지수는 토픽들이 서로 명확히 분류되는지 판단하는 지표로서 값이 작을수록 적합한 모델을 의미한다(Blei et al., 2003). 그러나 혼잡도 지수의 특성상 토픽 수가 늘어날수록 값은 작아지므로 일반적으로 기울기의 변화가 급격히 변화하는 시점을 선택한다(Inglis & Foster, 2018). Figure 3에서 볼 수 있듯이 KCI 논문은 8, SSCI 논문은 11을 기준으로 혼잡도 지수의 값이 가장 적거나 기울기가 급격하게 변화하므로 이를 적합한 K 값으로 설정하였다.

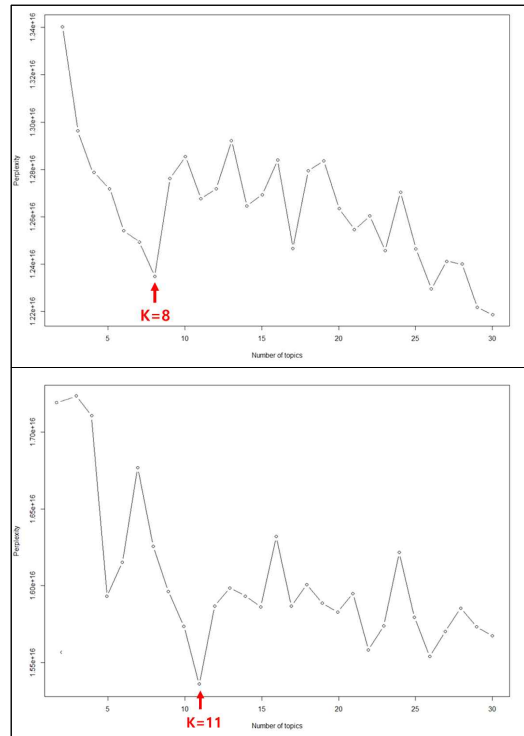


Figure 3. The perplexities for determining optimal number of topics(upper: KCI, lower: SSCI)

선정한 토픽 개수에 따라 국내외 수학교육 평가 연구에 대한 LDA를 수행하고 각 토픽에 해당하는 원문과 토픽을 구성하는 상위 10개의 단어를 검토하여 적절한 토픽 명을 부여하였다. 예를 들어, 추출된 토픽의 상위 10개 단어 집합이 {program, gifted, evaluation, underachiever, solving, factor, ability, skill, knowledge, item}로 출현한 경우 ‘영재 및 기초학력 프로그램 평가’로 토픽 명을 부여하였다.

수학교육 평가 연구의 동향을 살펴보기 위해 먼저 5년 단위로 키워드를 분석하고 전반적인 연구 흐름을 살펴보았다. 이 후 LDA를 수행하고 도출된 토픽들을 토대로 국내외 수학교육 평가 연구에서 유사한 토픽과 상이한 토픽을 분류하여 분석하였다. 또한 최근 평가 연구의 관심 영역을 살펴보기 위해 시계열 선형 회귀분석을 수행하고 Hot 토픽과 Cold 토픽을 도출하였다.

IV. 결과 분석

1. 시기별 키워드 분석

국내외 수학교육 평가 연구에 대한 전반적인

변화 추이를 살펴보기 위해 2000년부터 2020년까지의 연구들을 5년 단위로 구분하고 대표적인 키워드들의 출현 빈도를 살펴보았다. 핵심 키워드를 파악하기 위해 ‘mathematics’, ‘assessment’, ‘student’, ‘teacher’, ‘result’, ‘analysis’와 같이 수학교육 평가 연구에서 공통적으로 등장하고 다른 키워드들에 비해 지나치게 빈도가 높은 공빈출 단어는 삭제하였다. 시기별로 10개의 단어를 빈도순으로 정리한 내용은 Table 2와 같으며 시기별 100개의 단어를 추출하여 Word Cloud로 시각화한 결과는 Figure 3으로 제시하였다. Word Cloud에서 단어의 크기는 빈도를 의미한다.

Table 2와 Figure 4를 보았을 때, 시기별로 국내 연구에서 빈도수가 가장 많은 단어는 수행(performance), 프로그램(program), 수업(lesson), 교육과정(curriculum)이며, 국외 연구에서는 문제(problem)와 과제(task)가 꾸준히 최빈값으로 등장했다. 이러한 단어들은 넓은 의미에서 보았을 때 해당 시기의 수학교육 평가 연구를 대표하는 단어이며, 주요한 연구 대상이었음을 의미한다. 따라서 국내 수학교육 평가 연구는 수행평가, 프로그램 평가, 수업 평가, 교육과정 평가 순으로 평가 연구의 주된 대상이 변화되어 왔음을 알 수 있으며, 국외 수학교육 평가 연구는 시간의

Table 2. The top 10 most frequently used words in KCI and SSCI by period

| Rank | KCI | | | | SSCI | | | |
|------|------------------|----------------|-----------------|--------------------|---------------|-------------------|---------------|--------------|
| | 2000-2004 | 2005-2009 | 2010-2014 | 2015-2020 | 2000-2004 | 2005-2009 | 2010-2014 | 2015-2020 |
| 1 | performance | program | lesson | curriculum | problem | problem | task | problem |
| 2 | factor | lesson | item | lesson | conception | knowledge | problem | knowledge |
| 3 | level | method | program | item | strategy | understanding | knowledge | teaching |
| 4 | type | development | level | competency | development | teaching | teaching | task |
| 5 | problem | level | curriculum | problem | task | preservice | preservice | classroom |
| 6 | misunderstanding | problem | problem | program | practice | classroom | level | practice |
| 7 | curriculum | ability | ability | standard | knowledge | task | practice | learning |
| 8 | task | curriculum | standard | achievement | teaching | conception | learning | development |
| 9 | knowledge | learning | knowledge | textbook | understanding | development | conception | conception |
| 10 | system | performance | pisa | learning | classroom | learning | understanding | skill |

※ 이전 시기에는 언급되지 않았으나 시기별로 새롭게 출현한 키워드는 Bold 표시함

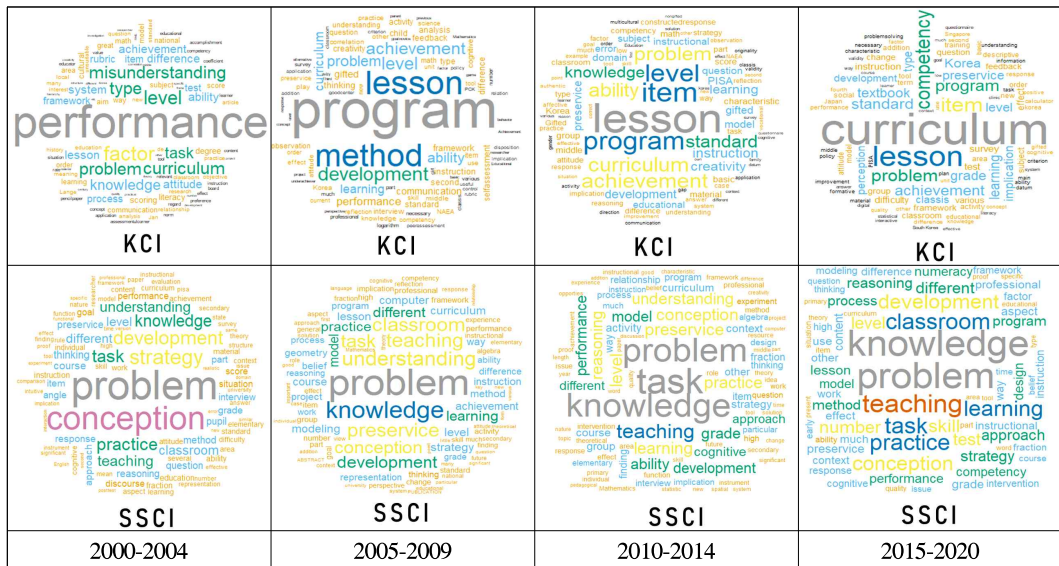


Figure 4. Word Clouds by period in KCI and SSCI

흐름과 무관하게 문제와 과제에 대한 평가 연구가 꾸준히 수행되어 왔음을 알 수 있다.

다음으로 이전 시기에는 언급되지 않았지만 시기별로 상위 10개 단어에 새롭게 출현한 키워드들에 대해 살펴보고, 관련 키워드가 출현한 시기의 원문을 검토하여 어떤 연구가 새로이 출현했는지 분석하였다. Table 2를 보았을 때, 국내 연구의 경우 2005-2009년에는 능력(ability), 수업(lesson), 방법(method), 프로그램(program)이 등장하였다. 2000-2009년에는 등장하지 않았지만 2010-2014년에 새로이 출현한 키워드는 문항(item), 기준(standard), PISA이며, 이는 국가수준학업성취도 평가와 국제 수학성취도 비교평가인 TIMSS나 PISA를 통해 우리나라 학생들의 수학성취 특성을 분석하려는 연구가 증가했음을 의미한다. 2010-2012년 국가수준 학업성취도 평가 결과에 나타난 중학교 수학과 성취수준별 학업성취 특성을 분석한 Lee & Jo(2014)의 연구가 그 예이다. 또한 단어 ‘기준’의 등장은 국내 수학 교육과

정의 특성을 국외 교육과정과 비교·분석하려는 연구가 증가했음을 의미하며, 수학적 과정을 중심으로 국외 수학과 교육과정의 경향을 비교·분석한 Shin(2011)의 연구를 예로 들 수 있다. 2000-2014년에는 등장하지 않았지만 2015-2020년에 새로이 출현한 키워드는 역량(competency), 교과서(textbook), 성취(achievement)로 나타났으며, 이는 역량 중심 교육을 강조한 2015 개정 교육과정의 영향으로 교과서와 성취기준에 대한 연구가 증가한 것으로 판단된다. 수업과 통합한 수학 교과 역량 중심의 과정중심평가 적용을 분석한 Jeon & Lee(2019)의 연구를 예로 들 수 있다.

국외의 경우 2005-2009년에 학습(learning), 예비교사(preservice)가 새롭게 출현하였다. 이는 학생의 학습에 영향을 미치는 수업 요인을 분석한 연구와 예비 교사의 인지적 정의적 특성을 분석하는 연구에 대한 관심이 증가한 것으로 해석된다. 교실 환경이 중학생의 수학 학습과 성취도에 미치는 영향을 분석한 Tarr, Reys, Reys, Chavez,

Shih, & Osterlind(2008)와 예비교사의 수업 분석 기술을 연구한 Morris(2006)의 연구가 그 예이다. 2000-2009년에는 등장하지 않았지만 2010-2014년에 새로이 출현한 키워드는 수준(level)이며, 이는 학습경로(learning trajectory)에 대해 증가된 관심과 함께 학생의 인지적 발달 수준을 세분화하려는 연구가 증가한 것으로 판단된다. 관련 연구로는 학생들의 길이 학습에 대한 경로를 평가한 Sarama, Clements, Barrett, Van Dine, & McDonel(2011)의 연구를 들 수 있다. 2015-2020년에는 기능(skill)이 새로 등장하였다. 주로 수리 기능(numeracy skill), 뺄셈 기능(subtraction skill), 기하학 기능(geometry skill) 등과 관련되며 유치원생과 저학년 학생들의 수학 관련 기능을 평가하는 연구가 주를 이루었다. 양적 역량(quantitative competency)과 관련하여 초등학생들의 비형식적, 형식적 수리 능력을 분석한 Van den Heuvel-Panhuizen & Elia(2020)의 연구가 그 예이다.

지속적으로 연구가 된 주제를 파악하기 위해 Table 2에서 전 시기에 걸쳐 3회 이상 등장하는 키워드를 추출하였으며 그 결과는 Table 3에 제시하였다. 국내외 시기별 공통 출현 키워드는 교육과정(curriculum), 수업(lesson), 수준(level), 문제

(problem), 프로그램(program)의 5개였으며, 국외는 교실(classroom), 개념(conception), 발달(development), 지식(knowledge), 학습(learning), 관행(practice), 문제(problem), 과제(task), 교수(teaching), 이해(understanding)의 9개 단어로 나타났다. 이는 국내 연구가 시계열에 따라 주요 키워드가 다소 변화하는데 반해, 국외의 경우 시간의 흐름에 관계없이 빈도수가 높은 키워드는 일정한 출현 경향을 보임을 의미한다. 또한 국외 연구에서는 수학적 이해와 개념, 발달, 지식과 같이 교사와 학생의 인지적 영역과 관련된 키워드가 주로 등장한데 비해, 국내에서는 교육과정, 수업, 수준, 문제, 프로그램에 관한 키워드가 독립적으로 출현하였다.

2. 국내외 유사한 토픽 분석 결과

LDA 수행을 통해 도출된 국내외 연구들의 토픽들을 비교한 결과 Table 4와 같이 5개의 유사한 토픽이 추출되었다. 토픽에 속하는 키워드들은 각 토픽에서 출현 확률이 가장 높은 상위 10개의 단어를 나열한 것이다. 유사한 토픽들을 구체적으로 살펴보면 ‘수학적 모델 평가’, ‘예비교사 평가’, ‘대규모 평가’, ‘교육과정 평가’ 학생의

Table 3. Commonly used words in KCI and SSCI, respectively(rearranged based on alphabet order)

| KCI | | | | SSCI | | | |
|------------|------------|------------|------------|---------------|---------------|---------------|-------------|
| 2000-2004 | 2005-2009 | 2010-2014 | 2015-2020 | 2000-2004 | 2005-2009 | 2010-2014 | 2015-2020 |
| curriculum | curriculum | curriculum | curriculum | classroom | classroom | | classroom |
| | lesson | lesson | lesson | conception | conception | conception | conception |
| level | level | level | | development | development | | development |
| problem | problem | problem | problem | knowledge | knowledge | knowledge | knowledge |
| | program | program | program | | learning | learning | learning |
| | | | | practice | | practice | practice |
| | | | | problem | problem | problem | problem |
| | | | | task | task | task | task |
| | | | | teaching | teaching | teaching | teaching |
| | | | | understanding | understanding | understanding | |

Table 4. The similar topics in KCI and SSCI

| KCI | | SSCI | |
|-----------|--|-----------|---|
| Topic | Keyword(top 10) | Topic | Keyword(top 10) |
| 수학적 모델 평가 | model , task, modeling , evaluation, attitude, performance, item, tool, factor, constructed | 수학적 모델 평가 | modeling , model , solution, problem, process, element, instructional, otl(opportunity to learn), activity, type |
| 예비교사 평가 | preservice , framework, pck(pedagogical content knowledge), perception, misunderstanding, professional , evaluation, program, literacy, tool | 예비교사 평가 | preservice , instructional, quality, flexibility, psts(pre-service teachers'), reasoning, conception, professional , belief, efficacy |
| 대규모 평가 | naea(national assessment of educational achievement), item , pisa , gender, score, level, gap, timss, related, social | 대규모 평가 | pisa , attitude, emotion, international, achievement, motivation, country, item , validity, comparison |
| 교육과정 평가 | textbook , evaluation, curriculum , difficulty, type, unit, item, question, achievement, standard | 교육과정 평가 | curriculum , textbook , national, reform, word, algebra, arithmetic, geometry, standard , trajectory |
| 학생의 역량 평가 | competency , evaluation, key, cognitive , perception, ability, solving, peer, student, tool | 학생의 역량 평가 | competency , cognitive , communication, interaction, boy, relationship, language, lesson, girl, knowledge |

※ 공통된 키워드는 Bold 표시함

역량 평가'로 나타났다. 토픽에 할당된 키워드와 원문을 검토하여 토픽들의 특징과 내용을 살펴 보았으며, 특정 문서가 해당 토픽에 할당될 비율인 'gamma' 값을 기준으로 토픽별 대표 논문을 선정하였다.

토픽별로 분류된 논문들의 구체적인 내용은 다음과 같다.

첫째, '수학적 모델 평가'에 대한 국내의 연구의 공통 키워드는 {model, modeling}으로 평가 절차나 방법에 대한 모델을 개발하거나 수업, 학생, 인지적 요인에 대한 모델링과 이에 대한 평가 논문들로 이루어져 있었다. 하위 주제에 대한 차이점을 살펴보면, 국내 연구만의 키워드는 {evaluation, attitude, performance, item, tool, factor, constructed}로서 수행평가, 태도 평가, 평가 도구, 구조화된 문항 평가 등에 초점이 맞추어져 있으며 Koh, Han, & Lee(2013)의 '교구를 활용한 수학적 과정의 평가모델 개발에 관한 연구-중학교 수학을 중심으로-'가 대표적인 논문으로 나타났다. 반면, 국외의 경우 {solution,

problem, process, element, instructional, otl (opportunity to learn), activity, type}와 같이 문제 해결 과정, 교수 요인, 학습 기회, 활동 유형 등에 대한 모델 평가에 초점이 맞추어져 있었으며, 예비교사들의 수학적 모델링 경험과 모델링 역량 기회를 제공하기 위한 모델링 모듈을 평가하는 Anhalt, Cortez, & Bennett(2018)의 논문이 대표적인 논문으로 나타났다.

둘째, '예비교사 평가'의 공통 키워드는 {preservice, professional}로 나타났으며 예비교사의 수학적 지식, 정의적 영역, 전문성을 평가하거나 예비교사가 문항, 수업, 동료 등을 평가하는 논문들이 주를 이루었다. 하위 주제에 대한 차이점을 살펴보면, 국내 연구만의 키워드는 {framework, pck, perception, misunderstanding, evaluation, program, literacy, tool}로서 평가를, 교육학적 내용 지식, 평가에 대한 인식, 오개념, 프로그래밍과 평가도구 등에 초점을 두고 있으며. 대표적인 논문은 Ko, Park, & Lee(2016)의 '초등 예비교사들의 수학교과에서의 평가에 대한 인식'

으로 나타났다. 반면, 국외 연구만의 키워드는 {instructional, quality, flexibility, reasoning, conception, belief, efficacy}로서 교수의 질, 유연성, 추론, 개념, 신념, 효능감 등 예비교사에 대한 인지적 혹은 정의적 영역의 평가에 초점을 두고 있었으며, 문제 해결 과정에서 문제 유형, 성취도, 예비 중등 수학교사들의 메타인지 행동 사이의 관계를 밝힌 Demircioğlu, Argün, & Bulut(2010)의 연구가 대표적인 논문으로 나타났다.

셋째, ‘대규모 평가’의 공통 키워드는 {pisa, item}으로 PISA, TIMSS 등의 대규모 평가를 통해 성별, 국가, 문항 등의 변인을 비교하는 연구들이다. 공통 키워드를 제외한 나머지 키워드를 살펴보면, 국내의 경우 {naea, gender, score, level, gap, timss, related, social}로서 국가수준학업성취도평가(NAEA)와 TIMSS 결과에서 성별, 수준, 사회적 요인 등의 평가에 초점을 두고 있었다. 대표적인 논문은 Lee & Lee(2011)의 ‘PISA수학 성취도 평가에 나타난 한국 학생들의 성차 추이의 배경 요인 연구’로 나타났다. 반면, 국외의 키워드는 {attitude, emotion, international, achievement, motivation, country, validity, comparison}로서 태도, 감정, 동기와 같은 정의적 영역과 국가 간 평가 결과를 비교하는 연구에 초점을 두고 있었다. 대표적인 논문은 PISA 2003과 TIMSS 2003의 평가 결과의 일부를 조망하고 수학 수행 능력과 실생활 맥락에서의 적용 능력 측정에 대한 시사점을 제시한 Ferrini-Mundy & Schmidt(2005)의 연구로 나타났다.

넷째, ‘교육과정 평가’의 공통 키워드는 {textbook, curriculum, standard}로 나타났으며, 교육과정을 기반으로 평가 문항이나 방향에 대한 반영 실태를 분석하거나 학생의 수학성취도에 미치는 영향을 평가하는 논문들로 구성되어 있다. 다만 ‘standard’의 의미는 조금 다르게 사용되

었는데, 국내는 성취기준(achievement standard)의 의미로 주로 활용되며, 국외는 CCSSM(Common Core States Standards for Mathematics)에서 사용된 수학교육과정 규준의 의미로 사용되고 있었다. 국내만의 키워드는 {evaluation, difficulty, type, unit, item, question, achievement}로서 난이도, 유형, 단원, 문항, 발문, 성취기준 등을 평가 하는데 초점을 두고 있었으며, 대표적인 논문은 Kang, Kim, Park, Lee, Cha, & Koh(2010)의 ‘2009 개정 교육과정에 따른 수학과 평가가 나아가야 할 방향’으로 나타났다. 반면, 국외의 키워드는 {national, reform, word, algebra, arithmetic, geometry, trajectory}로서 문장제, 대수, 기하, 학습 경로에 초점을 두고 있었으며, 네덜란드의 개정 교육과정이 수학적 사고 평가를 반영했는지 분석한 Drijvers, Kodde-Buitenhuis, & Doorman (2019)의 연구가 대표적인 논문으로 나타났다.

마지막으로 ‘학생의 역량 평가’의 공통 키워드는 {competency, cognitive}로 나타났으며, 역량 평가도구를 개발하거나 학생 역량을 평가하는 방법에 대한 연구들로 구성되어 있었다. 국내만의 키워드는 {evaluation, key, perception, ability, solving, peer, student, tool}로서 역량과 관련된 학생의 인식과 능력, 문제 해결과 평가도구에 초점을 두고 있었으며, 대표적인 논문은 Kim & Han(2018)의 ‘수학 문제해결 역량 평가도구 개발’로 나타났다. 반면, 국외만의 키워드는 {communication, interaction, boy, relationship, language, lesson, girl, knowledge}로서 언어, 수업, 지식, 관계, 의사소통, 상호작용 등 의사소통 역량에 초점을 두고 있었다. 대표적인 논문은 학생의 추론, 시공간 관계, 문제 해결과 문제 제시, 조합 추론 등에 관하여 학생의 역량을 평가하는 방법을 제안한 English(2016)의 연구로 나타났다.

이상의 결과를 종합하면, 국내외 수학교육 평

가 연구에서 서로 유사한 토픽이 나타나지만 연구의 초점은 서로 다르게 나타난다는 것을 알 수 있다. 국내 연구의 특징적인 키워드는 {program, tool, item, type, framework}, 국외 연구의 특징적인 키워드에서는 {belief, reasoning, conception, flexibility, motivation, interaction}과 같은 단어들이 여러 토픽에서 출현하였다. 이는 국내외 수학교육 평가에 관한 유사한 토픽이더라도 국내 수학교육 평가 연구는 평가도구나 평가틀, 프로그램 평가 연구에 초점을 두며, 국외 수학교육 평가 연구는 인지적·정의적 영역의 평가에 초점을 두고 있음을 유추할 수 있다.

3. 국내외 상이한 토픽 분석 결과

LDA 수행 결과, Table 5와 같이 국내외 연구에서 9개의 상이한 토픽이 나타났다. 국내 연구

에서 나타난 토픽은 ‘영재 및 기초학력 프로그램 평가’, ‘교사 평가 기준’, ‘평가 실태 및 활용’이며 국외는 ‘수학성취도에 미치는 요인’, ‘대학 수학 강의 평가’, ‘수업에서의 학생 평가’, ‘컴퓨터 도구 평가’, ‘교사 교육’, ‘수학성취도에 미치는 가정 요인 평가’로 나타났다. 이 중 ‘교사 평가 기준’과 ‘교사 교육 평가’는 교사의 역량이나 지식을 평가하다는 점에서 유사한 면이 있지만 국내 연구가 평가 기준(evaluation standard)의 요소(element)와 요인(factor)의 개발(development)에 초점을 맞추었는데 비해, 국외 연구는 교사 교육과 발달 연구(TEDS; Teacher Education and Development Study), 개입(intervention)과 피드백(feedback), 반응(response)과 같이 교사의 역량 개발에 초점을 두고 있다는 점에서 상이한 토픽으로 분류하였다.

나머지 토픽들은 키워드 구성에서 상이한 단

Table 5. The distinct topics in KCI and SSCI

| KCI | | SSCI | |
|-------------------|--|-----------------------|---|
| Topic | Keyword(top 10) | Topic | Keyword(top 10) |
| 영재 및 기초학력 프로그램 평가 | program, gifted, evaluation, underachiever, solving, factor, ability, skill, knowledge, item | 수학성취도에 영향을 미치는 인지적 요인 | achievement, cognitive, factor, skill, strategy, reasoning, intervention, representation, growth, score |
| | | 대학 수학 강의 평가 | proof, geometry, quality, university, lecture, logical, engagement, reasoning, college, quality |
| 교사 평가 기준 | evaluation, standard, instruction, development, element, question, knowledge, classroom, factor, process | 수업에서의 학생 평가 | lesson, children, skill, competency, reasoning, strategy, response, numeracy, item, diagnostic |
| | | 컴퓨터 도구 평가 | computer, technology, instrument, tool, reasoning, attitude, solution, formative, project, skill |
| 평가 실태 및 활용 | item, question, error, correct, response, survey, feedback, answer, application, descriptive | 교사 교육 평가 | teds(teacher education and development study), intervention, mck(mathematical content knowledge), reflection, feedback explanation, pck(pedagogical content knowledge), resource, dynamic, response |
| | | 수학성취도에 영향을 미치는 가정 요인 | home, parents, belief, achievement, discourse, perceived, attitude, perception, negative, solution |

어 분포를 보였다. 국내 연구에서는 ‘영재 및 기초학력 프로그램 평가’와 ‘평가 실태 및 활용’이 독립된 토픽으로 나타났다. 특히 키워드 ‘program’은 여러 토픽에서 공통적으로 출현하지만 ‘영재와 기초학력 학생에 관한 프로그램 평가’에 관한 연구가 국내 연구에서 하나의 독립된 토픽으로 나타난 점은 국외 연구와 차별화되는 점이다. 대표적인 논문은 Kim & Hwang(2010)의 ‘영재학급에서의 수학영재프로그램 평가에 관한 연구’로 나타났다. ‘평가 실태 및 활용’은 설문(survey)에 대한 답변(answer)을 바탕으로 적용(application) 실태와 피드백(feedback)에 관심을 둔 연구들로서 Kim, Cho, & Joo(2012)의 ‘서술형 평가에 대한 인식 및 실태에 관한 조사연구 -서울시 소재 초등학교를 중심으로-’가 대표적인 논문으로 나타났다.

국외 연구만의 특징적인 토픽들 중 ‘수학성취도에 미치는 인지적 요인’과 ‘수학성취도에 미치는 가정 요인’은 모두 수학성취도와 관련된 요인에 관한 연구이지만 그 요인이 인지적 영역인지 가정환경과 같은 외부 요인인지 여부에 따라 서로 다른 토픽으로 나타났다. ‘수학성취도에 영향을 미치는 인지적 요인’의 대표적인 논문은 학습 기회와 과제의 인지적 요구가 수학성취도에 미치는 영향을 평가한 Cueto, Ramirez, & Leon(2006)의 연구로 나타났으며, ‘수학성취도에 영향을 미치는 가정 요인’에 관한 대표적인 논문은 수학에 대한 부모의 관점이 학생의 수학적 내용 채택과 수행에 미치는 영향을 평가한 Hawighorst(2005)의 연구로 나타났다.

‘대학 수학 강의 평가’는 증명(proof)이나 기하학(geometry) 강의의 질(quality)과 대학생 참여(engagement)에 관한 연구로서 대표적인 논문은 미국 30개 대학의 유클리드, 비유클리드 기하학 과정의 내용과 설문 응답을 분석하고 평가한

Grover & Connor(2000)의 연구로 나타났다. ‘수업에서의 학생 평가’는 수업에서 학생의 기능(skill)과 역량(competency), 추론(reasoning), 전략(strategy), 반응(response) 등을 평가하는 연구들이 포함된다. 국내에서는 수업에서의 학생 평가가 독립적인 토픽으로 출현하지 않았는데, 이는 국내에서도 수업 중 학생의 수행평가나 과정평가에 관한 연구가 수행되고 있으나 다른 출현 토픽들에 비해 많은 연구가 수행되지 않았으며 국외에 비해 많은 관심을 받지 못하고 있음을 시사한다. ‘컴퓨터 도구 평가’는 컴퓨터(computer) 기술(technology)을 활용한 평가 도구(tool)와 장비(instrument)에 대한 연구로서 형성평가(formative assessment)와 관련하여 프로젝트(project)로 진행되고 있는 연구들을 포함한다. 컴퓨터로 구현한 피아제의 임상 인터뷰가 학생의 수학적 사고와 통찰력에 관한 형성평가에 적합한지를 평가한 Ginsburg, Lee, & Pappas(2016)의 연구가 대표적인 논문으로 나타났다.

4. 시계열 토픽 분석

가. 시계열에 따른 토픽 비중 분석

국내의 수학교육 평가 연구의 연도별 토픽 비중 변화를 파악하기 위해 2000년부터 2020년까지의 자료를 곡선추정 단순회귀선을 통해 살펴 보았으며 그 결과는 Figure 5와 같다. 가로축은 년도를 나타내며 세로축은 토픽이 해당 연도의 전체 연구에서 차지하는 비중을 의미한다.

먼저 국내 연구의 변화 추이를 살펴본 결과, 토픽 비중이 감소하는 토픽은 ‘수학적 모델 평가’, ‘영재, 기초학력 프로그램 평가’, ‘교사 평가 기준’에 관한 연구로 시간의 흐름에 따라 증감을 반복하지만 전체적으로 비중이 감소하는 추

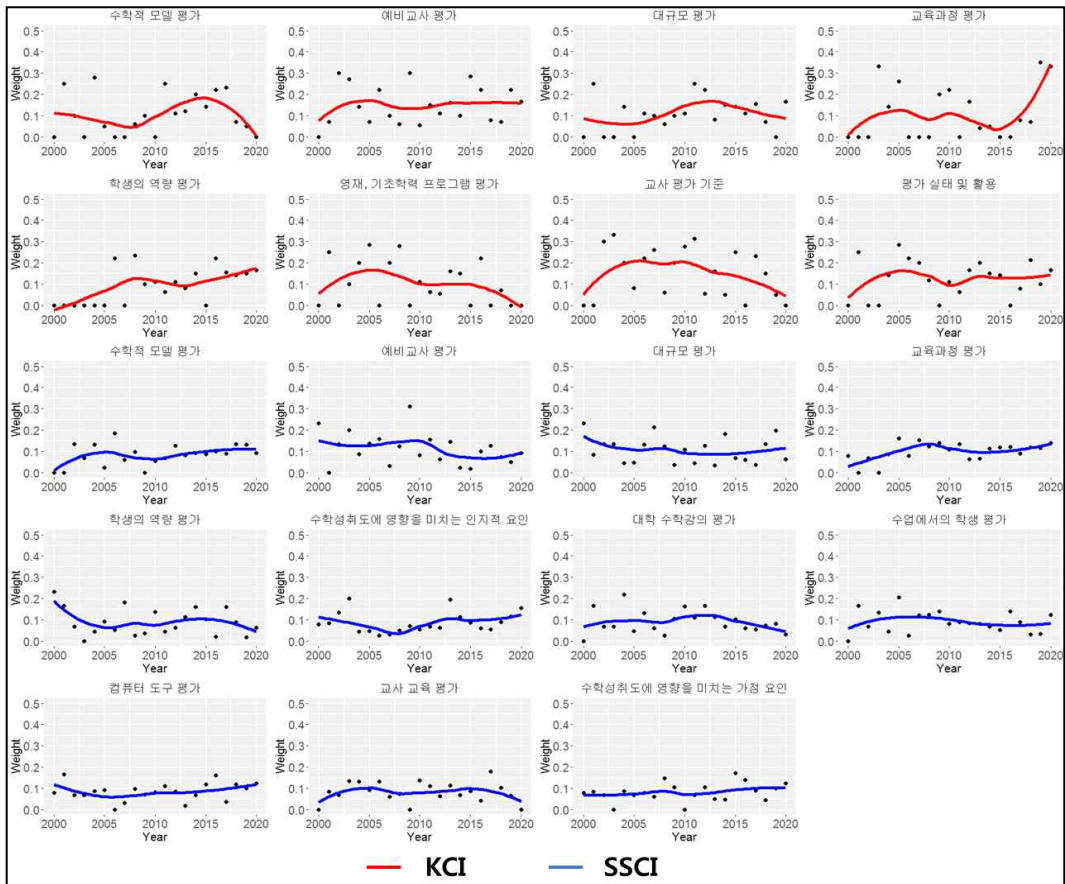


Figure 5. Research trends of KCI and SSCI journals

세를 보였다. 다음으로 토픽 비중이 증가하는 연구 주제는 ‘교육과정 평가’, ‘학생의 역량 평가’, ‘평가 실태 및 활용’으로 나타났다. ‘학생의 역량 평가’는 일부 구간에서 감소 경향을 보였지만 전체적으로 증가하는 경향을 보였으며, ‘평가 실태 및 활용’은 2005년까지 비중이 증가하다가 이후 유지되는 추이를 보였다. 이 중 가장 주목할 만한 토픽은 ‘교육과정 평가’인데 2015년 이후 그 비중이 급격하게 증가하는 경향을 보였다. 이는 2015 개정 교육과정이 고시됨에 따라 역량 중심의 다양한 평가 영역에 대한 연구가 증가했기 때문이라고 유추해 볼 수 있다. 토픽 비중의

변화가 크게 없는 연구들은 ‘예비교사 평가’와 ‘대규모 평가’이며 시간에 따른 증감 변화는 보이지만 대체로 일정한 경향을 나타냈다.

국외 연구의 경우 비중이 감소하는 토픽은 ‘학생의 역량 평가’, ‘예비교사 평가’이다. ‘학생의 역량 평가’는 2000년과 2015년을 기점으로 감소하는 경향을 보였으며 ‘예비교사 평가’는 전반적으로 연구 비중이 유지되는 경향을 보이지만 2010년을 기점으로 연구 비중이 하락하는 경향을 보였다. 증가 추이를 나타낸 토픽은 ‘수학적 모델 평가’와 ‘교육과정 평가’이며 일부 구간에서 감소 경향이 나타나지만 전반적으로 증가하

는 경향을 보였다. 그 외의 국외 연구들은 증감을 반복하며 그 비중이 조금씩 변화하였지만 전반적으로 연구의 비중이 유지되는 추이가 나타났다.

국내외 연구 추이를 살펴보았을 때, 주목할 점은 ‘수학적 모델’과 ‘학생 역량 평가’의 연구 경향이 상이하게 나타난다는 점이다. 국내 연구에서 ‘수학적 모델’은 감소, ‘학생 역량 평가’는 증가하는 경향을 나타낸 반면, 국외 연구에서는 이와 반대되는 경향을 보였다.

나. Hot and Cold 토픽

통계적으로 유의미한 시계열 변화를 살펴보기 위해 2000년부터 2020년까지 토픽별 비중에 대한 선형회귀분석을 수행하였다. 시계열 회귀분석의 유의확률(p-value) 값이 95% 유의수준에서 유의미하고 회귀계수 값이 양수(+)인 경우 Hot 토픽, 음수(-)인 경우 Cold 토픽, 유의미한 결과도 출되지 않은 토픽은 Neutral 토픽으로 분류하였다. Hot 토픽은 시계열에 따라 연구의 비중이 증가되고 있는 토픽, Cold 토픽은 연구의 비중이 감소되고 있는 토픽을 의미한다(Griffiths & Steyvers, 2004). Table 6은 국내외 토픽들의 시계열 회귀분석 결과이며 Hot 토픽은 ‘Hot’, Cold 토픽은 ‘Cold’, Neutral 토픽은 ‘.’로 표시하였다.

시계열 회귀분석 결과, 국내 연구의 ‘학생의 역량 평가’와 국외 연구의 ‘교육과정 평가’가 2개의 Hot 토픽으로 나타났으며 나머지 17개의 토픽은 Neutral 토픽으로 드러났다. 주목할 점은 국내 연구의 ‘교육과정 평가’의 경우, 시계열 분석에서 2015년 이후 급격한 증가 경향을 보였으나 통계적으로는 유의미한 결과를 나타내지 않았다는 점이다. 이는 일부 구간에서 연구 비중이 감소하였으므로 데이터의 값을 선형 회귀선으로

Table 6. Result of time series regression analysis

| | Topic | Regression coefficient | p-value | Hot/Cold/Neutral |
|------|-----------------------|------------------------|-------------|------------------|
| KCI | 수학적 모델 평가 | -.087 | .708 | - |
| | 예비교사 평가 | .033 | .888 | - |
| | 대규모 평가 | .019 | .339 | - |
| | 교육과정 평가 | .154 | .505 | - |
| | 학생의 역량 평가 | .595 | .004 | Hot |
| | 영재, 기초학력 | -.130 | .573 | - |
| | 프로그램 평가 | -.130 | .573 | - |
| | 교사 평가 기준 | -0.96 | .677 | - |
| | 평가 실태 및 활용 | .078 | .736 | - |
| | 수학적 모델 평가 | .359 | .11 | - |
| SSCI | 예비교사 평가 | -.313 | .167 | - |
| | 대규모 평가 | -.191 | .406 | - |
| | 교육과정 평가 | .455 | .038 | Hot |
| | 학생의 역량 평가 | -.195 | .398 | - |
| | 수학성취도에 영향을 미치는 인지적 요인 | .135 | .559 | - |
| | 대학 수학 강의 평가 | -.142 | .539 | - |
| | 수업에서의 학생 평가 | -.33 | .564 | - |
| | 컴퓨터 도구 평가 | .164 | .478 | - |
| | 교사 교육 평가 | -.023 | .923 | - |
| | 수학성취도에 영향을 미치는 가정 요인 | .296 | .192 | - |

나타냈을 때 유의미한 결과를 도출하지 못한 것으로 판단된다.

국내의 Hot 토픽은 ‘학생의 역량 평가’로서 2006년 이후부터 그 연구의 비중이 지속적으로 증가하는 추세를 보였다. 이러한 추세 변화의 원인을 살펴보면, 핵심역량을 규명하기 위해 수행된 OECD DeSeCo 프로젝트와 수학 교과의 핵심역량을 언급한 2009, 2015 개정 교육과정에 영향을 받은 것으로 보인다(Hwang & Kim, 2018). 또한 2015년 이후 연구 비중이 급격히 증가한 ‘교육과정 평가’ 연구의 경향과도 무관하지 않다(Figure 5 참조). 관련 연구로는 9개국의 수학과 교육과정에 반영된 핵심역량을 파악한 Kim et

al.(2015)을 예로 들 수 있다.

국외의 Hot 토픽 ‘교육과정 평가’는 2000년부터 2005년까지 비중이 급격하게 증가하다가 2010년까지는 감소하고 이 후 다시 증가하는 추세를 보였다. 즉, 2000년과 2010년을 기점으로 증가 추세를 보이는데, 이는 2000년과 2010년 설정된 기준의 영향을 받은 것으로 보인다. NCTM(2000)에서 학교수학의 원리와 기준을 발간함에 따라 기준에 대한 관심이 증가했는데, 이로 인해 기준과 관련된 다양한 교육과정 평가 연구가 수행되었다. 관련 연구로는 기준 기반 중학교 교육과정의 효과 평가(Reys, Reys, Lapan, Holliday, & Wasman, 2003), 교사의 지식과 기준 기반 교육과정에 의해 구성되는 학습 맥락 유형의 체계화(Chinnappan & Lawson, 2005)가 있다. 또한 수학에 대한 국가 교육 기준을 수립하기 위해 CCSSM이 2010년에 발표됨에 따라 기준에 관한 교육과정 연구들이 세계적으로 다시 주목을 받아왔으며, 이로 인해 수학과 교육과정을 비교하고 평가하는 연구 또한 새로이 연구자들의 관심을 받은 것으로 판단된다. CCSSM과 관련된 탐구 기반 학습의 잠재적 문제 탐색에 관한 연구(Schoenfeld & Kilpatrick, 2013)와 에스토니아, 핀란드, 스웨덴의 국가 수준 교육과정을 비교·분석한 연구(Hemmi, Bråting, & Lepik, 2020)를 그 예로 들 수 있다.

V. 결론 및 논의

본 연구는 2000년부터 2020년까지 수학교육 평가 연구를 여러 토픽으로 대별하고 이들이 어떤 의미를 지니는지, 그리고 시기별 변화가 어떠한지 살펴보았다. 그동안 연구 동향을 분석하고 그 전망에 대해 시사점을 제언한 여러 연구가

있었으나 수학교육 평가 연구의 동향을 시계열 변화에 따라 세밀하게 다룬 연구가 없다는 점에서 본 연구의 의의가 있다. 분석 결과를 토대로 도출한 결론은 다음과 같다.

첫째, 국내외 연구에서 공통된 토픽과 독자적인 토픽이 존재하며, 토픽별로 주로 연구되는 하위 주제는 상이한 경향을 보였다. LDA 수행 결과, 국내 수학교육 평가 연구에서는 8개의 토픽, 국외 수학교육 평가 연구에서는 11개의 토픽이 분류되었다. 분류된 토픽을 비교하였을 때 유사한 토픽 5개와 상이한 토픽 9개가 나타났는데, 먼저 유사한 토픽으로는 ‘수학적 모델 평가’, ‘예비교사 평가’, ‘대규모 평가’, ‘교육과정 평가’, ‘학생의 역량 평가’로 나타났다. 그러나 공통된 토픽일지라도 국내외에서 주목하고 있는 하위 주제는 서로 다르게 나타났다. 국내 수학교육 평가 연구의 경우 평가도구, 평가틀, 프로그램, 문항 등에 주로 연구의 초점이 맞추어진 반면, 국외 수학교육 연구에서는 신념, 추론, 개념, 인지적 유연성, 동기, 상호작용과 같이 인지적·정의적 영역의 평가에 보다 주목하고 있었다. 상이한 토픽을 살펴보면, 국내 연구에서는 ‘영재 및 기초학력 프로그램 평가’, ‘교사 평가 기준’, ‘평가 실태 및 활용’의 3개 토픽이 나타났다. 이는 국내 연구가 평가 기준을 마련하거나 개발한 프로그램을 평가하고 현장에서의 평가 결과를 환류하고 적용하는데 초점을 맞추고 있음을 의미한다. 반면, 국외에서는 ‘수학성취도에 영향을 미치는 인지적 요인’, ‘수학성취도에 영향을 미치는 가정 요인’, ‘대학 수학 강의 평가’, ‘수업에서의 학생 평가’, ‘컴퓨터 도구 평가’, ‘교사 교육 평가’의 6개 토픽이 독자적인 연구 주제로 나타났다. 이는 국외 연구에서 수학성취도와 관련된 요인 탐색과 평가, 컴퓨터 도구 활용과 수업에서의 학생 평가, 교사 교육 평가에 주목하고 있다

고 판단할 수 있다.

둘째, 국내 연구가 주목하고 있는 연구 주제는 시기별로 변화하는데 반해, 국외 연구의 연구 주제는 시간의 흐름과 무관하게 일관된 경향을 보였다. 키워드의 시기별 변화 양상에 주목하였을 때, 국내 연구는 수행 평가, 프로그램 평가, 수업 평가, 교육과정 평가 순으로 주된 연구 경향이 바뀌어져 왔으나 국외 연구에서는 문제와 과제에 관한 평가 연구가 꾸준히 수행되어온 것으로 나타났다. 또한 시기별로 공통적으로 출현하는 키워드를 추출하였을 때, 국내 연구에서는 10개의 키워드 중 5개의 키워드가 지속적으로 출현했으나 국외 연구에서는 9개의 키워드가 시간의 흐름과 무관하게 공통적으로 출현하였다. 이러한 결과는 국내 연구가 시기별 특정 이슈에 보다 민감하게 반응하는 경향이 있음을 시사한다. 물론 국외 연구 또한 특정 이슈와 관련된 연구가 수행되지만, 국내에 비해 상대적으로 일관된 연구를 수행하고 있다고 판단할 수 있다.

셋째, 수학교육 평가 연구의 특정 토픽들은 자연스러운 증감 패턴을 나타내기 보다는 특정 시기에 크게 변동하는 경향을 보였다. 시계열 토픽 분석 결과에 따르면, 국내 연구의 총 8개 토픽 중 비중이 감소하는 경향을 보이는 토픽은 ‘수학적 모델 평가’, ‘영재, 기초학력 프로그램 평가’, ‘교사 평가 기준’이며, 증가 경향을 보이는 토픽은 ‘교육과정 평가’, ‘학생의 역량 평가’, ‘평가 실태 및 활용’로 나타났다. 반면 국외 연구에서 비중이 감소하는 토픽은 ‘학생의 역량 평가’와 ‘예비교사 평가’이며 비중이 증가하는 토픽은 ‘수학적 모델 평가’와 ‘교육과정 평가’로 나타났다. 이 외 토픽들은 시간에 따른 일부 증감 변화를 보이지만 전체적으로 연구의 비중은 유지되는 경향을 보였다. 통계적으로 유의미한 경향 변화를 보인 토픽은 2개이며 국외 연구의

‘교육과정 평가’와 국내 연구의 ‘학생의 역량 평가’가 유의미한 증가 경향을 나타내는 Hot 토픽으로 나타났다.

본 연구의 결과를 통해 도출한 수학교육 평가 연구에 대한 시사점은 다음과 같다.

첫째, 수학교육 평가 연구가 명확한 방향성을 갖고 일관되게 수행될 필요가 있다. 시기별 키워드 분석 결과에 따르면, 국내 수학교육 평가 연구는 시기별로 집중된 키워드가 다르게 나타났으며 지속적으로 출현하는 연구 키워드도 국외에 비해 적게 나타나는 경향을 보였다. 이는 2000년부터 2020년까지 문제와 과제가 일관된 최상위 연구 키워드로 출현한 국외 연구와 대비되는 결과이며, 국내 연구에서 초점을 갖고 연구하는 주제가 빠르게 변화한다는 것을 시사한다. 물론, 시기별로 이슈가 되는 주제에 주목하고 이에 대한 연구를 수행하는 과정은 자연스러운 현상으로 볼 수 있다. 그러나 연구의 지속성과 깊이를 고려했을 때, 단기간의 집중된 연구도 중요하지만 한 분야에 대한 일관된 연구 수행도 함께 이루어질 필요가 있다.

둘째, 학생의 인지적·정의적 측면에 대한 평가 연구가 확산되어야 한다. LDA와 시기별 키워드 분석 결과에 따르면, 국내의 연구에서 5개의 유사한 토픽이 나타났으나 이에 따르는 상위 키워드들은 다르게 출현하였다. 특히, 국외 연구가 추론, 개념, 신념, 동기 등의 인지적·정의적 측면의 평가에 주목하는데 비해, 국내 연구는 평가 틀, 프로그램, 도구 평가에 초점을 맞추고 있었다. 이는 최근 국내 수학교육 평가 연구가 다양한 분야로 확장되고 있으나 전반적인 연구의 비중은 학생 성장의 모니터링이나 교수 결정을 위한 정보 제공의 목적보다 학생 성취도 측정과 프로그램의 효과성 판단에 보다 집중되었다는 것을 의미한다. 따라서 성취도 측정과 프로그램

효과성을 평가하려는 노력도 중요하지만 성장 참조 평가의 측면에서 수학적 가치, 동기, 이해, 추론과 같은 인지적·정의적 측면에 대한 평가 연구가 더욱 확산될 필요가 있다.

셋째, 교사의 전문성 개발을 위한 교사 평가 연구가 다양한 측면에서 수행될 필요가 있다. LDA 수행 결과, 국외 연구에서 ‘교사 교육 평가’가 특징적인 토픽으로 나타났으며 중재, 수학 내용지식, 교수학적 내용지식, 피드백, 반응 등 다양한 측면에서 교사 평가가 이루어지고 있었다. 반면, 국내 연구에서는 교사 평가 기준이나 지식에 관한 단어들만 도출되어 교사 평가에 관한 방법과 대상이 국외에 비해 제한적이었다. 이는 현직 교사 연구가 국내에 비해 국외에서 활발하게 이루어지고 있다고 보고한 Shin(2020)의 연구와 유사하다. 즉, ‘예비교사 평가’가 국내외 연구에서 유사한 토픽으로 도출된 것과 달리, 현직 교사의 전문성에 관한 연구는 국외에 비해 국내에서 많은 주목을 받고 있지 않음을 의미한다. 수업은 다양한 요인들의 복합적 과정이며 교사의 전문성은 수업에 대한 자기 연찬 과정에서 이루어진다는 점을 고려할 때, 예비교사의 전문성뿐만 아니라 현직 교사의 교과 지식, 학생과의 상호작용, 교사 교육 프로그램, 수업의 질, 수업 관행 등에 대한 다양한 평가와 피드백이 수행되어야 할 것이다.

넷째, 대학 수학 강의에 대한 연구가 수행될 필요가 있다. LDA 수행 결과, 기하, 추론, 증명과 관련된 대학 수학 강의와 질을 평가하는 ‘대학 수학 강의 평가’가 국외 연구의 독립적인 토픽으로 출현하였다. 이는 대학교 수학 강의를 예비 교사의 수학 교과 지식 발달에 중요한 역할을 함을 인식하고 강의의 질을 향상시키기 위한 연구가 꾸준히 이루어져 왔음을 시사한다. 반면, 국내에서는 예비교사 교육에 대한 토픽이 출현

하였으나 연구의 초점이 주로 초·중등 수학교육 평가에 치중되어 있고, 수학 강의와 같은 고등교육 평가 연구는 이에 비해 상대적으로 저조하므로 대학 수학 강의에 대한 심도 있는 평가 연구가 진행될 필요가 있다.

다섯째, 수학교육 평가에 컴퓨터를 활용하는 방안에 대한 연구가 필요하다. ‘컴퓨터 도구 평가’는 국내 연구에서는 나타나지 않았으나, 국외에서는 하나의 독립된 연구 분야로 나타났다. 이는 Pang et al.(2019)이 교육공학과 교구에 대한 국내 연구가 저조하다고 지적한 바와 같이, 평가 연구도 이러한 수학교육 연구 경향의 직·간접적인 영향을 받은 것으로 보인다. 컴퓨터를 활용한 평가는 채점자의 부담을 경감시켜 대규모 학생을 단기간에 평가할 수 있고 학생의 발달에 관해 보다 정확한 평가가 가능하다는 장점이 있다(Ginsburg et al., 2016). 이러한 생산성과 효율성으로 인해 그동안 국외에서는 컴퓨터를 활용한 평가 연구가 활발히 이루어져 왔다(e.g., Aldon & Panero, 2020; Foster, Anthony, Clements, Sarama, & Williams, 2016; Hoogland & Tout, 2018). 국내에서도 컴퓨터와 공학 기술을 활용하여 전통적 평가 방식의 한계를 극복하려는 노력이 시도되고 있으나 국외에 비해 연구의 관심은 저조한 실정이므로, 보다 효과적이고 체계적인 평가를 수행할 수 있도록 컴퓨터 도구 활용에 대한 평가 연구가 더욱 확산될 필요가 있다.

본 연구는 2000년부터 2020년 8월까지 국내외 수학교육 평가 연구의 동향을 분석하기 위해 토픽 모델링을 활용하여 연구 흐름을 조망하고 수학교육 평가 연구의 후속 과제를 제언하였다. 본 연구 결과가 추후 수학교육 평가 연구 발전에 유용한 도움이 되기를 기대한다.

참고문헌

- Aldon, G. & Panero, M. (2020). Can digital technology change the way mathematics skills are assessed? *ZDM*. <https://doi.org/10.1007/s11858-020-01172-8>
- Anhalt, C. O., Cortez, R., & Bennett, A. B. (2018). The emergence of mathematical modeling competencies: An investigation of prospective secondary mathematics teachers. *Mathematical Thinking and Learning*, 20(3), 202-221.
- Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, 50(1), 138-156.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Blei, D. M. & Lafferty, J. D. (2009). Topic models. In A. Srivastava, & M. Sahami (Eds.), *Text Mining: Classification, Clustering, and Applications* (pp. 71-94). Boca Raton, FL: Taylor & Francis Group.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3, 993-1022.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72, 1775-1781.
- Chae, S. H. (2001). Value relevance and conflict in educational evaluation. *Korean Journal of Educational Research*, 39, 27-50.
- 채선희(2001). 교육평가에 개입되는 가치간의 갈등유형 분석과 갈등해결을 위한 평가모형 탐색. *교육학연구*, 39, 27-50.
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computer & education. *Computers & Education*, 151, 103855.
- Chinnappan, M. & Lawson, M. J. (2005). A framework for analysis of teachers' geometric content knowledge and geometric knowledge for teaching. *Journal of Mathematics Teacher Education*, 8(3), 197-221.
- Cho, H. M., Kwon, O. N., Bae, Y. G., & Lee, A. R. (2014). Analyzing research trend of interaction in mathematics classroom in Korea. *Korean Association For Learner-Centered Curriculum And Instruction*, 14(6), 363-387.
- 조형미, 권오남, 배영곤, 이아란(2014). 한국 수학교실의 상호작용에 대한 연구 동향 분석, *학습자중심교과교육연구*, 14(6), 363-387.
- Cueto, S., Ramirez, C., & Leon, J. (2006). Opportunities to learn and achievement in mathematics in a sample of sixth grade students in Lima, Peru. *Educational Studies in Mathematics*, 62(1), 25-55.
- Demircioğlu, H., Argün, Z., & Bulut, S. (2010). A case study: assessment of preservice secondary mathematics teachers' metacognitive behaviour in the problem-solving process. *ZDM*, 42(5), 493-502.
- Drijvers, P., Kodde-Buitenhuis, H., & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathematics*, 102(3),

- 435-456.
- English, L. D. (2016). Revealing and capitalizing on young children's mathematical potential. *ZDM*, 48(7), 1079-1087.
- Ferrini-Mundy, J. & Schmidt, W. H. (2005). International comparative studies in mathematics education: Opportunities for collaboration and challenges for researchers. *Journal for Research in Mathematics Education*, 36(3), 164-175.
- Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. M. (2016). Improving mathematics learning of kindergarten students through computer-assisted instruction. *Journal for Research in Mathematics Education*, 47(3), 206-232.
- Ginsburg, H. P., Lee, Y. S., & Pappas, S. (2016). A research-inspired and computer-guided clinical interview for mathematics assessment: introduction, reliability and validity. *ZDM*, 48(7), 1003-1018.
- Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101, 5228-5235.
- Grover, B. W. & Connor, J. (2000). Characteristics of the college geometry course for preservice secondary teachers. *Journal of Mathematics Teacher Education*, 3(1), 47-67.
- Ha, S. H., Pang, J. S., & Ju, M. K. (2010). Research trends in elementary mathematics education: Focused on the papers published in domestic journals during the recent 5 years. *The Mathematical Education*, 49(1), 67-83.
- 하수현, 방정숙, 주미경(2010). 초등수학교육 연구동향: 최근 5년간 게재된 국내 학술지 논문을 중심으로. *수학교육*, 49(1), 67-83.
- Han, C. R. & Kwon, O. N. (2018). Domestic research trends and tasks on early algebra education: Focused on the elementary school mathematics. *Journal of Elementary Mathematics Education in Korea*, 22(2), 115-142.
- 한채린, 권오남(2018). 국내 초기 대수 교육 연구의 동향과 과제: 초등 수학을 중심으로. *한국초등수학교육학회지*, 22(2), 115-142.
- Hawighorst, B. (2005). Parents' views on mathematics and the learning of mathematics: An intercultural comparative study. *ZDM*, 37(2), 90-100.
- Hemmi, K., Bråting, K., & Lepik, M. (2020). Curricular approaches to algebra in Estonia, Finland and Sweden: A comparative study. *Mathematical Thinking and Learning*. Retrieved from <http://doi.org/10.1080/10986065.2020.1740857>
- Hoogland, K., & Tout, D. (2018). Computer-based assessment of mathematics into the twenty-first century: pressures and tensions. *ZDM*, 50(4), 675-686.
- Hwang, Y. H. & Kim, S. H. (2018). The recognition and needs of mathematics teachers for instruction and assessment to enhance mathematical competency, *The Journal of Educational Research in Mathematics*, 28(4), 651-669.
- 황윤희, 김선희(2018). 교과 역량 신장을 위한 수업과 평가 개선에 대한 수학 교사들의 인식과 요구. *수학교육학연구*, 28(4), 651-669.
- Inglis, M. & Foster, C. (2018). Five decades of mathematics education research. *Journal for Research in Mathematics Education*, 49(4), 462-500.
- Jeon, Y., & Lee, J. H. (2019). A practice of process-focused assessment for mathematical

- subject competency in elementary mathematics classroom. *Journal of the Korean School Mathematics Society*, 22(3), 199-219.
- 전영주, 이종학(2019). 초등 수학교실에서 역량 기반 과정중심평가의 적용 방안에 대한 연구. **한국학교수학회논문집**, 22(3), 199-219.
- Kang, M. W., Kim, S. H., Park, J. H., Lee, S. J., Cha, Y. W., & Koh, S. S. (2010). The direction to assessment of school mathematics in accordance with 2009 reformed curriculum. *Communications of Mathematical Education*, 24(2), 301-323.
- 강명원, 김성호, 박지훈, 이선준, 차용우, 고상숙(2010). 2009 개정 교육과정에 따른 수학과 평가가 나아가야 할 방향. **수학교육 논문집**, 24(2), 301-323.
- Kilpatrick, J. (1993). The chain and the arrow: From the history of mathematics assessment. In M. Niss (Ed.), *Investigations into assessment in mathematics education* (pp. 31-46). Dordrecht: Kluwer Academic Publishers.
- Kim, H. M. & Han, S. Y. (2018). A study on the development of the assessment tool for mathematical problem solving competency. *School Mathematics*, 20(1), 83-105.
- 김혜미, 한선영(2018). 수학 문제해결 역량 평가도구 개발. **학교수학**, 20(1), 83-105.
- Kim, J. H. & Hwang, W. H. (2010). Assessment study on educational programs for the gifted students in mathematics. *Communications of Mathematical Education*, 24(1), 235-257.
- 김정현, 황우형(2010). 영재학급에서의 수학 영재프로그램 평가에 관한 연구. **수학교육 논문집**, 24(1), 235-257.
- Kim, M. K., Cho, M. K., & Joo, Y. R. (2012). A survey of perception and status about descriptive assessment-focused on elementary school teachers in Seoul area. *Journal of Elementary Mathematics Education in Korea*, 16(1), 63-95.
- 김민경, 조미경, 주유리(2012). 서술형 평가에 대한 인식 및 실태에 관한 조사연구-서울시 소재 초등교사를 중심으로. **한국초등수학교육학회지**, 16(1), 63-95.
- Kim, S. H., Park, K. M., & Lee, H. C. (2015). An exploration of international trends about the core competencies in mathematics curriculum. *The Mathematics Education*, 54(1), 65-81.
- 김선희, 박경미, 이환철(2015). 수학과 교육과정에 반영된 핵심역량의 국제적 동향 탐색. **수학교육**, 54(1), 65-81.
- Kim, Y. K. & Pang, J. S. (2017). Research trends in elementary mathematics education: Focused on the papers published in domestic journals during the resent seven years. *Education of Primary School Mathematics*, 20(1), 19-36.
- 김유경, 방정숙(2017). 초등수학교육 연구동향: 최근 7년간 게재된 국내 학술지 논문을 중심으로. **초등수학교육**, 20(1), 19-36.
- Kim, Y. O. & Jung, S. H. (2012). An analysis of research trends on mathematics education evaluation. *East Asian mathematical journal*, 28(4), 487-504.
- 김영옥, 정성희(2012). 한국의 수학교육 평가 유형에 관한 연구동향. **영남수학회**, 28(4), 487-504.
- Ko, E. S., Park, M., & Lee, E. J. (2016). Prospective elementary teachers' perceptions on assessment in mathematics. *School Mathematics*, 18(1), 61-83.

- 고은성, 박민선, 이은정(2016). 초등 예비교사들의 수학교과에서의 평가에 대한 인식. **학 교수학**, 18(1), 61-83.
- Koh, S. S., Han, H. S., & Lee, C. Y. (2013). A study on the development of the model for the process-focused assessment using manipulatives -Focused on middle school mathematics-. *Communications of Mathematical Education*, 27(4), 581-609.
- 고상숙, 한혜숙, 이창연(2013). 교구를 활용한 수학적 과정의 평가모델 개발에 관한 연구 - 중학교 수학을 중심으로. **수학교육 논문집**, 27(4), 581-609.
- Ku, N. Y. & Lee, K. H. (2020). An analysis of research trends on mathematics teacher's curriculum material use in Korea and overseas. *Korean Association For Learner-Centered Curriculum And Instruction*, 20(9), 377-400.
- 구나영, 이경화(2020). 국내외 수학 교사의 교육과정 자료 사용 연구 동향 분석. **학습자 중심교과교육연구**, 20(9), 377-400.
- Lee, E. J., & Lee, K. H. (2011). A study on the factors influencing gender differences changes of Korean students in PISA mathematics assessment. *Journal of Educational Research in Mathematics*, 21(4), 313-326.
- 이은정, 이경화(2011). PISA 수학성취도 평가에 나타난 한국 학생들의 성차 추이의 배경 요인 연구. **수학교육학연구**, 21(4), 313-326.
- Lee, K. S. & Jo, Y. D. (2014). The characteristics of middle school mathematics achievement levels based on the results of the national assessment of educational achievement from 2010 to 2012. *School Mathematics*, 53(2), 219-237.
- 이광상, 조윤동(2014). 2010-2012 년 국가수준 학업성취도 평가에서 나타난 초등학교 성취 수준별 학업 특성. **수학교육**, 53(2), 219-237.
- Ministry of Education(2017). *Korean national elementary mathematics 1-1 teachers' guide*. Seoul: Chunjae education.
- 교육부. (2017). **수학 1-1 교사용 지도서**. 서울: 천재교육.
- Morris, A. K. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Mathematics Teacher Education*, 9(5), 471-505.
- Nabli, H., Djemaa, R. B., & Amor, I. A. B. (2018). Efficient cloud service discovery approach based on LDA topic modeling. *Journal of Systems and Software*, 146, 233-248.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- Niss, M. (1993). Assessment in mathematics education and its effects: An introduction. In M. Niss (Ed.), *Investigations into assessment in mathematics education* (pp. 1-30). Dordrecht: Kluwer Academic Publishers.
- Pang, J. S. et al. (2019). Domestic research trends of mathematics education: An analysis of journals published from 1963 to 2019. *The journal of educational research in mathematics*, 29(4), 709-739.
- 방정숙 외 9명(2019). 국내 수학교육 연구의 동향 분석-1963년부터 2019년까지 게재된 국내 수학교육 학술지 논문을 중심으로. **수학 교육학연구**, 29(4), 709-739.
- Park, S. Y., & Kim, W. K. (2011). A comparative analysis on research trends of secondary

- mathematics education between Korea and overseas. *The Mathematical Education*, 50(3), 285-308.
- 박선영, 김원경(2011). 국내외 수학교육 연구 동향 비교 분석. *수학교육* 50(3), 289-310.
- Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 34(1), 74-95.
- Sarama, J., Clements, D. H., Barrett, J., Van Dine, D. W., & McDonel, J. S. (2011). Evaluation of a learning trajectory for length in the early years. *ZDM*, 43(5), 667.
- Schoenfeld, A. H. & Kilpatrick, J. (2013). A US perspective on the implementation of inquiry-based learning in mathematics. *ZDM*, 45(6), 901-909.
- Shin, D. J. (2020). A comparative study of domestic and international research trends of mathematics education through topic modeling. *The Mathematical Education*, 59(1), 63-80.
- 신동조(2020). 토픽모델링을 활용한 국내외 수학교육 연구 동향 비교 연구. *수학교육*, 59(1), 63-80.
- Shin, J. S. (2011). A comparative study of mathematics curriculum in Finland. *Education of Primary School Mathematics*, 14(3), 225-236.
- 신준식(2011). 핀란드 수학과 교육과정 비교 분석. *초등수학교육*, 14(3), 225-236.
- Son, T. K. & Lee, K. H. (2020). An analysis of domestic research trends of mathematics curriculum research through topic modeling: Focused on domestic journals published from 1997 to 2019. *The Mathematical Education*, 59(3), 201-216.
- 손태권, 이광호(2020). 토픽모델링을 활용한 국내 수학과 교육과정 연구 동향 분석: 1997년부터 2019년까지 게재된 국내 수학교육 학술지 논문을 중심으로. *수학교육*, 59(3), 201-216.
- Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, 39(3), 247-280.
- Van den Heuvel-Panhuizen, M. & Elia, I. (2020). Mapping kindergartners' quantitative competence. *ZDM*, <https://doi.org/10.1007/s11858-020-01138-w>.
- Williams, S. R. & Leatham, K. R. (2017). Journal quality in mathematics education. *Journal for Research in Mathematics Education*, 48(4), 369-396.
- Wilson, L. D. (2007). High-stakes testing in mathematics. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1099-1110). Charlotte, NC: Information Age publishing.