

PISA 평가들에 따른 독서교육 연구 동향의 토픽모델링 분석

Topic Modeling Analysis of Research Trends in Reading Education Based on the PISA Assessment Framework

노 경 국 (Kyung-kuk Noh)* 김 은 혜 (Eun-Hae Kim)**
주 은 아 (Eun-a Ju)*** 최 희 주 (Hee-ju Choi)****

초 록

본 연구는 PISA 읽기 영역 평가들과 국내 독서교육 관련 연구 간의 연계 가능성을 탐색하고, 향후 연구 방향을 제안하는 것을 목적으로 한다. 이를 위해 2001년부터 2024년까지 발표된 국내 독서교육 관련 논문 1,204편을 수집하고, LDA 기반 토픽모델링 기법을 적용하여 전반적인 연구 동향을 분석하였다. 도출된 20개의 주요 토픽은 PISA 평가들의 세 차원(과정, 상황, 텍스트)과 비교하였으며, 그 결과 국내 연구는 '과정'과 '상황' 차원에 집중되어 있는 반면, 디지털 문해력이나 다중 텍스트 탐색 등 '텍스트' 차원에 대한 연구는 상대적으로 부족한 것으로 나타났다. 특히 도서관 서비스를 활용한 독서교육이 주요 주제로 부각되면서, 독서교육과 도서관의 연계 필요성을 확인하였다. 본 연구는 국내 독서교육 연구의 흐름을 국제적 평가 기준과 비교·분석하고, 실증적 데이터를 바탕으로 향후 연구 방향을 제시했다는 점에서 의의를 갖는다.

ABSTRACT

This study aims to explore the potential connections between the PISA reading assessment framework and domestic research on reading education in Korea, and to propose directions for future research. To this end, 1,204 Korean academic papers on reading education published between 2001 and 2024 were collected and analyzed using an LDA-based topic modeling method. The 20 major topics derived from the analysis were compared with the three dimensions of the PISA framework—process, situation, and text. The results indicate that domestic studies have primarily focused on the “process” and “situation” dimensions, while relatively few studies have addressed the “text” dimension, such as digital literacy and multiple-text comprehension. In particular, reading education utilizing library services emerged as a significant topic, underscoring the need for stronger integration between reading education and library programs. This study is significant in that it analyzes the trends of domestic reading education research in comparison with international evaluation standards and presents empirically grounded implications for future research.

키워드: 독서교육, 토픽모델링, PISA 읽기 영역 평가들, 디지털 문해력, 도서관 서비스
Reading Education, Topic Modeling, PISA Reading Assessment Framework, Digital Literacy,
Library Services

* 수원여자대학교 도서관 사서(libnoh@swc.ac.kr) (제1저자, 교신저자)

** 싱가포르 ISS 국제학교 강사(ibschoollibrarian@gmail.com) (공동저자)

*** 고양능곡중학교 사서교사(toyo93@korea.kr) (공동저자)

**** 논산여자고등학교 사서교사(memohee@naver.com) (공동저자)

논문접수일자 : 2025년 8월 20일 논문심사일자 : 2025년 8월 28일 게재확정일자 : 2025년 9월 1일
한국비블리아학회지, 36(3) : 173-200, 2025. <http://dx.doi.org/10.14699/kbiblia.2025.36.3.173>

※ Copyright © 2025 Korean Biblia Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구의 필요성과 목적

최근 초·중등학교를 중심으로 독서교육의 중요성이 지속적으로 강조되면서, 교육과정 개정, 수업 방법의 다양화, 독서 지도 프로그램 개발 등 다양한 실천적 노력이 이루어지고 있다. 그러나 『2023년 국민 독서 실태조사』에 따르면, 성인의 연간 독서율은 2013년 72.2%에서 2023년 43.0%로 급감하였고, 학생의 독서율도 96.8%에서 95.8%로 하락하는 등 전반적인 독서 활동 저하가 우려되고 있다(문화체육관광부, 2023). 특히 어휘력과 문해력 부족은 주요 교육 현안으로 꾸준히 지적되어 왔으며, 최근 10년 사이 기초학력 미달률이 최대 6배 가까이 증가하면서 교육 현장의 경각심을 불러일으키고 있다(통계청, 2025).

이러한 현실은 독서교육의 방향성과 실질적 효과에 대한 점검을 요구한다. 현재 국내 독서교육 연구는 양적 팽창에도 불구하고, 전통적인 문식성 중심의 접근에서 크게 벗어나지 못하고 있으며, 연구 주제가 특정 영역에 편중되어 있는 한계를 보인다(이지영, 2021, 603). 특히 21세기 사회가 요구하는 비판적, 통합적, 문제해결 중심의 읽기 역량을 독서교육과 어떻게 연결할 것인가에 대한 통합적 논의는 부족한 상황이다. 이와 같은 문제의식은 국내 독서교육 연구의 현황을 보다 구조적으로 분석하고, 독서교육의 방향성을 재조망하는 연구의 필요성을 시사한다.

한편, 읽기 역량의 국제적 기준을 제시하는 국제학업성취도평가(Programme for International

Student Assessment, PISA) 읽기 영역 평가틀(이하 PISA 평가틀)은 국내외에서 독서교육의 방향을 점검하고 재구성하는 데 중요한 준거로 활용되어 왔다. 기존 연구들은 주로 PISA를 비롯한 국제 및 국가 수준의 읽기 평가틀을 분석하여 국내 읽기 성취도 변화와 교육 정책에 대한 시사점을 도출해 왔다(노명완, 박창균, 2006; 정혜경, 조지민, 2013; 옥현진, 2014; 박혜영, 2017; 이인화, 2020). 이들 연구는 가정의 문화자본, 사회·경제·문화적 배경, 메타인지 전략, 독서 동기 등 다양한 요인이 읽기 성취에 영향을 미친다고 보고하며, 최근 성취도 하락과 하위 집단 확대를 교육 격차 심화의 신호로 해석하고 있다. PISA의 읽기 능력 개념은 시행 초기(2000)에는 목적 달성, 지식 및 잠재력 개발, 사회 참여를 위해 서면 텍스트를 이해·사용·반성하는 능력에 중점을 두었다(OECD, 2000). 그러나 2009년과 2015년 평가틀 개정에서는 디지털 텍스트 환경을 반영하고, 읽기에 대한 참여와 메타인지 요소를 포함하는 등 단순 독해를 넘어선 역량이 강조되었다(OECD, 2010; OECD, 2017). 특히 2018년 개정에서는 읽기 능력을 이해, 사용, 평가, 반영, 참여의 다차원적 과정으로 정의하며, 목표 지향적 읽기, 과제 수행, 정보 탐색 및 통합 능력 등 복합적 정보 처리 역량을 포괄하도록 확장하였다(OECD, 2019).

그러나 이러한 논의는 성취도 변화나 평가틀 자체에 집중되는 경향이 강하며, 국내 독서교육 연구 전반이 이 평가틀과 어떻게 맞물려 있는지를 구조적으로 분석한 연구는 부족하다. 이에 따라 PISA 읽기 평가틀을 기준으로 국내 독서교육 연구의 흐름과 주제 경향성을 정량적으로 진단하고, 국제 기준과의 연계 가능성을

탐색할 필요가 있다. 이러한 분석은 국내 독서교육 연구가 국제적인 읽기 역량 평가 기준과 얼마나 정합성을 이루고 있는지를 진단하는 데 중요한 기초자료가 될 뿐 아니라, 국내 연구의 편중된 경향이나 미진한 영역을 확인하여 균형 잡힌 연구·정책 방향과 교육 현장 개선 전략을 마련하는 근거가 된다.

본 연구는 PISA 평가들이 제시하는 읽기 역량의 구조와 특징을 분석하고, 최근 20여 년간 발표된 국내 독서교육 관련 학술 논문에 토픽 모델링을 적용하여 주요 주제의 분포와 흐름을 파악하고자 한다. 이를 바탕으로 국내 독서교육 연구가 PISA 평가들과 어떤 공통점과 차이를 보이는지 비교·분석하고자 한다. 이를 위한 연구 문제는 다음과 같다.

- 연구 문제 1. PISA 읽기 영역 평가들이 제시하는 읽기 역량의 구조와 특징은 무엇인가?
- 연구 문제 2. 국내 독서교육 연구는 PISA 읽기 영역 평가들과 어떤 공통점과 차이를 보이는가?

1.2 선행연구

본 연구에 앞서, 디지털 환경의 변화와 독서교육에 관한 연구와 텍스트 마이닝 및 토픽모델링 기법을 적용한 연구 사례를 살펴보았다.

먼저, 디지털 환경 변화에 따른 독서교육과의 관계를 중심으로 한 연구를 살펴보았다. 이인화 외(2021)는 학생들의 디지털 기기 사용 경향과 읽기 성취 간의 상관관계를 분석하여 비판적 디지털 기기 활용 교육의 필요성을 강

조하였다. 이지영(2021)은 ‘디지털 글쓰기’ 관련 연구를 대상으로 언어 네트워크 분석을 실시하여 주제의 빈도 및 네트워크 내 위치를 기반으로 연구 간 관계를 구조적으로 분석하였으며, 주제의 출현과 소멸 양상을 통합적으로 파악함으로써 연구 동향 분석의 새로운 가능성을 제시하였다. 정혜원 외(2021)는 랜덤 포레스트와 다층 모형을 적용하여 한국과 핀란드 학생의 읽기 성취에 영향을 미치는 변인을 비교·분석하여 핀란드와 비교해 한국 청소년의 인지적 및 정의적 역량을 함양하기 위한 시사점을 제안하였다. 강봉숙과 이승태(2024)는 학교도서관에서 디지털 기반의 독서교육을 위한 교육부와 교육청의 디지털 매체 제공 플랫폼 분석을 통해 디지털 기반 독서교육 활성화 방안을 도출하고 교육부 차원의 지원 필요성을 제시하였다. 김지혜와 이창희(2024)는 이주 배경 학생의 읽기 성취도와 배경 변인을 분석하면서 디지털 격차와 다문화 배경이 읽기 소양 형성에 미치는 영향을 밝혔다. 박주현과 강봉숙(2023)은 디지털 협력 수업이 학생들의 독서 리터러시 향상에 미치는 효과를 확인하고자 하였다. 연구 결과, 디지털 협력 수업은 고등학생의 표현 능력, 이해 능력, 정보 활용 능력에 통계적으로 유의미한 긍정적 영향을 미치는 것으로 나타났다. 이영호(2024)는 PISA 2022 데이터를 활용해 ICT(Information and Communications Technology) 활용 역량과 디지털 읽기 성취 간의 상관관계를 확인하였다.

다음으로 텍스트 마이닝 및 토픽모델링 기법을 적용한 연구 사례로, 김승주(2019)는 국어교육 분야를 중심으로 토픽모델링과 시계열 분석을 통해 국내 독서교육 연구의 주요 주제와

변화 양상을 분석하였다. 그 결과 독서교육이 교과서 텍스트, 교육과정, 문학교육에 편중되어 있음을 지적하였으며, 시간에 따른 주제 변화 추적의 가능성을 보여주었다. 정영주와 김혜진(2020)은 토픽모델링을 활용하여 학교도서관 관련 연구 동향을 분석한 결과, 27개의 주요 토픽을 도출하였다. 이 중 도서관 활용 수업과 정보 활용 관련 주제가 가장 많이 연구된 것으로 나타났다. 진성희와 김혜경(2022)은 교육 분야에서 수행된 텍스트 마이닝 활용 연구들을 심층 분석하여, 향후 다양한 교육 연구 주제에 텍스트 마이닝 기법이 유용한 연구 방법론으로 활용될 수 있음을 시사하였다. 임정훈(2022)은 물리학 독서 감상문 텍스트에 토픽모델링 기법을 적용하고, 도출된 주제어 간 관계를 연결 중심성, 매개 중심성, 위세 중심성과 같은 네트워크 분석 지표를 통해 정량적으로 분석하였다. 이 연구는 교육과정과 텍스트 간의 연계성을 구체적으로 규명하였다는 점에서 의의가 있다. 최숙기와 김세화(2022)는 국내 독서치료 분야의 연구 동향을 분석하기 위해 한국학술지인용색인을 활용하여 토픽모델링을 실시하였다. 그 결과를 바탕으로 독서치료 관련 연구의 주요 방향을 탐색하고, 향후 연구 방향을 제안하였다. 김은선(2025)은 토픽모델링 기법을 활용하여 국내 기초 문식성 연구의 전반적인 동향을 분석하고 향후 연구 방향을 제안하고자 하였다. 연구 결과 토픽의 분포는 교육과정 시기별로 차이를 보이는 것으로 나타났으며, 이는 향후 연구의 방향성을 제시하는 데 의의가 있다.

이처럼 기존 연구들은 읽기 성취도 분석, 디지털 전환, 교육 정책과의 연계 측면에서 다양한 논의를 전개해 왔다. 특히 디지털 환경 변화

에 따른 독서교육의 변화 양상을 분석하거나, 텍스트 마이닝 및 토픽모델링 기법을 활용해 특정 분야의 연구 동향을 정량적으로 분석하는데 주력해 왔다. 그러나 국내 독서교육 관련 연구가 국제학업성취도평가에서 제시하는 읽기 역량의 개념 및 구성 요소와 어떤 접점을 지니는지를 정량적으로 분석한 사례는 찾아보기 어렵다. 본 연구는 PISA 평가틀과의 연계성 관점에서 국내 독서교육 연구의 흐름을 체계적으로 조망하였다는 점에서 차별성을 지닌다.

2. 이론적 배경

2.1 PISA와 읽기 영역 평가틀

PISA는 2000년 약 30개국이 참여하며 시작되었으며, 이후 OECD 회원국뿐만 아니라 비회원국의 참여도 점차 확대되어, 8주기인 PISA 2022에는 총 81개국이 참여하였다. 이러한 참여 확대는 PISA가 학생, 학교, 교육 시스템의 특성을 체계적으로 점검하고, 국제적으로 비교 가능한 다양한 지표를 주기적으로 제공하기 때문으로 해석된다. 우리나라는 1주기인 PISA 2000을 시작으로 현재까지 전 주기에 지속적으로 참여해 왔다. 이를 통해 도출된 평가 결과와 장기적 추이 분석은 교육 성과를 모니터링하고 국가 교육 정책의 방향성을 설정하는 데 핵심적인 근거 자료로 활용되어 왔다.

PISA는 지난 20여 년간 다양한 평가 혁신을 주도하며 국제학업성취도평가의 표준을 제시해 왔으며, 현재 9주기인 PISA 2025 시행을 앞두고 있다(김성경 외, 2023, 19).

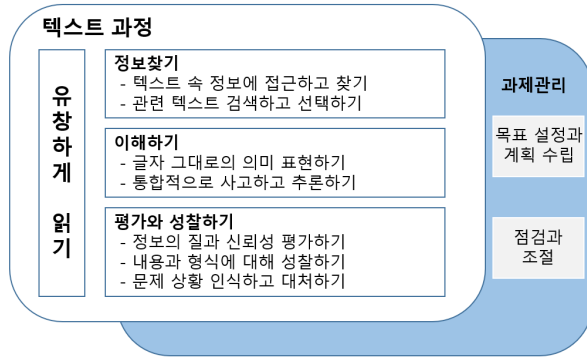
PISA 평가들은 PISA 2009, 2012, 2015 주기에서 동일한 평가들을 기반으로 읽기 영역을 측정하였으며, 평가들은 ‘양상’, ‘텍스트’, ‘상황’이라는 세 차원으로 구성되어 있다. ‘양상’ 차원에서는 ‘접근과 확인’, ‘통합과 해석’, ‘성찰과 평가’의 인지적 활동을 중심으로 구성되었고, ‘텍스트’ 차원에서는 텍스트의 체제 범주와 연속, 비연속, 혼합·다중 형식을 포함하며, 유형 범주에는 기술, 서사, 설명, 논증, 지시 등의 텍스트 종류가 포함되었다. ‘상황’ 차원은 텍스트가 사용되는 맥락에 따라 개인적, 공적, 교육적, 직업적 상황으로 구분되었다. 평가들은 PISA 2018을 기점으로 개정되었으며, PISA 2022에서도 해당 평가들과 함께 일부 추이 문항이 활용되었다. 이에 따라 PISA 2018에서 도입된 개념적 변화는 PISA 2022에 연속적으로 반영되었다. 이 시기의 읽기 소양은 ‘개인이 자신의 목표를 성취하고, 지식과 잠재력을 개발하며, 사회에 적극적으로 참여하기 위해 텍스트를 이해하고, 활용하며, 평가하고, 성찰하고, 교섭하는 능력’으로 정의되었다. PISA의 주기별 읽기 영역 평가들을 정리하면 <표 1>과 같다.

PISA 2022의 평가들은 2018년 평가들을 기반으로 하며, ‘과정’, ‘텍스트’, ‘상황’의 세 가지 차원으로 구성된다. 먼저, ‘과정’ 차원은 ‘텍스트 과정’, ‘과제 관리’, ‘유창하게 읽기’로 구성된다. ‘텍스트 과정’은 다시 ‘정보 찾기’, ‘이해하기’, ‘평가와 성찰하기’의 세 가지 하위 요소로 나뉜다. ‘정보 찾기’는 텍스트 내에 명시된 특정 정보나 세부 사항을 식별하고 선택하는 능력을 의미한다. ‘이해하기’는 명시적 또는 암시적인 정보와 아이디어를 해석하고, 이를 통합하거나 추론하는 능력을 포함한다. ‘평가와 성찰하기’는 텍스트의 신뢰성, 목적, 형식, 내용 등을 비판적으로 분석하고, 독자의 기존 지식이나 경험과 연결하여 성찰하는 능력을 의미한다. ‘유창하게 읽기’는 정확하고 자동화된 읽기를 통해 의미를 빠르게 구성하는 능력으로, 텍스트 처리의 효과적인 수행을 위한 전제 조건으로 간주된다. ‘과제 관리’는 읽기 활동 전반에서 요구되는 인지적 조절 능력과 자기조절 전략을 포함하며, 이는 목표 설정과 계획 수립, 이해 수준의 점검, 전략의 수정 및 보완 등 메타인지적 활동으로 구성된다. 이러한 ‘과정’ 차원은 읽기 활동 수행 시 복합

<표 1> PISA 주기별 평가들

PISA 2009 / 2012 / 2015			PISA 2018 / 2022		
차원	범주	종류	차원	범주	종류
양상	접근과 확인		과정	텍스트 과정	정보 찾기, 이해하기, 평가와 성찰하기
	통합과 해석			과제 관리	목표 설정과 계획 수립, 점검, 조정
	성찰과 평가			유창하게 읽기	
텍스트	체제	연속, 비연속, 혼합, 다중	텍스트	출처	단일, 다중
	유형	기술, 서사, 설명, 논증, 지시		조직과 탐색	정적, 동적
				체제	연속, 비연속, 혼합
유형	기술, 서사, 설명, 논증, 지시, 상호작용				
상황	개인적, 공적, 교육적, 직업적		상황	개인적, 공적, 교육적, 직업적	

※ 출처: 조성민 외, 2019, 42 재구성



〈그림 1〉 PISA 평가들의 '과정' 차원

※ 출처: OECD 2019, 33의 내용을 번역하여 인용함.

적으로 작동하는 인지적 과정을 설명한다. 과정 차원은 읽기 활동을 수행할 때 나타나는 인지적 과정으로 〈그림 1〉과 같이 복합적인 형태로 작동하는 것으로 볼 수 있다.

다음으로, '텍스트' 차원은 학습자가 읽는 텍스트의 성격에 따라 구성되며, 2018년 개정된 평가들에서는 '출처', '조직과 탐색', '텍스트 체제', '텍스트 유형'의 네 가지 하위 요소를 포함한다(OECD, 2019; 최인선 외, 2021, 307). '출처'는 단일 출처와 다중 출처 텍스트를 모두 포함할 수 있도록 설계되었다. '조직과 탐색'은 스크롤바, 검색 기능, 하이퍼링크 등 디지털 환경에서 자주 등장하는 동적 요소를 반영한다. '텍스트 체제'는 연속 텍스트, 비연속 텍스트, 혼합 텍스트로 구분된다. '텍스트 유형'은 텍스트의 기능이나 장르에 따라 기술, 서사, 설명, 논증, 지시, 상호작용 유형으로 분류된다. 또한, 정적 텍스트와 동적 텍스트의 구분은 정보 구조의 복잡성과 읽기 방식에 따른 학습자의 인지적 요구 차이를 반영한다. 이러한 구성은 디지털 환경을 포함한 다양한 텍스트 조건 속에서 학습자의 읽기 역량을 보다 정밀하게 측정하기

위한 것이다. 마지막으로, '상황' 차원은 읽기가 이루어지는 사회적·문화적 맥락을 반영하며, 평가 텍스트의 실제성과 타당성을 높이기 위해 고려된다. 이 차원은 '개인적 상황', '공적 상황', '교육적 상황', '직업적 상황'의 네 가지 범주로 구성된다. '개인적 상황'은 일상생활, 흥미, 취미 등과 관련된 읽기를 의미하며, '공적 상황'은 사회적 이슈나 공동체 활동 등 시민적 삶과 관련된 읽기를 포함한다. '교육적 상황'은 학교 학습이나 교육적 맥락에서 이루어지는 읽기를, '직업적 상황'은 직장이나 업무 수행과 관련된 읽기를 의미한다. 하나의 텍스트는 두 가지 이상의 상황 범주에 동시에 해당될 수 있으며, PISA는 이러한 맥락적 다양성을 반영하여 평가 문항을 구성함으로써 평가의 실제성과 타당성을 제고하고자 하였다(OECD, 2019, 46).

2.2 국내 연구 동향과 토픽모델링 기법

국내 독서교육 연구는 학습자의 읽기 능력 신장, 독서 동기 및 흥미 증진, 문해력 강화 등 다양한 측면에서 활발히 진행되어 왔다. 최근

디지털 매체의 확산과 교육과정 개정에 따라 독서교육 내용과 방법이 다양화되면서 연구 주제 또한 폭넓게 확장되고 있다. 예를 들어, 독서 흥미와 동기, 메타인지 전략, 디지털 환경에서의 문해력 등이 주요 관심사로 대두되고 있다(강봉숙, 이승태, 2024; 이인화, 2020; 노명완, 박창균, 2006). 그러나 방대한 연구 자료 속에서 전반적인 동향과 핵심 주제를 체계적으로 분석하기 위해서는 정성적 분석만으로는 한계가 있으므로, 정량적이고 체계적인 분석 도구의 도입이 필수적이다. 이에 최근 교육학 분야에서는 텍스트 마이닝과 토픽모델링 기법이 연구 동향 분석에 활발히 활용되고 있다(김지혜, 이창희, 2024; 진성희, 김혜경, 2022).

토픽모델링(Topic Modeling)은 확률 기반 알고리즘을 적용하여 방대한 문헌 집합으로부터 잠재적인 주제를 자동으로 추출하고, 문헌 간 주제의 유사성과 차이를 정량적으로 분석하는 텍스트 마이닝 기법이다(Blei et al., 2003). 주요 기법으로는 잠재 의미 색인(Latent Semantic Indexing, LSI), 이를 확장한 확률적 잠재 의미 색인(Probabilistic Latent Semantic Indexing, PLSI), 그리고 가장 널리 사용되는 잠재 디리클레 할당(Latent Dirichlet Allocation, 이하 LDA)가 있다(김희섭, 강보라, 2019). LDA는 이산형 데이터에 대한 확률적 생성 모델로, 문서와 단어 간의 동시 출현 정보를 바탕으로 전체 문서 집합의 잠재 구조를 추론한다(Blei et al., 2003; 박자현, 송민, 2013). 이를 통해 각 문서가 하나 이상의 주제를 어느 정도 비율로 포함하는지를 나타내는 주제 혼합 비율, 주제별 주요 단어 분포, 그리고 특정 단어가 특정 주제에 속할 확률 등을 산출할 수 있다. LDA는 토픽

모델링 과정에서 α 와 β 라는 하이퍼파라미터를 통해 주제 분포와 단어 분포의 밀도와 다양성을 조절하며(박중도, 2019), 사전 지식 없이도 적용 가능한 비지도 학습 기법으로, 분석자의 주관성을 최소화하고 주제 도출의 객관성을 확보할 수 있다는 점에서 다양한 분야에서 널리 활용되고 있다. 실제로 문헌정보학뿐 아니라 산업공학, 전산학, 사회과학 등에서 기술 동향 분석, 트렌드 탐지, 정책 분석 등에 효과적으로 사용되고 있다(김태경 외, 2016, 673).

국내 독서교육 분야에서도 토픽모델링을 적용한 연구가 증가하고 있으며, 이를 통해 연구 주제의 다변화와 주요 경향성을 객관적으로 진단하는 시도가 이루어지고 있다(진성희, 김혜경, 2022, 278). 이러한 분석은 연구 동향뿐 아니라 교육 정책 수립, 교수-학습 전략 개발, 맞춤형 교육 방안 마련 등에 유용한 기초 자료를 제공한다(박혜영, 2017, 278). 본 연구는 PISA 읽기 평가들을 기준으로 국내 독서교육 연구의 주제 분포와 변화를 토픽모델링 기법으로 분석함으로써 국제적 평가 기준과의 연계 가능성을 검토하고자 한다.

3. 연구 설계

3.1 분석 데이터 수집

본 연구는 국내 독서교육 관련 연구 동향을 살펴보고, PISA 읽기 영역의 평가 기준과 비교·분석함으로써 향후 독서교육 연구에 대한 시사점 도출을 목적으로 한다. 연구에 활용된 데이터는 한국과학기술정보연구원(KISTI)

이 운영하는 과학기술 지식인프라 통합서비스 ScienceON(<https://scienceon.kisti.re.kr>)을 통해 수집하였다. 검색 기간은 PISA가 처음 시행된 2000년을 제외하고, 그 이듬해인 2001년부터 2024년까지로 설정하였다. 먼저, 'PISA 읽기 영역'을 키워드로 검색하여 87편의 논문을 수집하였으며, 여기에 '독서교육' 키워드를 추가하여 1,117편의 논문을 추가로 확보하였다. 이에 따라 총 1,204편의 국내 발표 논문 데이터를 분석에 활용하였다.

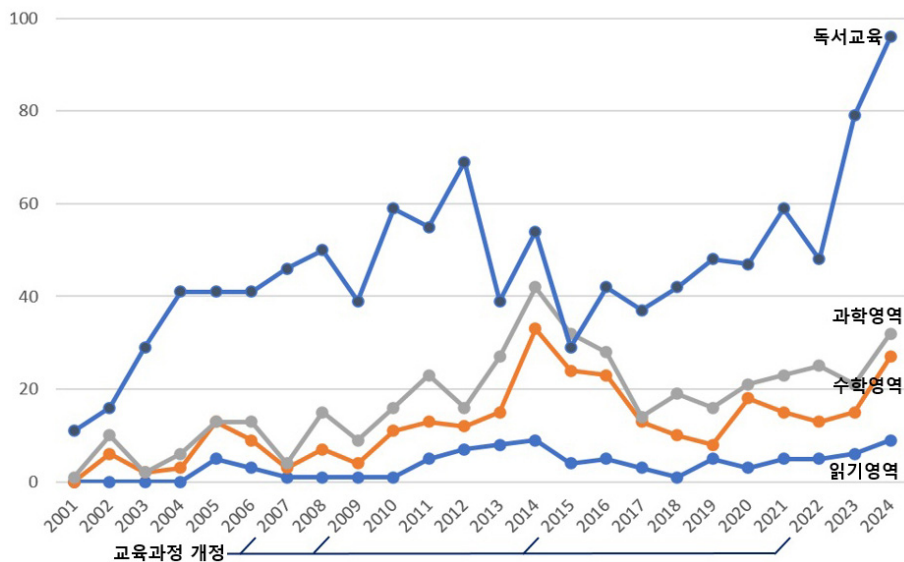
선정된 1,204편의 논문을 연도별 발행 수로 분석한 결과, 연도별 편차는 있으나 전반적으로 증가 추세를 보였다. <그림 2>는 연도별 '독서교육'과 PISA 각 영역 관련 연구 논문 수를 나타낸 것이다. 연도별로 다소 변동은 있으나, 전체적으로 관련 연구는 꾸준히 증가하고 있다. 그러나 <그림 2>에서 확인할 수 있듯이 다른 영역에 비하여 읽기 영역에 관한 연구는 다른 영

역에 비해 상대적으로 부족한 편임을 확인할 수 있다.

한편, 박민정과 성열관(2011)은 국내 연구는 국가 교육과정 고시 또는 개정 시점을 전후로 관련 연구가 집중되는 경향을 보인다고 하였다. 이러한 현상은 본 연구를 위해 확보한 독서교육 분야에서도 확인할 수 있다. 우리나라는 2007년, 2009년, 2015년, 2022년에 국가 교육과정이 개정되었는데(교육부, 2022; 국가교육과정정보센터, 2025) 이는 독서교육과 관련된 연구의 양적 증가에 직·간접적으로 영향을 미친 것으로 추정된다. 독서교육 및 PISA의 각 영역별 연구 동향은 <그림 2>와 같다.

3.2 데이터 전처리

본 연구는 국내에서 진행된 연구들의 주요 토픽을 분석하고자 텍스트 마이닝 기법인 넷마



<그림 2> 2001년 이후 PISA 및 독서교육 연구동향

이너 V4.5(www.netminer.com, 넷마이너)를 활용하여 텍스트 전처리 및 LDA 기반의 토픽 모델링을 수행하였다. 넷마이너는 텍스트 데이터를 기반으로 주요 키워드와 개념 간의 네트워크 분석을 지원하는 전문 소프트웨어로, 다양한 데이터 마이닝 기법을 활용할 수 있다. 특히, LDA 알고리즘을 통해 각 문서의 주제 분류와 함께 주제별 확률적 분석이 가능하다. 따라서, 본 연구에서는 LDA를 활용하여 국내에서 진행된 독서교육 관련 논문의 초록을 대상으로 텍스트 전처리와 토픽모델링을 실시하였다. 텍스트 전처리 과정은 다음과 같은 절차를 거쳐 수행되었다.

먼저, 분석의 초점을 명확히 하기 위해 형용사, 대명사, 동사 등 핵심 의미 도출에 불필요한 품사는 배제하고, 명사형 단어만을 추출하여 연구 주제와 관련된 핵심 명사들을 선별하였다. 이후 ‘문해력’, ‘문해능력’, ‘해석능력’과 같은 동의어나 유사어는 대표 단어로 통일하였고, ‘교원’, ‘교육자’, ‘교수자’ 등 유사한 의미를 지닌 단어들 또한 하나의 대표어로 정리하였다. ‘PISA’, ‘DLS’와 같은 축약어는 일관된 표기 방식으로 정비하였으며, ‘도서’, ‘전자책’, ‘그림책’, ‘소책자’ 등은 ‘장서’로, ‘초등학교’, ‘중학교’, ‘고등학교’, ‘대학교’, ‘대학원’ 등은 ‘학교’로 대표어를 설정하여 단어의 의미 통합과 분류의 일관성을 확보하였다. 또한, ‘독서교육’, ‘국어교육’, ‘학교도서관’과 같은 복합어는 의미 단위가 분리되지 않도록 복합 명사 처리를 적용하여 분석의 정확성을 높였다. 아울러, ‘최근’, ‘연구’, ‘목적’, ‘분석’ 등 학술 논문에서 자주 등장하지만 분석에 기여하지 않는 일반적 단어들과, ‘개’, ‘초’, ‘층’, ‘에’, ‘꼭’ 등 의미 없는 단일 음절 단어는 불용어

로 간주하여 분석에서 제외하였다. 이와 같은 정제 과정을 통해 분석 결과의 일관성과 신뢰성을 확보하고, 연구 주제와 관련된 핵심 개념의 추출 정확도를 제고하고자 하였다. 본 연구에서는 총 1,204편의 논문이 수집되었으나, 데이터 전처리 및 토픽모델링 과정에서 초록이 제공되지 않은 논문과 텍스트가 충분하지 않아 분석에 적합하지 않은 자료가 제외되어 최종적으로 1,039편이 토픽 분석에 활용되었다. 이상 데이터 전처리 작업의 사례는 <표 2>와 같다.

3.3 토픽 수 산정

본 연구에서는 대표적인 토픽모델링 기법인 LDA를 활용하였다. LDA 기반의 토픽모델링 결과를 정교하게 해석하기 위해서는 도출된 토픽의 품질과 적합성을 평가하는 과정이 필수적이며 본 연구에서는 토픽 일관성(topic coherence) 지표를 활용하여 최적의 토픽 수를 산출하였다. 토픽 일관성은 추출된 주제 내 단어들 사이의 의미상 유사한지를 평가하는 지표로, 대표적으로 c_v 방식과 UMass 방식이 활용된다(Roeder et al., 2024). UMass 방식은 문헌 기반의 단어 간 동시 출현 빈도를 기준으로 유사도를 측정하며, c_v 방식은 슬라이딩 윈도우 기반으로 평가한다. 일반적으로 c_v 값은 1에 가까울수록, UMass 값은 0에 가까울수록 해당 토픽의 일관성이 높다고 간주 된다. 이러한 지표는 모델이 생성한 토픽의 해석 가능성과 활용 가능성을 높이는 데 중요한 기준으로 작용한다(김지영, 노동조, 2023).

본 연구에서는 두 지표를 모두 고려하여 다양한 토픽 수에 대해 사전 검토를 실시하고, 그

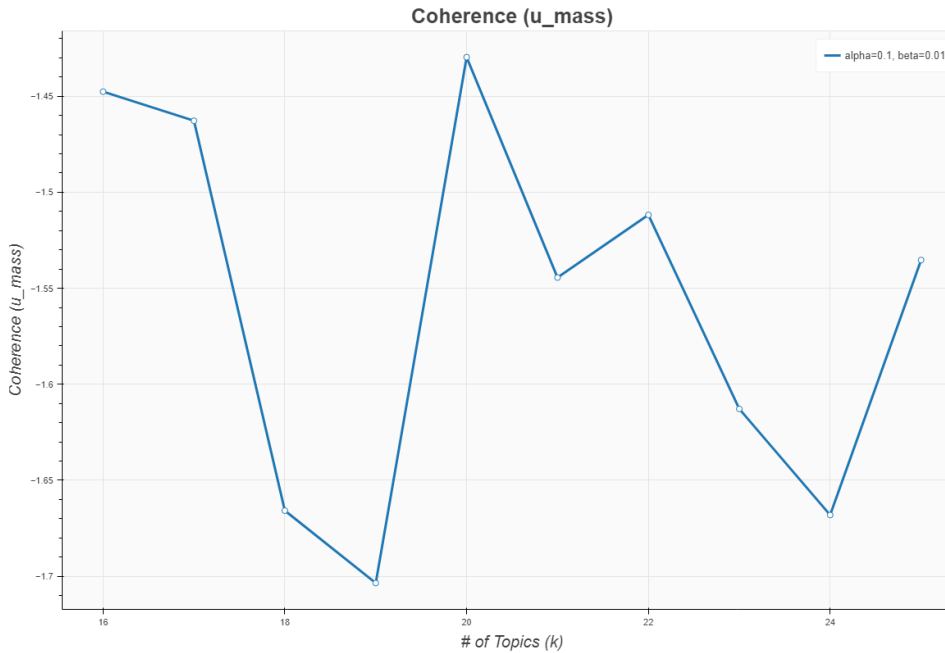
〈표 2〉 데이터 전처리 예시

No	사전	데이터 전처리 예시
1	유사 동일 용어	<ul style="list-style-type: none"> • 문해력 ← 문식력, 문해 능력, 해석 능력 등 • 교육 ← 공교육, 사교육, 통합교육, 평생교육 등 • 교사 ← 교원, 교육자, 교수자 등
2	축약어 통일	<ul style="list-style-type: none"> • PISA ← Programme for International Student Assessment, 국제학업성취도평가, 학업성취도평가 • DLS ← 독서교육종합지원시스템, 디지털 자료실, 디지털 자료실 지원센터, 디지털자료실, 디지털자료실지원센터 등 • WEB ← 웹페이지, 홈페이지, 인터넷 페이지 등
3	대표어 선정	<ul style="list-style-type: none"> • 장서 ← 도서, 전자책, 그림책, 소책자 등 • 학교 ← 초등학교, 중학교, 고등학교, 대학교, 대학원 등 • 학생 ← 초등학생, 중학생, 청소년, 대학생 등
4	복합어 정리	<ul style="list-style-type: none"> • 독서교육 ← 독서 교육 • 국어교육 ← 국어 교육 • 학교도서관 ← 학교 도서관
5	불용어 처리	<ul style="list-style-type: none"> • 고찰, 최근, 연구, 목적, 분석 등 반복적으로 등장하는 단어 • 길, 끝, 층, 쇼, 주 등 외글자로 의미가 없다고 판단되는 단어

결과를 비교·분석한 뒤 최적의 토픽 수를 선정하였다. 토픽 수의 결정은 하위 주제 도출과 결과 해석의 명확성에 큰 영향을 미치기 때문에 중요한 요소이다. LDA 모델은 사용자가 사전에 지정한 토픽 수에 따라 주제를 도출하는데 일반적으로는 혼잡도 등의 정량적 지표를 활용해 적절한 토픽 수를 판단하기도 한다. 그러나 혼잡도는 토픽 수가 증가함에 따라 지속적으로 0에 수렴하는 경향이 있어 해석 가능성과 실용성을 고려한 적정 토픽 수의 선정이 더욱 중요하다(Chang et al., 2009).

〈그림 3〉은 UMass 지표를 활용하여 토픽수(k)에 따른 모델 적합성을 비교한 결과를 제시한 것이다. 세로축은 토픽 일관성 점수, 가로축은 토픽 수를 의미하며, 그래프의 변화 양상을 통해 특정 구간에서 상대적으로 높은 일관성이 나타나는 지점을 확인할 수 있다. 본 연구에서는 토픽 일관성 지표를 중심으로 두 단계에 걸쳐 적절한 토픽 수를 탐색하였다. 이를 위해 먼

저 토픽 수를 5개에서 50개까지 5개 단위로 증가시키며 모델링을 수행하였고, 2단계에서는 10개, 20개, 25개, 30개 등 특정 구간을 중심으로 정밀한 비교·분석을 실시하였다. 그 결과, 16개, 20개, 22개의 후보군이 비교적 높은 일관성 점수를 보였으며, 이 가운데 k=20에서 가장 안정적이고 해석 가능한 토픽 구조가 도출되었다. 따라서 본 연구의 토픽모델링 분석에는 토픽 수 20개를 최종적으로 적용하였다. 이는 토픽 수가 지나치게 적을 경우(예: 10개) 하나의 토픽 내에 여러 주제가 혼합되어 모호성이 커지는 반면, 지나치게 많을 경우(예: 30개 이상) 세부 주제가 과도하게 분산되어 연구 맥락과의 연결성이 약화되는 현상이 발생하기 때문이다. 또한, 연구자의 주관적 개입을 최소화하고 분석의 객관성을 높이기 위해 정량적 평가 지표를 기반으로 토픽 수를 선정하였다. LDA 토픽모델링의 하이퍼파라미터는 α (alpha) 값 0.1, β (beta) 값 0.01로 설정하였다.



〈그림 3〉 토픽일관성(U_Mass)을 활용한 토픽모델 적합성 평가 결과

4. 연구 결과 및 분석

4.1 토픽모델링 결과

LDA 토픽모델링 수행 결과 도출된 20개의 토픽에 대한 토픽명은 토픽별 할당된 단어들의 분포를 기반으로 주제명을 부여하였다. 〈표 3〉에서 확인할 수 있듯이 대부분의 토픽명은 각 토픽의 상위 단어 10개의 조합으로 이끌어낼 수 있었다. 토픽모델링 결과 확인된 토픽과 선정된 주제명은 다음과 같다.

T1은 '교육', '실천', '지식', '과정', '자신' 등의 키워드를 중심으로, 교육의 실행 과정과 학습 전반에 대한 내용을 다룬 주제로 해석되어 주제명을 '교육과정'으로 설정하였다. T2는 '사회', '국가', '제도', '공간', '종교' 등과 같은 키워

드가 포함되어, 사회·문화적 제도 및 가치에 대한 비판적 사고를 요하는 평가 관점에서 '비판적 사고'로 명명하였다. T3은 '읽기', '문화', '프로그램', '활동', '지역' 등을 중심으로 독서 문화와 환경을 조성하는 활동을 나타내므로 주제명을 '독서문화'로 하였다. T4는 '주제', '분석', '분야', '정보', '문헌' 등의 키워드를 중심으로 정보의 해석과 통합적 이해 과정을 다루고 있어 '주제분석'으로 명명하였다. T5는 '학생', '학교', '교사', '결과', '분석' 등 학습자의 읽기 성과와 관련된 요소들을 포함하여, 주제명을 '학습결과'로 하였다. T6은 '운동', '교양', '읽기', '대중', '문화' 등의 키워드를 통해 일상적이고 실천적인 읽기 활동을 나타내며, 주제명은 '읽기활동'으로 해석되었다. T7은 '평가', '읽기', '문항', '능력', '전략' 등의 키워드가 읽기 능력

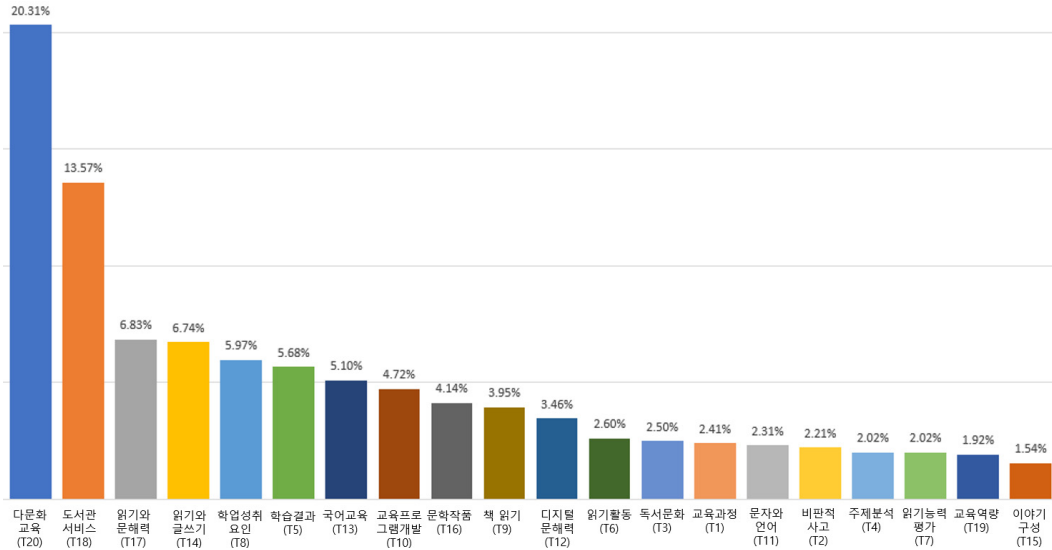
〈표 3〉 독서교육 관련 연구 토픽모델링 결과

구분	상위 키워드(10개)	키워드 주제	주제명	문헌수	분포
T1	교육, 실천, 지식, 과정, 자신, 학습, 이해, 목표, 사고, 경험	교육환경/제도적 틀	교육과정	25	2.41%
T2	사회, 국가, 제도, 공간, 종교, 일본, 전통, 시대, 말, 서당	사회·문화적 제도의 가치판단	비판적 사고	23	2.21%
T3	읽기, 문화, 프로그램, 활동, 지역, 정책, 사회, 인식, 생활, 환경	독서문화/환경	독서문화	26	2.50%
T4	주제, 분석, 분야, 정보, 문헌, 결과, 영역, 대상, 방법, 비교	정보의 해석 및 통합	주제분석	21	2.02%
T5	학생, 학교, 교사, 결과, 분석, 인식, 교육, 대상, 읽기, 영향	읽기목적/성과	학습결과	59	5.68%
T6	운동, 교양, 읽기, 대중, 문화, 경향, 책, 시기, 사회, 형성	일상적 독서	읽기활동	27	2.60%
T7	평가, 읽기, 문항, 능력, 전략, 수준, 검사, 개발, 결과, 도구	읽기능력 진단	읽기능력평가	21	2.02%
T8	학생, 성취, 영향, 과학, 학습, 학업, 수준, PISA, 수학, 학교	읽기성과 영향요인	학업성취요인	62	5.97%
T9	책, 읽기, 학생, 어린이, 활동, 자료, 교육, 참여, 활용, 내용	독서 습관	책 읽기	41	3.95%
T10	프로그램, 효과, 교육, 집단, 방법, 적용, 개발, 실험, 독서치료, 향상	독서프로그램 효과	교육프로그램개발	49	4.72%
T11	문자, 학문, 시대, 공부, 생활, 한글, 자신, 한문, 말, 문인	문자 체계	문자와 언어	24	2.31%
T12	정보, 문해력, 디지털, 기반, 활용, 자료, 시스템, 멀티미디어, 기술, 학습	디지털 텍스트의 정보 탐색	디지털 문해력	36	3.46%
T13	교육, 내용, 국어, 학습, 방법, 과정, 교육과정, 개발, 교과서, 교과	국어교과교육	국어교육	53	5.10%
T14	교육, 학습, 읽기, 글쓰기, 토론, 방법, 능력, 활동, 활용, 제시	읽기능력과 글쓰기	읽기와 글쓰기	70	6.74%
T15	서사, 이야기, 역사, 교육, 인물, 스토리텔링, 텍스트, 시조, 차제, 형태	서사적 읽기능력	이야기 구성	16	1.54%
T16	문학, 작품, 소설, 사회, 의미, 독자, 교육, 가치, 주제, 개념	문학 장르구분	문학작품	43	4.14%
T17	읽기, 독자, 텍스트, 방법, 문해, 과정, 의미, 사회, 변화, 책	읽기 정의와 탐구	읽기와 문해력	71	6.83%
T18	도서관 사서, 서비스, 분석, 정보, 공공, 교육, 프로그램, 운영, 방법	독서 지원 환경	도서관 서비스	141	13.57%
T19	교육, 역량, 과목, 교과, 과정, 영역, 운영, 요구, 사회, 선택	독자의 능력요인	교육역량	20	1.92%
T20	아동, 다문화, 동화, 가정, 활동, 자연, 교정, 번역, 이상, 가족	문화 다양성	다문화 교육	211	20.31%
합 계				1,039	100%

진단과 측정에 초점이 있으므로 판단하여 '읽기 능력평가'로 주제명을 정하였다. T8은 '학생', '성취', '영향', '학업', '과학' 등의 키워드를 중심으로 읽기 성과의 외부 요인을 다루었기 때문에 '학업성취 요인'으로 명명하였다. T9는 '책', '읽기', '학생', '어린이', '활동' 등의 키워드를 바탕으로 책 읽기 습관 형성과 관련된 연구로 해석되어 '책 읽기로 주제를 정하였다. T10은 '프로그램', '효과', '교육', '집단', '방법' 등의 키워드에서 독서 프로그램의 효과성을 분석한 연구로 판단되어 '교육프로그램개발'로 명명하였다. T11은 '문자', '학문', '시대', '공부', '생활' 등의 키워드가 텍스트와 문자 체계의 특성에 초점이 있음을 보여주어 '문자와 언어'로 설정하였다. T12는 '정보', '문해력', '디지털', '기반', '활용' 등의 키워드가 디지털 환경에서의 정보 탐색과 관련됨을 나타내어 '디지털문해력'으로 명명하였다. T13은 '교육', '내용', '국어', '학습', '방법' 등의 키워드에서 교과 중심의 국어교육 연구임을 보여주며 '국어교육'으로 설정하였다. T14는 '교육', '학습', '읽기', '글쓰기', '방법' 등의 키워드를 바탕으로 읽기와 글쓰기의 통합적 교수·학습에 초점을 두고 있어 '읽기와 글쓰기'로 주제를 설정하였다. T15는 '서사', '이야기', '역사', '교육', '인물' 등의 키워드를 중심으로 이야기 구성 요소와 서사 구조 이해에 관한 연구로 해석되어 '이야기 구성'으로 명명하였다. T16은 '문학', '작품', '소설', '사회', '의미' 등의 키워드를 통해 문학적 장르에 대한 이해를 나타낸다고 판단하여 '문학작품'으로 주제명을 정하였다. T17은 '읽기', '독자', '텍스트', '방법', '문해' 등의 키워드를 통해 읽기의 본질과 개념에 대한 탐구가 중심이므로 '읽기와 문해력'으로

설정되었다. T18은 '도서관', '사서', '서비스', '분석', '정보' 등의 키워드를 통해 독서 지원 기관과 시스템을 다룬 주제로 해석되어 '도서관 서비스'로 명명하였다. T19는 '교육', '역량', '과목', '교과', '과정' 등의 키워드가 독자의 역량과 관련된 교육적 요소를 나타내므로 '교육역량'으로 설정하였다. T20은 '아동', '다문화', '동화', '과정', '활동' 등의 키워드를 통해 다양한 문화적 배경을 지닌 독자를 위한 읽기 교육 연구를 다룬 것으로 판단하여 '다문화 교육'으로 주제명을 정하였다. 각 토픽에 따른 상위 주제명 10개와 문헌 수, 분포 등 세부 사항은 <표 3>과 같다.

토픽모델링 결과, 가장 큰 비중을 차지한 주제는 다문화 교육(T20, 20.31%)으로 나타났으며, 이는 독서교육 관련 연구에서 다문화적 배경과 학습자 다양성에 대한 관심이 꾸준히 확대되고 있음으로 해석할 수 있다. 두 번째로 높은 비율을 보인 도서관 서비스(T18, 13.57%)는 학교 도서관과 공공도서관을 중심으로, 학습자의 독서 환경 조성, 독서 지원 프로그램 운영, 사서교사의 역할, 그리고 디지털 자료 접근성 확보 등과 같은 주제가 집중적으로 다루어져 있음을 보여준다. 이는 도서관이 단순한 자료 제공 공간을 넘어 독서교육을 촉진하고 학습자의 읽기 경험을 확장하는 교육적 기반으로 기능하고 있음을 시사한다. 이외의 주제들은 모두 10% 미만의 분포를 보였는데, 읽기와 문해력(T17, 6.83%), 읽기와 글쓰기(T14, 6.74%), 학업성취요인(T8, 5.97%), 학습결과(T5, 5.68%), 국어교육(T13, 5.10%) 등이 비교적 고르게 분포하며 핵심 연구 영역을 구성하고 있었다. 토픽별 분포 추이는 <그림 4>와 같다.



〈그림 4〉 토픽별 문헌 분포

한편, LDA는 확률 기반 알고리즘으로 동일한 문서가 복수의 토픽에 기여할 수 있다. 따라서 일부 토픽 간에는 분포가 유사하게 나타나며, 이는 독서교육 연구의 여러 주제가 상호 연계되어 있음을 시사한다. 예를 들어, 국어교육 (T13, 5.10%)이나 디지털 문해력(T12, 3.46%)과 같은 주제는 각각 독립적으로 존재하지만, 읽기 능력, 교육과정, 비판적 사고와 같은 다른 주제들과도 긴밀히 연결되어 있다. 이러한 분포 양상은 독서교육 연구가 특정 분야에 국한되지 않고, 학습자의 읽기 능력 신장, 교육과정 개발, 디지털 환경 변화 등 다양한 맥락에서 종합적으로 탐구되고 있음을 보여준다.

4.2 PISA 평가들과 독서교육 토픽 비교·분석

본 연구는 PISA 평가들과 국내에서 수행된 독서교육 관련 연구를 대상으로 토픽모델링 기법을 적용하여 전반적인 연구 동향을 분석하

고, 이를 바탕으로 향후 국내 독서교육 연구의 방향성을 제안하는 데 목적이 있다. OECD는 읽기를 단순한 학문적 성과의 차원을 넘어, 학교 및 사회의 요구를 반영한 개념으로 재정의 하였으며, 읽기의 인지적 구성 요소를 ‘텍스트 처리’와 ‘과제 운용’의 두 차원으로 범주화하였다 (OECD, 2018).

본 연구에서는 토픽모델링을 통해 도출된 국내 독서교육 연구의 20개 주제를 PISA 읽기 영역의 세 가지 핵심 평가 요소인 ‘과정’, ‘텍스트’, ‘상황’을 기준으로 분류·분석하였다. 분석은 현장 사서 및 사서교사로서 10년 이상의 경력과 박사 과정 이상의 학문적 전문성을 갖춘 연구자 4인이 수행하였으며, 이를 통해 국내 독서교육 연구 주제가 국제적 읽기 평가 기준에 따라 어떠한 방식으로 체계화될 수 있는지를 탐색하였다. 매핑 과정은 PISA 평가들의 10개 하위 영역과 각 토픽을 구성하는 주요 키워드 간 의미적 유사성을 판단하여, 토픽을 해당 영

역에 배정하는 방식으로 이루어졌다. 먼저 상위 키워드 15~20개 중 3개 이상의 특정 평가나 영역의 정의, 설명 등과 직접적으로 일치 여부를 확인한 후 핵심 키워드 간 의미적 연관성이 평가 영역의 핵심 개념과 부합할 경우 해당 토픽을 그 영역에 할당하는 방식으로 진행하였다. 또한, 연구자 4인은 독립적으로 토픽-영역 매핑을 실시한 뒤 결과를 상호 비교하였으며, 의견이 불일치할 경우에는 다수 의견(3인 이상)을 우선 반영하였다. 다만 2:2로 의견이 갈린 경우에는 해당 토픽의 주요 키워드와 관련 문헌의 실제 활용 맥락을 추가 검토하여 최종 합의를 도출하였다. 이러한 절차를 통해 연구자의 주관적 개입을 최소화하고, 매핑 과정의 일관성과 신뢰도를 확보하고자 하였다.

분석 결과, 과정 차원의 텍스트 과정 범주에 서는 다음과 같은 매칭이 이루어졌다. '정보 찾기' 영역에는 도서관 서비스(T18)가 매칭되었는데, 이는 도서관을 활용하여 특정 정보를 탐색하는 과정과 높은 관련성을 가진다고 판단되었다. '이해하기' 영역에는 주제분석(T4), 디지털 문해력(T12), 읽기와 문해력(T17)이 매칭되었으며, 이는 텍스트에 담긴 정보와 아이디어를 해석하고 관계를 파악하여 의미를 구성하는 능력과 관련이 있다. '평가와 성찰하기' 영역에는 비판적 사고(T2), 학습 결과(T5), 읽기능력 평가(T7)가 매칭되었는데, 이는 텍스트의 내용, 형식, 목적, 신뢰성 등을 비판적으로 검토하고 이를 자신의 지식이나 경험과 연결하여 성찰하는 고차원적 읽기 능력과 관련된다. '유창하게 읽기'와 관련된 토픽으로는 읽기활동(T6), 책 읽기(T9), 읽기와 글쓰기(T14)가 매칭되었으며, 이는 정확하고 자동화된 읽기를 통해 의미

를 신속하게 구성하는 능력과 관련된다. 과제 관리 범주의 '목표 설정, 계획 수립, 점검 및 조정' 영역과 관련해서는 교육과정(T1)과 교육프로그램 개발(T10)이 매칭되었는데, 이는 읽기 전후의 과정을 조율하고 목표를 설정하며 자기 점검을 수행하는 기능과 밀접한 관련이 있다.

텍스트 차원과 관련하여, 출처(단일·다중 텍스트) 및 조직과 탐색(정적·동적 텍스트) 범주에서는 의미적으로 유사한 토픽이 확인되지 않았다. 그러나 텍스트 체제 범주(연속, 비연속, 혼합 텍스트)에서는 문자와 언어(T11)가 매칭되었는데, 이는 텍스트 형식과 구조에 따라 구분되는 체제와 유의미한 관련성을 보여준다. 또한 텍스트 유형(기술, 서사, 설명, 논증, 지시, 상호작용)과 관련해서는 국어교육(T13), 이야기 구성(T15), 문학작품(T16)이 매칭되었으며, 이는 다양한 텍스트 장르에 대한 이해 및 분석 능력과 관련된다.

상황 차원에서는 개인적, 공적, 교육적, 직업적 맥락과 관련된 토픽으로 독서문화(T3), 학업성취 요인(T8), 도서관 서비스(T18), 교육역량(T19), 다문화 교육(T20)이 매칭되었다. 이는 읽기 활동이 수행되는 맥락과 목적, 대상 등을 설명하거나 그에 영향을 미치는 요소와 밀접하게 관련된다.

전체 매칭 비중을 살펴보면, 과정 차원이 56.21%로 가장 높았고, 상황 차원이 44.27%, 텍스트 차원이 13.09%로 상대적으로 낮았다. 세부적으로, 과정 차원의 '정보 찾기'는 도서관 서비스(T18, 13.57%)와 연결되었으며, '이해하기'는 주제분석(T4, 2.02%), 디지털 문해력(T12, 3.46%), 읽기와 문해력(T17, 6.83%)과 매칭되었다. '평가와 성찰하기'는 비판적 사고(T2,

2.21%), 학습 결과(T5, 5.68%), 읽기능력평가(T7, 2.02%)와 연결되었으며, '유창하게 읽기' 범주는 읽기활동(T6, 2.60%), 책 읽기(T9, 3.95%), 읽기와 글쓰기(T14, 6.74%)가 매칭되었다. 과제 관리 영역의 '목표 설정, 계획 수립, 점검 및 조정'은 교육과정(T1, 2.41%)과 교육프로그램 개발(T10, 4.72%)과 관련되었다.

텍스트 차원에서는 총 13.09%의 비중을 차지하며, 텍스트 체제(연속·비연속·혼합)와 문자와 언어(T11, 2.31%)가 매칭되었고, 텍스트 유형은 국어교육(T13, 5.10%), 이야기 구성(T15, 1.54%), 문학작품(T16, 4.14%)과 관

련되었다.

상황 차원에서는 총 44.27%의 비중으로, 독서문화(T3, 2.50%), 학업성취 요인(T8, 5.97%), 도서관 서비스(T18, 13.57%), 교육역량(T19, 1.92%), 다문화 교육(T20, 20.31%)이 매칭되었다.

종합하면, 국내 독서교육 관련 연구는 PISA 평가들의 세 차원 중 '과정'과 '상황'에 상대적으로 높은 비중을 두고 있는 반면, '텍스트' 차원은 상대적으로 낮은 비중을 보이는 것으로 나타났다. PISA 평가들과 독서교육 관련 연구 토픽의 매칭을 정리하면 <표 4>와 같다.

<표 4> PISA 2022 평가들과 토픽 비교

PISA 2022 평가들			독서교육 연구 토픽 매칭	합계
차원	하위범주	종류		
과정	텍스트 과정	정보찾기	- 도서관 서비스 (T18, 13.57%)	56.21%
		이해하기	- 주제분석 (T4, 2.02%) - 디지털 문해력 (T12, 3.46%) - 읽기와 문해력 (T17, 6.83%)	
		평가와 성찰하기	- 비판적 사고 (T2, 2.21%) - 학습결과 (T5, 5.68%) - 읽기능력평가 (T7, 2.02%)	
		유창하게 읽기	- 읽기활동 (T6, 2.60%) - 책 읽기 (T9, 3.95%) - 읽기와 글쓰기 (T14, 6.74%)	
	과제 관리	목표설정과 계획수립 점검·조정	- 교육과정 (T1, 2.41%) - 교육프로그램 개발 (T10, 4.72%)	
텍스트	출처	단일, 다중	-	13.09%
	조직과 탐색구조	정적, 동적	-	
	텍스트 체제	연속, 비연속, 혼합	- 문자와 언어 (T11, 2.31%)	
	텍스트 유형	기술, 서사, 설명, 논증, 지시, 상호작용	- 국어교육 (T13, 5.10%) - 이야기 구성 (T15, 1.54%) - 문학작품 (T16, 4.14%)	
상황	개인적, 공적, 교육적, 직업적		- 독서문화 (T3, 2.50%) - 학업성취 요인 (T8, 5.97%) - 도서관 서비스 (T18, 13.57%) - 교육역량 (T19, 1.92%) - 다문화 교육 (T20, 20.31%)	44.27%

4.3 논의

본 연구는 PISA 평가들과 국내 독서교육 관련 연구의 흐름을 비교·분석하고, 두 영역의 주요 키워드 간 의미적 유사성을 기반으로 매칭함으로써 국내 연구의 경향성과 국제 평가 기준과의 연계 가능성을 탐색하였다. 분석 결과, 국내 연구는 '과정'(56.21%)과 '상황'(44.27%) 차원과 관련한 연구가 집중되고 있는 반면, '텍스트' 차원(13.09%)은 낮은 비중을 보였다. 특히 출처 유형(단일·다중 텍스트)과 조직·탐색 구조(정적·동적 텍스트)를 다룬 연구는 거의 이루어지지 않았다. 반면 디지털 문해력, 복합 텍스트 이해, 다중 출처 정보의 비교·통합, 동적 텍스트 구조의 이해와 같은 현대 문해력의 핵심 요소에 대한 연구는 상대적으로 부족하였다. 이는 디지털 기반 환경에서의 읽기 수행 역량을 중시하는 PISA 평가들의 방향성이 국내 독서교육 연구와 교육 현장에 충분히 반영되지 못하고 있음을 보여준다.

이러한 결과는 국내 독서교육이 전통적인 인쇄 매체 기반 텍스트와 학습 중심의 읽기 활동에 주로 초점을 두고 발전해 왔음을 시사하며, 이는 다음과 같은 거시적 요인으로도 연결된다. 첫째, 국가 교육과정 변화의 속도가 상대적으로 느리다는 점이다. 이는 독서교육이 여전히 전통적 인쇄 텍스트와 학습 중심의 읽기 활동에 머무르는 구조적 배경으로 작용하였다. 한국교육학술정보원(2023)은 2022 개정 교육과정에서 디지털 리터러시 교육 내용을 교과와 연계하려는 시도가 있었음에도, 실제 교수·학습 자료 보급과 현장 지원은 여전히 미비함을 지적하였다. 또한 교사 대상 조사에서는 2015

개정 교육과정의 디지털 리터러시 관련 변화를 인식하고 있음에도 불구하고, 실제 수업 적용 수준은 낮았으며, 교육과정 개정이 곧바로 현장 실천으로 연결되지 못하는 '실행 격차'가 확인되었다(노은희 외, 2019, 57).

둘째, 교사들의 디지털 기반 읽기 활동 인식이 '도구 활용 중심'에 머무르는 경향이 강하다. 정송과 노연경(2021)은 교실 내 ICT 활용이 주로 교사 주도의 단방향적 활동에 머무르고 있으며, 이로 인해 디지털 텍스트의 구조나 형식적 특성을 탐구하는 활동으로 발전하지 못하는 한계가 나타났다고 보고하였다. 또한 AI 디지털 교과서 관련 연구에서도 교사들은 기능적 측면(맞춤형 지원, 운영 편의성 등)에 주로 초점을 두어, 텍스트 장르나 다중양식적 특성을 심층적으로 탐구하는 논의는 상대적으로 미흡하였다(이승현, 이영호, 2024, 384).

셋째, 학술지 차원의 연구비 지원과 정책적 관심 편중 또한 연구 경향에 영향을 미쳤다. 강미희와 김정현(2014)에 따르면, 연구비 지원 논문 비율은 증가하였지만 대부분 교내 지원이었고, 정부 지원은 제한적이었다. 특히 연구 지원은 서비스·실천·정책 관련 주제에 집중되는 경향이 있으며, 텍스트 구조나 다중 출처 분석과 같은 기초적·형식적 연구는 상대적으로 소외되는 것으로 나타났다. 이러한 연구지원 및 정책 우선순위 구조는 독서교육 연구 방향을 제약하는 요인으로 작용했을 가능성이 있다.

5. 결론 및 제언

본 연구는 국제 학업성취도 평가(PISA) 읽

기 영역의 평가들을 중심으로, 국내에서 수행된 독서교육 연구의 경향성을 정량적으로 분석하고, 양자의 연계 가능성을 체계적으로 탐색함으로써 향후 독서교육 연구의 발전 방향을 제시하고자 수행되었다. 지금까지의 국내 독서교육 관련 연구들은 주로 정성적 분석이나 개별 사례 중심의 연구에 치중되어 있어, 독서교육 전반의 연구 지형을 종합적으로 조망하는데 한계가 있었다. 이러한 점을 보완하기 위해, 본 연구에서는 2001년부터 2024년까지 발표된 국내 학술지 논문 총 1,204편을 수집하여 LDA (Latent Dirichlet Allocation) 기반의 토픽모델링 기법을 적용한 정량적 분석을 실시하였다.

분석을 통해 도출된 20개의 주요 토픽은 PISA 평가들의 세 차원인 '과정', '상황', '텍스트' 차원에 따라 비교·분류되었으며, 이를 통해 국내 독서교육 연구가 어떤 문해력 요소에 초점을 두고 있는지를 객관적으로 파악하고자 하였다. 그 결과, '과정' 차원에 해당하는 텍스트 정보 처리 및 과제 관리와 관련된 토픽이 전체의 56.21%로 가장 높은 비중을 차지하였으며, '상황' 차원의 개인적·교육적·공적 맥락과 관련된 토픽은 44.27%를 차지하였다. 이러한 매칭 결과는 국내 독서교육 연구가 이 두 차원에 상대적으로 높은 관심을 기울여 왔음을 확인할 수 있었다.

반면, '텍스트' 차원, 특히 다중 출처 텍스트의 활용, 정보 간 비교 및 통합, 비선형 구조와 하이퍼링크 등의 디지털 탐색 방식과 같은 현대적 읽기 환경에서 요구되는 복합 문해력 요소와 관련된 토픽은 거의 나타나지 않았다. 이는 현재 국내 독서교육 연구가 여전히 전통적인 인쇄 기반 텍스트와 학습 중심의 문해력에

머무르고 있으며, 디지털 전환과 정보 다양성이 강조되는 국제 문해력 기준을 충분히 반영하지 못하고 있음을 시사한다. 특히, PISA가 강조하는 복합 텍스트 이해 역량, 다중 정보 출처 간 비판적 통합 능력 등은 향후 국내 독서교육 연구에서 보다 적극적으로 다루어져야 할 중요한 과제로 부상하고 있다. 연구 결과에서 주목할 만한 점은 '다문화 교육'(20.31%)과 '도서관 서비스'(13.57%)가 상대적으로 높은 비중을 차지하고 있는 현상이다. 이는 국내 독서교육 연구가 단순히 학습 성취 중심의 협의적 읽기 교육을 넘어, 사회적 맥락과 평생학습적 관점을 반영한 '보편적 문해력 교육'으로 점차 진화하고 있음을 보여준다. 이러한 흐름의 배경에는 다문화 사회로의 전환과 같은 사회적 요구, 독서진흥정책 및 도서관법 개정 등 제도적·정책적 지원이 자리하고 있다.

특히 다문화 교육에 관한 연구는 독서교육이 단순한 읽기 기술의 습득을 넘어, 사회적 포용과 다양성 존중이라는 가치를 실현하는 중요한 수단으로 기능함을 보여준다. 이는 PISA 평가들의 '상황' 차원, 즉 학습자가 속한 사회·문화적 맥락 속에서 읽기 역량을 발휘하는 능력과 직접적으로 맞닿아 있다. 한편, 도서관 서비스 연구는 학교 현장을 넘어 지역사회를 포괄하는 문해력 기반 인프라 구축의 필요성을 강조하며, 평생학습과 자율적 독서 문화를 지원하는 핵심 기반으로 자리매김하고 있다. 따라서 다문화 교육과 도서관 서비스의 비중 확대는 국내 독서교육이 개인적 차원을 넘어 사회적·정책적 차원으로 확장되고 있음을 시사하며, 이는 PISA가 지향하는 '보편적 문해력 교육'과도 맥락적으로 상통한다. 다만 이러한 연구 경향이 실제 수업

현장이나 디지털 기반 읽기 역량 연구와 어떻게 접목될 수 있는지는 여전히 후속 연구가 요구된다.

이러한 맥락에서 학교도서관은 단순한 자료 제공 기관을 넘어 독서교육의 실천적 공간으로 기능하며, 학습자의 문해력 신장에 실질적으로 기여하고 있다. 따라서 향후 독서교육 정책 수립과 프로그램 개발 과정에서는 학교도서관과 공공도서관 간의 연계 강화, 나아가 지역사회 인프라의 적극적 활용을 통한 도서관 중심 독서교육 생태계 구축 방안을 모색할 필요가 있다.

본 연구의 분석 결과를 토대로 다음과 같은 향후 연구 방향을 제안할 수 있다. 첫째, PISA 평가들의 '텍스트' 차원에서 상대적으로 미흡한 출처 유형(단일·다중 텍스트)과 조직·탐색 구조(정적·동적 텍스트)에 관한 연구를 강화해야 한다. 특히 디지털 및 복합 텍스트 환경에서 요구되는 다중 출처 정보의 비교·통합 능력과 동적 텍스트 구조의 이해를 심층적으로 탐구함으로써, 변화하는 읽기 환경에 부합하는 문해력 이론과 교육 방안을 개발할 필요가 있다. 둘째, 디지털 문해력 및 복합 텍스트 이해 능력과 관련된 교수·학습 전략 연구를 활성화해야 한다. 현재 국내 연구에서는 이 분야가 상대적으로 미흡하므로, 멀티모달 자료 해석, 온라인 정보 탐색, 정보 신뢰성 평가 등 디지털 리터러시 역량을 포함하는 통합적 문해력 교육 프로그램 개발과 실증적 효과 분석이 요구된다. 셋째, '상황' 차원에서 나타난 다문화 교육 및 사회적 소수자 관련 연구와 더불어, 직업적 맥락에서의 실용적 문해력 함양에 대한 연구가 확대되어야 한다. 이를 위해 직무 수행 능력 중심의 읽기 교육, 생애주기별 맞춤형 문해력 지

원, 직업계 고등학생과 성인 학습자를 포함한 다양한 학습자 집단에 대한 문해력 향상 방안 연구가 필요하다. 넷째, 독서문화, 읽기활동, 교육과정 등 개별 연구 토픽 간의 연계와 통합적 연구가 더욱 강화되어야 한다. 단편적인 활동 수준에 머무르지 않고, 복합적이고 체계적인 학습 시스템으로 독서교육을 발전시킴으로써, 읽기·쓰기·비판적 사고 등 다양한 언어 및 인지 역량의 통합적 발달을 도모할 수 있어야 한다. 마지막으로, 이러한 다양한 연구 방향을 바탕으로 현장 적용 가능하고 실효성 있는 교육 정책 및 프로그램 개발이 이루어질 수 있도록, 학계와 교육 현장, 정책 담당자 간의 긴밀한 협력 체계 구축이 요구된다. 이를 통해 국내 독서교육이 국제적 평가 기준과 조화되면서도 지역적 특성과 현실을 반영한 균형 있는 발전을 이룰 수 있을 것이다.

본 연구에서는 국내 독서교육 연구가 '과정' 및 '상황' 차원에 상대적으로 집중되어 있으며, 디지털 문해력, 다중 출처 정보 활용, 복합 텍스트 이해와 같은 '텍스트' 차원에서의 현대적 문해력 요소에 대한 연구가 미흡하다는 점을 확인하였다. 이러한 결과를 바탕으로 향후 연구는 다음과 같은 방향으로 확장될 필요가 있다. 첫째, 교육 현장의 실제 경험과 국제 평가들을 통합적으로 반영할 수 있는 혼합 방법론의 적극적인 도입이 요구된다. 둘째, 디지털 기반 텍스트와 복합 정보 구조를 반영한 교수·학습 전략과 평가 도구의 개발이 필요하다. 셋째, 학교도서관과 지역 도서관 간의 실질적 연계를 강화하여 지속 가능하고 실천 가능한 독서교육 체계를 구축해야 한다. 본 연구 결과는 이론적 측면에서는 국내 독서교육 연구의 구조와 방향

성을 총체적으로 파악하고, 이를 PISA 평가들과 비교함으로써 국제 기준에 비추어 국내 연구의 현주소를 진단할 수 있는 기초자료를 제공할 수 있다. 실천적 측면에서는 디지털 문해력 함양을 위한 교육 정책과 교수학습 방안 마련에 구체적 근거를 제시함으로써, 학교 현장과 교육 행정에 실질적인 기여를 할 수 있을 것으로 기대된다.

다만, 본 연구의 한계점은 다음과 같다. 첫째, PISA 평가들은 국제 비교를 목적으로 설계된 개념들이므로, 국내 교육 현장의 구체적 교수·학습 활동이나 문화적 특수성을 충분히 반영하지 못할 수 있다. 둘째, LDA 기반의 토픽모델링 분석은 문헌의 키워드 및 본문 내용에 기반하여 자동 추출된 주제를 바탕으로 이루어지기 때문에, 해석 과정에서 연구자의 주관이 개입될 수 있으며, 토픽과 PISA 차원 간의 일치 여부에 따라 해석의 정합성에 일정 수준의 한계

가 따를 수 있다. 셋째, 분석 대상이 국내 학술지 논문에 국한되어 있어, 실제 교육 현장의 실천 사례나 교사·학생의 인식과 같은 다양한 관점을 포괄하기 어려웠다는 점도 고려되어야 한다.

그럼에도 불구하고 본 연구는 국내 독서교육 연구의 전반적인 흐름을 계량적으로 진단하고, 국제 문해력 평가들과의 비교를 통해 향후 연구 및 정책 개발의 방향성을 제시하였다는 점에서 이론적·실천적 의의가 크다. 특히, 21세기 디지털 기반 사회에서 요구되는 새로운 문해력 역량을 반영한 독서교육 체계 구축을 위해 연구자, 교육자, 정책 입안자가 함께 협력할 필요성이 더욱 강조된다. 아울러, 향후 국제 데이터베이스 기반의 해외 학술지 논문을 대상으로 연구 동향을 분석하는 후속 연구가 이어질 것 기대한다.

참 고 문 헌

- 강미희, 김정현 (2014). “한국도서관·정보학회지”의 연구동향 분석. 한국도서관·정보학회지, 45(4), 457-482. <https://doi.org/10.16981/kliss.45.4.201412.457>
- 강봉숙, 이승태 (2024). 디지털 기반 독서 교육을 위한 학교도서관 지원 현황 분석: 교육부·교육청의 플랫폼 지원을 중심으로. 한국문헌정보학회지, 58(4), 395-413.
- 교육부 (2022). 2022 개정 교육과정 총론 고시(교육부 고시 제2022-194호). 교육부.
출처: <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N>
- 국가교육과정정보센터 (2025.5.11.). 교육과정 변천사.
출처: <https://ncic.re.kr/dwn/ogf/inventory.cs>
- 김성경, 김명화, 김인숙, 이신영, 백혜선 (2023). OECD 국제 학업성취도 평가연구: PISA 2022 결과

- 보고서 (연구보고 RRE 2023-10). 한국교육과정평가원.
- 김승주 (2019). 토픽모델링과 시계열 분석을 활용한 독서 연구의 동향과 전망: 최근 10년 간 국어교육학 관련 학술지 논문을 중심으로. *청람어문교육*, 70, 169-204.
<http://dx.doi.org/10.26589/jockle.70.201906.169>
- 김은선 (2025). 토픽 모델링을 활용한 국내 기초 문식성 연구 동향 분석. *독서연구*, 75, 135-173.
<http://dx.doi.org/10.17095/JRR.2025.75.4>
- 김지영, 노동조 (2023). 토픽모델링을 활용한 4차 산업혁명 분야의 국내 연구 동향 분석. *한국비블리아학회지*, 34(4), 207-234. <https://doi.org/10.14699/kbiblia.2023.34.4.207>
- 김지혜, 이창희 (2024). 이주배경학생의 읽기 소양 분석: PISA 2022 데이터의 '학생배경변인'을 중심으로. *국어교육연구*, 54, 31-66. <http://dx.doi.org/10.17313/jkorle.2024.54.31>
- 김태경, 최희련, 이홍철 (2016). 토픽 모델링을 이용한 핀테크 기술 동향 분석. *한국산학기술학회논문지*, 17(11), 670-681. <https://doi.org/10.5762/KAIS.2016.17.11.670>
- 김희섭, 강보라 (2019). 토픽 모델링을 활용한 도서관, 기록관, 박물관간의 연구 주제 분석. *한국도서관·정보학회지*, 50(4), 339-358.
- 노명완, 박창균 (2006). 읽기 평가들에 대한 비교 연구. *교육과정평가연구*, 9(1), 159-184.
- 노은희, 신호재, 이재진 (2019). 초·중학교 교사의 디지털 리터러시 교육에 대한 인식 분석. *교육과정평가연구*, 22(3), 31-60. <https://doi.org/10.22799/jce.2019.22.3.002>
- 문화체육관광부 (2023). 2023년 국민 독서실태 조사 (11-1371000-000154-10). 문화체육관광부.
- 박민정, 성열관 (2011). 교육과정 연구의 최근 동향 분석: 2000년 이후를 중심으로. *교육과정연구*, 29(4), 25-46. <http://dx.doi.org/10.15708/kscs.29.4.201112.002>
- 박자현, 송민 (2013). 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. *정보관리학회지*, 30(1), 7-32. <http://dx.doi.org/10.3743/KOSIM.2013.30.1.007>
- 박종도 (2019). 토픽모델링을 활용한 다문화 연구의 이슈 추적 연구. *한국문헌정보학회지*, 53(3), 273-289. <http://dx.doi.org/10.4275/KSLIS.2019.53.3.273>
- 박주현, 강봉숙 (2023). 디지털 협력수업이 독서 리터러시에 미치는 효과 연구. *한국문헌정보학회지*, 57(3), 211-230. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.211>
- 박혜영 (2017). PISA 읽기 평가들의 변화 특징 탐색. *국어교육*, 157, 259-283.
- 육현진 (2014). 성인문식성 교육과정 개발 방향 탐색: 작문 영역을 중심으로. *국어교육*, 146, 225-250.
- 이승현, 이영호 (2024). AI 디지털교과서 도입에 대한 초등 교사 인식 연구. *정보교육학회논문지*, 28(4), 377-386.
- 이영호 (2024). PISA 2022 기반 디지털 관련 변인이 디지털 읽기 역량에 미치는 영향 분석. *지능정보융합과 미래교육*, 3(3), 31-41. <http://dx.doi.org/10.59482/iicfe.2024.3.3.04>
- 이인화 (2020). PISA 2018 결과에 나타난 우리나라 학생들의 읽기 성취 특성 연구. *우리말글*, 86,

123-156.

이인화, 구남욱, 전지혜 (2021). PISA 2018에 나타난 한국 학생들의 디지털 기기 활용 경향과 읽기 성취의 특성: PISA 2009와의 비교를 중심으로. *독서연구*, 61, 91-135.

<http://dx.doi.org/10.17095/JRR.2021.61.3>

이지영 (2021). 디지털 글쓰기 관련 국내 연구 동향 분석: 언어 네트워크 분석과 토픽 모델링 분석을 중심으로. *문화와 융합*, 43(1), 603-624.

임정훈 (2022). 토픽모델링을 활용한 물리학 독서감상문 텍스트의 교육과정 연계성 분석. *한국도서관·정보학회지*, 53(2), 333-353. <http://dx.doi.org/10.16981/kliss.53.2.202206.333>

정승, 노연경 (2021). ICILS 2018을 통해 살펴본 우리나라 중등교사의 ICT 활용 및 인식 특성: 연령대와 과목을 중심으로. *교육문제연구*, 34(1), 25-50.

정영주, 김혜진 (2020). 토픽모델링을 활용한 학교도서관 연구동향 분석. *한국도서관·정보학회지*, 51(3), 103-121. <https://doi.org/10.16981/kliss.51.3.202009.103>

정혜경, 조지민 (2013). PISA 2009 읽기와 사회경제적 배경 변인에 기반한 역경 극복 학생 특성 분석. *교육과정평가연구*, 16(1), 161-184.

정혜원, 박소영, 김정인, 김아름 (2021). 청소년의 읽기 소양과 삶의 만족도의 영향변인 탐색: PISA 2018 한국·핀란드 국제비교. *교육과정평가연구*, 24(1), 123-152.

조성민, 구남욱, 김현정, 이소연, 이인화 (2019). OECD 국제 학업성취도 평가연구: PISA 2018 결과 보고서 (연구보고 RRE 2019-11). 한국교육과정평가원.

진성희, 김혜경 (2022). 텍스트 마이닝을 적용한 교육연구 동향 분석. *교육학연구*, 60(4), 267-297.

최숙기, 김세화 (2022). 토픽 모델링 분석을 활용한 독서치료에 관한 연구 동향 분석. *독서치료연구*, 14(1), 131-153. <http://dx.doi.org/10.35398/job.2022.14.1.131>

최인선, 김명화, 김수진, 김현정, 이신영 (2021). OECD 국제 학업성취도 평가 연구: PISA 2018 상위국 성취특성 및 교육맥락변인과의 관계 분석 (연구보고 RRE 2021-4). 한국교육과정평가원.

통계청 (2025). 기초학력미달률. 국가통계포털 지표누리.

출처: https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?cdNo=2&clasCd=10&idxCd=F0096&utm_source=chatgpt.com

한국교육학술정보원 (2023). 교육과정 연계 디지털 리터러시 교육 가이드라인 개발 연구 (CR 2023-1). 한국교육학술정보원.

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

Chang, J., Boyd-Graber, J. L., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: how humans interpret topic models. In in Proc. of NIPS '09. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada, 288-296.

- OECD (2000). Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy. <https://doi.org/10.1787/9789264181564-en>
- OECD (2010). PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science. <https://doi.org/10.1787/9789264062658-en>
- OECD (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy. <https://doi.org/10.1787/9789264281820-en>
- OECD (2018). PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science. <https://doi.org/10.1787/9789264305274-en>
- OECD (2019). PISA 2018 Assessment and Analytical Framework. <https://doi.org/10.1787/b25efab8-en>.
- Roeder, M., Both, A., & Hinneburg, A. (2024). Exploring the Space of Topic Coherence Measures. Available: <https://radimrehurek.com/gensim/models/coherencemodel.html>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Cho, Sung-Min, Koo, Nam-Wook, Kim, Hyun-Jung, Lee, So-Yeon, & Lee, In-Hwa (2019). OECD International Student Assessment Study: PISA 2018 Results Report (Research Report RRE 2019-11). Korea Institute for Curriculum and Evaluation.
- Choi, In-Seon, Kim, Myung-Hwa, Kim, Su-Jin, Kim, Hyun-Jung, & Lee, Shin-Young (2021). OECD International Student Assessment Study: Analysis of High-Performing Countries' Achievement Characteristics and Educational Context Variables from PISA 2018 (Research Report RRE 2021-4). Korea Institute for Curriculum and Evaluation.
- Choi, Suk-Gi & Kim, Se-Hwa (2022). An analysis of research trends on bibliotherapy using topic modeling. *Journal of Bibliotherapy Research*, 14(1), 131-153. <http://dx.doi.org/10.35398/job.2022.14.1.131>
- Jeong, Song & Noh, Eun-Gyeong (2021). A study on the ICT usage and perceptions of Korean secondary school teachers through ICILS 2018: focused on age group and subject. *Journal of Educational Research*, 34(1), 25-50.
- Jin, Seong-Hee & Kim, Hyek-Yung (2022). Analysis of educational research trends using text mining. *The Journal of Educational Research*, 60(4), 267-297.
- Jung, Hye-Kyung & Cho, Ji-Min (2013). Analysis of characteristics of resilience students based on PISA 2009 reading and socioeconomic background variables. *Journal of Curriculum*

- and Evaluation, 16(1), 161-184.
- Jung, Hye-Won, Park, So-Young, Kim, Jeong-In, & Kim, A-Reum (2021). Exploring factors influencing adolescents' reading literacy and life satisfaction: a Korea-Finland international comparison based on PISA 2018. *Journal of Curriculum and Evaluation*, 24(1), 123-152.
- Jung, Young-Joo & Kim, Hye-Jin (2020). An analysis of research trends on school libraries using topic modeling. *Journal of the Korean Library and Information Science Society*, 51(3), 103-121. <https://doi.org/10.16981/kliss.51.3.202009.103>
- Kang, Bong-Suk & Lee, Seung-Tae (2024). An analysis of school library support for digital-based reading education: focusing on platform support from the Ministry of Education and local education offices. *Journal of the Korean Society for Library and Information Science*, 58(4), 395-413.
- Kang, Mi-Hee & Kim, Jung-Hyun (2014). An analysis of research trends in the Journal of the Korean Society for Library and Information Science. *Journal of the Korean Society for Library and Information Science*, 45(4), 457-482. <https://doi.org/10.16981/kliss.45.4.201412.457>
- Kim, Eun-Seon (2025). An analysis of research trends in basic literacy in Korea using topic modeling. *Reading Research*, 75, 135-173. <http://dx.doi.org/10.17095/JRR.2025.75.4>
- Kim, Gi-Young & Noh, Dong-Jo (2023). An analysis of research trends in the fourth industrial revolution in Korea using topic modeling. *Journal of the Korean Biblia Society for Library and Information Science*, 34(4), 207-234. <https://doi.org/10.14699/kbiblia.2023.34.4.207>
- Kim, Hee-Sop & Kang, Bo-Ra (2019). An analysis of research topics among libraries, archives, and museums using topic modeling. *Journal of the Korean Library and Information Science Society*, 50(4), 339-358.
- Kim, Ji-Hye & Lee, Chang-Hee (2024). An analysis of reading literacy among students with immigrant backgrounds: focusing on "student background variables" in PISA 2022 data. *Journal of Korean Language Education Research*, 54, 31-66. <http://dx.doi.org/10.17313/jkorle.2024..54.31>
- Kim, Seung-Joo (2019). Trends and prospects of reading research using topic modeling and time-series analysis: focusing on journal articles in Korean language education over the past decade. *Cheongram Language and Literature Education*, 70, 169-204. <http://dx.doi.org/10.26589/jockle.70.201906.169>
- Kim, Sungk-Yung, Kim, Myung-Hwa, Kim, Mi-Sook, Lee, Shin-Young, & Baek, Hae-Seon (2023). OECD international student assessment study: PISA 2022 results report (Research

- Report RRE 2023-10). Korea Institute for Curriculum and Evaluation.
- Kim, Tae-Kyung, Choi, Hoe-Ryeon, & Lee, Hong-Chul (2016). Analysis of fintech technology trends using topic modeling. *Journal of the Korea Academia-Industrial Cooperation Society*, 17(11), 670-681. <https://doi.org/10.5762/KAIS.2016.17.11.670>
- Korea Education & Research Information Service (2023). Development of Guidelines for Curriculum-Linked Digital Literacy Education (CR 2023-1). Korea Education & Research Information Service.
- Lee, In-Hwa (2020). A study on reading achievement characteristics of Korean students in the PISA 2018 results. *Woorimalgeul*, 86, 123-156.
- Lee, In-Hwa, Koo, Nam-Wook, & Jeon, Ji-Hae (2021). Trends in digital device usage and characteristics of reading achievement among Korean students in PISA 2018: focusing on comparison with PISA 2009. *Reading Research*, 61, 91-135. <http://dx.doi.org/10.17095/JRR.2021.61.3>
- Lee, Ji-Young (2021). Analysis of domestic research trends on digital writing: focusing on language network analysis and topic modeling. *Culture and Convergence*, 43(1), 603-624.
- Lee, Seung-Hyun & Lee, Young-Ho (2024). A study on elementary school teachers' perceptions of the introduction of AI digital textbooks. *Journal of the Korean Society for Information Education*, 28(4), 377-386.
- Lee, Young-Ho (2024). An analysis of the effects of digital-related variables based on PISA 2022 on digital reading competency. *Intelligent Information Convergence and Future Education*, 3(3), 31-41. <http://dx.doi.org/10.59482/iicfe.2024.3.3.04>
- Lim, Jeong-Hoon (2022). An analysis of curriculum relevance in physics book report texts using topic modeling. *Journal of the Korean Society for Library and Information Science*, 53(2), 333-353. <http://dx.doi.org/10.16981/kliss.53.2.202206.333>
- Ministry of Culture, Sports and Tourism (2023). 2023 Survey on the reading habits of Koreans (11-1371000-000154-10). Ministry of Culture, Sports and Tourism.
- Ministry of Education (2022). 2022 Revised National Curriculum: General Guidelines (Ministry of Education Notification No. 2022-194). Ministry of Education. Available: <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=141&boardSeq=93458&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=040401&opType=N>
- National Curriculum Information Center (2025, May 11). Curriculum revision history. Available: <https://ncic.re.kr/dwn/ogf/inventory.cs>
- Noh, Eun-Hee, Shin, Ho-Jae, & Lee, Jae-Jin (2019). An analysis of primary and secondary

- school teachers' perceptions of digital literacy education. *Journal of Curriculum and Evaluation*, 22(3), 31-60. <https://doi.org/10.22799/jce.2019.22.3.002>
- Noh, Myung-Wan & Park, Chang-Kyun (2006). A comparative study on reading assessment frameworks. *Journal of Curriculum and Evaluation*, 9(1), 159-184.
- Ok, Hyun-Jin (2014). Exploring directions for developing adult literacy curriculum: focusing on writing domain. *Korean Language Education*, 146, 225-250.
- Park, Hye-young (2017). Exploring characteristics of changes in the PISA reading assessment framework. *Korean Language Education*, 157, 259-283.
- Park, Ja-Hyun & Song, Min (2013). A study on the research trends in library & information science in Korea using topic modeling. *Journal of Korean Society for Information Management*, 30(1), 7-32. <http://dx.doi.org/10.3743/KOSIM.2013.30.1.007>
- Park, Jong Do (2019). A study on issue tracking on multi-cultural studies using topic modeling. *Journal of the Korean Society for Library and Information Science*, 53(3), 273-289. <http://dx.doi.org/10.4275/KSLIS.2019.53.3.273>
- Park, Ju-Hyeon & Kang, Bong-Suk (2023). A study on the effects of digital collaborative classes on reading literacy. *Journal of the Korean Society for Library and Information Science*, 57(3), 211-230. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.211>
- Park, Min-jung & Sung, Yeol-Gwan (2011). An analysis of recent trends in curriculum research: focusing on studies since 2000. *Journal of Curriculum Studies*, 29(4), 25-46. <http://dx.doi.org/10.15708/kscs.29.4.201112.002>
- Statistics Korea (2025). Underachievement Rate in Basic Academic Skills. KOSIS (Korean Statistical Information Service). Available: <https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?cdNo=2&clasCd=10&idxCd=F0096>

[부록] 데이터 전처리 단어 리스트

1. 불용어 사전 리스트 (가, 나, 다 순)

각, 고찰, 구분, 국내, 국외, 기본, 기준, 길, 끝, 날, 노동, 다각, 동안, 동일, 때, 모두, 모색, 목적, 사용, 선, 속물, 쇼, 순차, 시사점, 시작, 약, 안, 암시, 압도, 앞, 약점, 양, 연간, 연구, 연구물, 연구원, 연구자, 예방, 예시, 예전, 오류, 외부, 원, 의, 이때, 이면, 이미지, 이행, 이후, 일제, 일환, 임의, 저조, 적, 적극, 적기, 적성, 적시, 전, 점점, 제출, 조각, 조사, 조언, 존치, 종강, 종료, 주, 주가, 주관식, 주력, 지배, 지속, 지시문, 지연, 지인, 지침서, 지향, 직, 총, 최근, 출발점, 트랙, 트렌드, 파동, 파악, 파일, 판, 평정, 학술지, 학회, 한편, 할당, 해석, 해설, 해외, 후반부, 훼손 등 총 1,252개의 단어 불용어 처리

2. 복합명사 리스트 (가, 나, 다 순)

개정교육과정, 고전문학, 교과연계, 교육과정, 교육정책, 국어교육, 다문화가정, 독서감상문, 독서교육, 독서치료, 동화책, 디지털교육, 디지털독서, 문해능력, 성취기준, 성취평가, 언어교육, 웹사이트, 자기주도학습, 정보활용, 주제전문사서, 중등교육, 하위범주, 학교도서관, 학습과정, 학습목표, 학습역량, 협동수업, 협력수업 등 총 55개

3. 동의어 및 대표 단어 처리 리스트 (가, 나, 다 순)

- **논문**: 학술논문, 학술지, 학위논문, 학회지
- **교육**: 공교육, 사교육, 통합교육, 평생교육
- **교사**: 교원, 교육자, 교수, 교수자, 선생님
- **도서관**: 대학도서관, 라이브러리, 마을문고, 전문도서관, 학교도서관, library
- **다문화**: 결혼이민자, 다문화가정, 다문화가족, 다문화부모, 이주가정, 국제결혼
- **대학교**: 고등교육기관, 공대, 대학, 인문대학, 전문대학, 전문대학교, 캠퍼스
- **문자**: 글, 글감, 글자, 단어, 문장, 본문, 설명문, 소감문, 활자
- **문해력**: 독해능력, 리터러시, 문식력, 문해, 읽기능력, 읽기소양, literacy
- **사서**: 주제사서, 주제전문사서, 전문사서, 참고사서
- **스마트폰**: 핸드폰, 전화기, 통신기기
- **장서**: 고전, 교양서적, 그림책, 단행본, 도서, 동화책, 베스트셀러, 전자책, 책
- **학교**: 고등학교, 교육기관, 배움터, 초등학교, 초·중등, 중학교

- **학생**: 초등학생, 중학생, 청소년, 대학생
- **DLS**: 독서교육종합지원시스템, 디지털 자료실, 디지털 자료실 지원센터, 디지털자료실, 디지털 자료실지원센터, 독서지원시스템, 학교독서교육시스템
- **PISA**: Programme for International Student Assessment, 국제학업성취도평가, 학업성취도 평가, OECD 성취도평가, OECD학업성취도평가
- **WEB**: 웹페이지, 홈페이지, 인터넷 페이지, 인터넷사이트, 블로그

등 총 69개 동의어, 대표단어 처리