

# LDA 토픽모델링을 활용한 고교학점제 뉴스기사 토픽 분석

전은정 (이화여자대학교 교육학과, 강사)\* · 최윤정 (이화여자대학교 교육학과, 부교수)†

## ■ 요약 ■

본 연구는 고교학점제 전면 도입을 앞둔 시점에서 뉴스기사가 다루는 고교학점제 관련 토픽은 무엇인지 분석하여 해당 정책에 대한 시사점을 제공하고자 하였다. 이를 위해 R 프로그램(버전 4.2.2)을 활용하여 교육부의 고교학점제 도입 발표 시기(2017. 11. 27.)부터 2022년까지의 고교학점제 관련 뉴스기사 7,115건을 대상으로 LDA 토픽모델링 분석을 분석하고, 토픽의 연도별 변화를 살펴보았다. 분석 결과 총 16개의 토픽이 도출되었으며, 기존 빅데이터를 활용한 선행연구와 유사한 주제 외에 '교육감 후보의 고교학점제 관련 공약', '외고, 국제고, 자사고 등 일반고 전환 추진'과 '지정 취소 위법 판결', '학령인구 감소 및 중등교원 양성' 등의 토픽이 추가로 도출되었다. 이는 고교학점제 도입 및 정착 과정에 여러 이해관계자들의 다양한 시각을 종합적으로 고려해야 함을 시사한다. 더불어 연도별 토픽 비중 변화를 통해 '고교학점제의 실제 운영 방식 및 현황'에 대한 토픽이 증가 추세 있음과 '대입제도', '고교유형' 등의 토픽이 감소 추세에 있음이 드러났다. 이러한 결과를 통해 고교학점제의 도입 및 정착이 적극적으로 진행되고 있으며 하나의 독립된 제도로 자리잡고 있음을 알 수 있다.

**주제어** : 고교학점제, 학점제, LDA, 토픽모델링, 텍스트마이닝

\* 제1저자: 전은정, 강사, 이화여자대학교, 서울 서대문구 이화여대길 52, euneun1202@ewhain.net

† 교신저자: 최윤정, 부교수, 이화여자대학교, 서울 서대문구 이화여대길 52, younjengchoi@ewha.ac.kr

• 이 논문은 제1저자 전은정의 박사학위논문 일부를 수정·보완한 것임.

• 논문 접수: 2023년 7월 31일 / 수정본 접수: 2023년 9월 3일 / 게재승인: 2023년 9월 5일

## I. 서론

고교학점제는 2025년 전면 도입을 목표로 한다. 고교학점제가 도입되면 학생들은 본인의 적성 및 진로설계에 따라 과목을 선택하여 수강하게 되고, 이수기준 충족 시 부여받은 학점을 누적하여 졸업하게 된다. 고교학점제 도입에는 공간지원, 과목개설, 교사수급, 학업설계지도, 교육과정 운영 등 여러 측면에서 큰 변화가 필요하다. 교육부는 고교학점제 연구·선도학교를 지속적으로 확대하여 2022년을 기준 일반계고의 84%를 지정하였고 2023년에는 모든 일반계고가 연구·준비학교를 운영하도록 하여 고교학점제 전면 도입을 준비할 계획이다. 연구·선도학교의 경우 학생의 선택과목 이수단위 및 선택과목 수가 연차적으로 확대되는 경향을 보이고 있으며, 학생의 수요를 반영한 교육과정 편성에 노력하고 있다. 더불어 88.7%의 연구·선도학교가 온·오프라인 공동교육과정을 운영하고 있으며, 온라인 공동 교육거점센터 역시 22개교로 확대 운영되고 있다. 또한 교육과정 이수지도팀을 구성하여 개인 맞춤형 진로·학업설계 지도가 이루어지고 있으며 다양한 진로·학업설계 지원 프로그램, 최소 성취수준 보장 지도가 이루어지고 있다(한국교육과정평가원, 2023). 그러나 일부 선행연구를 통해 고교학점제 실시에 대한 교사 간, 교사와 학생 간 인식 차이가 있는 것으로 나타났다(신윤범, 원효현, 2020; 신철균, 주주자, 2021; 장창곡, 2021; 최정희, 2019). 또한 학생이 과목을 선택하는 동기 등에 대해서도 서로 다른 연구 결과가 존재한다(서봉언, 2020; 이상은, 장덕호, 2019; 정윤경 외, 2021). 뿐만 아니라 언론을 통해 일부 전문가들의 고교학점제 관련 문제 및 과제 제시(한유경, 2022), 일부 국민 및 특정 지역 교사 과반 이상의 고교학점제 도입 반대(박재연, 2023; 이성기, 2021), 고교학점제 도입에 대한 일부 학부모 단체와 교사 집단의 이견(이종익, 2021), 교원 운용과 관련하여 교원 수급 가능성 및 세부 제도에 대한 부정적 의견(장성환, 2021) 등의 내용을 확인할 수 있다.

위와 같은 내용을 통해 교육현장에 큰 변화를 가져올 고교학점제 도입과 관련하여 여러 사회 집단의 다양한 의견이 혼재되어 있음을 알 수 있다. 따라서 고교학점제와 관련하여 다루어지는 빅데이터로부터 사회적 토픽을 거시적 시각에서 전반적으로 파악함으로써 시사점을 발견하고 고교학점제가 안정적으로 안착할 수 있도록 도울 필요가 있다.

이를 위한 연구대상으로 대표적 빅데이터인 뉴스 기사를 다룰 수 있다. 빅데이터 분석 방법을 통해 고교학점제를 연구한 선행연구 대부분은 포털사이트 게시물 및 댓글 또는 학술 논문을 연구 대상으로 하고 있다. 신문 기사를 연구대상으로 한 선행연구에는 김륜희, 지은림(2022)의 연구와 권충훈(2020)의 연구 두 건이 존재한다. 그러나 김륜희, 지은림(2022)은 고교학점제로 인한 대입제도 이슈를 주제로 다루고 있으며 권충훈(2020)은 연도별 기사 발행 수와 주요 키워드 간 상관관계를 주로 다루고 있어 고교학점제 전반에 대해 언론이 다루는 토픽을 파악하기는 쉽지

않다. 뉴스 기사가 생성하는 담론은 다루는 주제가 가진 목적 및 방향성에 대해 특정한 해석을 담고 있는 경우가 많다. 또한 독자에게 해당 주제를 인식하게 하고 특정 목표에 도달하게끔 실천을 독려하는 측면이 있다(설양환, 2017). 뉴스 기사에서 다루는 고교학점제와 관련된 토픽을 분석하는 것은 협소한 시각에서 벗어나 사회적으로 쟁점이 되는 사안이 무엇인지 폭넓게 파악할 수 있게 한다. 따라서 본 연구에서는 고교학점제 전면 도입을 앞둔 시점에서 뉴스기사에서 다루는 고교학점제 관련 주제는 무엇인지 분석하여 해당 정책에 대한 시사점을 제공하고자 하였다.

본 연구에서는 고교학점제 도입 발표(2017. 11. 27.) 이후 2022년까지 고교학점제를 주제로 보도된 뉴스 기사를 대상으로 LDA 토픽모델링 분석을 실시하여 뉴스기사에서 다루는 고교학점제 관련 토픽이 무엇인지 살펴보고자 하였다. 이에 따른 연구문제는 다음과 같다.

- 연구문제 1. 뉴스 기사를 통해 분석된 고교학점제 토픽은 무엇인가?
2. 뉴스 기사 외 자료가 다룬 고교학점제 토픽과 차이가 있는가?
3. 연도의 흐름에 따라 토픽은 어떤 변화를 보이는가?

## II. 이론적 배경

### 1. 고교학점제

고교학점제가 도입되면 학생들은 본인이 갖춘 학업능력을 바탕으로 하여 스스로 진로 및 적성에 따른 과목을 선택하게 된다. 선택한 과목을 수강하고 이수 기준에 도달하면 해당 학점을 취득하게 되는데, 누적된 취득학점 등 졸업 요건을 충족한 학생들은 졸업하게 된다(교육부, 2021a). 즉 고교학점제는 규격화 및 획일화된 기존의 교육체제 문제점 개선 및 개별 맞춤형 교육체제로의 변화 필요성에 맞추어 도입하게 된 제도이다(교육부, 2020). 이와 같은 고교학점제는 2017년 11월, 교육부의 ‘고교학점제 추진 방향 및 연구학교 운영계획 발표’를 시작으로 2025년 전면 도입을 목표로 하고 있다.

고교학점제의 도입은 4차 산업혁명 시대가 도래함에 따라 단순히 지식이나 기술을 습득하는 능력이 아닌 창의력, 융합적 사고력, 문제해결능력 등이 부각될 것이며 학생 수 감소라는 현재 교육 여건을 미래형 교육 실현의 기회로 활용할 수 있다는 데 배경이 있다(교육부, 2017). 또한 고교 교육은 대입 및 수능에 얽매어 획일적 교육과정, 암기식 수업, 서열화 강조 등의 문제에 직면해 있어 변화가 요구된다. 이러한 상황에서 고교학점제 도입은 입시중심의 교육에서 벗어

나 진로 개척 역량, 자기주도적 학습 능력 등을 기르도록 지원하여 학생 성장이 중심이 되도록 한다. 더불어 입시 및 수능 준비에 초점을 두던 교육에서 벗어나 학생이 개개인에게 필요한 교과목을 스스로 선택할 수 있도록 하는 유연하고 개별화된 교육이 가능하게 한다(교육부, 2017).

고교학점제가 시행되면 수업일수가 기준이 되던 졸업 요건에 학업성취율 등이 적용되게 되며 학생들은 졸업을 위해 총 192학점 이상 취득해야 한다. 과목 선택과 관련하여서는 교과 구분 없이 다양한 과목을 학생들이 스스로 선택할 수 있으며 학교에서는 학생의 수요를 기준으로 과목을 편성하게 된다. 또한 대학 진학 시 내신 성적이나 수능 점수 등 성적에 따라 진학하던 것에서 본인이 설계한 진로에 따라 수강한 과목이 중심이 되게 된다. 공간의 측면에서 학급별 교실이 주된 공간이었던 것에서 학생의 개별 시간표에 따라 교실을 이동하게 되고 스터디룸, 복합공간 등 새로운 공간이 마련된다. 수업 장소 역시 인근에 위치한 타 학교 또는 지역사회, 공동교육과정에 의한 온라인 등으로 확장된다(교육부, 2021b).

고교학점제 도입 및 연구·선도학교 지정에 따라 학생 선택과목이 확대되고 선택과목 개설에 대한 별도 규정이 마련되는 등 ‘교육과정 편성 및 운영’에 많은 변화가 생겼다. 교육과정 이수지도팀이 구성되고 학업설계지도 프로그램이 별도 운영되는 등 ‘진로 및 학업설계지도’ 역시 활발해지고 있으며, ‘평가’의 경우에도 대부분의 연구학교에서 과정중심평가가 활성화되는 등 여러 영역에서 많은 변화가 일어나고 있다(한국교육과정평가원, 2020). 또한 인접한 학교 간 협력하여 공동교육과정을 운영하거나 대학, 기업, 연구기관, 공공기관 등 지역사회기관 즉, 학교 밖 자원과 연계한 교육활동이 확대되는 등 지역 내 고교교육 공동체 역시 활발히 구축되고 있다(교육부, 2022).

## 2. 고교학점제 관련 빅데이터 분석 연구

고교학점제를 주제로 빅데이터 분석기법을 활용한 선행연구의 경우 고교학점제에 대한 사회적 인식을 확인하기 위한 연구(김소영, 홍후조, 2021; 김태현, 김석우, 2019; 박지혜, 2020; 조효제, 길혜지, 2022; 서추현, 2022)가 대부분이었고 그 외 이슈 및 과제를 제시하거나(김륜희, 지은림, 2022; 박지한, 김아영, 2021), 정책적 및 학술적 시사점을 제시하기 위한 연구(전은정, 최윤정, 2023; 조성범, 신하영, 2022)가 있었다. 연구 대상으로는 포털사이트 게시물 및 댓글(김소영, 홍후조, 2021; 김태현, 김석우, 2019; 박지혜, 2020; 조효제, 길혜지, 2022; 서추현, 2022) 혹은 신문 기사(김륜희, 지은림, 2022; 권충훈, 2020), 학술논문(전은정, 최윤정, 2023; 조성범, 신하영, 2022), 연구학교 운영 결과보고서(박지한, 김아영, 2021) 등이 활용된 것으로 나타났다. 분석 방법으로는 n-gram, 네트워크 분석, 토픽모델링, 감성분석 등이 활용되고 있었

다. 위 선행연구를 각 연구 별로 나누어 살펴보면 다음과 같다.

김소영, 홍후조(2021)는 2020년 1월~10월까지 주요 포털 내 고교학점제 관련 게시물을 분석하여 대중이 고교학점제를 비교적 정확히 인식하고 있음과 과목선택, 학점 이수 기준 등에 높은 관심이 있음을 밝혔다. 더불어 수업, 대학, 교사의 연결중심성이 높게 나타난 것에 비추어 대중이 고교학점제의 실질적 측면을 인식하기 시작한 것으로 보았다. 또한 연구 대상이 된 게시물은 진로 맞춤형 교육과정, 고교학점제 도입 기반마련, 고교학점제 전면 도입, 고교학점제 운영 등 4개의 군집으로 나타났음을 언급하였다.

김태현, 김석우(2019)는 고교학점제 도입 과정에 전문가들의 다양한 관점 및 사회적 의견을 반영하고자 포털사이트 및 SNS 등으로부터 수집한 데이터를 군집화한 결과 선택 중심 교육과정 운영, 고교학점제 전면 도입을 위한 준비, 고교학점제 정책 이해도 향상 방안, 다양한 교과목 개설로 나뉘었음을 밝혔다.

박지혜(2020)는 2018년 1월~2020년 7월까지 포털사이트 내 작성된 학교 간 공동교육과정 관련 텍스트 데이터를 수집하여 분석한 결과 상위 60개 키워드는 교육과정 정책 추진, 정보 공유, 대입 준비, 공동교육과정 운영 사례라는 네 개의 군집을 형성하는 것으로 나타났다.

조효제, 길혜지(2022)는 2017년 2월~2021년 8월까지 포털사이트 내 고교학점제 관련 기사의 댓글을 대상으로 영역별 쟁점을 정리한 결과 고교학점제의 기반 조성 영역에는 현장의견 반영, 교육정책비판 및 개선요구, 공교육 강화가 포함됐고, 운영과정 영역에는 교원임용 및 역량 강화, 고교학점제 선택권이 포함되었으며 운영결과 영역에는 수능 및 입시제도가 포함된 것으로 나타났다.

서추연(2022)은 2019년~2021년까지 포털사이트 내 고교학점제를 키워드로 하는 텍스트 데이터를 수집하여 분석한 결과 2021년 데이터에서 홍보-소통, 소통-미흡, 미흡-지적 등 부정적 견해 및 반응이 증가한 것으로 나타났는데 고교학점제 전면 도입이 가까워지고 있는 상황과 코로나-19로 인한 온라인수업 진행이 맞물리면서 사회 전반에 불안감이 증가한 것을 원인으로 보았다.

김륜희, 지은림(2022)은 고교학점제가 발표된 2017년 11월 27일부터 2022년 4월까지의 신문기사 내 고교학점제와 대입제도를 키워드로 하는 신문 기사를 분석한 결과 정시확대 이슈로 수능, 정시, 확대 등의 단어 빈도가 높았고 수능과 학생부종합전형 연관어로 절대평가와 공정성 빈도가 높게 나타났으며 수능과 관련하여 절대평가, 과목, 영향력 등의 출현 빈도가 높게 나타났다. 토픽 분석 결과 1) 고교학점제의 도입 취지 및 시행, 2) 고교학점제의 도입과 대입 개편 방안, 3) 수능전형 비중 확대와 학생부종합전형 축소, 4) 정시 확대와 교육계 찬반 논란, 5) 미래교육과 대학입시 변화 등이 도출되었다.

전은정, 최윤정(2023)은 2018년~2022년 8월까지 고교학점제를 키워드로 하는 학술지 논문 토픽을 분석하였다. 그 결과 ‘학교문화(학생, 학부모, 교원 인식 및 문화)’, ‘정착조건 및 운영방향’, ‘환경시설(수업공간, 시스템 지원 등)’, ‘학생 과목선택(교육과정 측면)’, ‘선택과목(교과교육) 확대 및 개편’, ‘학생(진로선택 및 학업설계역량, 공강시간 활용 등)’ 영역을 다루는 토픽이 많은 것으로 나타났다.

박지한, 김아영(2021)은 교육부가 지정한 31개 고교학점제 1차 연구학교(2018년~2020년) 운영 결과보고서의 결론부에 기술된 고교학점제 관련 쟁점의 텍스트를 분석하였다. 텍스트 분석을 통해 ‘개인별 시간표 운영’, ‘자원 수급’, ‘성취평가제 도입’, ‘교사역할 변화’라는 네 가지 영역으로 쟁점을 범주화하고, 영역별로 ‘시간표상 공백 운영’, ‘지역 격차’, ‘미이수 관리’, ‘생활지도 약화’를 구체적으로 논의되어야 할 주요 과제로 도출하였다.

권충훈(2020)은 고교학점제 관련 기사의 변화 양상을 분석하여 정책이 발표되는 때 기사 건수가 급증함을 밝혔고 키워드를 분석하여 고교학점제 정책이 채택, 결정 단계를 거쳤으며 현재 집행 단계로 가고 있음을 밝혔다.

조성범, 신하영(2022)은 2017년 5월부터 2022년 1월까지 고교학점제 관련 논문 분석 결과 주요하게 다룬 주제어는 ‘고교 교육과정’, ‘선택 과목’, ‘교과 교육과정’, ‘학생 선택형 교육과정’, ‘진로 과목’으로 나타났고, 응집 구조 분석 결과, 1) 진로 교육 및 교육과정 편성 운영 관련 연구, 2) 학생 선택형 교육과정 및 교과목 이수체계와 졸업 요건 관련 연구, 3) 교과 교육과정 사례를 중심으로 한 교육과정 재구조화 관련 연구, 4) 대입 제도와 평가 제도 관련 연구라는 주제어 클러스터가 나타났다.

빅데이터를 활용한 고교학점제 관련 선행연구들은 데이터 분석 결과를 내용에 따라 군집화하여 제시하고 있음을 알 수 있다. 연구 대상이 된 데이터들이 공통적으로 포함하고 있는 내용은 ‘도입 기반 마련’, ‘도입 및 시행’, ‘고교학점제 운영’, ‘교육과정 개편’, ‘학교 문화’, ‘환경 시설’, ‘대입’ 등의 영역이다(〈표 1〉 참조).

연구 대상이 된 데이터 대부분은 학생이 수강할 과목을 선택할 수 있도록 제도를 운영하는 것과 관련된 ‘고교학점제 운영’ 영역을 포함하는 것으로 나타났다. 선택과목 확대개설 및 개편에 대한 내용을 다룬 ‘교육과정 개편’ 영역과 고교학점제 도입 기반 마련에 필요한 내용을 다룬 ‘도입 기반 마련’ 영역의 내용을 포함한 데이터 역시 많은 것으로 나타났다. 고교학점제도를 도입 및 추진한다는 내용을 다룬 ‘도입 및 시행’ 영역, 고교학점제와 연계한 대학 입시 관련 내용을 다룬 ‘대입’ 영역 관련 내용 역시 데이터에서 다루고 있음을 확인할 수 있다. 일부 연구를 통해 정책에 대한 이해도 및 학교 관계자 인식과 관련된 ‘학교문화’ 영역과 수업 공간, 제반 시스템 지원 등과 관련된 ‘환경시설’ 영역 관련 내용의 데이터도 드러났다.

〈표 1〉 빅데이터 활용 고교학점제 관련 선행연구

저자 \ 영역	영역	도입 기반 마련	도입 및 시행	고교학점제 운영	교육과정 개편
김소영, 홍후조(2021)		고교학점제 도입 기반 마련	고교학점제 전면 도입	고교학점제 운영	진로맞춤형 교육과정
김태현, 김석우(2019)		고교학점제 전면 도입을 위한 준비		선택중심 교육과정 운영	다양한 교과목 개설
박지혜(2020)			교육과정 정책 추진	공동교육과정 운영 사례	
조효제, 길혜지(2022)		현장 의견 반영, 교육정책 비판 및 개선 요구, 공교육강화		교원임용 및 역량강화, 과목선택권	
김륜희, 지은림(2022)			고교학점제 도입 취지 및 시행		
전은정, 최윤정(2023)		정착 조건 및 운영 방안		학생 과목선택	선택과목 확대 및 개편
조성범, 신하영(2022)				학생 선택형 교육과정 및 교과목이수체계, 졸업요건연구	교육과정 편성 운영 및 재구조화 연구
저자 \ 영역	영역	학교 문화	환경 시설	대입	
김소영, 홍후조(2021)					
김태현, 김석우(2019)		정책 이해도 향상 방안			
박지혜(2020)				대입준비	
조효제, 길혜지(2022)				수능 및 대입제도	
김륜희, 지은림(2022)				고교학점제 도입과 대입 개편 방안, 수능 전형 비중 확대와 학생부종합전형 축소, 정시 확대와 교육계 찬반 논란, 미래교육과 대학입시 변화	
전은정, 최윤정(2023)		학생, 학부모, 교원 인식 및 문화	수업공간, 시스템 지원 등		
조성범, 신하영(2022)				대입제도와 평가제도 연구	

※ 서추연(2022), 박지한, 김아영(2021), 권충훈(2020)의 연구는 위 범주로의 구분에 적합하지 않아 제외

## II. 연구방법

### 1. 분석대상

본 연구에서는 고교학점제를 주제로 하는 뉴스 기사를 연구 대상으로 하였다. 뉴스 기사는 한국언론진흥재단이 운영하는 빅카인즈 사이트를 통해 수집하였다. 기간은 고교학점제 도입이 처음 발표된 날인 2017년 11월 27일부터 2022년 9월 30일까지로 하고, 언론사는 빅카인즈에서 제공하는 모든 언론사를 대상으로 하였다(〈표 2〉 참조).

〈표 2〉 연구 대상으로 선정한 뉴스 기사 발행 언론사

구분	언론사
전국일간지	경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보
경제일간지	매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제
지역일간지	강원도민일보, 강원일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보
방송사	KBS, MBC, OBS, SBS, YTN
전문지	디지털타임스, 전자신문

위와 같이 검색 기간과 언론사를 설정한 후 제목 및 본문이 고교학점제라는 주제어를 포함하는 기사를 검색하여 8,517건의 결과를 얻었다.

빅카인즈는 교육과 관련된 기사의 경우 통합분류를 기준으로 사회 영역 내 교육·시험으로 분류하므로 해당 분류에 포함되지 않은 기사는 우선 제외한 후 제목이 고교학점제를 정확히 포함하는 기사는 복원하여 분석 대상으로 하였다. 그 결과 총 7,115건의 기사가 분석에 활용되었다.

### 2. 분석절차 및 도구

전체 문서 토픽 분석 절차는 다음과 같다. 7,115건의 뉴스기사는 csv파일로 저장해 R 소프트웨어 버전 4.2.2.로 분석하였다. 텍스트 자료는 자연어처리(natural language processing; NLP)



가 가능하도록 가공해야 하는데 먼저 의미의 최소 단위인 형태소로 문장을 바꾸는 작업이 필요하다. 이후 자료 정제(data cleaning) 과정을 통해 특수문자 및 불용어(stopword) 등을 제거하는 전처리(preprocessing)작업이 필요하다.

본 연구에서는 빅카인즈가 제공하는 뉴스 중 키워드 항목을 분석에 활용하였다. 빅카인즈는 키워드 항목에 대해 본문 내에서 추출된 단어 중 단순 숫자, 이메일 주소, 시간을 뜻하는 단어(밤, 낮, 새벽 등)를 제외한 결과라 설명하고 있다. 실제 데이터 검토 결과 용언은 대부분 제거되어 있으나 숫자, 특수문자, 명사 이외의 품사 등이 다수 혼재되어있는 것을 확인하였다. 이에 2글자 이상의 명사를 명확히 추출하고자 R 프로그램의 tidytext 패키지(Silge & Robinson, 2016), KoLNP 패키지(Jeon & Kim, 2016), stringr 패키지(Wickham, 2015), dplyr 패키지(Wickham et al., 2015) 등을 통해 텍스트 데이터를 토큰화하고 형태소를 분석하여 보통명사를 추출하고 명사 이외의 형태소, 특수문자, 숫자 등은 제외하였다.

형태소 분할 및 자료 정제 등 전처리를 마친 후 단어빈도(term frequency; TF)와 단어 빈도-역 문서 빈도(term frequency-inverse document frequency; TF-IDF)값을 산출하고 해당 값을 고려하여 모든 문서에서 큰 의미를 가지지 않고 반복적으로 사용되는 단어를 불용어 처리하였다. 불용어 처리한 단어는 ‘교육, 학교, 학생, 학생들, 고교, 고교학점제’이다.

토픽 분석을 위한 topicmodels R 패키지(Grün & Hornik, 2011)의 LDA() 함수는 DTM(Document-Term\_matrix)구조를 기본 입력구조로 가지므로 R 프로그램의 tidytext 패키지(Silge & Robinson, 2016)를 활용하여 tidy format을 dtm format으로 변환하였다.

LDA분석을 실시하기 위해서는 먼저, 적절한 토픽 수를 결정하는 과정을 거쳐야 한다(유주영, 2022). LDA라는 명칭에서 알 수 있듯이 잠재되어 있는 토픽을 디리클레(dirichlet) 분포를 통해 얻어내는 것이기에 토픽을 알기 전 토픽 수를 미리 지정해주어야 한다. 토픽 분석 시 어려운 과제 중 하나는 토픽의 개수를 정하는 것이다. 이론에 근거하여 전적으로 연구자가 토픽 수를 추정할 경우 연구자 자의성과 관련하여 비판받을 수 있다(Hall et al., 2008). 이에 많은 경우 Arun et al.(2010), Cao et al.(2009), Deveaud et al.(2014), Griffiths et al.(2004) 등 네 개 지수에 근거하여 토픽 수의 적합도를 판단하는 하이퍼파라미터 튜닝 과정을 거친다. Griffiths et al.(2004)는 깁스샘플링의 로그 우도(log likelihood) 조화 평균(harmonic mean method)이 최대가 되는 값을 토픽 수로 한다. 하나의 문서집합을 토픽으로 나누어감에 따라 초반에는 형성된 토픽 간의 거리 평균값이 증가한다. 하지만 일정 수 이상의 토픽이 구성되면 평균보다 짧은 토픽 간 거리들이 증가하게 되고 평균값은 점차 낮아진다. 따라서 조화 평균이 낮아지기 직전에 해당하는 토픽의 수로 문서집합을 나눌 경우 분포 균형이 가장 잘 맞아진다고 본다.

Cao et al.(2009)은 토픽 분포 간 코사인 유사도가 가장 작아지는 값을 토픽 수로 선택한다. 토픽 구조의 안정성 측정을 위해 코사인 거리를 이용하여 토픽 간 상관관계를 측정하고 평균 거리가 작을수록 더욱 안정적으로 본다. 토픽 간 상관관계가 작다면 토픽의 유사성이 떨어지며, 독립성을 가진다는 점에 착안하여, 모든 토픽 쌍 사이의 평균 코사인 거리를 구하고 최적의 토픽 수를 추출하는 방식을 제안한 것이다.

Arun et al.(2010)은 토픽-단어 행렬에서 구한 특이값(singular values) 분포의 쿨백-라이블러 거리(Kullback-Liebler divergence)가 최소화되는 지점에서  $k$ 를 정한다. 쿨백-라이블러 거리(Kullback-Liebler divergence)는 서로 다른 확률 분포의 차이를 측정하는 척도이다. 하지만 대칭적(symmetric)이지 않다는 문제가 있다. 이에 Deveaud et al.(2014)은 토픽 분포 간 젠슨-샤논 거리(Jensen-Shannon distance)가 최대화되는 값을  $k$ 로 선택하는 방법을 제안한다. 즉, 모든 토픽 쌍 사이의 차이를 계산할 때 발생할 수 있는 문제를 피하기 위해 쿨백-라이블러 거리(Kullback-Liebler divergence)를 대칭적일 수 있게 개량한 대칭버전인 젠슨-샤논 거리(Jensen-Shannon distance)를 활용하는 것이다. 위와 같은 이론들을 종합하면, Cao et al.(2009)와 Arun et al.(2010)는 그 값이 최소화 되는 지점을 최적의 토픽 수로 보고, Griffiths et al.(2004)와 Deveaud et al.(2014)는 그 값이 최대화 되는 지점이 최적의 토픽수로 본다. 하지만 이러한 절차를 수행하더라도 토픽 수 결정 시 연구자의 주관에 완전 배제하기는 어렵다. 먼저 위와 같은 지수는 텍스트의 특성을 고려하거나 해석하여 얻은 결과가 아니라는 점에서 그러하다. 또한 알고리즘에 의해 추출된 4개의 지수들은 해당 이론에 따른 제각각의 수치를 보이는 경우가 많아 연구자의 종합적 판단이 필요하기에 자의적 해석을 제거하기 어렵다. 선행연구를 통해 살펴보면, 해석의 용이성을 고려하여 토픽을 일정 수 미만으로 제한하거나(이서진, 민경석, 2022), 너무 적은 수의 토픽은 위험성이 있다고 판단하여 제외하거나(Emidi & Galán, 2022), 모델 성능 지표의 변화가 최소화되는 지점의 토픽 수를 선정하거나(조경원, 우영운, 2019), 그래프가 완만해지는 지점을 판단하여 토픽 수를 선정하는(Buhin Pandur et al., 2021; 이정, 2022; 권나현 외, 2022; 홍준우, 홍태호, 2021) 등으로 문제를 해결한다. 이러한 문제로 인해 다수 연구들은 하이퍼파라미터 튜닝 방법을 통해 토픽 수 후보를 정한 뒤 텍스트를 해석하여 가장 적절한 토픽 수를 선택하고 있다(백영민, 2019; 이소정, 2020). 본 연구에서도 하이퍼파라미터 튜닝을 통해 토픽 수를 정한 후 토픽별 분류된 단어 및 문서의  $\beta$ 와  $\gamma$ 를 고려하여 최종 토픽을 정하였다. 토픽 수 범위는 교육부 정책 연구자료와 선행연구 등을 고려하여 2개에서 40개 사이의 토픽 수를 고려하도록 설정하였다.

이후 R 프로그램의 topicmodel 패키지(Grün & Hornik, 2011)의 LDA() 함수를 활용하여 토픽 분석을 실시하였다. 토픽 모형 생성의 반복수행 횟수와 관련하여 선행연구들에서 보통

1,000에서 2,000사이의 값을 권고하나(안주영 외, 2016), 보다 정밀한 값을 산출하기 위해 5,000으로 반복횟수를 설정하기도 한다(권순보, 2018). 본 연구에서는 토픽 생성 반복수행 횟수를 5,000으로 설정하였다.

LDA 토픽 모델 생성을 위해 하이퍼파라미터인  $\alpha$ 와  $\eta$ 를 설정해주어야 한다.  $\alpha$ 값이 커질수록 모든 문서에 있어서 토픽 배합이 비슷하게 되고  $\eta$ 값이 커질수록 한 토픽 내에서 모든어가 고르게 분포하게 된다. 각 문서가 적은 수의 토픽과 연관되어 있다고 판단하면  $\alpha$ 를 작게 설정하여야 하고, 한 토픽 내 주요 단어가 적다고 판단하면  $\eta$ 를 작게 설정해야 한다. 본 연구의 경우 문서별 토픽 및 토픽 내 주요 단어를 차별적으로 분석하고자 하였다. 또한 일부 선행연구에서  $\alpha$ 는 0.1,  $\eta$ 는 0.001로 설정하여 분석한 바 있어(이민철 김혜진, 2018; 한지영, 허고은, 2021), 본 연구에서는  $\alpha$ 는 0.1,  $\eta$ 는 0.001로 설정하였다.

LDA 분석결과를  $\beta$ 와  $\gamma$ 를 제공한다.  $\beta$ 파일에는 각각 다른 토픽들과 해당 토픽에 대한 단어 관련 벡터가 제공되는데, 각 단어의 가중치는 특정 주제에 해당 단어가 존재할 확률을 나타낸다.  $\gamma$ 파일에는 알고리즘에 입력된 문서집합의 각 문서에 대한 벡터가 포함되어 있는데, 각 주제별 하나씩의 벡터가 제공된다. 각 문서에 있는 각 주제의 가중치는 해당 문서에 대한 주제의 중요성을 나타낸다(Fyffe, 2009). 즉, 각 토픽에 포함되는 단어와 해당 단어가 각 토픽에 포함될 확률을 의미하는  $\beta$ 와 특정 문서가 특정 토픽을 담고 있는 확률을 의미하는  $\gamma$ 를 제공하는 것이다(유주영, 2022). 토픽모델링은 분석 결과 자체로 해당 토픽의 맥락적 의미를 제시하지 않고, 주제어만으로 정확한 의미를 이해하는데 한계가 있어(유주영, 2022) 많은 연구에서 토픽명을 정할 때  $\beta$ 와  $\gamma$ 를 함께 고려한다. 조서연 외(2021)는 토픽 수 선정 시 주제별 상위 확률값( $\beta$ )을 갖는 단어와 주제별 상위 원문( $\gamma$ )의 성격을 검토하여 결정하였다. 손태권, 황성환(2020)은 특정 문서가 해당 토픽에 할당될 비율인  $\gamma$ 를 기준으로 토픽별 대표 논문을 선정하였다. 이대영, 이현숙(2021)은 선정된 토픽 수의 적절성 판단에  $\gamma$ 와  $\beta$ 를 활용했다. 본 연구에서는 tidytext R 패키지(Silge & Robinson, 2016)를 이용하여  $\beta$ 와  $\gamma$ 를 추출하고 토픽별 주요 단어와 주요 기사별 토픽을  $\beta$ 와  $\gamma$ 를 통해 정리한 후 적절한 토픽명을 부여하는데 활용하였다.

이후 토픽 분석 결과에 대한 연도별 비중 변화를 분석하였다. 고교학점제 도입 계획이 발표된 해가 2017년이므로 2022년까지 연도별로 분석할 경우 6개 연도의 데이터를 추출할 수 있다. 6개 연도별 토픽 비중 비교를 통해 분석된 토픽이 시간적 흐름에 따라 어떤 변화를 보이는지 알아봄으로써 고교학점제 정착 과정 및 고교학점제와 관련된 외부 상황의 변화 양상을 파악해볼 수 있다. 즉, 시간적 흐름에 따른 토픽의 변화를 통해 시사점을 얻을 수 있다. 분석을 위해 각 문서의 연도 정보를 1년 단위로 입력하였다. 이후 reshape2 R 패키지(Wickham, 2017)

와 ggplot2 R 패키지(Wickham et al., 2016)를 활용하여 16개 중 하나의 토픽으로 분류된 문서들의 토픽별 비중을 연도를 기준으로 그래프로 시각화하였다. 이를 통해 시간의 흐름에 따라 특정 토픽으로 분류된 문서의 수가 늘어남과 줄어듦을 확인할 수 있도록 하였다.

### III. 분석결과

#### 1. 단어 빈도 분석

단어빈도(term frequency; TF)는 특정한 단어가 문서집합 내에 얼마나 자주 등장하는지를 나타내는 값으로(홍원의 외, 2014) 그 값이 높을수록 빈번히 등장하는 단어라고 생각할 수 있다(권순보, 2018; 정근하, 2011). 단어빈도 분석 결과 자주 등장하는 단어는 운영, 교육부, 과목, 대학, 학점, 수업, 지원, 교사, 확대, 선택 등의 순으로 나타났다(<표 3> 참조).

〈표 3〉 단어 빈도(TF) 분석 결과

순위	단어	빈도	순위	단어	빈도
1	운영	15,319	11	진로	9,023
2	교육부	14,182	12	대입	8,632
3	과목	13,182	13	수능	7,817
4	대학	12,383	14	자사고	7,797
5	학점	12,099	15	정책	7,710
6	수업	11,851	16	지역	7,605
7	지원	10,534	17	정부	6,859
8	교사	10,335	18	정시	6,538
9	확대	9,678	19	미래	6,508
10	선택	9,609	20	추진	6,325

출현빈도가 높을수록 중요한 단어라 생각할 수 있으나, 특정 단어가 모든 문서 내에서 자주 사용되는 경우, 해당 단어는 상대적 중요성을 가진다기보다 흔하게 등장한다는 것을 의미한다. TF-IDF는 어떤 단어가 흔하지 않지만 특정한 문서 내에서는 빈번히 등장하는 정도를 나타내는 지표이다. DF(document frequency)는 단어가 사용된 문서의 빈도를 의미하므로 이 수치가 클

수록 여러 문서 걸쳐 흔히 사용된 단어라 할 수 있다. IDF(inverse document frequency)는 전체 문서 수에서 DF가 차지하는 비율을 구한 값의 역수에 로그를 취한 값이다. IDF가 크면 문서에 따라 드물게 사용되는 단어라 할 수 있다. TF-IDF는 TF(term frequency)와 IDF(inverse document frequency)를 곱한 값이다. 즉 흔하지 않은 단어이나 특정한 텍스트에는 자주 사용되면 TF-IDF값은 커진다. 이를 수식(3.1)으로 나타내면 다음과 같다.

$$TF-IDF = TF \times \log \frac{N}{DF} \quad (3.1)$$

TF-IDF에 의한 빈도의 경우 자사고, 과목, 수능, 학점, 정시, 수업, 교육부, 대입, 일반고, 교육감 등의 순으로 나타났다 (<표 4> 참조). 이 중 자사고, 수능, 정시, 대입 등의 단어는 단순 단어빈도(TF) 분석 시 10위권 밖으로 나타났던 단어이다. 또한 일반고, 교육감, 공동 등의 단어는 단순 단어빈도 분석 시 20위권 내에 들어있지 않던 단어로 흔하지 않지만, 특정 문서에서 빈번히 사용되는 단어임을 의미한다.

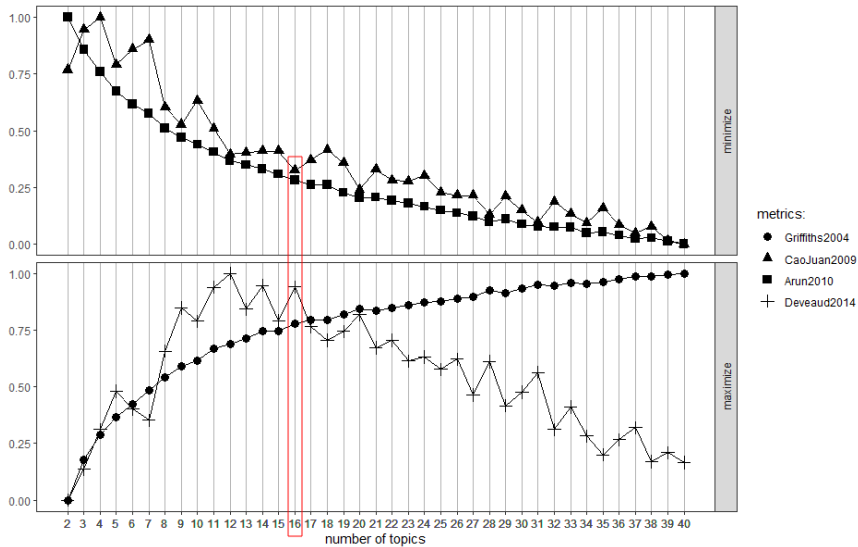
<표 4> TF-IDF 분석 결과

순위	단어	빈도	TF-IDF	순위	단어	빈도	TF-IDF
1	자사고	7,797	62.82	11	운영	15,319	43.15
2	과목	13,182	54.09	12	대학	12,383	43.03
3	수능	7,817	48.51	13	진로	9,023	42.21
4	학점	12,099	48.27	14	교사	10,335	42.00
5	정시	6,538	46.95	15	지원	10,534	40.82
6	수업	11,851	46.21	16	교원	4,931	39.99
7	교육부	14,182	45.49	17	선택	9,609	39.10
8	대입	8,632	45.31	18	공동	3,443	38.11
9	일반고	6,048	44.60	19	지역	7,605	37.92
10	교육감	6,053	44.33	20	전환	5,303	37.10

## 2. 토픽 분석

토픽 수 결정을 위해 Arun et al.(2010), Cao et al.(2009), Deveaud et al.(2014), Griffiths et al.(2004) 지수를 참조하고 토픽별 분류된 단어 및 문서의  $\beta$ 와  $\gamma$ 를 고려하여 최종 토픽 수를 결정하였다.

Cao et al.(2009)와 Arun et al.(2010)는 그 값이 최소화되는 지점을 최적의 토픽 수로 보고, Griffiths et al.(2004)와 Deveaud et al.(2014)는 그 값이 최대화되는 지점을 최적의 토픽수로 본다. Griffiths et al.(2004)의 경우 몇 차례 눈에 띄는 상승을 보이지만 16개의 토픽 수 전 15개 토픽 수에서 다소 주춤거리는 모습을 보이다가 16개 토픽 수에서 크게 상승하였다. 작은 수치를 보일 때 적절한 토픽 수임을 의미하는 Cao et al.(2009)의 경우 13-15개 토픽 수에서 주춤거리는 모습을 보이다가 16개 토픽 수에서 눈에 띄게 하락하는 모습을 보인다. Deveaud et al.(2014)의 경우 큰 수치를 나타낼수록 적절한 토픽 수임을 나타내는데 16개 토픽 수에서 급상승하는 모습을 보이며 이후 Griffiths et al.(2004)와 교차하는 모습을 보인다. 보통 Griffiths et al.(2004)와 교차하기 전 Deveaud et al.(2014) 지수가 최대일 경우의 토픽 수를 적절한 토픽으로 본다(양지연, 2019). Arun et al.(2010)은 다소 단조적 감소 추세를 보여 채택할만한 명확한 토픽 수를 확인하기 어려웠다([그림 1] 참조). 이에 토픽 수는 16개로 정하였다.



[그림 1] 하이퍼파라미터 튜닝에 의한 모델 성능 지표

아래 <표 5>는 분류된 문서가 많은 순서대로 토픽 번호를 나열하고 각각의 토픽에 포함된 주요 단어를  $\beta$ 순으로 10개씩 나열한 것이다. 많은 문서가 특정 토픽으로 분류되었다는 것은 해당 토픽을 다루고 있는 문서가 많음을 의미한다.

<표 5> 토픽별 주요 단어 분석 결과(분류된 문서 수 기준)

토픽 번호	주요 단어	문서 수
11	운영, 학점, 지원, 지역, 학부모, 역량, 강화, 정책, 미래, 진로	845
8	학점, 과목, 선택, 교육부, 수업, 시행, 이수, 전면, 운영, 교사	818
15	정시, 수능, 대입, 확대, 교육부, 개편, 대학, 정부, 비율, 전형	685
1	과목, 선택, 진로, 수업, 운영, 개설, 교사, 교과, 온라인, 공동	529
10	수업, 미래, 공간, 지원, 학습, 운영, 혁신, 교실, 온라인, 구축	487
5	교육부, 부총리, 대학, 정책, 정부, 장관, 추진, 국가, 사회, 유은혜	471
16	운영, 공동, 대학, 지역, 수업, 연계, 온라인, 강좌, 협력, 학점	443
9	교육감, 후보, 정책, 학력, 공약, 선거, 지역, 미래, 확대, 추진	415
14	자사고, 일반고, 전환, 외고, 국제, 교육부, 정부, 폐지, 일괄, 특목고	400
13	진로, 프로그램, 운영, 활동, 진학, 진행, 인재, 교장, 과학, 학습	378
7	대학, 전형, 대입, 수능, 입시, 정시, 학생부, 학년, 수시, 입학	316
3	교원, 교사, 교육부, 학급, 중등, 감소, 양성, 정원, 임용, 감축	310
2	지원, 운영, 사업, 선정, 계획, 교육부, 일반고, 예산, 지역, 확대	290
4	교사, 생각, 수업, 아이들, 사회, 변화, 학부모, 공부, 중요, 사교육	285
6	과목, 개정, 교육과정, 수학, 교육부, 수업, 교과, 선택, 중학교, 학년	267
12	자사고, 취소, 지정, 전환, 결과, 일반고, 서울, 서울시교육청, 교육부, 소송	176
합계		7,115

토픽 번호별  $\beta$ 지수 상위 단어는 다음 <표 6>와 같으며 지면 관계상 상위 5개 단어만을 제시하였다.  $\beta$ 는 단어가 해당 토픽에 속할 확률로, 각 토픽을 구성하는 단어들이 토픽 간에 상호 배타적으로 구성되는지 확인 가능하다(Blei et al., 2003). 문서 내 단어들은  $\beta$ 를 가지고 있고  $\beta$ 를 기준으로 상위에 위치한 단어를 통해 토픽별 내용을 유추할 수 있다.

〈표 6〉 토픽별 상위 5개 단어와  $\beta$ 지수

	토픽1		토픽2		토픽3		토픽4	
	단어	$\beta$	단어	$\beta$	단어	$\beta$	단어	$\beta$
1	과목	0.0457	지원	0.0550	교원	0.0497	교사	0.0168
2	선택	0.0386	운영	0.0375	교사	0.0426	생각	0.0152
3	진로	0.0328	사업	0.0234	교육부	0.0186	수업	0.0107
4	수업	0.0300	선정	0.0154	학급	0.0160	아이들	0.0090
5	운영	0.0280	계획	0.0146	중등	0.0131	사회	0.0080
	토픽5		토픽6		토픽7		토픽8	
	단어	$\beta$	단어	$\beta$	단어	$\beta$	단어	$\beta$
1	교육부	0.0297	과목	0.0305	대학	0.0462	학점	0.0481
2	부총리	0.0210	개정	0.0221	전형	0.0292	과목	0.0423
3	대학	0.0199	교육과정	0.0197	대입	0.0261	선택	0.0239
4	정책	0.0175	수학	0.0196	수능	0.0246	교육부	0.0202
5	정부	0.0170	교육부	0.0144	입시	0.0196	수업	0.0190
	토픽9		토픽10		토픽11		토픽12	
	단어	$\beta$	단어	$\beta$	단어	$\beta$	단어	$\beta$
1	교육감	0.0433	수업	0.0243	운영	0.0292	자사고	0.0606
2	후보	0.0279	미래	0.0205	학점	0.0261	취소	0.0209
3	정책	0.0137	공간	0.0169	지원	0.0235	지정	0.0162
4	학력	0.0125	지원	0.0127	지역	0.0181	전환	0.0138
5	공약	0.0104	학습	0.0124	학부모	0.0122	결과	0.0126
	토픽13		토픽14		토픽15		토픽16	
	단어	$\beta$	단어	$\beta$	단어	$\beta$	단어	$\beta$
1	진로	0.0225	자사고	0.0646	정시	0.0380	운영	0.0293
2	프로그램	0.0166	일반고	0.0470	수능	0.0366	공동	0.0275
3	운영	0.0143	전환	0.0360	대입	0.0345	대학	0.0248
4	활동	0.0132	외고	0.0233	확대	0.0302	지역	0.0236
5	진학	0.0097	국제	0.0231	교육부	0.0209	수업	0.0156



각 토픽별로  $\gamma$ 가 높게 나타난 상위 5개 기사의 주요 내용은 다음과 같이 정리할 수 있다(〈표 7〉 ~ 〈표 22〉 참조).  $\gamma$ 는 문서가 특정 토픽으로 분류될 확률을 의미하는 것으로, 1일 때 문서가 해당 토픽에 속할 확률이 1임을 의미한다(Blei et al., 2003). 따라서 주제별로 속해있는 문서들의  $\gamma$ 가 1에 가까운 값이 나올수록 분류가 잘 되었다고 판단할 수 있다. 문서가 속하는 토픽에 대한  $\gamma$ 는 1에 가까운 반면 문서가 속하지 않는 토픽에 대한  $\gamma$ 는 0에 가깝게 나타난다. 토픽 분석 시  $\gamma$ 가 높은 문서들을 고려할 수 있다. 이 경우 기준이 되는  $\gamma$  수치 혹은 문서의 수는 없으며 연구자가 토픽 해석에 도움이 되는 정도까지 활용하게 된다. 대부분의 선행연구에서는 독자의 이해를 돕기 위해  $\gamma$ 를 기준으로 상위 5개 문서를 제시하고 있다(조서연 외, 2021; 류병준, 하용훈, 2022; 이소정, 2020; 주보혜 외, 2019. 조성은 외, 2018). 본 연구에서 역시  $\gamma$ 가 높게 나타난 상위 5개 기사의 주요 내용을 제시하였다.

토픽명의 경우 토픽별 단어 포함 확률인  $\beta$ 와 문서의 토픽 분류 확률인  $\gamma$ 를 동시에 고려하였으며 즉, 토픽별로 분류된 상위 단어 및 상위 문서 내용을 종합적으로 검토한 결과이다.

〈표 7〉의 경우 토픽 1에 포함된 주요 문서의 내용이다. 토픽 1의 주요 단어는 과목, 선택, 진로, 수업, 운영, 개설, 교사, 교과, 온라인, 공동 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 1의 토픽명은 ‘과목 확대 개설 및 진로에 따른 선택 수업 운영’으로 정하였다.

〈표 7〉 토픽 1(과목 확대 개설 및 진로에 따른 선택 수업 운영)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
1	2899	광주시교육청 미래 진로 카드 제작 광주시교육청 고교학점제 조성 학생 맞춤 교육 지원	0.9962
1	3968	대구시교육청 고교학점제 대비 온라인 공동 교육 캠퍼스 거점 센터 운영	0.9957
1	6881	울산시교육청 중3 학부모 대상 고교학점 진로 학업 설계 상담 지원	0.9945
1	1666	고교 학점 선도 학교 지정 **고 학교 인근 공동교육과정 운영 학생 만족 선택 교과목 확대 과목 선택	0.9771
1	4964	진로 1학기 교육학 선택 수업 거점센터 경북도교육청 공동 교육 포항**고 개설 교육학 온라인 활용	0.9731

〈표 8〉의 경우 토픽 2에 포함된 주요 문서의 내용이다. 토픽 2의 주요 단어는 지원, 운영, 사업, 선정, 계획, 교육부, 일반고, 예산, 지역, 확대 등 이다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 2의 토픽명은 ‘교육청, 고교학점제 선도학교 지정 및 지원’으로 정하였다.

〈표 8〉 토픽 2(교육청, 고교학점제 선도학교 지정 및 지원)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$r$
2	2256	경기도교육청 2023년 도내 자율 공립 일반고 전환 **고, **고 시작 11개 전환 방침	0.9953
2	233	서울 지역 고교학점제 선도학교 선정 서울시교육청 공모	0.9886
2	247	대구시교육청 고교 고교학점제 선도학교 지정 ... 연구학교 선정 시행 고교학점제 요건 제안 실천 모델 역할	0.8924
2	3553	시교육청 학교 일반고 역량 강화 사업 학교 지정 선정 자율 학교 고교학점제 학교 지정 우선권 부여	0.8873
2	6700	고교학점제 구축 세종창의적 교육과정운영	0.8844

〈표 9〉의 경우 토픽 3에 포함된 주요 문서의 내용이다. 토픽 3의 주요 단어는 교원, 교사, 교육부, 학급, 중등, 감소, 양성, 정원, 임용, 감축 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 3의 토픽명은 ‘학령인구 감소 및 중등 교원 양성’으로 정하였다.

〈표 9〉 토픽 3(학령인구 감소 및 중등 교원 양성)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$r$
3	6864	학령 인구 감소 교사 교육 경제적 논리 고교학점제 교사들 업무 상황 교원 연일 감축 학생들 피해	0.9973
3	917	중등 1인 OECD 국가 평균 교사 학생 13.1 도달 고교학점제, 자유학년제 교육 수요 대응 OECD 국가 평균 개선 11명 유지	0.9973
3	6637	교육부 학령 인구 감소 수급 모델 교원 채용 설명 교육부 관계자 고교학점제 수급 변수 학령 인구 감소세 상쇄 상황	0.997
3	2065	학생 교육 다양화 개별화 추구 진로 생활 지도 내실화 교원 확충 교총 학령 인구 감소 교원 정부 정규 축소 고교학점제 기조 고교학점 배치	0.9967
3	7072	학령인구 감소 교사 인원 감축 아이들 교육 환경 제공 가능성 과밀 학급 해소 고교학점 시행 현장 교사 수요 상황	0.9967

〈표 10〉의 경우 토픽 4에 포함된 주요 문서의 내용이다. 토픽 4의 주요 단어는 교사, 생각, 수업, 아이들, 사회, 변화, 학부모, 공부, 중요, 사교육 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 4의 토픽명은 ‘고교학점제 도입 취지 및 학교 적응 현황’으로 정하였다.

〈표 10〉 토픽 4(고교학점제 도입 취지 및 학교 적응 현황)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
4	5464	그간 학교 역할 지식 전달 수준 교사 경험 방법 역할 수행 객관식 측정 지식 가치 일방 지식 전달 발견 해결 역량 고교학점제 미래 교육 희망 자리매김	0.9707
4	945	개인 창의성 위주 입시 암기 위주 반성 초등학교 객관식 시험 폐지 독서 교육 자유학기 민주시민교육 고교학점제 취지 진행	0.941
4	1350	고교학점제 고등학생 대학생 수업 선택 ... 고교학점제 선생님 업무 수업 운영 부담 ... 교사 가치 고교학점제 성공 사람들 자율성 회복 다원화 급변	0.9324
4	4917	고교 학점 인식 조사 학생들 학교 자신 진로 설계 정의 ... 고교학점제 교사 부족 대입 제도 공간 지적들 국가 주도 커리큘럼 탈피 학생 선택 패러다임 전환 시도	0.9239
4	116	고교 학점 학부모들 학생들 혼란 격변 교과과정 불안감 상황들 생각 기본 충실 학습 생각	0.9063

〈표 11〉의 경우 토픽 5에 포함된 주요 문서의 내용이다. 토픽 5의 주요 단어는 교육부, 부총리, 대학, 정책, 정부, 장관, 추진, 국가, 사회, 유은혜 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 5의 토픽명은 ‘교육부, 고교학점제 도입에 따른 대입제도 개편’으로 정하였다.

〈표 11〉 토픽 5(교육부, 고교학점제 도입에 따른 대입제도 개편)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
5	6891	고교학점제 준비 고교학점 개편 고교 체제 차질 교육부 당장 연말 고시 교육과정 개정 교육 연말 시안 마련 개편	0.9779
5	6294	정부 국정 추진 고교학점 보완 미래 사회 수요 대입 제도 개편 마련 교육부 고교학점제 대입 제도 개편 사항 철저히 추진 당부	0.9975
5	918	고교학점제 준비 고등 교육 역량 대학 혁신 지원 사업 추진	0.9967
5	6561	정부 교육 개혁 중요성 강조 추진 고교학점 보완 수요 미래 교육 사회 변화 반영 대입 제도 개편	0.9959
5	2583	부총리 고교체제 혁신 대입제도 단순화 고교학점제 교육체제 시작 ... 교육다양화 대학교육 미래 사회 변화 선도	0.9775

〈표 12〉의 경우 토픽 6에 포함된 주요 문서의 내용이다. 토픽 6의 주요 단어는 과목, 개정, 교육과정, 수학, 교육부, 수업, 교과, 선택, 중학교, 학년 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 6의 토픽명은 ‘과목 선택권 보장을 위한 교육과정 개정’으로 정하였다.

〈표 12〉 토픽 6(과목 선택권 보장을 위한 교육과정 개정)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
6	5403	미래 대입 교육과정 개정 교육 고교학점제 중심 교육과정 전반적 개정 반영 평가 체제 현행 수능 지속 부실화	0.9986
6	5364	전면 고교학점제 고등학교 교과 전반적 변화 경제 정치 고교 일반 선택 과목 진로선택과목 배치 고교학점제 개편안	0.9982
6	6980	교육부 공개 시안 개정 교육 2024년 연차적 적용 교육과정 개정 교육 고교학점제 시수 고등학교 수업 조정 과목	0.9982
6	5172	사회교과군 과목 고등학교 일반 선택 현행 4개 권고 고교 학점 체제 학생들 진로 적성 과목 선택 실질적 확보	0.9981
6	6998	고교학점제 취지 학생 과목 선택 보장 현행 공통 과목 교과목 체계 공통과목 일반선택 진로선택 융합선택 개편	0.998

〈표 13〉의 경우 토픽 7에 포함된 주요 문서의 내용이다. 토픽 7의 주요 단어는 대학, 전형, 대입, 수능, 입시, 정시, 학생부, 학년, 수시, 입학 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 7의 토픽명은 ‘고교학점제에 적합한 대입 전형 모색’으로 정하였다.

〈표 13〉 토픽 7(고교학점제에 적합한 대입 전형 모색)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
7	3321	학교 교육 공부 대입 반영 정시 모집 교과 평가 전면 시행 고교학점제 고교 현장 안착 긍정적 역할	0.9973
7	2431	고교학점제 신교육정책 반영 수능 체제 마련 시행 방침 ... 가칭 사회 통합 전형 신설 법제화... 교과 성적 위주 권고	0.9967
7	5659	통합 수능 전년도 입학 결과 학생들 지원 상황 3년 여파 2022 교육 고교학점제 혼란	0.996
7	6240	입시 전형 변화 고교학점제 시도 고교 교육 정상 반영 정시모집 선발 학생들 자퇴 이탈 비율 수시 요소 반영 시각	0.9565
7	3322	전면 고교학점제 고교 현장 안착 긍정적 역할 설명 교육부 대입 공정 강화 방안	0.93

〈표 14〉의 경우 토픽 8에 포함된 주요 문서의 내용이다. 토픽 8의 주요 단어는 학점, 과목, 선택, 교육부, 수업, 시행, 이수, 전면, 운영, 교사 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 8의 토픽명은 ‘과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내’로 정하였다.

〈표 14〉 토픽 8(과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
8	1560	고교학점제 내년 마이스터 시행 과목 선택 이수 졸업 평가제도	0.9979
8	3773	3년 192학점 졸업 대학생처럼 과목 시간표 짤다 전면 고교학점제 토대 진로 탐색	0.9977
8	3705	고교학점제 종합 추진 계획 학기 최소 학점 수강 조기 졸업 불가 ... 이수 과목 만큼 학점 졸업 인정 제도	0.9974
8	1581	교육부 21일 정부서울청사 마이스터 고교학점제 방안 학생 진로 과목 선택 이수 누적학점 기준 도달 졸업 제도	0.9972
8	3771	2025년 고등학교 교과목 선택 고교 학점 전면 시행...교육 국정 입시 위주 교육과정 학생들 진로 적성 수업 취지	0.9971

〈표 15〉의 경우 토픽 9에 포함된 주요 문서의 내용이다. 토픽 9의 주요 단어는 교육감, 후보, 정책, 학력, 공약, 선거, 지역, 미래, 확대, 추진 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 9의 토픽명은 ‘교육감 후보 고교학점제 관련 공약’로 정하였다.

〈표 15〉 토픽 9(교육감 후보 고교학점제 관련 공약)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
9	6519	교육감 진보 성향 더불어민주당 소속 도지사 교육감 보수 성향 추진 정부 고교 학점 탄력 예상 고교학점제 찬성 후보 당선	0.9987
9	6435	전남 교육감 비서실장 교육 행정 분야 경험 주장 후보 문제인 정부 추진 윤석열 정부 기초 계승 고교학점제 시기상조 견해	0.9976
9	6070	충북 교육 후보님 정책 긍정 평가 추상 정책 현실 반영 정책 염려 고교학점제 입장 경험 학교 현장 중요 고교학점제 찬성	0.9952
9	6305	고교학점제 공방 후보 맞출 교육 정착 고교 학점 조기 심혈 보수 성향 **** 후보 학력 저하 지적 고교 학점 신중 접근 제기	0.9908
9	5621	강원도 교육감 내년 안전 학교 구축 교육 기초 학력 책임 고교학점제 준비 최선	0.9631

〈표 16〉의 경우 토픽 10에 포함된 주요 문서의 내용이다. 토픽 10의 주요 단어는 수업, 미래, 공간, 지원, 학습, 운영, 혁신, 교실, 온라인, 구축 등이다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 10의 토픽명은 ‘수업 지원을 위한 혁신적인 미래 공간 혁신적 미래형 학교공간 조성’으로 정하였다.

〈표 16〉 토픽 10(혁신적 미래형 학교공간 조성)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$r$
10	3559	그린스마트 미래 학교 고교학점제 확대 고도화	0.9968
10	3577	교육공간 건물 형태 물론 실제 학생 활동 설계 특징 ..., 고교학점제 학생 중심 교육과정 공간들 심사위원들 우수 평가	0.9958
10	3486	온라인 수업 친숙 콘텐츠 만큼 디지털 활용 수업 확산 전망 과목 공동 교육 운영 고교학점제 현장감 원격수업 확대 예상	0.9784
10	4164	도서관 로비 구조적 연결 교사 학생 소통 중심 공유 공간 ... 공간 혁신 사업 고교 학점제 상설 학교 모델학교	0.9742
10	4152	전국 고등학교 시설 탈바꿈 규격화 교실 수강 인원 수업 상황 다목적 활용 고교학점제 학습 ... 학생들 개방 공간 휴식 교류 정서 함양	0.9692

〈표 17〉의 경우 토픽 11에 포함된 주요 문서의 내용이다. 토픽 11의 주요 단어는 운영, 학점, 지원, 지역, 학부모, 역량, 강화, 정책, 미래, 진로 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 11의 토픽명은 ‘고교학점제 운영을 위한 지역별 역량 강화 노력’으로 정하였다.

〈표 17〉 토픽 11(고교학점제 운영을 위한 지역별 역량 강화 노력)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$r$
11	4104	고교 학점 정책 프로그램 준비 ... 학업 관리 역량 강화 교육 주체 형성 고교 학점 공감대 방안 주도 역할 마련 수요자 정책	0.9966
11	2277	대전교육정책 홍보 전시 홍보관 10개 부스 고교학점제	0.9962
11	4270	교육 지원청 고교학점제 2025년 전면 시행 조성 음성교육협력센터 음성 고교 발전 협의체 선포식	0.9962
11	4105	대전교육청 고교학점제 조성 계획 교육 공동체 역량 강화 종합 추진 계획	0.9959
11	5510	충남교육청 고교학점제 교육 소외 지역 교감 협의체 고교학점제 교육 소외 지역 지역 공유형 협력 체제 구축	0.9959

〈표 18〉의 경우 토픽 12에 포함된 주요 문서의 내용이다. 토픽 12의 주요 단어는 자사고, 취소, 지정, 전환, 결과, 일반고, 서울, 서울시교육청, 교육부, 소송 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 12의 토픽명은 ‘서울시교육청 자사고 지정 취소 위법 판결’로 정하였다.

〈표 18〉 토픽 12(서울시교육청 자사고 지정 취소 위법 판결)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
12	4667	자사고 외국어고 국제고 일반고 일괄 전환 패소 ... 학생들 과목 선택 보장 고교학점제 정부 학교 선택 모순 지적	0.9987
12	4658	자율행사립고 자사고 지위 회복 학교 법인 **학원 자사고 지정 취소 처분 *** 경기도 교육감 상대 제기 소송 승소	0.9986
12	4447	법원 자사고 취소 위법 10곳 지위 회복 **고, **고 승소...고교 서열 고교학점 자사고 일반고 전환 논리	0.9983
12	4446	자사고 교장단 ***교육감 행정 법원 판결 항소 철회 요구 이날 서울행정법원 **고, **고 자사고 지정 취소 처분 취소 소송 학교 판결 승소	0.9981
12	5767	서울시교육청 항소 취하 2025년 예정 교육부 자사고 외교 국제고 일반 정책 일괄 전환 직접적 영향	0.9977

〈표 19〉의 경우 토픽 13에 포함된 주요 문서의 내용이다. 토픽 13의 주요 단어는 진로, 프로그램, 운영, 활동, 진학, 진행, 인재, 교장, 과학, 학습 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 13의 토픽명은 ‘진로 프로그램 특화 운영(직업계고)’로 정하였다.

〈표 19〉 토픽 13(진로 프로그램 특화 운영)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
13	7053	***여상 기업 활용 교육과정 NCS 중심 교육 고교학점제 전공 심화 코스제 교육 전공 지식 학과 교육과정 학과 교육 운영 학생들 취업	0.9987
13	4472	성공적 고교학점제 운영 방과 수업 중화 대상 학과체험 제공 전국 상업 경진 대회 참여 우수 성적 ...사무행정 호텔식음료서비스 금상 전자상거래 회계실무 세무 실무 수상	0.9984
13	2878	산학 일체 도제 사업 특성화 ... 고교학점제 선도 학교 시대 산업 혁명 변화 반영 학생 중심 직업 교육 편성	0.998
13	5550	경기도 도제 학교 운영 학생 기업 실무 능력 학습 고교 학점 선도 학교 학생 진로 탐색 기회 학과 코스제	0.9973
13	4428	학생들 체험 각자 소질 잠재력 개발 목적 진로활동 고교학점제 이해 증진 주체 고교생활 진로 탄력 함양	0.9965

〈표 20〉의 경우 토픽 14에 포함된 주요 문서의 내용이다. 토픽 14의 주요 단어는 자사고, 일반고, 전환, 외교, 국제, 교육부, 정부, 폐지, 일괄, 특목고 등으로 나타난 바 있다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 14의 토픽명은 ‘교육부 외교, 국

제고, 자사고 일반고 전환 추진'으로 정하였다.

〈표 20〉 토픽 14(교육부 외고, 국제고, 자사고 일반고 전환 추진)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
14	2167	고교 서열 해소 일반 방안 교육 역량 강화 교육부 중등교육법 시행령 개정 고교학점제 자사고 국제고 일반고 전환	0.9985
14	2162	자사고 외고 국제고 시한부 선고 자율형사립고 자사고 특수목적고 특목고 외국어고 국제고 79개 일반고 일괄 전환	0.9977
14	1937	고교학점제 ... 고교 서열 해소 고교학점제 이행 교육 공약 모양새 교육 정부 자사고 외고 국제고 일반고 일괄 전환	0.9976
14	2314	자사고 국제고 일반고 일괄 전환 단계 ... 이날 장관 이날 고교학점제 안정 준비 일반고 역량 강화 방안 실질적 추진	0.997
14	2147	외국어 국제고 2025년 일반고 전환 골자 고교 서열 해소 일반 방안	0.9968

〈표 21〉의 경우 토픽 15에 포함된 주요 문서의 내용이다. 토픽 15의 주요 단어는 정시, 수능, 대입, 확대, 교육부, 개편, 대학, 정부, 비율, 전형 등으로 나타났다. 주요 단어들과 표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 15의 토픽명은 '정시 확대 정책과 고교학점제'로 정하였다.

〈표 21〉 토픽 15(정시 확대 정책과 고교학점제)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
15	670	정시 전형 비중 현행 확대 권고 ... 수능 고교 체제 개편 고교학점제 종합적 공론 결과	0.9979
15	1842	정시 확대 하한선 방안 논의 호응 모양새 전문가들 정시 확대 고교학점제 정부 역점사업 배치 측면 교육 해소 근본적 해결책 신중 접근 조언	0.9978
15	1871	정부 대표적 교육 정책 고교학점제 해소 고교 서열 수능 내신 절대평가 정시 확대 수능 변별력 중요성	0.9977
15	516	수능 vs 학종 비율 논의 여부 수시 정시 통합 여부 논란 논의 가능성 교육부 이첩 가능 ...고교학점제 경험 학생들 학년도 대입 전형 개편 논의	0.9976
15	634	정시 확대 가단 절대평가 절충안 수능 절대 평가 전제 고교학점제 조치 학교 교육 정상 정부 교육정책 타격 가능성	0.9971

〈표 22〉의 경우 토픽 16에 포함된 주요 문서의 내용이다. 토픽 16의 주요 단어는 운영, 공동, 대학, 지역, 수업, 연계, 온라인, 강좌, 협력, 학점 등으로 나타난 바 있다. 주요 단어들과



표와 같이 나타난 주요 문서를 종합적으로 고려하여 토픽 16의 토픽명은 ‘지역 대학 및 인근 고교 연계 온라인 공동교육과정’으로 정하였다.

〈표 22〉 토픽 16(지역 대학 및 인근 고교 연계 온라인 공동교육과정)에 포함된 문서 주요 내용

토픽	문서번호	주요내용	$\gamma$
16	1172	고등학생들 과목 선택 기회 제공 오프라인 공동교육과정 지역 연계 공동 교육 미래 공동교육과정 온라인 공동 교육 포함	0.9964
16	6821	고교학점제 꿈키움 캠퍼스 운영 ...고교 공동교육과정 대학 연계 공동 교육 ... 전남교육청 교육감 고교학점제 안정 정착	0.9959
16	4722	**교육청 구축 고교 학점 플랫폼 대학 지역 공동 교육 시작	0.9957
16	6568	***대 **교육청 교육과정 대학 연계 공동 교육 종강 3월 교과 참여 고등학생 교육과정 대학 연계 공동 교육 모습	0.9957
16	4788	대학 연계 공동 교육 꿈키 프로그램 운영 **교육청 시행 고교 학점제 전면 대비 ...학생 개개인 진로설계 지원	0.9957

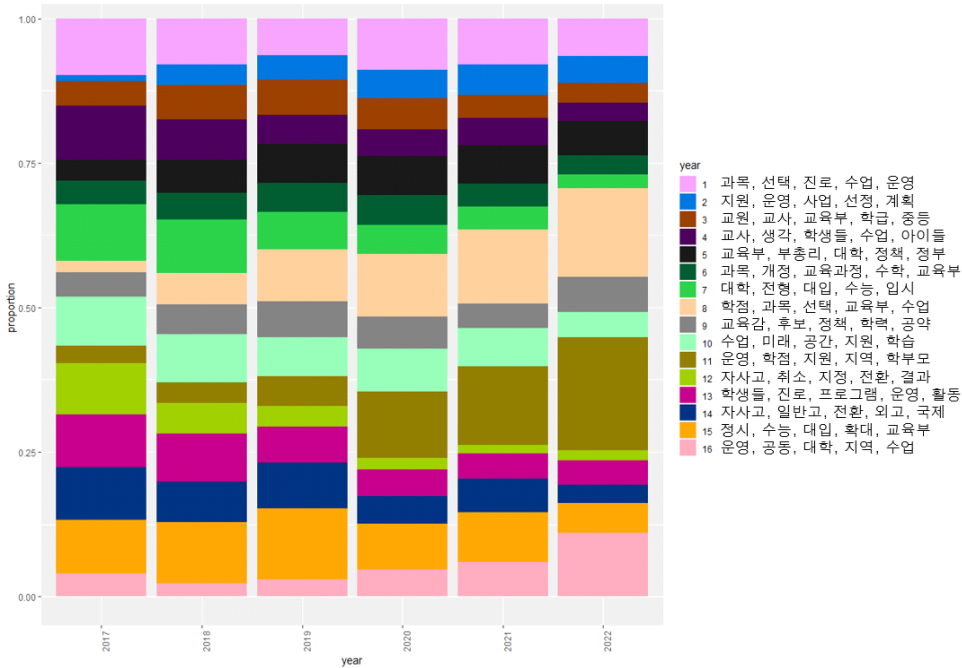
최종적으로 토픽명을 정리하면 〈표 23〉와 같다. 표의 가장 오른쪽 열의 ‘문서 수’는 해당 토픽으로 분류될 확률이 높은 문서의 갯수를 의미한다. 많은 문서가 분류된 토픽일수록 해당 토픽을 다루는 문서가 많음을 의미한다. 토픽 분석 결과 분류된 문서가 가장 많은 토픽은 ‘고교학점제 운영을 위한 지역별 역량 강화 노력(845건)’으로 나타났다. 다음으로 분류된 문서가 많은 토픽은 ‘과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내(818건)’로 나타났다. 더불어 ‘정시 확대 정책과 고교학점제’, ‘과목 확대 개설 및 진로에 따른 선택 수업 운영’, ‘혁신적 미래형 학교공간 조성’, ‘교육부, 고교학점제 도입에 따른 대입제도 개편’, ‘지역 대학 및 인근 고교 연계 온라인 공동교육과정’, ‘교육감 후보 고교학점제 관련 공약’, ‘교육부 외교, 국제고, 자사고 일반고 전환 추진’, ‘(직업계고) 진로 프로그램 특화 운영’, ‘고교학점제에 적합한 대입 전형 모색’, ‘학령인구 감소 및 중등 교원 양성’, ‘교육청, 고교학점제 선도학교 지정 및 지원’, ‘고교학점제 도입 취지 및 학교 적응 현황’, ‘과목 선택권 보장을 위한 교육과정 개정’, ‘서울시교육청 자사고 지정 취소 위법 판결’ 등이 토픽으로 등장하였다.

〈표 23〉 전체 문서의 토픽 분석 결과(분류된 기사가 많은 토픽 순)

순위	토픽번호	토픽명	문서 수
1	11	고교학점제 운영을 위한 지역별 역량 강화 노력	845
2	8	과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내	818
3	15	정시 확대 정책과 고교학점제	685
4	1	과목 확대 개설 및 진로에 따른 선택 수업 운영	529
5	10	혁신적 미래형 학교공간 조성	487
6	5	교육부, 고교학점제 도입에 따른 대입제도 개편	471
7	16	지역 대학 및 인근 고교 연계 온라인 공동교육과정	443
8	9	교육감 후보 고교학점제 관련 공약	415
9	14	교육부 외고, 국제고, 자사고 일반고 전환 추진	400
10	13	(직업계고) 진로 프로그램 특화 운영	378
11	7	고교학점제에 적합한 대입 전형 모색	316
12	3	학령인구 감소 및 중등 교원 양성	310
13	2	교육청, 고교학점제 선도학교 지정 및 지원	290
14	4	고교학점제 도입 취지 및 학교 적응 현황	285
15	6	과목 선택권 보장을 위한 교육과정 개정	267
16	12	서울시교육청 자사고 지정 취소 위법 판결	176

### 3. 시간 흐름에 따른 토픽 변화

시간의 흐름에 따라 각 토픽을 다루는 뉴스기사 비중이 변화하는 모습을 파악하고자 연도를 기준으로 분석하였다. [그림 2]는 고교학점제 도입을 처음 발표한 해인 2017년부터 2022년까지 연도별로 각 토픽 비중의 변화를 보여주고 있다. 그래프 오른쪽 범례는  $\beta$ 기준 상위 5개 단어이며 토픽명을 쉽게 비교해볼 수 있도록 〈표 24〉과 같이 번호 순으로 토픽을 정렬하여 제시하였다. 토픽 8번인 ‘과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내’와 토픽 11번인 ‘고교학점제 운영을 위한 지역별 역량 강화 노력’, 그리고 토픽 16번인 ‘지역 대학 및 인근 고교 연계 온라인 공동교육과정’이 증가하고 있음을 확인할 수 있다. 반면 토픽 7번인 ‘고교학점제에 적합한 대입 전형 모색’과 토픽 15번인 ‘정시 확대 정책과 고교학점제’는 감소하는 추세를 보이고 있다. 또한 토픽 12번인 ‘서울시교육청 자사고 지정 취소 위법 판결’, 토픽 14번인 ‘교육부 외고, 국제고, 자사고 일반고 전환 추진’ 역시 감소하는 모습을 보인다.



[그림 2] 연도에 따른 토픽 변화

〈표 24〉 토픽 분석 결과(토픽 번호순)

토픽번호	토픽명
1	과목 확대 개설 및 진로에 따른 선택 수업 운영
2	교육청, 고교학점제 선도학교 지정 및 지원
3	학령인구 감소 및 중등 교원 양성
4	고교학점제 도입 취지 및 학교 적응 현황
5	교육부, 고교학점제 도입에 따른 대입제도 개편
6	과목 선택권 보장을 위한 교육과정 개정
7	고교학점제에 적합한 대입 전형 모색
8	과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내
9	교육감 후보 고교학점제 관련 공약
10	혁신적 미래형 학교공간 조성
11	고교학점제 운영을 위한 지역별 역량 강화 노력
12	서울시교육청 자사고 지정 취소 위법 판결
13	(직업계고) 진로 프로그램 특화 운영
14	교육부 외교, 국제고, 자사고 일반고 전환 추진
15	정시 확대 정책과 고교학점제
16	지역 대학 및 인근 고교 연계 온라인 공동교육과정

#### IV. 결론

본 연구는 고교학점제를 주제로 쓰여진 뉴스기사의 토픽이 무엇인지 분석하여 고교학점제의 도입 및 정착에 시사점을 제공하고자 하였다. 이를 위해 교육부의 고교학점제 도입 발표 이후 2022년까지 고교학점제와 관련된 뉴스기사 7,115건을 대상으로 LDA 기반 토픽모델링 분석을 실시하였다.

고교학점제 관련 뉴스기사 토픽 분석 결과는 다음과 같다.

먼저 토픽 분석에 앞서 데이터 내 TF(단어 빈도)와 TF-IDF(단어 빈도-역 문서 빈도)를 분석하였다. 그 결과 TF 분석 시 상위 10위권 밖으로 나타났던 자사고, 수능, 정시, 대입 등의 단어와 20위권 내 들지 않았던 교육감, 공동 등의 단어가 TF-IDF 분석을 통해 드러남을 확인할 수 있었다. 이들 단어들은 자사고 등의 일반고 전환 추진, 정부의 대입제도 개편 발표, 교육감 후보의 고교학점제 관련 공약, 공동교육과정 운영 등의 토픽을 대표하는 단어들이라 할 수 있다. 따라서 TF-IDF 분석을 통해 실제 토픽분석에 앞서 의미있는 토픽을 예측해볼 수 있음을 알 수 있다. 특히 공동이라는 단어를 제외한다면 고교학점제를 직접적으로 대표하는 단어라기보다 특정 상황에 의해 일정 기간 동안 이슈가 된 토픽에 포함되는 단어가 드러났다고 할 수 있다. 일부 선행연구는 TF-IDF 이용에 주의를 기울여야 함을 강조한다. TF값이 매우 큰 경우 IDF로도 교정이 불가능할 수 있고(이성직, 김한준, 2009), 고유명사 및 새로운 개념 분석 시 단어의 중요도가 과소평가되거나 과대평가될 수 있어(안예지, 김동일, 2020; 임화진, 박성현, 2015) TF를 중심으로 TF-IDF를 보완적으로 사용하는 것이 더욱 적절하다고 본 것이다(임화진, 박성현, 2015). 본 연구 결과 분석된 TF-IDF 상위에 위치한 단어는 빅데이터를 활용해 고교학점제를 분석한 선행연구에서 다루지 않은 새로운 토픽을 적절히 대표하고 있었다. 그러나 TF와 TF-IDF 상위 10개 단어 중 중복 단어인 과목, 학점, 수업, 교육부를 제외할 경우 TF는 운영, 대학, 지원, 교사, 확대, 선택 이라는 단어가 남는 반면 TF-IDF는 자사고, 수능, 정시, 대입, 일반고, 교육감 이라는 단어가 남는다. 즉, 고교학점제라는 제도 자체의 본질적인 성격을 파악하는 데는 오히려 TF가 유리할 수 있는 반면, 데이터 내 흔하지 않은 주제의 문서 및 단어를 파악하는데는 TF-IDF가 유리한 것으로 보인다. 따라서 TF와 TF-IDF를 상호 보완적으로 활용해야 할 것으로 판단된다. 특히 불용어 선정 있어 TF-IDF를 기준으로 하위 30개 단어를 설정하거나(Schutze et al., 2008) 중위수를 사용해야 한다는 가이드가 제공되기도 하는데(Grun & Hornik, 2011) 이와 같은 방법으로 불용어를 선정하는 경우 신중해야 할 것으로 보인다.

고교학점제 관련 기사를 대상으로 토픽 분석한 결과 1) 고교학점제 운영을 위한 지역별 역량 강화 노력, 2) 과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내, 3) 정시 확대 정책과

고교학점제, 4) 과목 확대 개설 및 진로에 따른 선택 수업 운영, 5) 혁신적 미래형 학교공간 조성, 6) 교육부, 고교학점제 도입에 따른 대입제도 개편, 7) 지역 대학 및 인근 고교 연계 온라인 공동교육과정, 8) 교육감 후보 고교학점제 관련 공약, 9) 교육부 외교, 국제고, 자사고 일반고 전환 추진, 10) (직업계고) 진로 프로그램 특화 운영, 11) 고교학점제에 적합한 대입 전형 모색, 12) 학령 인구 감소 및 중등교원 양성, 13) 교육청, 고교학점제 선도학교 지정 및 지원, 14) 고교학점제 도입 취지 및 학교 적응 현황, 15) 과목 선택권 보장을 위한 교육과정 개정, 16) 서울 시교육청 자사고 지정 취소 위법 판결로 나타났다. 분류된 문서가 많은 순으로 상위 5개 토픽을 구체적으로 살펴보면 다음과 같다.

첫 번째, 분류된 문서가 가장 많은 토픽은 고교학점제 운영을 위한 지역별 역량 강화 노력으로 나타났다. 위 토픽으로 분류된 주요 기사를 살펴보면 각 교육청에서 고교학점제 정책 도입을 위해 준비, 홍보, 공감대 형성 노력, 협의체 구성 등의 노력을 하는 내용을 담고 있다. 이러한 결과는 본 연구 대상이 된 신문기사 출처에 지역일간지를 모두 포함한 것도 하나의 원인일 수 있다. 지역일간지의 경우 해당 지역 지자체, 공공기관 등의 사업 추진 내용을 다루는 경우가 많다. 고교학점제 도입 정책에 따라 각 지역에서 다양한 형태로 이를 지원 및 홍보하고 있음을 알 수 있다.

두 번째 토픽은 과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 안내에 대한 토픽이다. 관련 기사들은 고교학점제가 시행될 것임을 안내하며 선택과목 이수, 학점 취득, 졸업요건 등 여러 영역에서 고교학점제가 가진 특징을 다루고 있다. 뉴스기사는 대중에게 지식이나 정보를 전달하는 역할을 한다. 정부는 맡은 임무 완수를 위해 업무 내용을 국민에게 알리고 이해와 협력을 얻어야 하며(윤정길, 2000), 이 과정에서 대다수 국민은 언론을 통해 정부 활동을 알게 된다(이강웅, 2002). 또한 언론은 공공문제에 대해 정확한 정보를 제공하고자 한다(이강웅, 2002). 즉, 고교학점제라는 새로운 제도를 도입하게 되면서 해당 제도와 관련한 정책홍보 및 정보전달을 위한 기사가 상당 부분을 차지했음을 알 수 있다.

세 번째 토픽은 정시 확대 정책과 고교학점제이다. 위 토픽으로 분류된 기사들은 정시 전형 비중을 확대하겠다는 정부 정책이 고교학점제의 방향과 배치되는 측면이 있음을 주로 다루고 있다. 2019년 입시 불공정이 이슈가 된 것을 계기로 정시 비중 상향에 대한 논의가 일어났다. 고교학점제는 학생이 선택한 진로에 따라 이수할 과목을 선택하고 그 연장선상에서 대학에 진학하게 된다. 따라서 고교학점제 취지에 더욱 부합하는 전형은 정시보다 학생부종합전형이라 할 수 있다. 고교학점제 도입과 정시 확대가 동시에 정책으로 제시되어 상호 충돌하는 부분이 이슈가 된 것이라 볼 수 있다.

네 번째 토픽은 과목 확대 개설 및 진로에 따른 선택 수업 운영으로 나타났다. 위 토픽에 포

합된 기사는 공동교육과정운영, 온라인 활용 과목 개설, 진로 맞춤 교육 지원, 진로·학업설계 등에 대한 내용을 담고 있다. 고교학점제 취지에 부합하도록 교과 선택이 가능하게 하기 위해서는 학생들이 진로에 맞게 필요한 과목을 수강할 수 있어야 한다. 하지만 재학 중인 학교 내에 원하는 과목이 개설되지 않을 수 있으므로 인근 고등학교와 공동교육과정을 운영하여 타교에서 수강한 과목 학점을 인정해주거나 온라인 강의를 지원하거나, 지역사회에서 개설한 과목 이수 학점을 인정해줌으로써 교과목 선택권을 보장해주고 있다. 이와 같은 수업 운영 방식은 학생의 입장에서 과목 선택권이 확대됨을 의미하는 동시에 학교 내부적으로 행정업무, 교사 수급, 성적처리 등 여러 면에서 큰 변화가 필요하며 갈등 발생 요소가 될 수 있음을 의미하기에 세밀한 관찰이 필요하다.

다섯 번째 토픽은 혁신적 미래형 학교공간 조성이다. 해당 토픽에 포함된 주요 기사들은 학생 활동 및 교육과정을 고려한 교육공간 설계, 다목적 활용 및 휴식, 교류, 정서함양을 통한 개방공간 등 학점제에 적합한 공간으로의 혁신에 대한 내용을 담고 있다. 고교학점제 운영을 위해서는 학생들의 교실 이동이 용이해야 하고, 수업 간 공강시간에 대비하여 자율학습 등 다양한 활동을 할 수 있는 공간이 필요하다. 또한 진로 활동, 동아리 모임 등을 위한 다용도 공간과 온라인 공동교육과정에 참여할 수 있도록 PC를 비롯한 각종 기자재가 갖추어진 공간 역시 필요하다. 고교학점제 도입을 위해서는 제도가 운영될 수 있는 공간이 갖추어져야 함이 필수적이고 이에 따라 해당 토픽이 도출된 것으로 볼 수 있다. 이운서, 이주연(2022)은 고교학점제의 변화 범위는 고교 교육 전반에 해당하는 것이며 학교 공간 등의 시설 및 인프라 준비가 되어있지 않으면 새로운 변화 이행이 어렵거나 의도한 변화를 제한적으로만 수행하는 결과를 가져올 수 있음을 지적하였다. 학교 공간 조성은 고교학점제 운영의 전제가 될 정도로 중요성을 가지는 것으로 간과하지 않아야 할 것을 의미한다.

토픽의 연도별 등장 비율 변화를 살펴본 결과 몇 가지 시사점을 발견할 수 있었다. 첫 번째, 고교학점제 도입이 진행될수록 과목선택, 이수, 학점취득, 졸업기준 등 고교학점제 운영 방식을 안내하는 성격의 토픽이 꾸준히 증가함을 확인할 수 있다. 고교학점제 전면 도입이 가까워지고 이를 현장에서 체감하면서 해당 제도가 실제로 어떻게 운영되는 것인지 등 관련 방식에 대한 구체적 내용을 알기를 원하는 수요가 있었던 것이라 이해할 수 있다. 더불어 고교학점제의 도입 취지와 구체적인 내용을 알림으로써 해당 정책이 현장에 제대로 자리잡게 하기 위한 정부의 노력이 있음을 예상할 수 있게 하고 동시에 언론이 자신의 역할에 충실히 임하고 있음을 알게 한다. 두 번째, 고교학점제 도입이 진행될수록 이와 관련된 지역의 노력에 대한 기사가 증가하였다. 해가 거듭될수록 실제 추진되는 사업이 늘어나고 사업 운영 현황에 대한 뉴스거리 역시 늘어나는 것이라 이해할 수 있다. 고교학점제가 제대로 정착하고 운영될 수 있게 지역이

많은 노력을 기울이고 있으며 실제 많은 내용이 진행되고 있을 것으로 기대할 수 있다. 지방자치에 영향을 주는 중앙부처의 정책과정에 대한 지자체의 참여 및 협력 보장은 필수적 전제이다(임현, 2020). 중앙정부가 전국을 대상으로 정책을 실시할 경우 대부분 실제 실행은 지방정부를 통하게 되는데, 안중기(2023)의 연구에 의하면 지자체장은 예산집행을 결정 시 대통령과 자신 간 정치적 가치관 일치 여부 및 지방정부의 중앙정부 의존도를 고려한 것으로 나타났다. 따라서 뉴스기사 등을 통한 각 지자체별 고교학점제 도입에 대한 반응 등을 통해 진행 과정의 특성 및 결과의 차이점 등을 파악해볼 수 있을 것이다. 세 번째, 인근 대학, 고교, 지자체, 지역 내 유관기관 등이 연계하는 공동교육과정의 개설 및 운영을 다루는 기사가 늘어남을 확인할 수 있다. 한 학교 내에서 학생의 요구를 수용하는 모든 과목을 개설하는 것이 현실적으로 불가능하기에 지역 내 대학, 고교, 기관 등이 연계하여 해당 과목을 공동으로 개설 및 운영하고 있다. 대학 및 전문기관이 참여하므로 개설되는 과목 내용의 전문성 역시 보장될 것을 기대할 수 있다. 고교학점제 시행을 위해 도입된 새로운 방식인 공동교육과정이라는 제도가 해를 거듭할수록 실제 현장에서 시행되는 비중이 높아지고 있음을 예상할 수 있다. 그러나 공동교육과정 운영에는 공간, 강사, 교육과정 등에 대한 복합적 고려가 요구되며 대입공정성과 관련이 있어 신중해야 한다는 연구(임중현, 우선영, 2022)와 행정업무가중, 교사과업, 강사 채용기준 등에 있어 추가적인 작업이 요구되거나 학생들의 반응임을 언급한 연구(박명희, 2021) 등이 존재한다. 따라서 고교학점제 의도에 부합하는 좋은 제도가 제대로 자리잡을 수 있도록 현장의 목소리를 반영한 제도 개선 및 보완이 필요할 것으로 보인다. 네 번째, 고교학점제와 관련된 대입 전형 기사는 최근 들어 감소하는 추세를 보였다. 2018년 8월 교육부는 대입제도 개편안을 내놓으면서 고교학점제가 2025년 전면 도입될 것임을 밝혔다. 이에 기사들은 고교학점제 시행에 따른 대입제도 변화는 2028년에 적용될 것으로 예상하는 내용을 다루었다. 이후 2019년 9월, 입시 비리 등의 사건과 함께 공정성 문제가 불거지면서 대입 전형이 다시 화두에 오르며 고교학점제가 함께 언급되었고, 10월에는 정시 전형 확대가 발표되면서 관련 기사가 다소 늘어났다. 그러나 2020년 이후로는 관련 기사가 점차 줄어들어 2022년에는 해당 토픽 관련 뉴스기사 비중이 현저히 적음을 확인할 수 있었다. 또한 자사고, 외고, 국제고 등에 대한 일반고로의 전환과 관련된 뉴스기사가 2020년부터 급격히 감소함을 확인할 수 있었다. 고교학점제 도입이 발표될 당시 고교 서열화에 대한 해소 기대가 있었기에 정부 정책과 맞물려 타 유형의 고등학교를 일반고로 전환하려는 움직임이 있었다. 그러나 2018년 서울시 교육청의 자사고 지정 취소에 대한 위법 판결이 났고 2019년에 교육부로 자사고 지정취소가 요청되었는데, 뉴스기사가 이러한 흐름에 따른 내용을 다루고 있음을 알 수 있다. 그러나 2020년 이후 해당 토픽의 기사 비중이 줄어들음을 확인할 수 있다. 즉, 제도 도입 발표 직후에는 대입제도와 고교유형 등의 이슈가 고교

학점제와 함께 묶여 논의되었으나 시간이 흐르고 고교학점제가 자체적으로 힘을 가진 하나의 제도로 자리잡게 되면서 별개의 영역으로 다루어지는 것으로 해석할 수 있다. 종합해보면 고교학점제는 사회적으로 하나의 독립적인 제도로 자리잡고 있는 것으로 보인다.

고교학점제를 주제로 빅데이터 분석을 한 선행연구와 비교했을 때, 본 연구에서의 토픽 분석 결과는 고교학점제 전면 도입을 위한 준비, 정착 조건 및 운영방향 등 고교학점제 도입 기반 마련, 고교학점제 전면 도입 및 시행, 교육과정 정책 추진 등 도입 및 시행, 선택중심 교육과정, 공동교육과정 운영, 학생 선택형 교육과정 등 고교학점제 운영, 진로 맞춤형 교육과정, 다양한 교과목 개설, 선택과목 확대 및 개편, 교육과정 재구조화 연구 등 교육과정 개편, 학생, 학부모, 교원 인식 및 문화, 정책 이해도 향상 방안 등 학교 문화, 수업공간, 시스템 지원 등 환경 시설, 대입전형 및 대입 제도, 대입 준비, 평가제도 등 대입으로 나뉘는 선행연구(김소영, 홍후조, 2021; 김태현, 김석우, 2019; 박지혜, 2020; 전은정 최윤정, 2023; 조성범, 신하영, 2022) 카테고리 빠짐없이 모두 포함하는 것으로 나타났다.

한편, 선행연구와 일치하는 공통 주제 외에 ‘교육감 후보 고교학점제 관련 공약’, ‘외고, 국제고, 자사고 등 일반고 전환 추진 및 지청 취소 위법 판결’, ‘학령인구 감소 및 중등교원 양성’ 등의 사회적 이슈가 되는 토픽이 추가로 분석되었음을 확인할 수 있었다. 이는 본 연구가 빅데이터를 활용한 고교학점제 관련 선행연구들과는 차별점을 갖는 부분이라 할 수 있다. 먼저, 교육감 후보의 고교학점제 공약 및 자사고, 외고, 국제고 등 타 유형 고교의 일반고로 전환 등의 토픽은 진영싸움이 될 소지가 있음을 의미할 뿐 아니라 제도 도입에 반대 세력이 적지 않을 수 있음을 의미한다. 다만, 자사고 등의 일반고 전환 관련 토픽은 연도별 변화 분석을 통해 최근 그 비중이 낮아졌음을 확인하였다. 고교학점제 정착에 미칠 영향력이 줄어들었을 것임을 예상할 수 있으나 그럼에도 지속적으로 관심을 두어야 할 것이다. 다음으로, 학령인구 감소 및 중등교원 양성의 토픽은 각각의 영역으로 나누어 살펴볼 필요가 있다. 학령인구 감소에 따라 개별 학생 한 명씩에 집중하는 교육 서비스 제공이 필요해진 사회적 변화를 수용할 수 있는 제도로 고교학점제가 활용될 수 있음을 보여주고 있다. 중등교원 양성의 경우 두 가지 시각에서 바라볼 수 있다. 고교학점제 시행으로 학생 감소에 따른 여유 교사 활용이 가능할 것으로 보는 입장과, 필요한 중등교원의 수가 더욱 늘어나 증원이 필요하다는 입장이 맞물려 있는 것이다. 이러한 결과를 통해 관계자들은 고교학점제 도입에 있어 더욱 폭넓은 시각으로 접근해야 함을 알 수 있다. 단순히 제도의 운영에만 집중하는 등 협소한 시각으로 접근하는 것은 연쇄적으로 엮여 있는 다른 문제가 발생하는 것을 간과하게 할 우려가 있다. 따라서 고교학점제 시행을 위해 전제되어야 할 조건들과 이러한 조건들이 정치적으로 갖는 의미, 고교학점제 시행을 통해 득실이 나뉘는 집단의 목소리, 사회환경의 변화 등을 모두 함께 고려해야 한다. 이를 통해 필요한



제반 여건을 탄탄히 해야 성공적인 고교학점제 도입 및 정착을 가져올 수 있다.

연구의 한계 및 후속 연구를 위한 제언은 다음과 같다.

첫 번째는 뉴스기사 제공 언론사에 관한 것이다. 본 연구는 고교학점제 도입이 발표된 2017년 11월 27일 이후 기사를 연구 대상으로 설정하였다. 빅카인즈는 54개 언론사의 뉴스 기사를 수록하고 있으며 그 중 53개 언론사는 본 연구에서 수집하고자 설정한 기간인 2017년 11월 27일 이전부터 빅카인즈에 기사를 제공하고 있다. 그러나 조선일보의 경우 2018년 1월 1일 이후로 기사를 제공하고 있어 2017년 11월 27일부터 12월 31일까지의 조선일보 기사는 본 연구 대상에 포함되지 못하였다는 한계점이 있다. 또한 전국일간지와 지역일간지를 구분하지 않고 모두 포함하여 지역 사업에 초점을 맞춘 뉴스기사 비중이 높게 나타났다는 한계점이 있다. 따라서 후속 연구에서는 2018년 이전 기사를 제공하지 않는 조선일보 기사는 개별 수집하여 연구 대상에 포함하고 전국일간지와 지역일간지를 나누어 분석한다면 또 다른 시사점을 얻을 수 있을 것이라 기대된다.

두 번째는 토픽명 부여에 관한 것이다. 본 연구에서는 연구자가 분석 결과 얻은  $\beta$ 와  $\gamma$ 값을 근거로 각 토픽의 토픽명을 부여하였다. 토픽별 단어 및 문서의 가중치를 고려하여 신중하게 부여한 토픽명이기는 하나, 전문가 의견을 반영한 타당화 과정이 생략되었다는 한계가 존재한다. 따라서 후속 연구에서는 토픽명에 대한 타당성을 더욱 확보하기 위해 연구 대상이 되는 문서 내용의 전문가 의견을 수렴하여 반영할 수 있을 것이다.

## 참고문헌

- 교육부. (2017). 고교학점제 추진 방향 및 연구학교 운영계획. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=C&s=moe&m=020402&opType=N&boardSeq=72593>
- 교육부. (2020). 고교학점제 연구학교 운영안내서(연구자료 ORM 2020-22). 교육부.
- 교육부. (2021a). 포용과 성장의 고교교육 구현을 위한 고교학점제 종합 추진계획. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=83590&lev=0>
- 교육부. (2021b). 2025년 고교학점제 전면 적용을 위한 단계적 이행 계획(안)(2022-2024). <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=88188>
- 교육부. (2022). 나의 진로와 적성에 맞는 수업, 학교 간 공동교육과정에서 찾았어요. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=93470>
- 권나현, 오종민, 하은희. (2022). 텍스트 마이닝을 활용한 COVID-19에 대한 대중의 관심 주제와 정서 분석. *The Ewha Medical Journal*, 45(2), 46-54.

- 권순보. (2018). 빅데이터 분석 기법을 활용한 수능 절대평가 여론 텍스트데이터 분석. [석사학위논문, 한국교원대학교 교육대학원]
- 권충훈. (2020). 고교학점제 관련 언론사 기사들의 주요 변화 분석. 韓國컴퓨터情報學會論文誌, 25(7), 183-191.
- 김륜희, 지은림. (2022). 고교학점제와 연계한 대입 제도 방안에 관한 키워드 및 토픽 분석. 교육연구논총, 43(4), 77-106.
- 김소영, 홍후조. (2021). 빅 데이터를 통한 고교학점제의 대중인식 분석. 교육방법연구, 33(1), 179-203.
- 김태현, 김석우. (2019). 고교학점제 도입에 대한 사회적 인식의 빅데이터 활용 특징적 요소 분석. 교육혁신연구, 29(2), 225-242.
- 류병준, 하용훈. (2022). 토픽모델링 이용한 한·미·일·중 해양무인체계 연구 동향 분석. 한국산학기술학회 논문지, 23(11), 395-403.
- 박명희. (2021). 공동교육과정 운영 실태 및 개선안:대구지역 고등학교를 중심으로. 교육과정평가연구, 24(4), 31-58.
- 박재연. (2023. 2. 22). 국민 54.6% “고교학점제 도입 반대”. 뉴스토마토. <http://www.newstomato.com/ReadNews.aspx?no=1178484&inflow=N>
- 박지한, 김아영. (2021). 고교학점제 안착을 위한 부분도입기의 과제: 고교학점제 1차 연구학교 운영 결과에 대한 텍스트 분석을 중심으로. 학습자중심교과교육연구, 21(24), 309-331.
- 박지혜. (2020). 학교 간 공동교육과정에 관한 사회적 인식 연구: 빅데이터 분석을 활용하여. 교육문화연구, 26(5), 85-104.
- 백영민. (2019). R을 이용한 텍스트마이닝. 한울아카데미.
- 서봉언. (2020). 고교학점제 도입과 학교의 변화. 교육사회학연구, 30(3), 55-79.
- 서추연. (2022). 텍스트 마이닝을 활용한 고교학점제의 사회적 인식 변화 분석. 학습자중심교과교육연구, 22(21), 111-127.
- 설양환. (2017). 국내 신문 기사에 나타난 영재교육 및 영재에 대한 담론 분석. [박사학위논문, 건국대학교 대학원]
- 손태권, 황성환. (2020). 토픽모델링을 활용한 국내외 수학교육 평가 연구 동향 분석. 수학교육학연구, 30(4), 601-624.
- 신윤범, 원효현. (2020). 고교학점제에 대한 교사 인식의 현상학적 분석. 水産海洋教育研究, 32(1), 109-117.
- 신철균, 주주자. (2021). 교육지원청 수준에서의 고교학점제 운영 사례 : 학생과 교사의 인식을 중심으로. 열린교육연구, 29(2), 139-168.
- 안주영, 안규빈, 송민. (2016). 텍스트 마이닝을 이용한 매체별 에볼라 주제 분석: 바이오 분야 연구 논문과 뉴스 텍스트 데이터를 이용하여. 한국문헌정보학회지, 50(2), 289-307.
- 안중기. (2023). 지방정부의 중앙정부에 대한 협력요인 연구: 중앙정부의 예산집행률 제고 정책을 중심으로. 行政論叢, 61(2), 73-97.
- 유주영. (2022). 학생인권에 관한 사회적 이슈 분석: 토픽모델링을 활용한 국내 언론 보도기사 분석을 중심으로. 教育學研究, 60(1), 205-234.
- 윤정길. (2000). 관리와 PR. 대영문화사
- 이대영, 이현숙. (2021). LDA 토픽 모델링의 적정 토픽 수 결정 방법 탐색 : 혼잡도와 조화평균법 활용을 중심으로. 교육평가연구, 34(1), 1-30.

- 이민철, 김혜진. (2018). 텍스트 마이닝 기법을 적용한 뉴스 데이터에서의 사건 네트워크 구축. *지능정보연구*, 24(1), 183-203.
- 이상은, 장덕호. (2019). 고교학점제 학생 선택형 교육과정 편성 과정에 나타난 쟁점과 과제. *학습자 중심교과교육연구*, 19(14), 109-136.
- 이소정. (2020). 잠재디리클레할당 분석을 이용한 ‘노인일자리’ 관련 신문기사 토픽분석. *디지털융복합연구*, 18(10), 537-546.
- 이서진, 민경석. (2022). 텍스트 자료에 대한 네트워크 분석과 토픽 모델링 : 대학 역량기반 교육을 중심으로. *교육평가연구*, 35(1), 165-188.
- 이성기. (2021. 7. 22). 충북 고교 교사 66.2% 고교학점제 반대…“원점 재검토 필요”. *뉴스1*. <https://www.news1.kr/articles/?4380270>
- 이윤서, 이주연. (2022). 고교학점제 도입에 따른 학교공간 재구조화의 방향. *청소년시설환경*, 20(2), 37-45.
- 이정. (2022). 대선 관련 인터넷 뉴스의 댓글과 대댓글 간 비교를 통해 살펴본 온라인 토론의 진행 가능성. *지능정보연구*, 28(2), 33-55.
- 이종억. (2021. 7. 27). 고교학점제 놓고 학부모단체·전교조 이견. *충북일보*. <https://www.inews365.com/news/article.html?no=676064>
- 임종현, 우선영. (2022). 일반고 교육과정 다양화를 위한 학교 밖 교육 운영 방향 탐색. *미래교육학연구*, 35(3), 51-74.
- 임현. (2020). 중앙정부와 지방자치단체 간 협력제도의 법적 과제. *지방자치법연구*, 20(4), 295-317.
- 장성환. (2021. 2. 17). 교사도 없는데… ‘고교학점제’ 2025년 전면 도입키로. *뉴데일리*. <https://www.newdaily.co.kr/site/data/html/2021/02/17/2021021700179.html>
- 장창곡. (2021). IPA 기법을 활용한 고교학점제 연구학교 운영에 관한 중요도와 실행도 분석. *학습자 중심교과교육연구*, 21(8), 103-121.
- 전은정, 최윤정. (2023). 고교학점제 관련 논문 토픽 분포 분석. *교육발전*, 42(3), 895-915.
- 정윤경, 류지은, 안유진, 박초롱. (2021). 고교학점제 도입에 따른 고등학교 진로교육의 현황과 과제. *농업교육과 인적자원개발*, 53(1), 105-127.
- 조경원, 우영운. (2019). 텍스트 마이닝을 이용한 4 차 산업 연구 동향 토픽 모델링. *한국정보통신학회 논문지*, 23(7), 764-770.
- 조서연, 임다미, 함은혜. (2021). 원격수업 운영을 위한 교사 지원 요구 분석. *한국교원교육연구*, 38(3), 171-194.
- 조성범, 신하영. (2022). 고교학점제 연구 경향 분석. *문화와 융합*, 44(3), 149-168.
- 조성은, 송철중, 노범래, 민기채, 이철수, 허신행, 김예슬. (2018). 남북 보건복지의 공적 의제 분석과 전략 개발에 관한 연구(연구보고서(수시) 2018-05). *한국보건사회연구원*.
- 조효제, 길혜지. (2022). 텍스트 마이닝을 활용한 고교학점제 쟁점 분석. *교육연구논총*, 43(2), 299-328.
- 주보혜, 임정미, 노범래, 신영규, 임새아, 정희선. (2019). 국제기구와 주요 선형 국가의 고령화 대응 정책 분석(연구보고서 2019-28). *한국보건사회연구원*.
- 최정희. (2019). 문·이과 통합 및 고교학점제에 대한 교사들의 인식과 대학 교육에 주는 시사점: 연구학교 및 선도학교를 중심으로. *교육문화연구*, 25(1), 421-447.
- 한국교육과정평가원. (2020). 고교학점제 연구학교 운영 현황 분석(연구보고 CRC 2020-11). *한국교육과정평가원*.

- 한국교육과정평가원. (2023). 2022학년도 고교학점제 성과발표회 자료집(연구자료 ORM 2023-7). 한국교육과정평가원.
- 한유경. (2022. 8. 11). 교육 정책 혼돈 속 고교학점제 ‘성취평가제’, ‘미이수제’ 보완 · 지적사항 쏟아져. 충청투데이. <https://www.cctoday.co.kr/news/articleView.html?idxno=2166330>
- 한지영, 허고은. (2021). Analyzing Students' Non-face-to-face Course Evaluation by Topic Modeling and Developing Deep Learning-based Classification Model. *한국문헌정보학회지*, 55(4), 267-291.
- 홍준우, 홍태호. (2021). 코로나 19 팬데믹 상황에서 감성분석을 이용한 미국, 중국, 한국 여행자의 온라인 리뷰 비교 분석. *한국 IT 서비스학회지*, 20(5), 159-176.
- Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Hyderabad, India, June 21-24, 2010. Proceedings. Part I 14* (pp. 391-402). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-13657-3\\_43](https://doi.org/10.1007/978-3-642-13657-3_43)
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022. <https://doi.org/10.1162/jmlr.2003.3.4.-5.993>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. <https://doi.org/10.1145/2133806.2133826>
- Buhin Pandur, M., Dobša, J., Beliga, S., & Meštrović, A. (2021). Topic modelling and sentiment analysis of COVID-19 related news on Croatian Internet portal. In *Proceeding of Conference on Data Mining and Data Warehouses 2021*. Slovenian KDD Conference on Data Mining and Data Warehouses. <https://ailab.ijs.si/dunja/SiKDD2021/Papers/Panduretal.pdf>
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing* 72(7-9), 1775-1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numerique* 17(1), 61-84. <https://doi.org/10.3166/DN.17.1.61-84>
- Emidi, C., & Galán, S. (2022). *Prospectus Content as Predictor of IPO Outcome: A topic model approach* [Master's thesis, Lund University]. <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9083567&fileId=9083574>
- Fyffe, M. E. (2009). *Post-processing of LDA for cluster quality analysis* [Doctoral dissertation, Rensselaer Polytechnic Institute]. [https://dspace.rpi.edu/bitstream/handle/20.500.13015/906/16739\\_rpithes-short-cleaned.pdf?sequence=3](https://dspace.rpi.edu/bitstream/handle/20.500.13015/906/16739_rpithes-short-cleaned.pdf?sequence=3)
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl\_1), 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Grun, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, 40(13), 1-30. <https://doi.org/10.18637/jss.v040.i13>
- Hall, D., D. Jurafsky, & C. Manning. (2008, October). Studying the history of ideas using topic models. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 363-371). <https://aclanthology.org/D08-1038.pdf>

- Jeon, H., & Kim, T. (2016). KoNLP: Korean NLP package. R package version 0.80.1. <https://CRAN.R-project.org/package=KoNLP>
- Nikita, M. (2016). ldatuning: Tuning of the latent dirichlet allocation models parameters. R package version 1.0.2. URL <https://CRAN.R-project.org/package=ldatuning>.
- Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, 1(3), 37. <https://doi.org/10.21105/joss.00037>
- Wickham, H. (2015). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.0.0.9000. <https://CRAN.R-project.org/package=stringr>
- Wickham, H., Francois, R., Henry, L., & Muller, K. (2015). dplyr: A grammar of data manipulation. R package version 1.1.2. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Chang, W., Wickham, M. H., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2016). ggplot2: Create elegant data visualisations using the grammar of graphics. R package version 3.4.3.
- Wickham, H. (2017). reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package. R package version 1.4.3. <https://cran.r-hub.io/web/packages/reshape2/reshape2.pdf>

## ABSTRACT

### Topic analysis of news articles related to high school credit system using LDA topic modeling

Eunjeong Jeon (Ewha Womans University, Lecturer, Seoul)  
Youn-Jeng Choi (Ewha Womans University, Associate Professor, Seoul)<sup>†</sup>

This study aimed to provide policy implications by analyzing topics related to the high school credit system covered by news articles before the full introduction of the high school credit system. The R program(version 4.2.2) was used to analyze the Latent Dirichlet Allocation(LDA) topic modeling of 7,115 news articles related to the high school credit system from the announcement of its introduction by the Ministry of Education(November 27, 2017) to 2022. Changes were examined by year. As a result of the analysis, a total of 16 topics were derived. In addition to common themes consistent with previous studies using big data: pledges related to the high school credit system by candidates for superintendent of education; promotion of conversion to general high schools such as foreign language high schools, international high schools, and private high schools and cancellation of designated illegal judgment; School-age population decline; and secondary teachers' topics were further analyzed. In addition, through the change in proportion of topics by year, it was revealed that the actual operation method and status of the high school credit system is on the rise, and topics such as the college entrance system and type of high school are on the decline. Through these results, it can be confirmed that the introduction and settlement of the high school credit system are actively progressing and that it is in the process of establishing itself as an independent system. In addition, in the process of introducing and settling the high school credit system, various perspectives of various stakeholders should be comprehensively considered, and follow-up studies are suggested.

**Key words** : high school credit system, credit system, LDA, topic modeling, text mining

---

<sup>†</sup>Corresponding author: Youn-Jeng Choi, Associate Professor, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, younjengchoi@ewha.ac.kr