

이슈페이퍼 2017-14

누리과정 정책수요에 대한 빅데이터 분석: 워드클라우드와 의미망 분석을 중심으로

박 창 현 | 부연구위원

1. 연구의 필요성 및 목적
2. 빅데이터를 활용한 여론 분석 방법
3. 워드 클라우드와 의미망 분석 결과
4. 연구의 시사점 및 제언

누리과정 정책수요에 대한 빅데이터 분석 : 워드클라우드와 의미망 분석을 중심으로*

박 창 현 | 부연구위원

요 약

- 본 연구의 목적은 빅데이터 기법을 활용하여 미디어상에 나타난 국민들의 여론을 분석하여 누리과정 정책의 수요와 인식을 파악하는 것임.
- 2013년 1월부터 2016년 6월까지 누리과정 관련 온라인 뉴스를 추출하였다. 네이버 검색 엔진을 통해 ‘누리과정’, ‘유아교육’, ‘보육’, ‘보육과정’, ‘유아교육과정’으로 검색된 전체 수집 기사는 총 58,441개였으며, 17,651건의 유효 기사를 중심으로 워드 클라우드 분석과 의미망 분석을 실시하였음.
- 워드 클라우드 분석 결과, 누리과정과 관련하여 높은 빈도를 보인 키워드는 누리과정 예산, 정부, 지원, 보육교사, 교육청, 편성, 교육감, 유아, 프로그램 순으로 나타났음.
- 의미망 분석 결과, 누리과정, 누리과정 예산, 경제부총리가 중심성이 높은 노드로 나타났으며, 예산편성과 시도교육청은 누리과정과 누리과정 예산을 매개하는 문지기(gatekeeper) 또는 브로커(broker) 역할을 하였음.
- 빅데이터에 나타난 국민들의 요구는 누리과정으로 이미지화 되는 유아교육과 보육 서비스의 전반의 질 관리와 공공성 강화, 누리과정 예산의 안정성으로 파악되었음.

1 연구의 필요성 및 목적

- 우리나라는 2011년 5세 누리과정과 2013년 3-5세 연령별 누리과정 도입하였으며, 누리과정 정책을 실행한지 6년차에 접어들고 있음. 그간 누리과정 정책은 한국의 유아교육과 보육 현장에 상당한 양적, 질적 변화를 이끌어왔음.

* 본 원고는 육아정책연구소 2016년 일반과제인 박창현, 양미선, 조혜주(2016). 누리과정의 질 제고를 위한 운영 다양화 방안 보고서를 중심으로 재구성하였음.

- 이윤진, 김지현, 이민경(2016)의 연구에서는 무상교육비의 보편적 지원, 공통교육과정을 통한 취약계층 유아발달 지원, 교사의 질적 수준 향상 등의 긍정적 성과를 제시하였음. 이진화, 박창현, 윤지연(2015)는 원장, 교사, 행정가들의 의견을 중심으로 초등이전 공교육으로서의 유아교육 대국민 이미지 상승, 유아교육과 보육기회의 균등, 연령별 교육·보육 과정 체계적 운영이 가능한 점을 긍정적 성과로 보았음.
- 부정적 성과로는 학부모의 경제적 부담 경감이 미흡한 점과 누리과정 재정의 불안정성(이윤진, 김지현, 이민경, 2016), 교사의 업무 증가(이진화, 박창현, 윤지연, 2015), 3-5세 교육·보육과정의 획일화(박창현, 양미선, 조혜주, 2016; 이진화, 박창현, 윤지연, 2015), 교사들의 학급 교육·보육과정 운영 자율성 약화(박창현, 양미선, 조혜주, 2016) 등으로 요약됨.
- 최근 연구의 동향을 살펴보면 빅데이터 기법을 활용하여 국민들의 여론과 동향을 파악하여 정책 개발과 예측에 활용하는 연구들이 늘고 있음.
 - 향후 누리과정의 정책의 방향을 설정하기 위해 누리과정 정책에 대한 효과와 방향성 검토는 기본적으로 정책 수요자인 부모와 유아의 요구를 파악하는 것에서 시작할 필요가 있음.
 - 누리과정 정책 분석을 위해 국민들의 인식과 요구를 파악하고, 국민들의 인식의 지형도를 그려보는 일은 정책개발의 기초가 될 수 있음.
- 본 연구에서는 언론에 나타난 누리과정에 대한 주요 논점을 파악하고, 국민들의 인식을 파악하기 위해 국내 온라인 뉴스를 추출하여 빅데이터 기법을 활용하여 미디어상에 나타난 네티즌들의 의견을 분석하여 누리과정 정책 수요를 다각적으로 파악하고자 하였음. 본 연구에서는 워드 클라우드와 의미망 분석에 나타난 네티즌의 누리과정에 대한 인식의 지형도를 파악하는데 중점을 두었음.

2 빅데이터를 활용한 여론 분석 방법

가. 분석 대상

- 분석대상은 누리과정 키워드로 네이버 검색 엔진을 통해 검색된 온라인 뉴스와 기사 댓글이며, 2013년 1월 1일부터 2016년 6월 30일 사이의 기사로 한정하였음.

나. 자료수집방법

- 빅데이터 분석 전문기관이 자체 개발한 Crawler를 활용하여 분석을 위한 데이터를 수집하였으며, 네이버 뉴스 검색엔진¹⁾에서 검색어로 ‘누리과정’, ‘유아교육’, ‘보육’, ‘보육과정’, ‘유아교육과정’으로 검색된 전체 수집 기사는 총 58,441개였으며, 삭제기사는 40,789개였음.
 - 유효기사의 수는 17,651건으로 유효기사의 비율은 전체의 30.2%였음.
 - 네이버 뉴스에서 누리과정을 다룬 언론사는 일간지 20개, 방송통신 32개, 경제 및 IT 72개, 인터넷 신문사 51개, 지역지 67개, 전문지 81개였다. 총 기사수는 일간지 1,859개, 방송통신 5,893개, 경제 및 IT 뉴스 3,619개, 인터넷신문 1,457개, 지역지 3,724개, 전문지 1,099개였음.
 - 삭제 키워드는 홍보성 기사와 키워드에 의해 자동 수집된 뉴스들 중에서 신문사 이름이나, 실명이 거론된 인명, ‘별누리교실’ 등과 같이 조사 목적에 부합하지 않는 키워드는 삭제하였음.

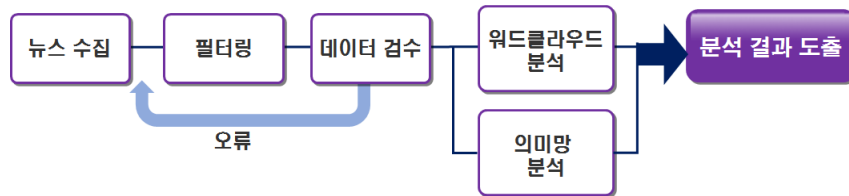
〈표 1〉 전체 수집글과 유효 게시글 수

전체 수집글	삭제글	유효 게시글	유효율
58,441	40,789	17,652	30.2%

1) 네이버 검색엔진은 타 검색엔진에 비해 해당기간동안 평균 점유비율이 82.63%로 가장 높게 나타나 대표적인 포털 검색엔진으로 선정하였음. 인터넷트렌드(<http://www.internettrend.co.kr>)에서 2013년 5월 1일부터 2016년 5월 26일까지 검색엔진 점유비율을 검색한 결과 네이버 82.63%, 다음 13.58%, 구글 1.85%순으로 나타났다.

다. 분석 절차 및 방법

- 최종적으로 필터링된 데이터를 대상으로 연도별 워드클라우드 분석과 의미망 분석을 실시하였음.
 - 워드 클라우드 분석은 화제어²⁾ 추출 솔루션을 통해 추출된 화제어를 분석하여 산출된 빈도 값의 크기에 따라 워드 클라우드를 구성하였음. 이는 단어의 크기에 따라 키워드³⁾의 빈도수를 가시성 있게 분석하는 기법임.
 - 본 연구에서는 분석의 유의미성을 고려하여, 연구진과 빅데이터 분석가의 협의 하에 화제어를 분석 키워드 빈도수를 최대 200순위까지 수집하였음.
 - 의미망 분석은 필터링 된 데이터를 대상으로 넷마이너를 활용하여 사회연결망 분석(social network analysis)를 실시하였음.
- 조사의 정확도를 향상시키기 위해 수집된 데이터 중 본 조사와 관련이 없는 불필요한 데이터에 대한 1차 필터링(filtering)을 진행하였음.
 - 홍보성, 광고성, 사행성 관련 글에 공통적으로 포함된 키워드를 일괄 삭제한 후 남은 글들을 모두 확인하여 삭제작업을 실시하였으며, 데이터의 정확도를 높이기 위해 1차적으로 필터링 된 데이터를 검수하고 재필터링 작업을 실시하였음.
 - 분석절차는 다음과 같음.



[그림 1] 빅데이터 분석절차

2) 화제어는 게시글 내 키워드들을 분석 전문기관의 내부 추출 알고리즘에 따라 빈도 및 가중치를 적용하여 산출된 키워드를 의미함. 키워드는 게시글 내 문단 및 문장을 구성하고 있는 명사, 형용사, 용언 등으로 나열된 단어들을 의미함. 본 연구에서는 화제어를 최대 200순위까지 수집하였으며, 1순위인 '누리과정'은 빈도수가 4515회였으며, 200순위인 '우려하다'는 118회의 빈도를 나타냈음. 자료의 유의미성과 가독성을 고려하여 빈도수 1~200순위까지를 기술하였음.

3 워드 클라우드와 의미망 분석 결과

가. 워드 클라우드 분석에 나타난 네티즌의 누리과정에 대한 인식 (2013년 1월-2016년 6월)

- 2013년 1월 1일에서 2016년 6월 30일까지 높은 등장빈도를 보인 키워드는 ‘누리과정’, ‘예산’, ‘누리과정예산’, ‘정부’, ‘지원’, ‘보육교사’, ‘교육청’, ‘편성’, ‘새누리당’, ‘교육감’, ‘유아’, ‘프로그램’ 순으로 키워드가 나타났음.

 - ‘누리과정’은 전체기간 동안 2014년을 제외하고 가장 높은 등장빈도를 나타낸 키워드였음.
 - 워드 클라우드 분석에서 예산과 정부, 지원, 교육청, 편성에 관한 키워드의 빈도가 높은 것은 누리과정 예산 갈등의 문제를 반영하고 있음.
 - 누리과정 정책은 이명박, 박근혜 정권 시기에 실행된 주된 육아정책이므로 당시 여당이었던 ‘새누리당’의 역할과 입지가 영향을 미친 것으로 파악됨. 반면, 유아와 프로그램이 상대적으로 낮았고, 국민들의 인식 속에 누리과정은 ‘교육과정’보다 ‘프로그램’으로 인식하고 있는 것으로 파악됨.
 - 누리과정을 검색하면 나오는 ‘누리놀이’와 같은 수많은 사교육 프로그램들의 영향, 정부의 누리과정 대국민 홍보 미약, 국가수준의 유아교육과정에 관한 관심 부족 등의 영향이 있었을 것으로 추정해볼 수 있음.
 - 보육교사에 대한 관심이 높은 것은 아동학대사건과 연결되면서 교사의 인성과 자격에 대한 국민들의 우려가 높기 때문인 것으로 분석됨(박창현, 양미선, 조혜주, 2016).
- 2013년도에는 누리과정과 관련하여 ‘강조하다’, ‘확대하다’, ‘지급하다’, ‘학부모’의 키워드들도 나타났음.

 - 이는 5세 누리과정 3-4세로 확대되고, 새정부가 적극적으로 추진하려는 의지를 보인 시기였기 때문에 누리과정에 대한 기대가 담겨있는 키워드로 볼 수 있음.
- 2014년에는 새누리당, 여야, 새정치민주연합이라는 키워드와 합의라는 키워드들도 나타났음.

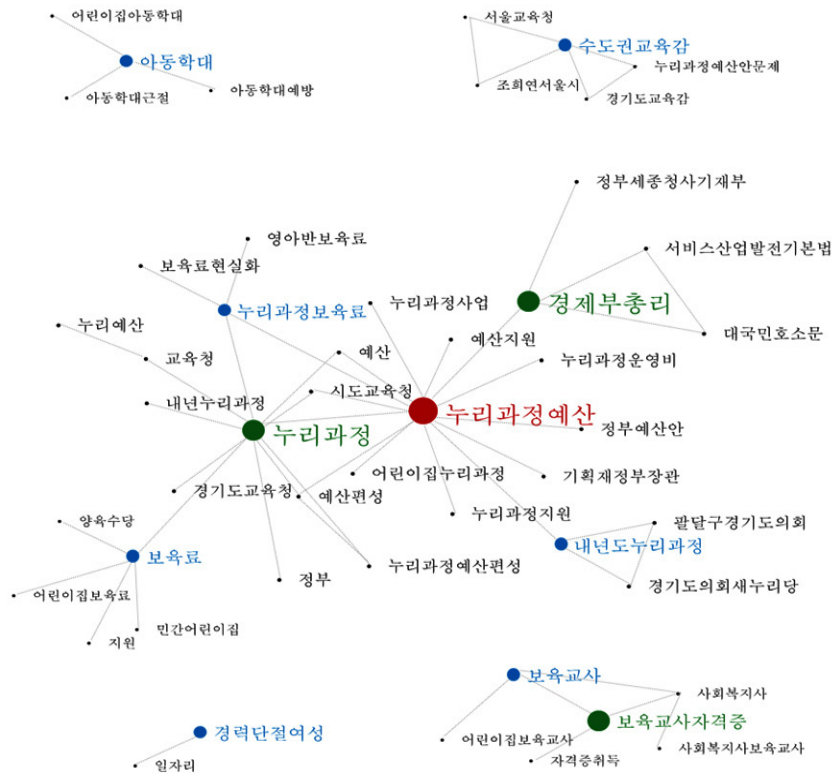
 - 이는 예산안 처리와 밀접한 관련이 있는 것으로 분석되었으며, 보육교사 키워드도 지속적으로 높게 나타났음.

- | 순위 | 화제어 | 빈도 |
|----|--------|------|
| 1 | 누리과정 | 4515 |
| 2 | 예산 | 3528 |
| 3 | 누리과정예산 | 1757 |
| 4 | 정부 | 1567 |
| 5 | 편성하다 | 1541 |
| 6 | 지원 | 1429 |
| 7 | 보육교사 | 1222 |
| 8 | 교육청 | 1104 |
| 9 | 편성 | 1040 |
| 10 | 서울 | 990 |
| 11 | 강조하다 | 963 |
| 12 | 새누리당 | 939 |
| 13 | 교육 | 932 |
| 14 | 교육감 | 926 |
| 15 | 발표하다 | 887 |
| 16 | 교육부 | 841 |
| 17 | 도교육청 | 781 |
| 18 | 유아 | 700 |
| 19 | 실시하다 | 678 |
| 20 | 프로그램 | 629 |

〔그림 2〕. 〈표 2〉 워드 클라우드 검색 빈도 분석 결과 (2013년 1월-2016년 6월)

나. 의미망 분석에 나타난 네티즌의 누리과정 인식의 지형도 (2013년 1월-2016년 6월)

- 누리과정, 누리과정 예산, 경제부총리가 중심성이 높은 노드로 등장하였음. 누리과정이 연구의 핵심 주제라는 점에서 누리과정 자체에 대한 노드의 중심성이 높게 나타났음.
- 중심성이 높은 누리과정, 누리과정 예산, 경제부총리 노드별로 다른 특성을 보이고 있음.
 - 누리과정과 누리과정 예산은 약한 연대(weak tie)⁴⁾를 하는 반면, 경제부총리는 누리과정이나 누리과정 예산보다는 강한 연대(strong tie) 형태의 네트워크 특성을 나타냈음.



[그림 3] 누리과정에 대한 의미망 분석 결과(2013-2016 전체년도)

4) 약한 연대는 중심성 높은 노드가 다른 노드와 1 대 1의 연결망을 주로 맺고 있는 형태로, 방사형의 모습을 보인다. 약한 연대를 갖는 경우 중심성 높은 노드의 연결망 빈도는 높지만, 다른 노드는 중심성 높은 노드와의 연결선만 보유하고 다른 노드와의 연결선은 매우 낮은 빈도를 갖는다. 약한 연대의 네트워크망은 다양한 가치와 정보를 모으거나 연결시키기 위한 장점을 갖는다(Gronovetter, 1973; Scott, 2000).

- 누리과정과 누리과정 예산이 약한 연대의 네트워크를 갖는다는 것은 개방성이 높은 네트워크를 갖고 있다는 것을 의미하며, 제한된 정보나 가치로 한정되기 보다는 다양한 가치와 정보 그리고 가치와 정보를 갖고 있거나 빅데이터 담론에 제공하는 집단이나 개인과의 개방적 연계를 가질 수 있는 것으로 해석됨.
- 지방자치단체, 교육과정, 보육비 등은 누리과정 및 누리과정 예산과 연결되어 개방적 네트워크를 구성하는 주요 노드였음.
- 반면, 부총리로 대표되는 강한 연대의 네트워크망은 주로 보육료, 중앙정부 부처 및 관계자, 지방자치단체 간의 그물형의 네트워크를 구성하여 상대적으로 폐쇄적인 구조를 보이고 있음.
 - 이런 구조는 정부부처에서의 누리과정에 대한 접근방식이 지자체 관련 부처 간의 논의를 상대적으로 개방성이 부족하게 또는 누리과정과 관련된 다양한 주제를 반영하지 못하는 방식으로 진행되고 있음을 시사함.
- 누리과정에 대한 노드는 누리과정 예산을 연결하고, 지체와 보육료를 연결하는 ‘문지기(gatekeeper)’나 브로커(broker) 역할을 하는 것으로 나타났음. 이렇게 중심성이 높은 2개의 노드를 연결하는 매개노드는 네트워크 연구에서 ‘문지기(gatekeeper)’나 브로커(broker)로 개념화됨(김용학, 2007, 현영섭, 신은경, 이향란, 2011; Kalish, 2008).
- 특히 누리과정과 어린이집 누리과정이 다르게 개념화되고 있는 것으로 나타났는데, 이는 어린이집 누리과정 예산 편성이 보육대란과 연결되면서 이슈화되었기 때문인 것으로 해석할 수 있음.
- 누리과정을 둘러싼 중앙정부와 시도교육감의 정치적 갈등은 어린이집 누리과정비 지원을 불투명하게 만들었고, 국민들의 인식 속에는 어린이집 누리과정을 분리해서 생각하게 만드는 원인으로 작용했을 가능성이 있음.
- 이는 누리과정이 애초 기대했던 취지와 목표인 유치원과 어린이집에 다니는 3-5세 모든 유아에게 평등한 교육·보육기회를 제공하자는 원칙에(교육과학기술부, 보건복지부, 2013) 어긋나는 결과이므로 추후 누리과정 정책 입안과 추진과정에서 증세나 국비지원 등을 통해 안정적이고 지속가능한 누리과정 정책을 마련해 나갈 필요가 있을 것임(박창현, 2016).
- 네트워크 분석에서 중심부와 주변부로 확연하게 구분되며, 주변부는 독립적 네트워크 구조를 갖고 있는 것으로 나타났음. 이런 구조는 결국 누리과정과 관련된 핵심적 화두와 새로운 화두 또는 다양한 화두 간의 연계성이 부족하다는 점을 보여주고 있음.

- 네트워크 분석에서 중심부를 보면 누리과정, 누리과정 예산, 경제부총리의 노드를 중심으로 하는 의미망이 중앙에 위치하고 있으며, 주변부에는 지방자치단체, 보육교사자격증, 아동학대, 경력단절여성, 교육감 등에 대한 네트워크가 존재하였음.
- 지방자치단체와 교육감 등은 지방자치단체 수준에서 누리과정 예산의 필요성에 대한 입장을 표명하거나, 지역의 정치세력에 따라 갈등되는 양상을 보여주는 노드였음.
 - 광역시·도 단위의 교육감이나 지방자치단체의 갈등의 담론이 등장하였으며, 특히 수도권 지방자치단체의 경우 예산안에 대한 적극적인 담론 형성과 관련되어 있는 것으로 나타났다.
 - 수도권 지방자치단체는 네트워크 분석의 하단에 별도의 독립된 방사형 네트워크를 구축할 정도로 독립성을 갖고 있으면서 동시에 예산과 관련된 담론을 형성하고 있음. 따라서 지방자치단체는 누리과정의 운영에서 지배적인 영향을 갖고 있으며, 그 지배력이 어린이집이나 유치원 단위의 누리과정에도 영향을 줄 수 있다고 판단됨.
 - 보육교사자격증의 경우 사회복지 분야 자격 등과 관련하여 보육교사의 자격취득과정, 보육교사자격증 취득 조건 등에 대한 담론이 관련되었음. 누리과정은 어린이집이나 유치원 대상의 예산지원과도 관련되기 때문에 예산에서 중요한 부분을 차지하는 교사 인건비와 관련된 노드들이었음.
 - 네트워크 분석의 좌측에는 아동학대와 경력단절여성을 중심으로 하는 네트워크 구조가 존재하였음.
 - 누리과정의 운영과 관련하여 아동학대와 경력단절여성은 누리과정의 직접적인 결과나 요소라고 하기는 어려우나, 교육과 보육의 질의 측면에서 아동학대가 논의되었으며, 누리과정의 기대목표에 경력단절여성의 재취업을 고려했다는 점을 살펴볼 때 관련성이 있다고 해석할 수 있음 (국무총리실, 기획재정부, 교육과학기술부, 행정안전부, 보건복지부, 2012).

4

연구의 시사점 및 제언

- 본 연구에서는 국민들의 여론이 누리과정을 우리나라 유아교육과 보육 서비스 전체로 이해하고 있다는 점을 볼 때 누리과정의 질 관리와 누리과정 예산의 안정성은 한국의 육아정책에서 보다 중요하게 다루어질 필요가 있음.
- 본 연구에서는 예산의 한계가 있어 2013년 1월부터 2016년 6월까지로 연구를 제한하였으므로 2011년 만 5세 누리과정 도입 시기부터의 여러 담론의 형성이 3-5세 연령별 누리과정과 어떻게 연결되는지에 대해서는 다루지 못하였음. 추후 시행 6년차에 접어든 현 시점에서 지속가능한 누리과정 정책 유지를 위해 정부차원에서 누리과정에 관한 빅데이터 분석을 위한 신디케이트(syndicate)를 마련하고 꾸준히 분석해나갈 필요가 있음.
- 본 연구에서 2013년부터 2016년 전체 데이터에 대한 의미망 분석을 실시하였으나, 연도별로 실시하지는 못하였으므로, 추후 연구에서는 데이터들을 보다 잘 구축하고 정련하여 연도별로 나타난 여론의 지형도를 살펴보고 추후 정책이 나아가야할 방향을 예측해나갈 필요가 있음.
- 모든 3-5세를 위한 평등한 교육과 보육을 강조했던 애초 누리과정의 목표를 고려할 때 유치원과 어린이집 환경에서 나타나는 누리과정의 차이 또는 격차가 누리과정의 질과 어떤 관련성을 나타내는지 추후 연구가 필요함.
- 누리정은 국가수준의 유아교육과정 정책일 뿐만 아니라, 무상교육 및 유보통합 정책 등의 다양한 의미를 담고 있는 기표임. 누리과정에 대한 다양한 의견과 비판을 겸허히 수용하고, 유·보 통합의 관점에서 누리과정 정책을 어떻게 지속가능하게 발전시켜 나갈 수 있을지 더 많은 고민이 필요한 시점임.

참 고 문 헌

- 교육과학기술부·보건복지부(2013). 3-5세 연령별 누리과정: 해설서.
- 국무총리실·기획재정부·교육과학기술부·행정안전부·보건복지부(2012. 1. 18). 3-4세 누리과정 도입 계획. 관계부처 합동 보도자료.
- 김병규(2015. 3. 23). 어린이집, CCTV 보다 교사 늘려 정서적 친밀감 높여야. 연합뉴스, <http://www.yonhapnews.co.kr/bulletin/2015/03/23/0200000000AKR20150323062200017.HTML>/에서 2017년 4월 26일 인출.
- 김세관(2015. 4. 1). 어린이집 CCTV설치, 실사가 감시에 제동. 머니투데이, <http://the300.mt.co.kr/newsView.html?no=2015040118537671349>/에서 2017년 4월 26일 인출.
- 김용학(2007). 사회 연결망 분석. 서울: 박영사.
- 박창현(2016). 신문기사에 나타난 누리과정 정책 효과 인식 연구: 사회연결망분석을 활용하여. 유아교육연구, 36(5), 399-422.
- 이진화·박창현·윤지연(2015). 3-5세 누리과정 이용실태와 정책 효과 분석. 서울: 육아정책연구소.
- 최인봉·이채희·이은경 (2015). 소셜 빅데이터를 활용한 학업성취도 평가 이슈 분석. 서울: 한국교육과정평가원.
- 손현성(2015. 1. 17). 보육교사 아동학대에 원장 동료들 입 다물면 은폐 무방비. 한국일보, <http://hankookilbo.com/v/3f4cd0ce66194de3821c7c10fbfb32b0>/에서 2017년 4월 26일 인출.
- 조규희·서미선·이정우(2016. 12. 1). 여야, 누리과정 합의...3년한시 특별회계 설치: 중앙정부 일반회계 전입금 규모 1조원 안팎 될 듯. 뉴시스 1. <http://news1.kr/articles/?2846178>/에서 2017년 4월 26일 인출.
- 박창현·양미선·조혜주(2016). 누리과정 질 제고를 위한 운영 다양화 방안. 서울: 육아정책연구소.
- 이윤진·김지현·이민경(2016). 누리과정 정책 성과 분석. 서울: 육아정책연구소.

현영섭·신은경·이향란(2011). 다문화 학습 동아리의 구조적 공백에서 신뢰와 결혼이주 구성원 브로커의 역할. 평생교육학연구, 17(2), 1-31.

Granovetter, M. S. (1973). The strength of weak tie. America Journal of Sociology, 78, 1360-1380.

Kalish, Y. (2008). Bridging in social networks: Who are the people in structural holes and why are they there?. Asian Journal of Social Psychology, 11. 53-66.

Scott, J. (2000). 김효동, 김광재 역(2012). 소셜 네트워크 분석. 서울: 커뮤니케이션북스.