The SNU Journal of Education Research
December 2017, Vol.26, No.4, pp.17-42.

Keyword Network Analysis on 'Free Semester Policy' with Korean Newspaper Articles*

Anna Shin
Seoul National University

Sun-Geun Baek**
Seoul National University

Ye-Lim Yu
Korean Educational Development Institute

Yun-Kyung Kim
Seoul National University

ARTICLE INFO

Article history:
Received Dec 4 2017
Revised Dec 24 2017
Accepted Dec 27 2017

Keywords:
Free Semester Policy, big data analysis, keyword network analysis

ABSTRACT

This study aims to explore issues emerged from Korea's 'Free Semester Policy' through keyword network analysis on newspaper articles. Using web-scraping with Python, the published articles from 11 major daily newspapers between 2013 and 2017 were collected. After preprocessing the collected data, keyword frequency analysis and keyword network analysis were conducted in each of the five phases of the policy process in order to identify the keywords and to ascertain their association. Throughout all phases of the policy, promoting opportunities for career exploration and improving instruction methods via national and local support and cooperation were at the core of the policy. The analysis showed that there were visible changes as the policy progressed from one phase to another. Having identified the core issues surrounding Free Semester Policy, this research will be a stepping stone to analyze other education policies.

* An earlier version of this article was presented at the 18th international conference on education research, Educational Research Institute, Seoul National University, Korea, October 19th, 2017.
** Corresponding author, dr100@snu.ac.kr

I. Introduction

'Free Semester Policy', a major middle school education policy in Korea, has been planned and implemented since 2013. This education policy was designed to provide students with opportunities to seek their dreams and talents, while lessening the academic pressure for one semester in middle school. After pilot operated with 42 schools in 2013, the policy was extended to 811 pilot or volunteered schools in the following year, and to 2,551 schools in 2015. And as of 2016, all middle schools in Korea participated in the policy.

During the initial phase of the policy, various issues had surfaced. Since all middle schools were bound to follow, the merits and demerits of the policy were debated through the media. Given the extensive public interests and concerns about the policy, identifying core issues and their changes in each phase is necessary to facilitate an effective implementation in the future. In this respect, keyword network analysis using newspaper articles has the advantage of getting a better insight into the hidden issues present within a text and better understanding of its narrative structure (Paranyushkin, 2011). Also, it has the advantage of improving traditional content analysis and organizing aspects of communication in the policy process in that newspaper articles tend to emphasize specific semantic associations through the use and arrangement of words (Nam & Park, 2007).

Furthermore, the analysis of newspaper articles is crucial in enhancing understanding of Free Semester Policy due to its unusual process of policy formation. The policy started out as somewhat abstract and indeterminate ideas such as 'education for exploring dreams and potentials' and 'education for seeking happiness'. Then, details of the policy were gradually determined through interactions with policy stakeholders in a pilot operation (Kim, 2017). Therefore, in order to detect the issues in the process of policy formation, it is important to explore through newspaper articles how the involvement of multiple stakeholders such as Ministry of Education, provincial education offices, school units, local organizations, parents, teachers, and students formulated the policy.

To meet the needs, this study explored the core issues emerged from Free Semester Policy through keyword network analysis on newspaper articles—3,329 articles in 11 major daily newspapers published between January 2013 and July 2017 were scraped using Python. After preprocessing and conducting morphological analysis on the articles, top keywords related to the policy were extracted using keyword frequency analysis. Then, networks among top keywords were constructed in each phase of the policy process, and the changes across the phases were investigated.

II. Literature review on Free Semester Policy

'Free Semester Policy' is expected to enrich students' talents and characters, to maximize happiness and satisfaction about school lives, and to regain the trust toward public education (Ministry of Education, 2013). This policy allows middle schools to operate one semester of their curriculum free of term exams so that students can explore their dreams and talents without a burden to prepare for term exams. During the semester, students take core subjects such as Korean, Mathematics and English and spend the rest of their school hours engaged in 1) career exploration activities, 2) elective theme activities, 3) art and sports activities, and 4) club activities (Ministry of Education, 2015). This policy is leading changes in education by innovating classes and evaluations. The changes include educational outcomes such as an improved teacher-student relationship, an increased student participation in class, an enhanced core competencies and career competencies of the students (Kim et al., 2016; Kim, 2017), the expansion of peer network and supportiveness (Shin, 2017), and satisfaction toward school (Kim, 2017).

The policy was first announced on November 21th of 2012, when a presidential candidate Park Geun-Hye introduced the policy as a part of her campaign promises. After she was elected and her administration set sail on February 25th, 2013, the Ministry of Education officially introduced the policy and its action plan as a major project of the year. To minimize undesirable spinoffs, 42 pilot schools were assigned to pilot operate the policy during the fall semester of 2013. In 2014, 38 pilot schools and 732 volunteered schools (the total of 811 schools, which was 25% of middle schools nationwide) adopted the policy. In 2015, 2,551 pilot or volunteered schools joined in. Then finally in 2016, the policy was adopted in all middle schools throughout the nation (total of 3,204 middle schools) (Ministry of Education, 2015). The next president Moon Jae-In who took office in 10th of May, 2017, has announced to continue to actively support the policy, which implies that the policy aligns with the needs of the times regardless of one's political stance. Detailed process of policy implementation is provided in <table 1>.

<Table 1> Phases of Free Semester Policy (Koo, 2017; Shin et al., 2015)

Phase	Process	Date
Introduction of policy (-2013.08)	Announcement of electoral commitment introducing Free Semester Policy	2012.11.21
	Announcement of key policy 'Life with creativity education and culture' by The Commission on Presidential Transition	2013.02.21
	Introduction of the policy as a major project of the Ministry of Education	2013.03.28
	System constructed to operate Free Semester Policy	2013.04
	Proposal of pilot operation plan of the policy	2013.05.29
	Opening ceremony of Free Semester Policy pilot schools	2013.06.04

1st year of pilot operation (2013.09-2014.02)	Operation of 42 pilot schools	2013.09 -
	Assignment of volunteered schools	2014.02
2nd year of pilot operation (2014.03-2015.02)	Operation of pilot and volunteered schools	2014.03 -
	Operation of volunteered schools	2014.09 -
3rd year of pilot operation (2015.03-2016.02)	Expansion of volunteered schools	2015 -
	Constructing partnerships with organizations that provide infrastructure for students' experiences	2015 -
	Enacting 'Career Education Act'	2015.06
	Operating support groups for career practicum of Free Semester Policy	2015.07
	Proclamation of revised enforcement ordinance of 'Elementary and Secondary Education Act'	2015.09.15
	Announcement of '2015 National Educational Curriculum Revision'	2015.09.23
	Confirmation and announcement of 'middle school Free Semester Policy implementation plan'	2015.11.25
Full implementation (2016.03-present)	Full implementation of Free Semester Policy	2016.03
	Announcement of action plan of the Ministry of Education	2017.01.06

Evaluating Free Semester Policy had employed a specific model or framework and collected data through literature review, survey, interview, and case study. Kang and An (2015) evaluated the policy using CIPP (Context, Input, Process, and Output) model. The researchers surveyed 550 students, 550 parents, and 110 teachers from 11 schools (7.8% of all pilot schools in Seoul). The result revealed parents' concerns that Free Semester Policy would deprive their children of chances to be admitted to selective prestigious high schools and beyond. Meanwhile, Koo (2017) developed an evaluation scale for Free Semester program using CIPP model and verified the reliability and validity of the scale, using confirmatory factor analysis and Rasch model. He collected data through literature reviews, expert interview, delphi research, pilot tested on 220 teachers in Gyeongnam province, 217 teachers in Seoul, 224 teachers in Daejeon, 278 teachers in Busan.

Using an analytical framework of Cooper, Fusarelli and Randall (2004), Park, Joo and Ko (2014) reviewed Free Semester Policy, identified its challenges, and proposed future directions. Park (2015) also employed critical discourse framework to analyze as to how the Ministry of Education, teachers, students, and parents respond differently and create tension regarding Free Semester Policy. He reviewed literature published by the Ministry of Education and interviewed teachers, students, and parents of pilot schools in Daejeon and Jeonbuk to understand the structure and the issues surrounding the policy and to analyze the daily language of the interest groups that display and process discourse.

Shin and Park (2015) carried out a case study on three schools to present optimistic and pessimistic views on the policy. Hopeful aspects were 'chances of internal reflections and exploration of new possibilities', 'formulating new relationships', 'teacher-driven execution of policy', and 'improved career awareness motivating students to learn'. Unpromising aspects, on the other hand, were 'inconsistency of policy', 'misconceptions on the policy', 'perspectives toward school education in variance', and 'difficulty of career practicum and evaluation'.

Kim (2017) wrote a memoir identifying significant controversy of the policy. First, unlike the original policy intention of class innovation, an aspect of career education was hugely highlighted as the core of the policy through the media. Second, as terms exams are not carried out during the free semester, there were concerns about how students' GPA will be determined for high school entrance. Third, there was disagreement over whether the policy target should be middle school students or high school students. While some expected the policy effect to apply positive leverage to following years, others argued that the policy will put a crimp on students' school lives thereafter.

In the earlier studies, the model or framework researchers employed reflects their presupposition about the policy, thus presenting the likelihood of inhibiting a fair assessment. Moreover, the data researchers used is not only drawn from limited areas, schools, and stakeholders (such as students, parents, and teachers) but is also affected by specific features differ by the method of collection. Thus, big data, which is less likely to be affected by any particular interest group, regional interests, frames, or method, can provide a reliable and comprehensive evidence to analyze the policy.

III. Method

A. Data collection

The data used in this study are articles addressing 'Free Semester Policy' from 11 major daily newspapers published between January 2013 and July 2017 in Korea. From the NAVER news website (http://news.naver.com), the articles searched by the term 'free semester' were scraped and saved with URL, title, date, press name, and body text using 'BeauifulSoup4' and 'Selenium' library with Python. BeautifulSoup4 is a Python library for pulling data out of HTML and XML files, which provides idiomatic ways of navigating, searching, and modifying a parse tree (Richardson, 2017). Selenium is a library to control web-browser automatically (Hunt et al., 2018), which is useful for scraping contents from websites developed in JavaScript. Duplicate articles were excluded based on URL, and

irrelevant articles were also removed according to the following criteria; 1) the articles not containing the term ‘free semester’ in title or body text, 2) the articles that merely reflected administrative issues such as promotion, transfer, etc. After filtering them, the remaining 3,329 articles were used for an analysis. The number of newspaper articles per phases of the policy is presented as <table 2>

<Table 2> The number of newspaper articles per phases of the Free Semester Policy

Phase		**number of newspaper articles**	**number of terms**
Policy introduction		276	8,656
Pilot operation	1st year	132	4,945
	2nd year	426	11,110
	3rd year	962	17,905
Full implementation		1,533	23,684
All phases		**3,329**	**35,611**

B. Data cleaning

The data cleaning is a process of converting the collected text data into a form that can be easily analyzed later using Natural Language Processing (NLP) technique (Yu & Baek, 2017). In this study, data cleaning consists of two stages: 1) preprocessing of text data and 2) morphological analysis. In the preprocessing step, to remove unimportant expressions, a list of stopwords was composed. Stopwords refer to the words that appear repeatedly but do not contribute to the main theme of the articles. And then, nouns were normalized by terms whose meaning basically synonymous.

To extract keywords from each article, KoNLP (Jeon, 2016) package in R program was used. KoNLP is a representative Korean natural language processing tool based on Hannanum analyzer, a morphological analyzer developed by Semantic Web Research Center (SWRC) at KAIST. It breaks down a sentence into the smallest units of meaning, called morphemes, to helps extract and analyze from text data useful information. We used ‘Sejong’ and ‘Woorimalsam’ dictionary in KoNLP, adding new words relevant to Free Semester Policy into the dictionary. Then from the articles, nouns were extracted and saved as the term-document matrix using tm (Feinerer & Hornik, 2017) package.

C. Data analysis

1. Keyword frequency analysis

Keyword frequency analysis is a method calculating the frequency of terms in documents. Both Term Frequency (TF) and normalized Term Frequency-Inverse Document Frequency (normalized TF-IDF) were used for keyword frequency analysis. The number of times a term occurs in an entire document is called *TF*. Higher TF value indicates that the word is relatively more important in the document. TF is the simplest way to figure out which term is considered important in a whole document. TF is not, however, enough to signify the importance of a term in a document in that words with large TF value tend to appear frequently in all documents in the corpus but simultaneously are likely to contribute little to understanding the theme of the documents. To eliminate the impact of frequent terms that exist in almost all documents, *normalized TF-IDF* can be applied (Lertnattee & Theeramunkong, 2004). Normalized TF-IDF was obtained by multiplying normalized TF by IDF, which is the inverse frequency of the document. To control the length of the document, normalized TF-IDF was calculated by the formula provided below (Feinerer & Hornik, 2017).

$$\begin{aligned}\text{Normalized TF} - \text{IDF}_{\text{i,j}} &= \text{Normalized TF}_{\text{i,j}} \times \text{IDF}_{\text{i}} \\ &= \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log_2 \frac{|D|}{|\{d|t_i \in d\}|}\end{aligned}$$

where, $n_{i,j}$: the frequency of term i in document j

$\sum_k n_{k,j}$: sum of frequencies of all terms in document j

$|D|$: the number of documents in a corpus

$|\{d: t_i \in d\}|$: the number of documents containing term i

In this study, the top 10 keywords were selected per policy phrases based on TF and normalized TF-IDF for network analysis using tm package in R.

2. Keyword network analysis

Keyword network analysis is a technique identifying linkages between extracted keywords based on the co-occurrence frequency of each pair of keywords (He, 1999). In this study, keyword networks of Free Semester Policy were constructed using igraph (Csardi, 2015) package in R. In each network graph, keywords were represented as nodes and linked by edges if they co-occurred in an article. Also, the frequency of co-occurrence among keywords was applied as the weight of the link in the network. In the igraph package,

this weight was calculated by summing up the edge weights of the adjacent edges for each node as the formula below, which is called *strength* (Barrat et al., 2004; Csardi, 2015).

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij} \ .$$

Where a_{ij} : the degree between node i and j
w_{ij} : the weight of edge between node i and j

Keyword networks are typically visualized with force-directed algorithm or spring embedders (Yu & Baek, 2017). The idea of a force-directed algorithm is to consider a force between any two nodes. In this algorithm, the nodes are represented by steel rings and the edges are springs between them. The basic idea is to minimize the energy of the system represented attractive force and repulsive force by moving the nodes and changing the forces between them. In this study, standard force-directed algorithm was employed as a basis, that is, the *Fruchterman and Reingold algorithm* (Fruchterman & Reingold, 1991). The main advantage of this algorithm is in its speed, achieved by using a very simplistic model of interacting forces in the network graph (Klapka & Slaby, 2016).

In addition, a *community analysis* was conducted to detect the structure of the keyword network. Community analysis is the method decomposing the network into sub-units named 'community'. Each community is divided so that the highly-interconnected nodes can be grouped together, and the less-connected nodes can be categorized into different communities. And to detect communities in the network, an optimal modularity is calculated, using *Louvain methods*—a fast and easy method to calculate modularity in weighted network graph (Blondel et al., 2008). The modularity is a scalar value ranging from –1 to 1, and positive and larger value of modularity indicates the presence of community structure (Newman, 2006). For the weighted networks, modularity is calculated by comparing the link density inside communities with the links between communities as provided below (Blondel et al., 2008).

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where A_{ij} : the weight of the edge between i and j
$k_i = \sum_j A_{ij}$: the sum of the weights of the edges attached to node i
c_i : the community to which node i is assign
δ : delta function. δ(u,v)=1 if u=v, otherwise 0
m : $1/2 \sum_{ij} A_{ij}$.

IV. Result

A. Keyword frequency analysis

1. Keyword frequency in all phases

Throughout all phases, the most frequently appearing keywords were provided in <Table 3>. When looking at the combined result of TF and normalized TF-IDF, 'career', 'subject matter', 'teacher', 'practicum', and 'instruction' were the words with the highest frequencies. According to normalized TF-IDF, however, 'private tutoring', 'provincial education office', and 'term exam' appeared as highly ranked top 10 keywords, while TF highly ranked 'operation' and 'learning' in newspaper articles.

<Table 3> Top 10 keywords for all phases

No	TF		normalized TF-IDF	
	Keyword	**Value**	**Keyword**	**Value**
1	Career	15,256	Career	44.32
2	Subject Matter	15,053	Subject Matter	38.31
3	Teacher	7,522	Teacher	36.35
4	Practicum	6,662	Practicum	36.28
5	Instruction	5,952	Instruction	34.80
6	Program	5,237	Private Tutoring	32.69
7	Activity	5,107	Provincial Education Office	32.37
8	Operation	4,636	Program	32.12
9	Learning	4,351	College	27.29
10	Parents	4,343	Term Exam	26.55

To control the length of documents, normalized TF-IDF was used to analyze keyword frequencies and their networks for the rest of the research.

2. Keyword frequency per policy phase

For a further analysis, 10 keywords were selected in each phase, summing up to the total of 22 keywords as provided in <Table 4>. And the wordclouds of top 22 keywords in each phase are provided in Figure 1. Also, to identify the relative importance of keywords in each phase, the ratio of normalized TF-IDF was calculated—dividing the value of normalized TF-IDF by the maximum value of normalized TF-IDF of each phase as is shown in Figure 2.

First of all, 'transition year' appeared most frequently during the introduction phase but sharply decreased in the following phase. In the 1st year of pilot operation, normalized

TF-IDF values of ‘visit’, ‘event’, ‘pilot school’ and ‘research’ had the highest frequency; this can be accounted for by the former President Park’s visit to one of the pilot schools. ‘Career’ and ‘subject matter’ appeared as the most important keywords from the 2nd year of pilot operation. In fact, the keyword ‘career’ reached its peak during the 2nd year of pilot operation and full implementation phase, whereas ‘subject matter’ was the most important keyword during the 3rd year of pilot operation.

Also, keywords such as ‘academic achievement’, ‘high school entrance’, and ‘application’ appeared frequently during the policy introduction phase but decreased in the following phase. Presumably, the announcement that the academic achievement during the free semester is not reflected in high school admission is the possible explanation. On the contrary, the frequency of the ‘private tutoring’ decreased in the 1st year of pilot operation but began to increase from the 2nd year. It can be inferred that ‘private tutoring’ received little attention du+ring initial phase as the academic achievement was decided not to be reflected in high school admission. However, by emphasizing the importance of accelerated learning, private tutoring reappeared.

<Table 4> Top 22 Keywords by normalized TF-IDF

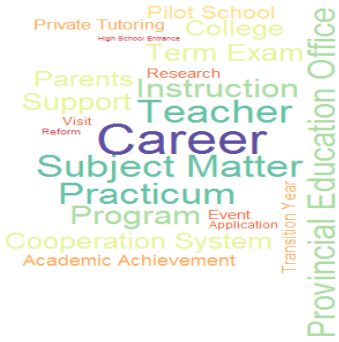
Keyword	**Phase**					**All Phases**
	Policy introduction	Pilot operation			Full implementation	
		1st year	2nd year	3rd year		
Transition Year	**3.61**	0.19	2.04	2.62	2.00	12.66
Career	**2.72**	**1.75**	**6.45**	**11.54**	**20.67**	**44.32**
Academic Achievement	**2.62**	1.01	2.33	5.18	8.89	20.74
Provincial Education Office	**2.55**	1.26	**4.41**	**9.37**	**14.66**	**32.37**
Subject Matter	**2.51**	**1.49**	**4.74**	**12.61**	**16.90**	**38.31**
High School Entrance	**2.50**	0.24	0.82	2.55	5.63	12.86
Pilot School	**2.42**	**2.63**	2.75	4.36	5.42	21.44
Term Exam	**2.30**	0.99	**3.58**	7.76	10.35	**26.55**
Practicum	**2.28**	**1.56**	**4.62**	**10.29**	**17.54**	**36.28**
Application	**2.21**	0.37	1.49	2.80	5.28	12.75
Visit	0.56	**2.83**	1.69	2.84	4.70	13.30
Event	0.91	**2.61**	1.88	3.94	7.23	16.78
Research	1.11	**2.58**	1.89	3.22	6.49	15.33
Instruction	1.85	**1.79**	**4.13**	**9.22**	**17.33**	**34.80**
Program	1.46	**1.55**	**3.99**	**9.69**	**15.10**	**32.12**
Teacher	2.16	**1.44**	**4.93**	**10.48**	**16.99**	**36.35**
Support	1.46	1.13	**3.54**	7.59	10.80	24.80
Cooperation System	1.21	1.10	**3.36**	6.35	8.14	20.24
Private Tutoring	1.62	0.58	2.18	**9.25**	**19.18**	**32.69**

Reform	0.40	0.36	1.24	**8.55**	5.00	16.61
College	1.92	0.78	3.30	**8.24**	**13.05**	**27.29**
Parents	1.76	0.87	3.26	7.28	**12.48**	25.61

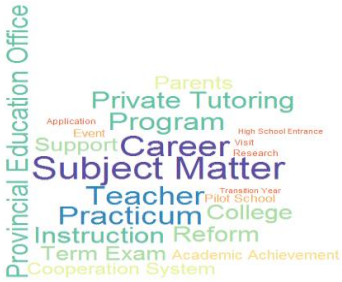
* Values in **bold** indicate highly ranked top 10 keywords in each phase

(a) policy introduction

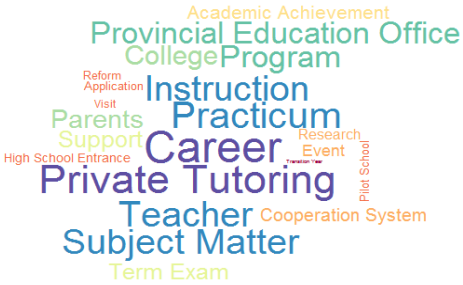
(b) 1st year of pilot operation



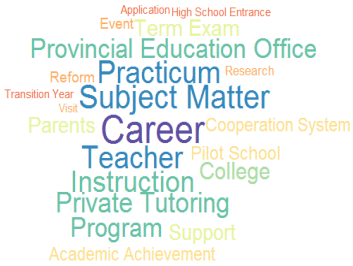
(c) 2nd year of pilot operation



(d) 3rd year of pilot operation



(e) full implementation



(f) all phases

Figure 1. Wordclouds of top 22 keywords in each phase

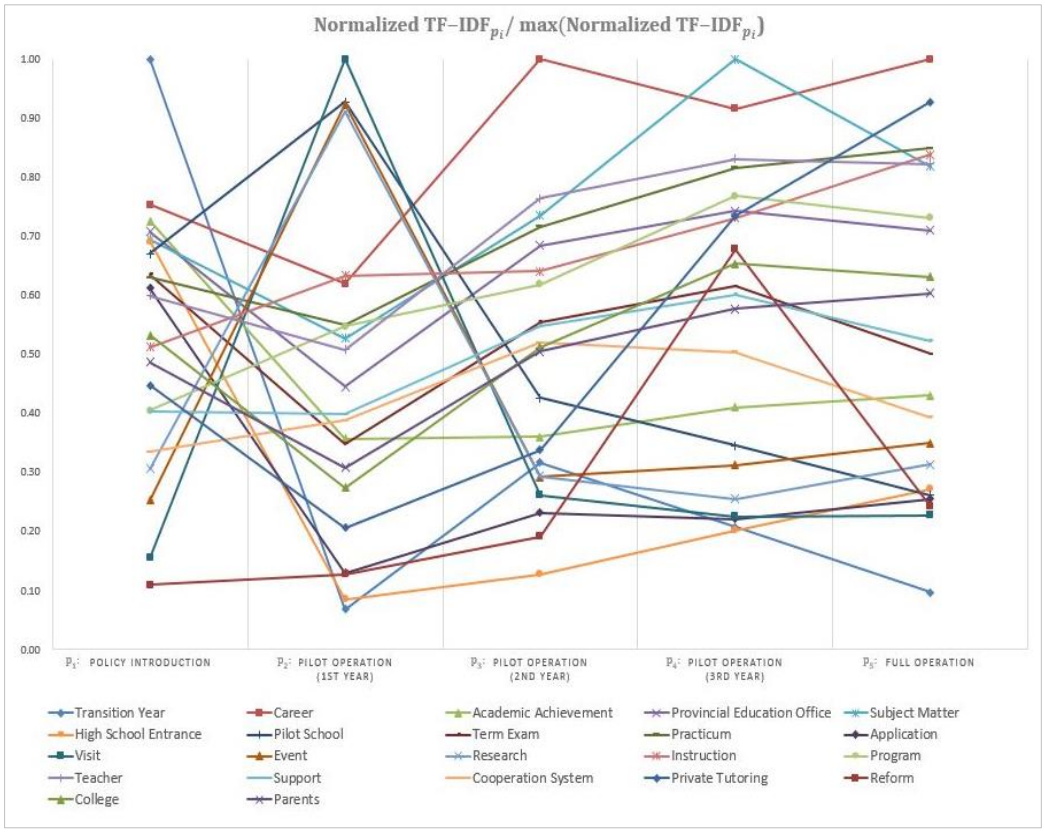


Figure 2. The relative importance of keyword per phase

B. Keyword network analysis

1. Network analysis of all phases

The keyword network graphs of the top 22 keywords throughout all phases are shown in Figure 3. In the network, all nodes are fully interconnected, thus resulting in equal degree centralities between keywords. By using the co-occurrence frequency between keywords as weights, we calculated the strength of node which is the sum of link weights connected to the node. The size of the nodes in the graph corresponds to the strength of the nodes. The bold lines between nodes indicate that the weight of the link is higher than that of other links. And according to the network structures, three major issues each of which is displayed in red, blue, or yellow could be narrowed down, representing different community memberships (modularity=0.083).

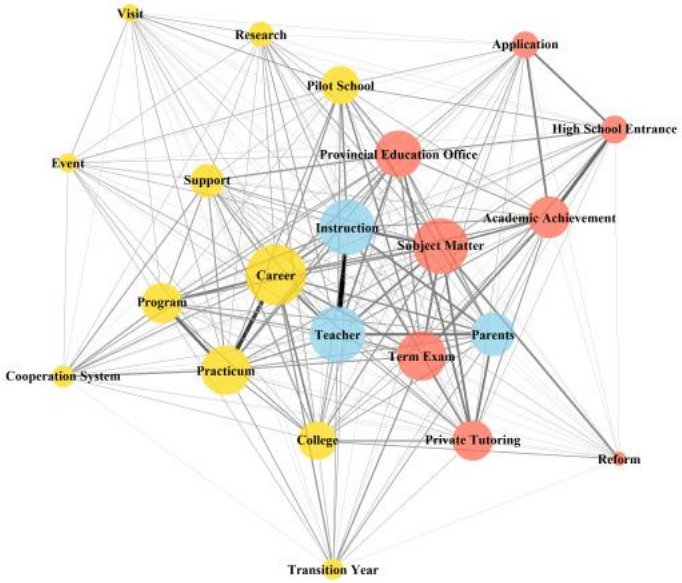


Figure 3. Keyword network graph of all phases

<Table 5> Keywords of Free semester policy throughout all phases

No	Keyword	Strength	No	Keyword	Strength
1	Career	350.22	12	College	223.06
2	Instruction	322.64	13	Pilot School	222.19
3	Teacher	322.26	14	Support	194.22
4	Subject Matter	321.68	15	High School Entrance	163.89
5	Practicum	286.31	16	Application	156.84
6	Term Exam	285.99	17	Research	144.64
7	Provincial Education Office	271.22	18	Cooperation System	128.87
8	Parents	253.24	19	Transition Year	124.63
9	Academic Achievement	241.93	20	Event	112.94
10	Private Tutoring	234.47	21	Visit	103.54
11	Program	233.54	22	Reform	84.67

First, the keyword 'career' had the highest strength, and the word constructed a community with 'practicum', 'program', 'support', and 'cooperation system' and more. Looking at this community, providing career practicum fully supported by cooperation system was at the center of public attention.

Second, the keywords ‘instruction’ and ‘teacher’ had the second and the third highest strengths, and the words formed a community with ‘parents’. Improving teaching-learning activity was considered a primary concern and objective of the policy. Also, various programs and meetings for teachers and parents were offered to help them better understand the policy.

Third, ‘subject matter’ and ‘term exam’ had relatively high strengths and constructed a community with keywords like ‘academic achievement’, ‘private tutoring’ and ‘high school entrance’. Since Free Semester Policy was designed to lessen academic pressure, thus obviating term exams during the semester, whether or not academic achievement results would be reflected in high school admission received much attention. By contrast, a growing concern for an increased reliance on private tutoring during free semester had been brought up.

2. Network analysis of per policy phase

a. Policy introduction

The keyword network graph of top 22 keywords from policy introduction phase is shown in Figure 4. In this phase, three major issues emerged (modularity=0.207). The first issue included ‘high school entrance’, ‘academic achievement’, and ‘application’ and had the highest strength in the network. The strong link weight of the keywords demonstrates that the decision to obviate academic achievement collected in Free Semester from the high school admission was highlighted. There had been controversies over the influence of such decision: while some worried about the decline of students’ academic achievement level, others believed Free Semester Policy will develop competencies of students such as a social skill and a self-directed learning ability, which are far beyond knowledge-based academic achievement in its importance (Shin & Park, 2015).

The second issue involved the keywords: ‘career’, ‘transition year’, and ‘practicum’. The idea of free semester is taken from an education policy of Ireland— transition year. It offers one-year program of career exploration for students who have completed the junior cycle, expecting to enter the senior cycle of their post-primary education.

The third issue showed that ‘pilot school’ forming a community with other keywords like ‘cooperation system’ and ‘event’. To successfully launch a pilot operation, the demand for a cooperation system consisting of private and public institutions made appearance. Accordingly, the Ministry of Education, Center for Free-Semester Program at Korean Educational Development Institute, and related organizations signed Memorandum of Understanding (MOU).

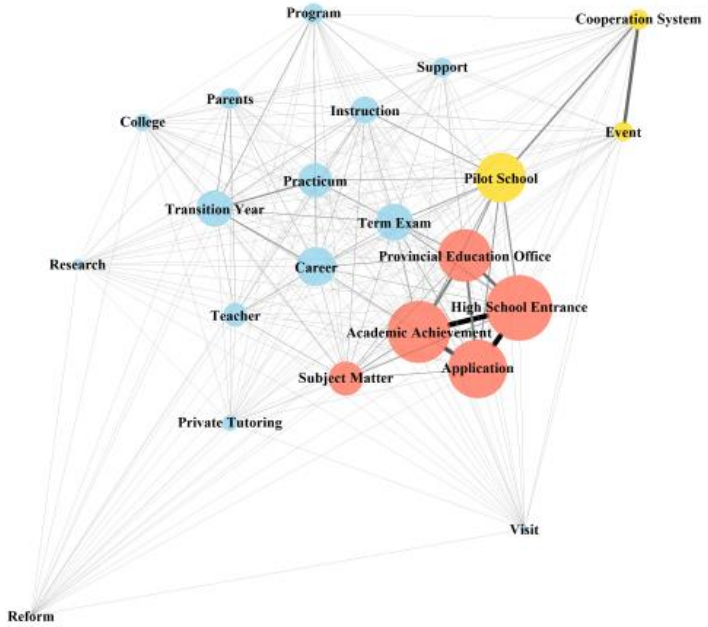


Figure 4. Keyword network graph of policy introduction phase

<Table 6> Keywords of Free Semester Policy in policy introduction phase

No	Keyword	Strength	No	Keyword	Strength
1	High School Entrance	0.76	12	Teacher	0.28
2	Academic Achievement	0.72	13	Parents	0.24
3	Application	0.68	14	Program	0.24
4	Provincial Education Office	0.61	15	Cooperation System	0.24
5	Pilot School	0.57	16	Support	0.23
6	Career	0.46	17	Event	0.22
7	Term Exam	0.43	18	College	0.20
8	Transition Year	0.42	19	Private Tutoring	0.18
9	Subject Matter	0.39	20	Research	0.12
10	Practicum	0.39	21	Visit	0.09
11	Instruction	0.32	22	Reform	0.03

b. 1st year of pilot operation

The keyword network graph of the top 22 keywords during the 1st year of pilot operation shown in Figure 5 (modularity=0.209). In this phase, the keywords with high strength were ‘pilot school’, ‘visit’, ‘research’ and ‘event’, constructing one community. As 42 pilot schools took initiative, the pilot operation piqued the public interest. Also, the former President Park Geun-hye’s visit to one of the pilot schools, Dongjak middle school, was widely reported in the press. Meanwhile, ‘instruction’, ‘subject matter’, and ‘teacher’ formed another community colored in blue. During this phase, various manuals and guidebooks were developed and distributed mainly by Ministry of Education and Korean Educational Development Institute to improve instructions and evaluation in pilot schools by promoting teachers’ competencies (Choi et al, 2014; Ji et al, 2014).

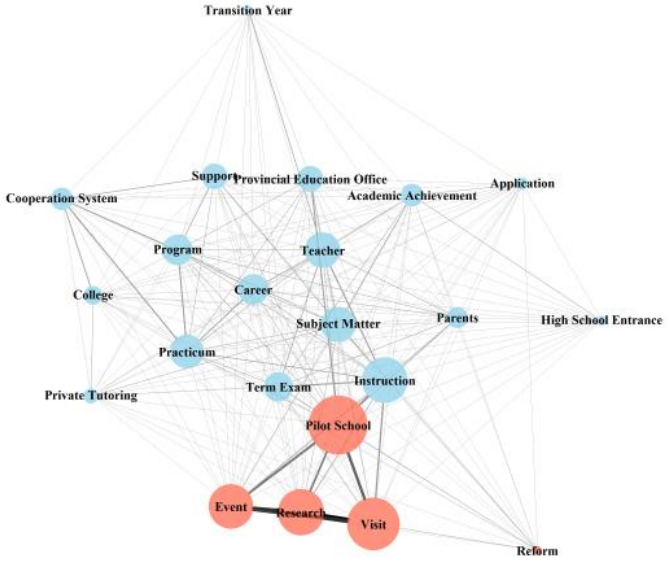


Figure 5. Keyword network graph of 1st year of pilot operation

<Table 7> Keywords of Free Semester Policy of 1st year of pilot operation

No	Keyword	Strength	No	Keyword	Strength
1	Pilot School	21.87	12	Support	9.66
2	Visit	19.53	13	Provincial Education Office	9.66
3	Research	17.31	14	Cooperation System	8.53
4	Instruction	17.10	15	Academic Achievement	8.46
5	Event	16.61	16	Parents	7.96
6	Teacher	13.25	17	College	7.07

7	Subject Matter	13.07	18	Private Tutoring	5.63
8	Practicum	12.56	19	Application	4.58
9	Program	11.49	20	High School Entrance	3.75
10	Career	11.39	21	Reform	3.34
11	Term Exam	10.94	22	Transition Year	3.05

c. 2nd year of pilot operation

The keyword network graph of the top 22 keywords during the 2nd year of pilot operation is provided in Figure 6. In this phase, there were two communities in the network (modularity= 0.077). As 38 pilot schools and 732 volunteered schools added numbers to the total participation, the intent of Free Semester Policy—providing opportunities for career exploration, improving instructional methods, and lessening the academic pressure by obviating term exams—repeatedly came to the public attention. It is worthy of note that the keywords that mixed up and belonged to the same community in the previous phase (1st year of pilot operation) were divided into two communities during this phase. The co-existence of these two communities represents the issues on what the ultimate objective of the policy is: improving class instruction or providing career education. While the former is an original purpose of policy (Kim, 2017), the latter was both emphasized by media and attracted attention of the public.

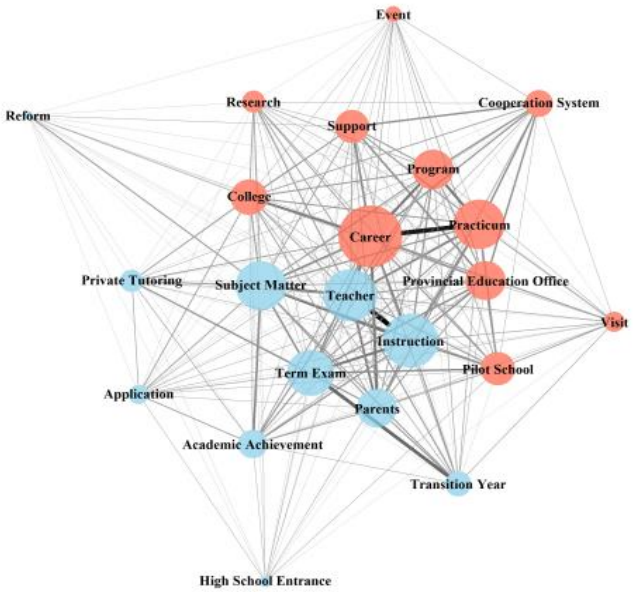


Figure 6. Keyword network graph of 2nd year of pilot operation

The keywords—‘career’ and ‘practicum’—showed the highest strength in one community that also included ‘program’, ‘provincial education office’, ‘support’, and ‘cooperation system’. This composition demonstrates the need for the provincial education offices to devise a support system, which would utilize local resources to raise the accessibility to career practicum programs and to ensure its quality. For instance, provincial education office of Seoul developed and distributed a manual and workbook for seventeen different elective theme activities.

The keywords—‘instruction’, ‘teacher’, ‘subject matter’, and ‘term exam’—had the highest strengths in the other community. Here, the strength of the keyword ‘parents’ is relative high in this community, and this is a visible change compared to the previous phases, indicating an increased interest of the parents in the policy followed by an effort communicating the purpose and the contents of the policy.

<Table 8> Keywords of Free Semester Policy of 2nd year of pilot operation

No	Keyword	Strength	No	Keyword	Strength
1	Career	49.25	12	Pilot School	26.62
2	Instruction	42.97	13	Academic Achievement	23.05
3	Teacher	41.49	14	Cooperation System	21.72
4	Practicum	39.67	15	Transition Year	20.50
5	Subject Matter	39.41	16	Private Tutoring	18.13
6	Term Exam	36.87	17	Research	17.77
7	Program	31.93	18	Visit	16.44
8	Provincial Education Office	31.17	19	Application	15.58
9	Parents	30.93	20	Event	12.86
10	College	28.62	21	High School Entrance	9.16
11	Support	26.63	22	Reform	7.09

d. 3rd year of pilot operation

The keyword network graph of top 22 keywords during the 3rd year of pilot operation is provided in Figure 7. In this phase, three communities turned up (modularity= 0.100). First, one community primarily consisted of the keyword ‘subject matter’, with the highest strength. Here, ‘subject matter’ appeared frequently with ‘private tutoring’ and ‘parents’ which demonstrates evidently higher strength. This indicates that the reliance on private tutoring during the free semester was a primary concern for parents in this phase.

Second, exhibiting similar pattern to the previous phase, another community consisted of keywords of second and third highest strength respectively, 'teacher' and 'instruction'. Since the number of volunteered schools increased, an interest in the changes free semester caused gained momentum in the 3rd year of pilot operation.

The last community was made up of main keyword 'career' together with 'practicum', 'program', 'provincial education office', 'support', and 'cooperation system'. As the number of volunteered schools increased to encompass 70% of all middle schools in Korea, the call for building infrastructures for career practicum programs grew urgent. The Ministry of Education enacted the 'Career Education Act' and established partnerships with various organizations. Also, regional offices of education launched and operated 'support groups for career practicum of Free Semester Policy' in cooperation with local organizations. Regarding the keyword 'provincial education offices', there have been concerns that the quality of Free Semester programs will vary significantly from regions. As rural areas, particularly, lack infrastructure for students' practicum, Free Semester Policy could possibly worsen regional disparities in quality of education. Thus, financial support in a government level for such less-equipped regions is crucial to ensure effective implementation of Free Semester Policy in all regions (Shin & Park, 2015).

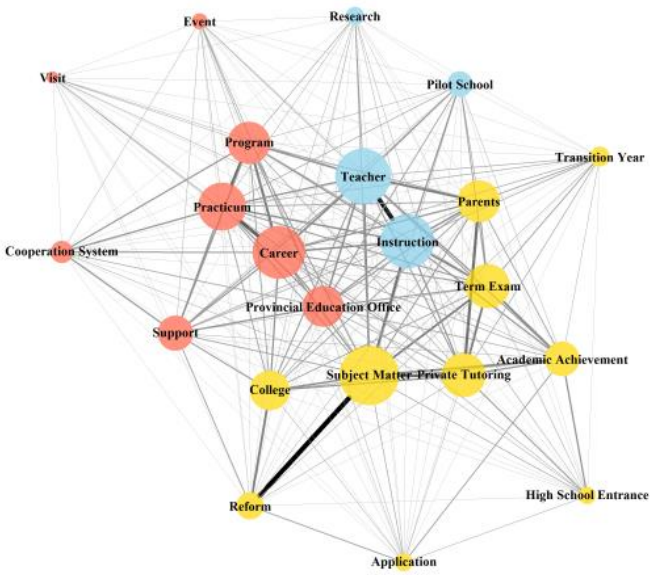


Figure 7. Keyword network graph in 3rd year of pilot operation

<Table 9> Keywords of Free Semester Policy in 3rd year of pilot operation

No	Keyword	Strength	No	Keyword	Strength
1	Subject Matter	101.83	12	Support	60.31
2	Teacher	95.03	13	Academic Achievement	59.66
3	Instruction	92.48	14	Reform	46.96
4	Career	90.47	15	Pilot School	43.81
5	Practicum	82.37	16	Cooperation System	38.65
6	Term Exam	76.25	17	Transition Year	34.53
7	Private Tutoring	75.44	18	Research	31.75
8	Program	73.02	19	Application	31.01
9	Parents	72.54	20	High School Entrance	30.25
10	Provincial Education Office	70.40	21	Event	28.23
11	College	68.51	22	Visit	21.33

e. Full implementation phase

The keyword network graph of top 22 keywords during the full implementation phase is provided in Figure 8.

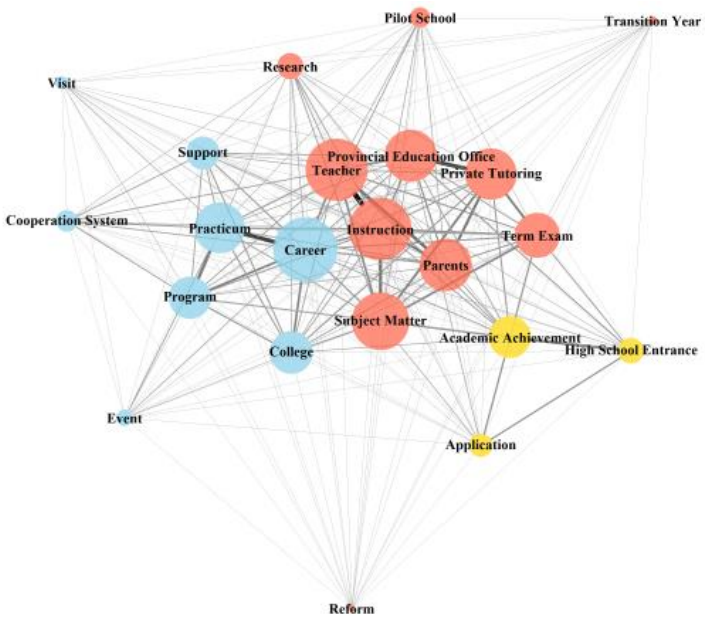


Figure 8. Keyword network graph of full implementation phase

<Table 10> Keywords of Free Semester Policy in policy introduction phase

No	Keyword	Strength	No	Keyword	Strength
1	Career	149.99	12	Academic Achievement	98.38
2	Instruction	145.59	13	Support	77.89
3	Teacher	144.91	14	Research	61.36
4	Subject Matter	134.36	15	High School Entrance	59.97
5	Parents	122.50	16	Application	54.83
6	Provincial Education Office	121.64	17	Cooperation System	50.87
7	Private Tutoring	119.73	18	Pilot School	47.69
8	Practicum	118.57	19	Event	39.19
9	Term Exam	105.34	20	Visit	30.55
10	College	100.33	21	Transition Year	20.43
11	Program	99.03	22	Reform	16.04

In this phase, a minimal change of the network structure occurred compared to previous phases (modularity= 0.108). One community primarily consisted of 'instruction', 'teacher' and 'subject matter', whereas another community incorporated 'career', 'practicum', and 'support'. The last community was constructed based on 'academic achievement', 'application', and 'high school entrance', which could be found in the introduction phase. As the possibility of applying Free Semester outcome to high school levels took place, the Ministry of Education announced a seminal plan to revise the high school admission system.

V. Conclusion and discussion

This study explored issues emerged from Korea's 'Free Semester Policy' through keyword network analysis on newspaper articles. Using web-scraping method of Python, the published articles from 11 major daily newspapers between 2013 and 2017 were collected. After preprocessing the collected data, keyword frequency analysis and keyword network analysis were conducted in each of the five phases of the policy process in order to identify the keywords and to ascertain their association.

According to the result of keyword frequency analysis, following keywords with high frequency: 'career', 'subject matter', teacher', 'practicum', and 'instruction'. According to the normalized TF-IDF results of the keyword network analysis, the rankings of keywords were slightly different across the policy phases. During the introduction phase, 'high school

entrance', 'academic achievement', and 'application' were the keywords with the highest strength, indicating that the press focused on how the accomplishments from the Free Semester would affect high school admission. In the 1st year of pilot operation, 'pilot school', 'visit', 'research', and 'event' were the main keywords, indicating public interest in the pilot operation. In the 2nd year of pilot operation, the keyword 'career' appeared the most, whereas, the keyword 'subject matter' had the highest strength in the 3rd year of pilot operation. Then, in the full implementation phase, 'career' again emerged, showing the highest strength, followed by 'instruction'.

Throughout all phases, the most important issues of Free Semester Policy were providing opportunities for career exploration via national and local cooperation and support system, improving teaching and learning methods, and addressing the problem of private tutoring or high school entrance exam. This represents that the newspaper articles have successfully delivered and distributed the main ideas of the policy –innovation of class instruction and introduction of career education to middle school (Ministry of Education, 2013, 2015) – to the public, which might have contributed to public's better understanding, interest, and support toward the policy.

Furthermore, the result also manifests how those issues have dynamically changed to alter its emphasis over time. For example, when first proposed as a presidential candidate pledge, the primary goal of Free Semester Policy was an innovation of class instruction in middle schools. Nevertheless, as politicians and the press focused on career education – which had been a small part of the original draft– there was confusion about the primary goal of the policy among the Ministry of Education and provincial education offices in the beginning of the policy (Kim, 2017). In this regard, as a result of this study, 'career' appeared as a more important keyword than 'subject matter' in the 2nd year of pilot operation, while result was in reverse in the 3rd year of pilot operation. This depicts how the two main goals of the policy have coexisted, having a somewhat competitive relationship in forming the policy.

Besides, the result is worthwhile in that it reveals not only the intended consequences but also the unintended side effects of the policy. The intended aspect of the policy is presented by keywords 'instruction' along with 'subject matter', 'teacher', 'reform', and 'career' along with 'practicum' and 'program', representing two main goals of the policy. Meanwhile, the unintended aspect of the policy is unveiled. Keywords 'high school entrance', 'academic achievement', and 'application' that emerged during policy introduction phase reappeared during full implementation phase. Keywords of 'parents' and 'private tutoring' grew in its relative importance during 2nd and 3rd year of pilot operation respectively. These keywords disclose problems that have occurred during policy implementation and thus identifying them would contribute to successful implementation of the policy in the future.

These results demonstrate how big data analysis can effectively monitor, visualize, and accumulate long-term time series data using massive amount of data. Particularly because major issues surrounding the policy alter according to its phases, in order to accomplish successful implementation of the policy, it is crucial to utilize big data analysis to examine newspaper articles and provide adequate feedback in a timely manner. The feedback might include supplementing or amending current state of the policy. What is more, big data analysis was particularly useful in examining structure or pattern from newspaper articles of unstructured and nonlinear properties (Park, 2016).

This research revealed major issues of Free Semester Policy, and the keyword network analysis used in this study can also be applied to analyze other education policies. However, this study has following limitations. This research has drawn data only from 11 newspapers with a nationwide subscription; however, to get a closer look at how Free Semester Policy was facilitated by local supports, future research may invite more local newspapers into the analysis. Particularly, since execution of Free Semester Policy is highly dependent on local resources, comparison of local newspaper articles and nationwide newspaper articles will likely reveal meaningful implications. Moreover, newspaper articles can also be compared against official documents published from the Ministry of Education. Meanwhile, this study focused on identifying keywords and their co-occurrence patterns, while leaving out the questions as to what topics are formed by the keywords. Therefore, applying an advanced big data analysis method like *topic modeling* and performing content analysis of representative articles on each topic might be of use. In addition, opinion mining such as *sentiment analysis* might also be helpful in analyzing the sentiments or emotions in documents including nouns and verbs.

References

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America, 101*(11), 3747-3752.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment, 2008*(10), 2-12.

Choi, S. D., Shin, C. G., Hwang, Y. H., Lee, S. E., & Kim, Y. E. (2014). *Operation Manual of Free Semester Policy*. Republic of Korea: Korean Educational Development Institute.

Csardi, G. (2015). *Igraph : Network Analysis and Visualization. R package version 1.0.0*. Retrieved from https://cran.r-project.org/web/packages/igraph/igraph.pdf.

Feinerer, I., & Hornik, K. (2017). *tm : Text Mining Package. R package version 0.7-2.* Retrieved from https://cran.r-project.org/web/packages/tm/tm.pdf.

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience, 21*(11), 1129-1164.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library trends, 48*(1), 133-159.

Hunt, D., Inman-Semerau, L., May-Pumphrey, M. A., Sussman, N., Grandjean, P., Newhook, P., Suarez-Ordonez, S., Stewart, D., & Kumar, T. (2018). *Selenium Documentation*. Retrieved from http://www.seleniumhq.org/docs.

Jeon, H. (2016). *KoNLP : Korean NLP Package. R package version 0.80.1*. Retrieved from https://cran.r-project.org/web/packages/KoNLP/KoNLP.pdf.

Ji, E. L., Won, H. H., Min, G. S., Son, W. S., Kim, T. I., Lee, S. G., & Kang, C. H. (2014). *A study on evaluation methods for middle school free semester system.* Republic of Korea: Korean Educational Development Institute.

Kang, Y. H., & An, S. H. (2015). An evaluation of 2014 Seoul free semester system based on CIPP model: focusing on middle school career education and guidance on high school entrance. *The Journal of Research in Education, 28*(4), 195-232.

Kim, D. S. (2017). Difference of career maturity, core competencies, and school satisfaction according free learning semester. *The Journal of Curriculum and Evaluation, 20*(3), 101-121.

Kim, E. Y., Kim, K. A., Kim, B. K., Yang, H. J., Lee, S. E., Choi, S. D., & Cha, S. H. (2016). *Mid- and long-term approach to development of the free-semester program: a focus on expanding the program to all school levels*. Republic of Korea: Korean Educational Development Institute.

Kim, J. C. (2017). Reflections on the free semester program. *Korean language education research, 52*(3), 5-35.

Klapka, O., & Slaby, A. (2016). nVidia CUDA platform in graph visualization. In S. Kunifuji, G. A. Papadopoulos, A. M. Skulimowski, & J. Kacprzyk (Eds.), *Knowledge, Information and Creativity Support Systems: Selected Papers from KICSS' 2014-9th International Conference* (pp. 511-520). Limassol: Cyprus.

Koo, K. H. (2017). *The development of evaluation scale for middle school free semester activity program*. Unpublished doctor's thesis, Busan University, Busan, Republic of Korea.

Lertnattee, V., & Theeramunkong, T. (2004). Effect of term distributions on centroid-based text categorization. *Information Sciences, 158*, 89-115.

Ministry of Education. (2013). *Middle school free semester policy pilot operation plan*.

Ministry of Education. (2015). *Middle school free semester policy operation plan(draft).*

Nam, I. Y., & Park, H. W. (2007). Network Analysis of Headlines in the Newspaper Articles on the Prospective Presidential Candidates and their PR Strategy in Korea. *Korean Party Studies Review, 6*(1), 79-108.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences, 103*(23), 8577-8582.

Paranyushkin, D. (2011). *Identifying the pathways for meaning circulation using text network analysis*. Berlin: Nodus Labs. Retrived http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis.

Park, D. M. (2016). Automated time series content analysis with news big data analytics: analyzing sources and quotes in one million news articles for 26 years. *Korean Journal of Journalism & Communication Studies, 60*(5), 353-407.

Park, H. Y. (2015). A critical discourse analysis regarding the tensions between the identities on the application of the free semester program. *The Journal of Curriculum Studies, 33*(3), 1-27.

Park, K. Y., Joo, Y. H., & Ko, Y. J. (2014). The direction and task of the free semester system. *Korean Journal of Educational Administration, 32*(1), 25-51.

Richardson, L. (2017). *Beautiful Soup Documentation: Release 4.4.0*. Retrieved from https://media.readthedocs.org/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf.

Shin, C. K. (2017). An analysis of the social network change characteristics of students with a free-semester program experience. *The Journal of Yeolin Education, 25*(2), 43-63.

Shin, C. K., & Park, M. J. (2015). The case study of three schools with free-semester policy. *Korean Journal of Educational Administration, 33*(1), 309-338.

Shin, C. K., Kim, E. Y., Hwang, E. H., Song, K., & Park, M. (2015). *A study on improvement measures for the free semester program: based on the linkage between free semester and regular semester*. Republic of Korea: Korean Educational Development Institute.

Yu, Y. L., & Baek, S. G. (2017, June). *Text network analysis on '2015 revised national curriculum' policy*. In M. H. Kim (Chair), 2017 Annual Conference conducted at the meeting of the Korean Society for Educational Evaluation and the Korean Educational Research Association, Daegu, Republic of Korea.

Authors

Shin, Anna
Seoul National University, 1st author
anna017@snu.ac.kr

Baek, Sun-Geun
Seoul National University, corresponding author
dr100@snu.ac.kr
Yu, Ye-Lim
Korean Educational Development Institute
yelimyu@kedi.re.kr

Kim, Yun-Kyung
Seoul National University
nadia1324@snu.ac.kr