

# Big-data Analytics: Exploring the Well-being Trend in South Korea Through Inductive Reasoning

Younghan Lee<sup>1</sup>, Mi-Lyang Kim<sup>2\*</sup>, and Seoyoun Hong<sup>2</sup>

<sup>1</sup> Mississippi State University  
Mississippi State, MS 39762 USA  
[e-mail: yl690@msstate.edu]

<sup>2</sup> Soonchunhyang University  
Asan-shi, Choongnam, South Korea  
[e-mail: mlkim@sch.ac.kr; ghdhsy0617@naver.com]

\*Corresponding author: Mi-Lyang Kim

*Received March 10, 2021; revised May 4, 2021; accepted May 26, 2021;  
published June 30, 2021*

---

## Abstract

To understand a trend is to explore the intricate process of how something or a particular situation is constantly changing or developing in a certain direction. This exploration is about observing and describing an unknown field of knowledge, not testing theories or models with a preconceived hypothesis. The purpose is to gain knowledge we did not expect and to recognize the associations among the elements that were suspected or not. This generally requires examining a massive amount of data to find information that could be transformed into meaningful knowledge. That is, looking through the lens of big-data analytics with an inductive reasoning approach will help expand our understanding of the complex nature of a trend. The current study explored the trend of well-being in South Korea using big-data analytic techniques to discover hidden search patterns, associative rules, and keyword signals. Thereafter, a theory was developed based on inductive reasoning – namely the *hook, upward push, and downward pull* to elucidate a holistic picture of how big-data implications alongside social phenomena may have influenced the well-being trend.

---

**Keywords:** Apriori algorithm, Big-data analytics, Degree of visibility, Inductive reasoning, Keyword emergence map, Well-being.

---

This research was supported by a research grant from the National Research Foundation of Korea Grant, the Korean government [MOE][NRF-2017S1A2A2039405]. This research was partially funded by Soonchunhyang University.

## 1. Introduction

The current study aims to unravel the force that influenced the trend in people's interest and involvement related to well-being in South Korea. To achieve this goal, we examine big-data using the Apriori algorithm that is used to extract association rules for discovering knowledge [1] [2] and the Degree of Visibility (DoV) to detect the momentum of keywords [3] [4]. The implications are then drawn from the empirical data based on an inductive reasoning approach, that is, in rather complicated situations and beyond a certain level of complexity of information, as in the case of big-data [5], such approach as one of the predictable methods is used for more effective argumentation and evaluation of the knowledge gained [6] [7]. The implications of the current study contribute to the body of knowledge on several occasions.

First, the significance of well-being is well noted in terms of how it leads to the overall quality of life (e.g., health outcomes and life satisfaction) for individuals [8] [9]. The quality of life is a major issue in South Korea since the country ranks at the bottom tier in each of the work-life balance, perceived health, and life satisfaction category amongst the Organizations for Economic Co-operation and Development (OECD) countries [10]. Therefore, it is important to understand what people associate with well-being when searching for relevant information online [11]. Equally important is figuring out what particular social phenomenon caused a change in search behavior over time that can be portrayed as continued or discontinued interest and involvement related to well-being. These efforts may help policymakers and community leaders to effectively implement tactics and strategies to sustain continued levels of interest and involvement of well-being that leads to improving the overall quality of life.

Second, the application of inductive reasoning utilizing big-data has groundbreaking potential for observing, monitoring, and detecting anomalies and patterns of social activities related to well-being. Today's societies are faced with complex and changing social environments, wherein it becomes extremely difficult to detect social phenomena that impact the trend in searching for well-being information online. The availability of big-data potentially avoids such difficulties by using a data-driven selection of relevant predictors (i.e., inductive inquiry) that either push the trend upward or pull downward. This predictive nature of inductive reasoning may further provide grounds for anticipating future trends in well-being.

Third, and finally, the inductive reasoning approach employed in this study may contribute to the theory-building process of investigating the well-being phenomenon in which induction occurs before theory framing. That is, the study outcomes obtained through induction can later be tested deductively for theory development as a result of its foundation in data [6] [12]. This interplay between theory and data, which is a form of systematic qualitative inquiry [13] [14] can generate more accurate and interesting results related to well-being studies in general.

## 2. Method

### 2.1 Keyword search protocol

It was determined that well-being emerged as a popular topic in early 2000 after a preliminary analysis and several rounds of in-depth discussions between scholars and industry experts within the domains of leisure, recreation, and sport management. As a result, the period considered to search for well-being was established between 2000 to 2019. The next step was to form the baseline of keywords associated with well-being to generate a smaller set of

primary keywords that are most frequently searched while searching for well-being. Based on the expert discussions, a total of eight keywords (i.e., LOHAS, wellness, sohwakhaeng, YOLO, hygge, happiness, healing, and health) were selected as the baseline for initial crawling. The crawling results generated a total of thirty-eight additional keywords for further crawling and frequency analysis.

## 2.2 Data collection

A simultaneous crawling thread was conducted by searching the application programming interface (API) of Naver.com utilizing Chrome web-driver, and BeautifulSoup and Selenium techniques based on python programming. Web contents (i.e., Blogs, Knowledge Q&As) that contained any of the thirty-eight keywords from the initial crawling result were collected based on the published dates. Thereafter, a numerical value of 1 was assigned to the content that included any of the thirty-eight keywords and 0 for the content that had none of the associated keywords. A total of 2,672,692 data were collected that consisted of forty-four categories based on whether the web contents included any of the thirty-eight keywords, and the five primary keywords for each year in addition to the URL, publish date, search type, blog title, and Knowledge Q&A title.

### 2.2.1 Experimental environment

The following information explains the hardware and library package that were used to analysis the data. Computer specifications; Windows 10 (operating system), AMD Ryzen 7-2700 eight-core processor 3.20 GHz, 16 GB RAM, 64-bit system, and NVIDIA GeForce GTX 1050 Ti. Library and package; R (3.6.0 version) for Apriori algorithm (arules), Interactive plot (visNetwork), Parallel coordinates plot (ggplot), Degree of Visibility (ggplot), and Keyword emergence map (ggplot).

## 2.3 Analysis

### 2.3.1 Apriori algorithm

One of the most well-known algorithms to find the association rules between two or more keywords is the Apriori algorithm due to its exploratory nature and ease of execution [15] [2]. This approach is considered effective in terms of discovering the frequent itemsets and thus generating the association rules in a two-step process. The first step requires the algorithm to scan the database to identify the itemsets that satisfy the predefined minimum support. Thereafter, the association rules are generated based on the predetermined minimum confidence. The *support* indicates how frequently a combination of antecedent and consequent of a rule appears together in the database [16]. The *confidence* portrays the strength of the rule by estimating the probability  $P(A|B)$ , that is the portion of cases wherein the consequent appears given that the antecedent has appeared [17]. The following equations depict the estimation of *support* and *confidence*.

$$\text{support}(A \rightarrow B) = P(A \cap B) = \frac{|A \cup B|}{|D|}, \quad (1)$$

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \rightarrow B)}{\text{support}(A)} = \frac{P(A \cap B)}{P(A)} \quad (2)$$

The association rule needs to satisfy the predetermined minimum thresholds  $\alpha$  and  $\beta$  given that  $\text{support}(A \rightarrow B) \geq \alpha$  and  $\text{confidence}(A \rightarrow B) \geq \beta$ . In this study, the minimum threshold for *support* and *confidence* is set at  $\alpha = .01$ , and  $\beta = .7$ . That is, only the rules above the 1% threshold for  $\alpha$ , and 70% threshold for  $\beta$  will be observed [18].

In such a case that the algorithm generates a large set of rules that satisfies the predetermined thresholds for both  $\alpha$  and  $\beta$ , *confidence* fails to accept the baseline frequency of the

consequent, thus making it inadequate [19]. To overcome such limitations, a measure known as the *lift* is used. The *lift* indicates the probability of how much  $\beta$  will increase if A occurs [19]. When  $lift(A \rightarrow B) > 1$ , then a positive interdependence exists between the antecedent and consequent, which the rule is deemed valuable. When  $lift(A \rightarrow B) < 1$ , then a negative interdependence exists between the antecedent and consequent. When  $lift(A \rightarrow B) = 1$ , then A and B are independent with no correlation. The higher the *lift* value, the more meaningful the generated rules.

### 2.3.2 Degree of visibility (DoV)

The degree of visibility is the statistical representation of the signal level of a future sign that measures the degree of a keyword in a data set based on its occurrence frequency [6]. To establish the degree of visibility, a value is set by the ratio of the number of occurrences and the total number of articles. Thereafter, a multiplying factor (i.e., weight) is applied to the most recent occurrence, contributing a different weight for every period [4]. The DoV of keyword  $i$  in period  $j$  is defined as:

$$DoV_{ij} = \left( \frac{TF_{ij}}{NN_j} \right) \times \{1 - tw \times (n - j)\}$$

$TF_{ij}$  is the total occurrence frequency of a keyword  $i$  in the period  $j$ ,  $NN_j$  is the total number of articles in the period  $j$ , whereas  $n$  is the number of periods, and  $tw$  is the time-weight [20].

### 2.3.3 Keyword emergence map (KEM)

The visualization of the DoV can be accomplished through keyword emergence mapping. This technique is used to detect weak and strong signals in which the x-axis is represented by the average time-weighted increasing rate and the y-axis is represented by the absolute average term frequency of each keyword [6]. The KEM is divided into four quadrants based on the median value. The high-left (A) and high-right (B) quadrants indicate *weak signals* and *strong signals*, respectively [6]. According to [21], weak signals are topics that have an abnormal pattern and are also rarely exposed to the receivers. In contrast, strong signals have the potential to become a trend since the topic pattern is relatively more stable and further exposed. This entails that weak signals may not be considered as part of a future trend, however, a clue for identifying how a trend will be formed in the future [22].

The remaining bottom quadrants; low-left and low-right, refer to *latent signals* and *well-known but not so strong signals*, respectively [22]. In essence, Keywords with *latent signals* are considered to have below-average share and growth rate, whereas the ones that are well known but with mediocre signals have above-average share and low growth rate.

### 2.3.4 Inductive reasoning

While deductive reasoning (i.e., top-down, theory-driven) relies on testing a single theory for empirical adequacy, that is testing a priori hypothesis [23], inductive reasoning is a bottom-up, data-driven, approach based on drawing general inferences from particulars or cases of empirical data [24] — in other words, drawing inferences about underlying patterns based on observations is the distinguishing trait of inductive reasoning. Likewise, the application of big-data analytics is defined by evaluating information based on identifying relations between measured variables [7].

Specifically, big-data analytics typically does not involve theories or hypotheses about the underlying relations. Rather, the use of observed patterns in a data-driven approach guides future decisions [7]. The depth and wealth of information in big-data, that is, a significantly

large number of variables of unspecified theoretical establishment, make big-data analytics fall within the framework of inductive inquiry [7]. The theoretical models explored in an inductive reasoning approach can then be further tested using deductive methods.

### 3. Results and Discussion

#### 3.1 Frequent itemset mining based on apriori algorithm

The apriori algorithm analysis was conducted to determine the association rules of the keywords related to well-being. Based on a total of 60,743 items from the web contents, 142 rules were discovered in terms of support and reliability thresholds of above .01 and .7, respectively. **Tables 1 and 2** indicate the top 10 association rules for each period. The analysis results entail that, in general, people associate health with well-being at a 100 percent confidence rate regardless of the period. The keyword with the highest support and reliability was also *health*. The lift values of 2.31 and 2.55 for both periods further imply that the probability of people thinking about well-being when associated with health is approximately two times higher than when it is not. Other noticeable keywords with similar values about *health* include *exercise* and *culinary*. Meanwhile, the third rule in **Table 1** indicates that people imagine well-being almost twice as much when three keywords, *health*, and *happiness* are associated together than when they are not. The remaining rules can be interpreted similarly.

**Table 1.** Association rules analysis results (Phase I)

No	LHS	RHS	Support	Confidence	Lift	Count
1	Health	Well-being	.24	1	2.31	3431
2	Health; Recent	Well-being	.051	1	2.31	786
3	Health; Happiness	Well-being	.06	1	2.31	910
4	Health; Mind	Well-being	.06	1	2.31	886
5	Health; Consume	Well-being	.06	1	2.31	883
6	Health; Pursue	Well-being	.07	1	2.31	957
7	Health; Exercise	Well-being	.07	1	2.31	942
8	Health; Heart	Well-being	.06	1	2.31	915
9	Health; Nature	Well-being	.07	1	2.31	938
10	Health; Interest	Well-being	.08	1	2.31	1063

Note. LHS = Left hand side Antecedent; RHS = Right Hand side Consequent

Overall, the keyword with the highest support and confidence turned out to be *health* between 2000 and 2019. In other words, people searched for well-being about 2 times more when associated with *health* alone than when it was not (or associated with more than two keywords) for two decades. Interestingly, between 2011 and 2019, keywords such as *culinary*, *exercise*, *idea*, and *community* emerged as new associations to well-being without being part of the *health* keyword combination. Further information is indicated in [Table 2](#).

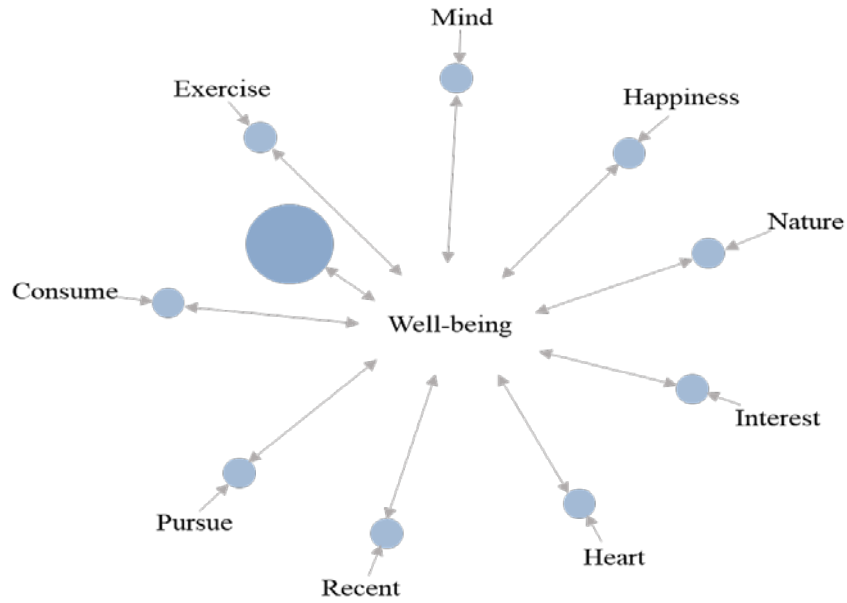
**Table 2.** Association rules analysis results between (Phase II)

No	LHS	RHS	support	confidence	Lift	count
1	Health	Well-being	.19	1	2.55	8802
2	Health; Culinary	Well-being	.07	1	2.55	3494
3	Health; People	Well-being	.06	1	2.55	2967
4	Health; Ourself	Well-being	.07	1	2.55	3498
5	Health; Community	Well-being	.08	1	2.55	3760
6	Health; Idea	Well-being	.08	1	2.55	3834
7	Culinary; Community	Well-being	.05	.92	2.36	2552
8	Idea; Culinary	Well-being	.06	.91	2.31	2769
9	Exercise	Well-being	.08	.90	2.31	3710
10	Culinary	Well-being	.13	.88	2.25	5957

Note. LHS = Left hand side Antecedent; RHS = Right Hand side Consequent

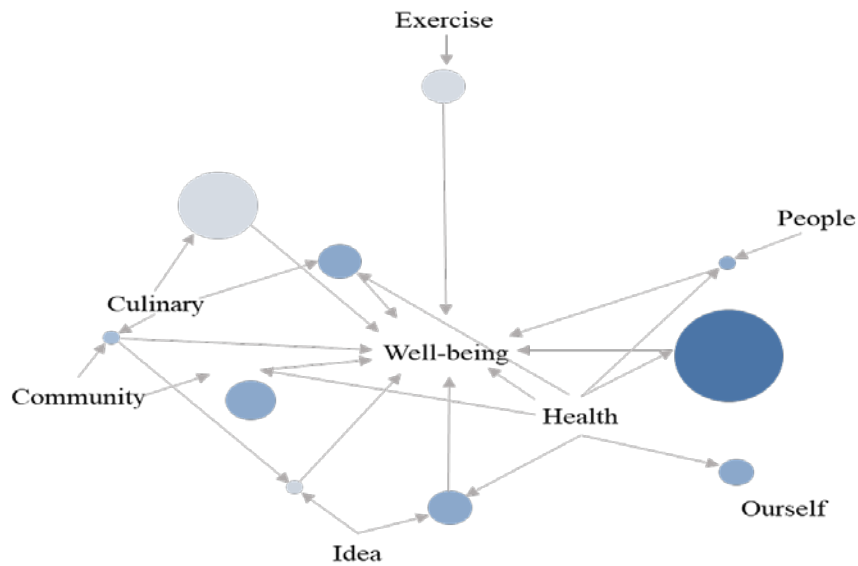
### 3.2 Interactive plot

[Fig. 1 and 2](#) are constructed to visualize the associative rules based on the top 10 rules. The circle size indicates support, the color depth represents the lift, and the distance between the circles refers to the degree of association, meaning the closer the distance the stronger the association. In this study, the depth (2.42375) of the circle colors is identical and the largest circle represents *health* followed by *culinary* and *exercise*. This entails that, there is a 10.86% chance that *health* will appear with well-being and people think of well-being twice as more when associated with *health* than when it is not. *Health* is also connected to four other circles, which means that this keyword is affiliated with four rules. While [Fig. 1](#) depicts a relatively normal plot pattern, [Fig. 2](#) displays a rather abnormal pattern in which the scattered keywords associated with well-being represent a relatively more unpredictable support and lift structure.



Note. Size = support (.056-.246); Color = lift (2.314-2.314)

Fig. 1. Interactive plot for top 10 rules (Phase I)

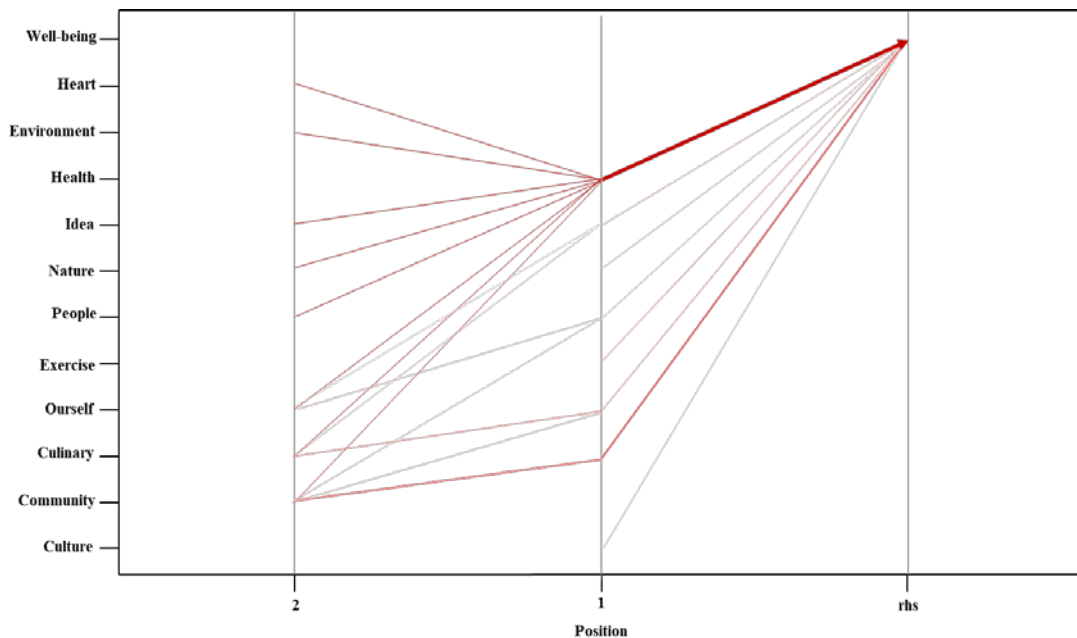


Note. Size = support (.055-.188); Color = lift (2.247-2.551)

Fig. 2. Interactive plot for top 10 rules (Phase II)

### 3.3 Parallel coordinates plot

**Fig. 3** depicts the parallel coordinates plot for the top 20 rules in which the width and color of the line between positions 1 and rhs represent the support and reliability of a particular keyword in association with the referenced keywords within the plot; the wider the width the stronger the support, and the darker the color the higher the reliability. The domain between positions 2 and 1 indicates the association between the keywords; more associations lead to a wider and darker line from position 1 to rhs. For example, *health* consists of three associated keywords (i.e., *feeling*, *ourself*, and *culinary*), which generated the widest and the darkest line implying that it has the strongest support and the highest reliability of the associated keywords toward well-being. Other noticeable keywords that showed relatively strong support and high reliability are *environment*, *idea*, *nature*, *culinary*, and *community*.



**Fig. 3.** Parallel coordinates plot for 20 rules

### 3.4 Degree of visibility (DoV)

**Table 3** depicts the average increase rate and the average frequency of each keyword involved in the search between 2004 and 2019. The information in **Table 1** indicates that the average increase rate for the majority of the keywords is a negative number. This entails that the trend in searching for well-being is generally in a decline. However, the effect on the overall trend may not be as significant as the numbers indicate due to the minuscule overall decline rate.

Specifically, the change in the average increase rate and the average frequency each year between 2004 and 2010 for every keyword except *culinary*, are modest; thus, the trend was relatively steady during that period. Meanwhile, a relatively more drastic trend pattern was recognized between 2010 and 2019 as the variance of each keyword in terms of the degree of visibility is quite significant. The keywords associated with well-being, in particular, increased at a high rate starting from 2015 until it hit a turning point in 2019 and started to decline. The fluctuating trend during this period implies that significant social events that are both relevant and irrelevant to well-being must have occurred that affected the overall trend.



**Table 3.** Degree of visibility, increasing rate, and average term frequency (top 10 keywords)

Keyword	Degree of Visibility (2004-2019)																Inc. Rate	Avg. frequency
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019		
Well-being	11714	8232	4404	2985	2938	4240	5110	5768	5490	12050	11186	13271	18577	18617	28090	7729	-.03	10025
Health	2709	2319	1200	1000	984	1079	1709	1575	1683	3024	3031	3603	4645	4442	5944	2827	-.02	2610
People	2185	1588	1122	834	741	1003	1574	1183	1035	2018	1754	2336	2392	2870	5393	1928	-.03	1872
Idea	1508	1379	1276	919	729	989	1462	1488	1099	2557	2426	2864	2899	3457	6852	2397	.02	2143
Exercise	1429	1028	638	598	635	422	576	503	413	1028	682	956	2170	2799	3392	2595	.04	1241
Mind	1374	745	260	186	202	221	399	242	217	436	391	556	668	781	714	677	-.07	504
Nature	1292	663	460	291	385	471	672	694	636	1106	1175	1628	1878	1567	3008	882	-.03	1050
Culture	1252	700	807	640	596	749	471	553	478	718	752	821	946	1344	1354	793	-.04	810
Ourselves	1228	1010	903	692	550	1201	1479	1054	1089	2060	2171	2346	2223	2978	5050	1073	.00	1694
Pursue	1225	788	491	235	343	196	459	191	255	363	231	316	307	644	1105	548	.01	481

### 3.5 Keyword emergence map (KEM)

Figures 4 and 5 represent the keyword emergence map generated based on the degree of visibility (DoV) equation for keyword  $i$  during period  $j$ , where  $NN$  denotes the total number of references,  $n$  the length of the period, and  $tw$  a pre-determined time weight (.05). The purpose of the map is to detect the keywords that are gaining momentum and those that are disappearing from the discourse. **Table 4** is the summary of the keyword signals in terms of strength (i.e., weak or strong) and state (i.e., latent or stagnate).

**Table 4.** Summary of keyword future signals

Phase	Strong (Q1)	Weak/Emerging (Q2)	Latent (Q3)	Well-known/Stagnant (Q4)
I	Ourselves, culinary, sustainable environment, LOHAS.	Happiness, space, travel, service, sports, experience, feeling, card, camping, healing.	Interest, recent, style, life, staying fit, daily life.	Well-being, health, people, mind, exercise, idea, nature, culture, merchandise, environment, heart, consume.
II	Travel, exercise, feeling, interest, space, staying fit, camping, card, healing.	Merchandise, happiness, consume, pursue, style, daily life, sports, YOLO.	Sustainable environment, service, environment, life, LOHAS, mind recent.	Well-being, health, idea, culinary, people, ourselves, experience, nature, culture, heart.

**Fig. 4** shows a graph view of the KEM with the keywords detected between 2000 and 2010. The most noticeable keywords with strong signals include *ourselves*, *culinary*, *sustainable environment*, and *LOHAS*; whereas, *happiness*, *travel*, *sport experience*, *camping*, and *healing* are considered to have weak signals. While the keywords in the strong signal quadrant may drive the current trend in well-being, the ones in the weak signal quadrant such as *camping*,

*healing*, *sport*, and *experience* possess emerging signals that can potentially become strong signals that contribute to the future trend. The main keywords in the fourth quadrant are *well-being*, *health*, *exercise*, *nature*, *culture*, and *environment*, which are well-known but with stagnant signals, that is, these are pre-established words with little room for growth because of the significant amount of time they have been exposed. In contrast, the keywords with latent signals such as *daily life*, *up-to-date*, and *style* have the least amount of exposure. These keywords are somewhat concealed or dormant until circumstances become suitable for development in the future in terms of being part of the well-being trend.

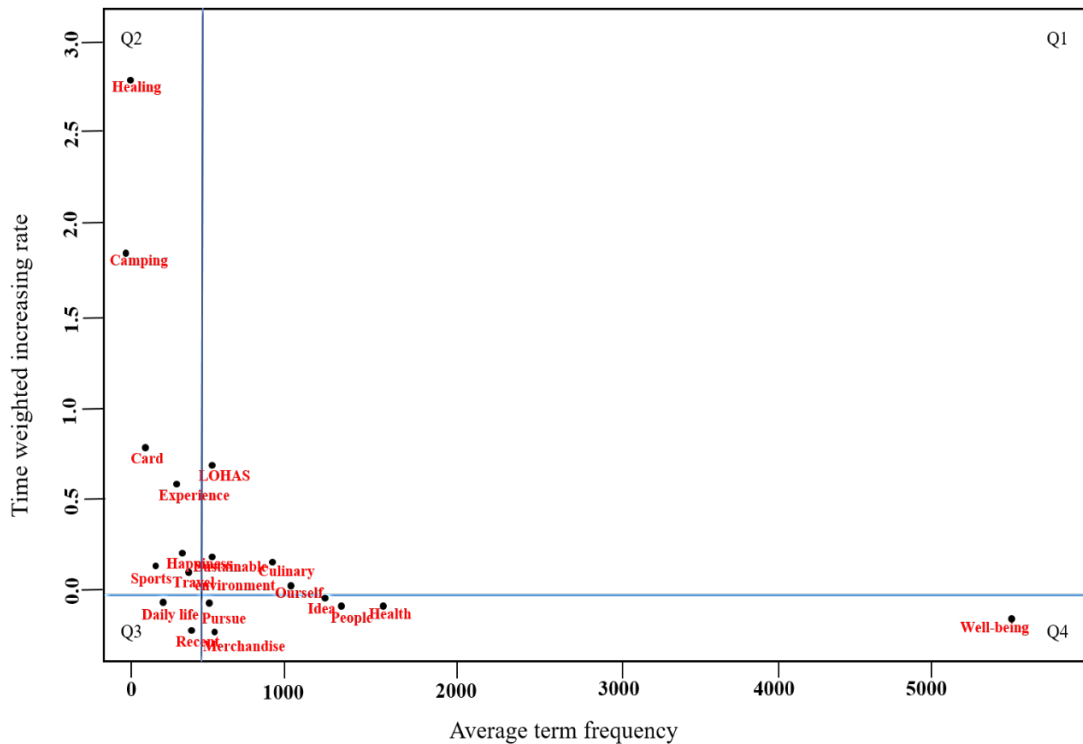
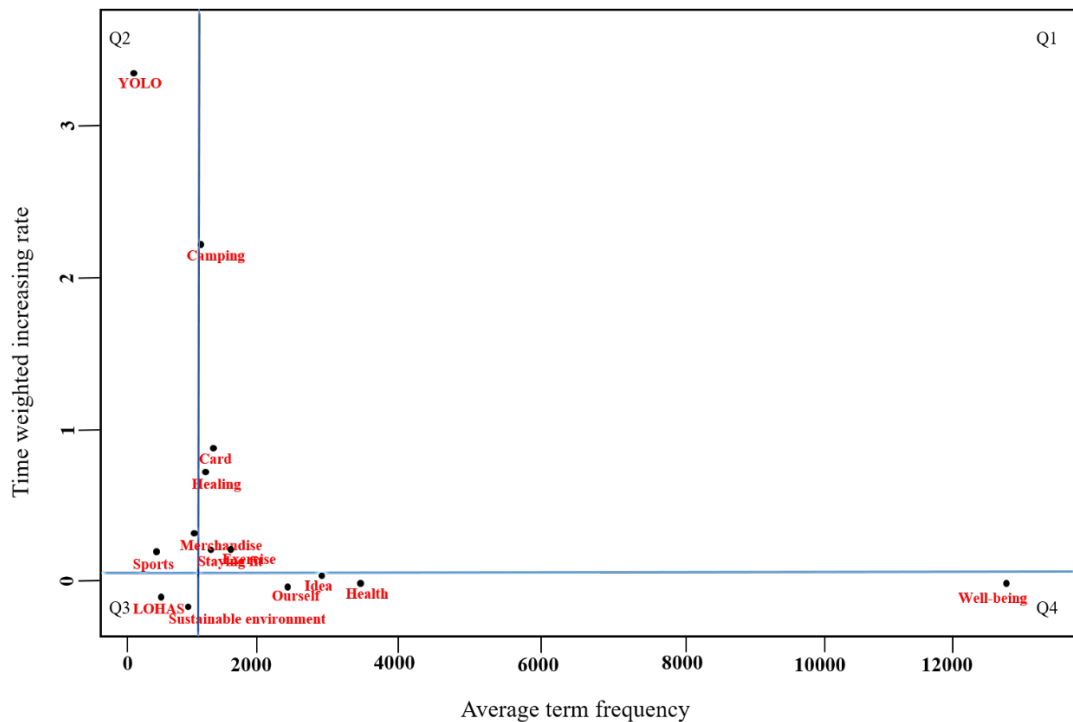


Fig. 4. Keyword Emergence Map (Phase I)

Fig. 5 depicts the KEM based on the keywords detected between 2011 and 2019. This figure also helps understand the flow of the keywords from one quadrant to the other compared to the earlier period represented in Fig. 4. The most notable change is that several keywords that were located in the weak signal quadrant in Fig. 4 such as *camping*, *healing*, and *travel* have moved to the strong signal quadrant in Fig. 5. This supports the notion that weak signal keywords have the potential to become strong signals over time [22] [23]. Another noteworthy change is that new keywords with emerging signals have appeared in Fig. 5, that is *YOLO* and *style* have the potential to become a major part of the well-being trend in the near future. Also of note is that there is little change in the composition of the keywords in the fourth quadrant regardless of the periods. This implies that the keywords well-known with stagnant signals are indeed pre-established with a strong footing in the formation of an already established trend [22] [23]. In this sense, the root of the well-being trend can be represented by several keywords such as *health*, *nature*, *exercise*, and *culture*.



**Fig. 5.** Keyword Emergence Map (Phase II)

### 3.6 Inductive reasoning

During the initial stage of the inductive reasoning process, that is; observing the data analysis results, the researchers noticed that the data can be divided into two phases based on the big-data implications. The researchers then further observed the social phenomena that occurred during each phase to make sense of which incidents may have influenced the different approaches in searching for well-being. Thereafter, a theory was developed that can effectively explain the well-being trend in South Korea.

#### 3.6.1 Phase I: The conceptual establishment of well-being

In the earlier phase between 2000 and 2010 the relatively more frequently searched keywords associated with well-being such as *health*, *nature*, *culture*, and *happiness* are rather abstract and conceptual terminologies that insinuate well-being was being understood as a concept that relates to the quality of life. As can be seen in the keyword emergence map during this phase (see [Fig. 4](#)), the first and second quadrants are mostly filled with primary keywords such as *happiness*, *space*, *experience*, *LOHAS*, *sustainable environment*, *feeling*, and *ourself* that depict an abstract concept of well-being. The information in the interactive plot also supports this phenomenon. As noted in [Fig. 1](#), the primary keywords that are rather conceptual terminologies such as *health*, *happiness*, *mind*, *nature*, and *mental* form a normal plot pattern centered on well-being, which reflects a relatively more stable structural support.

#### 3.6.2 Phase II: The utilitarian approach in achieving well-being

During the second phase, people began searching about the elements of well-being not only to understand its concept but also to practice well-being with a utilitarian approach, that is, people

were now interested in the knowledge and skills required to improve well-being more effectively. This shift in search behavior from simply understanding the meaning of well-being to forming a lifestyle to successfully achieve well-being is recognized by the keyword search patterns (e.g., *exercise*, *culinary*, *travel*, and *camping*) in association to well-being that indicate action tendencies toward practicing well-being.

Another evidence that supports the utilitarian approach during this phase can be found in the KEM, **Table 4**, in which many of the keywords that describe a certain action related to well-being have moved to either first or second quadrants in Phase II from where they were originally identified as either latent, weak, or stagnant signals in Phase I. For example, *daily life* in the third quadrant that used to be latent in Phase I is now located in the second quadrant as one of the emerging keywords. Also, a few keywords with weak or stagnant signals in Phase I are now found in the first quadrant in Phase II such as *staying fit*, *travel*, *camping*, *exercise*, and *healing* that have emerged as keywords with strong signals. These newly recognized strong signal keywords are especially important to highlight since the ones in the first quadrant are known to drive a trend in a particular direction [22] [23]. As in the case of well-being, it seems that people were moving toward understanding well-being as something that can be achieved through a variety of life activities.

### 3.6.3 The Hook, Upward Push, and Downward Pull

The *hook* refers to a noteworthy social event or incident that facilitates the beginning of a search trend for a particular topic. In this sense, the South Korean government hinted at the possibility of implementing the five-day workweek policy in early 2000, which was gradually officialized in 2006 [25]. During this time was when discussions throughout the nation took place about utilizing the extra leisure time toward improving well-being that perhaps led to the active search behaviors of people trying to grasp the meaning of the well-being concept for the first time; hence, the frequented conceptual terminologies related to well-being. This government policy acted as the *hook* (see **Fig. 4**) in phase I that initially induced people's interest toward understanding the well-being concept and created a foundation to push the search trend upward in search of knowledge and skills to achieve the state or condition of well-being at the start of phase II.

The *upward push* is established upon a series of coherent events that extends the search behaviors initiated by the *hook* and pushes the momentum of the search trend upward to a certain extent. As in the case of well-being, the search trend gained significant upward momentum when the Korean government subsequently implemented the five-day school week policy in 2012, which allowed students of all ages to have more leisure time that can be dedicated toward improving their well-being [26]. Furthermore, when the *upward push* is triggered, it is more likely that the nature of the search behavior is gradually elevated from being aware of a particular social phenomenon to forming a higher level of interest toward the elements of the phenomenon (see **Fig. 6**).

Meanwhile, the *downward pull* refers to the condition when the upward momentum stalls and triggers a downward momentum, that is to say, people losing interest in continuously searching for information related to the topic originally hooked on. The *downward pull* is caused by a series of whether distressful (e.g., national crisis), amusing (e.g., celebrity gossip), alarming (e.g., a pandemic outbreak of a disease) or entertaining (e.g., Olympic games) events significant enough to suppress the upward momentum and instantly start a downward momentum in a search trend. For instance, in 2014, the ferry Sewol carrying 476 people on board sank off on the Southwest coast of South Korea, killing more than 300 people [27]. Also, the Middle East Respiratory Syndrome (MERS) outbreak took place in 2015 [28]. These

events amongst many others that occurred during phase II, when people were interested in finding specific information related to improving their well-being with a utilitarian approach, triggered a *downward pull* that undermined the upward momentum to practice well-being (see Fig. 6).

On the other hand, an *upward push* was triggered soon after the South Korean government announced a plan to promote leisure activities nationwide in 2016 and later enacted the national citizen leisure development Act in 2017. The trend in the search behaviors related to well-being peaked in 2018 as the government implemented the 52 hour-workweek policy and once again a *downward pull* was triggered starting in 2019 (Fig. 6).

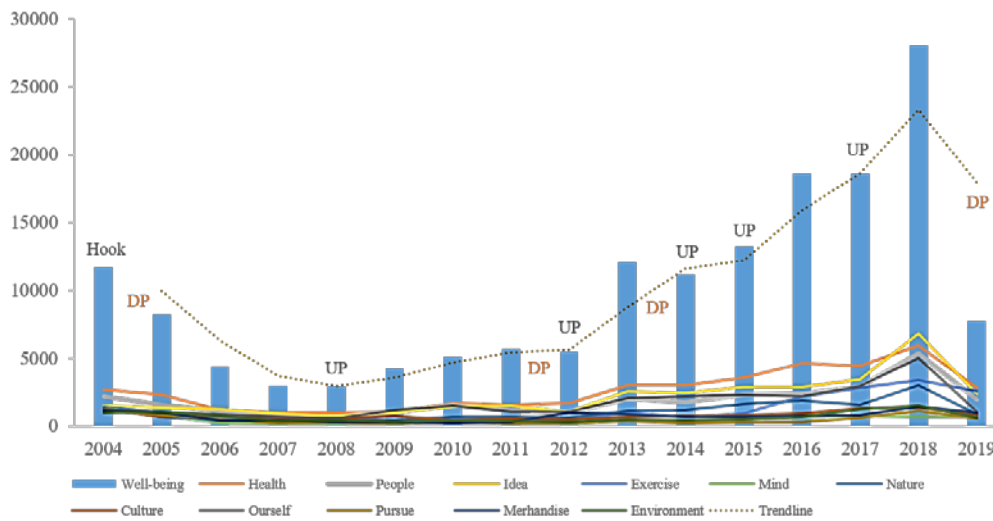


Fig. 6. The holistic view of the well-being trend

#### 4. Conclusion and Future Research

We explored the internet search behaviors related to well-being to discover hidden patterns, association rules, degree of visibility, and potential relationships between big-data implications and social phenomena to explain the well-being trend in South Korea. Big-data analysis results indicated that the trend could be divided into two phases. The first phase represents the period in which well-being was being established as a concept whereas the second phase was about pursuing well-being with a utilitarian approach. The implications were further investigated by additionally observing social phenomena during each phase. A theory was then developed to provide a holistic view on understanding the nature of the well-being trend from its establishment to the revolving up and down movements within the path. The implications of the current study may expand beyond the boundary of South Korea.

First, while former studies have made noble attempts in defining the concept and qualities of well-being, not much work has been done on a holistic level. For instance, the associative keywords identified in the current research imply that perhaps well-being is not merely a unidimensional concept but also a utility that consists of various qualities or attributes. Therefore, future studies can use this information to redefine or revise how we understand well-being with a more comprehensive approach and revisit the following questions. What is well-being, what factors consist of well-being, and how it can be practiced? This effort will

expand the body of knowledge in terms of understanding the full spectrum of the well-being concept.

Second, we derived the conclusion of the current research based on the inductive reasoning process, that is supporting big-data with further observations. This bottom-up, data-driven, approach allowed us to identify patterns in big-data without forming hypotheses. In return, new perspectives and hidden insights were gained that otherwise might not have been possible if based on a deductive process [24]. We encourage scholars to reevaluate the research methods associated with big-data. That is to say, utilizing inductive reasoning can add new dimensions of interpretations to the study outcome.

Finally, in terms of recognizing and matching patterns, we took the liberty to simplify the problems of complication by constructing a schema to work with, which a model was formed to describe certain patterns, actions, and behaviors; hence, the establishment of the *hook*, *upward push*, and *downward pull*. We used this simple model to fill the gaps in our understanding related to the well-being trend in South Korea. Future studies are warranted in terms of empirically testing the generalizability of the current model in a variety of contexts. That includes the application of an integrated search system in which various internet-based data platforms such as Twitter, Wikipedia, and Bing Web are utilized for semantic searching. This effort will help researchers to make more reasonable inferences about the exploratory nature of analyzing big-data [29] [30]. We also encourage to refine the current model or seek alternative theories by testing the study outcomes deductively to provide more meaningful insights related to well-being studies in general.

### Acknowledgement

The authors would like to acknowledge the research support provided by the Research Institute for Sport Convergence (RISC) at Mississippi State University.

### References

- [1] J. Singh, H. Ram, and D. J. Sodhi, "Improving efficiency of apriori algorithm using transaction reduction," *International Journal of Scientific and Research Publications*, vol. 3, no. 1, pp. 1-4, 2013. [Article \(CrossRef Link\)](#)
- [2] Al-Maolegi, Mohammed, and Bassam Arkok, "An improved Apriori algorithm for association rules," *International Journal on Natural Language Computing*, vol. 3, no. 1, pp. 21-29, Feb. 2014. [Article \(CrossRef Link\)](#)
- [3] Griol-Barres, Israel, Sergio Milla, and José Millet, "System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques," in *Proc. of ICAART 2019 - Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, vol. 2, 631–638, 2019. [Article \(CrossRef Link\)](#)
- [4] Yoon, Janghyeok, "Detecting Weak Signals for Long-Term Business Opportunities Using Text Mining of Web News," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12543–12550, 2012. [Article \(CrossRef Link\)](#)
- [5] Y. S. Kim and S. R. Jeong, "Opinion-Mining Methodology for Social Media Analytics," *KSII Transactions on Internet and Information Systems*, vol. 9, no. 1, pp. 391-406, 2015. [Article \(CrossRef Link\)](#)
- [6] McAbee, Samuel T., Ronald S. Landis, and Maura I. Burke, "Inductive Reasoning: The Promise of Big Data," *Human Resource Management Review*, vol. 27, no. 2 pp. 277–290, 2017. [Article \(CrossRef Link\)](#)

- [7] Fogel, David B., Kumar Chellapilla, and Peter J. Angeline, "Inductive Reasoning and Bounded Rationality Reconsidered," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 142–146, 1999. [Article \(CrossRef Link\)](#)
- [8] F. A. Huppert, and J. E. Whittington, "Evidence for the independence of positive and negative well-being: Implications for quality of life assessment," *Br J Health Psychol*, vol. 8, no. 1, pp. 107–122, 2003. [Article \(CrossRef Link\)](#)
- [9] Howell, Ryan T., Margaret L. Kern, and Sonja Lyubomirsky, "Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes," *Health Psychology Review*, vol. 1, no. 1, pp. 83–136, 2007. [Article \(CrossRef Link\)](#)
- [10] How's Life? 2020: Measuring Well-being [Online]. Available: <https://www.oecd-ilibrary.org/economics/how-s-life/volume-issue-9870c393-en>.
- [11] H. S. Hwang, "The Influence of personality traits on the Facebook Addiction," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 30, pp. 1626–1638, 2017. [Article \(CrossRef Link\)](#)
- [12] Eisenhardt, Kathleen M., and Melissa E. Graebner, "Theory Building from Cases: Opportunities and Challenges," *Academy of Management Journal*, vol. 50, no. 1, pp. 25–32, 2007. [Article \(CrossRef Link\)](#)
- [13] Glaser, Barney G., and Anselm L. Strauss, *Discovery of grounded theory: Strategies for qualitative research*, Routledge, 2017.
- [14] O'Reilly, Kelley, David Paper, and Sherry Marx, "Demystifying grounded theory for business research," *Organizational Research Methods*, vol. 15, no. 2, pp. 247–262, 2012. [Article \(CrossRef Link\)](#)
- [15] Hong, Jungyeol, Reuben Tamakloe, and Dongjoo Park, "Discovering Insightful Rules among Truck Crash Characteristics Using Apriori Algorithm," *Journal of Advanced Transportation*, vol. 2020, 2020. [Article \(CrossRef Link\)](#)
- [16] S. Kotsiantis, & D. Kanellopoulos, "Association rules mining: A recent overview," *GESTS International Transactions on Computer Science and Engineering*, vol. 32, no. 1, pp. 71–82, 2006. [Article \(CrossRef Link\)](#)
- [17] Pande, Anurag, and Mohamed Abdel-Aty, "Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool," *Safety science*, vol. 47, no. 1, pp. 145–154, 2009. [Article \(CrossRef Link\)](#)
- [18] Chen, Guoqing et al, "A New Approach to Classification Based on Association Rule Mining," *Decision Support Systems*, vol.42, no. 2, pp. 674–689, 2006. [Article \(CrossRef Link\)](#)
- [19] Lee, Sangdeok et al., "Application of Association Rule Mining and Social Network Analysis for Understanding Causality of Construction Defects," *Sustainability*, vol. 11, no. 3, 2019. [Article \(CrossRef Link\)](#)
- [20] J. Song, Y. Han, K. Kim, and T. M. Song, "Social big data analysis of future signals for bullying in South Korea: Application of general strain theory," *Telematics and Informatics*, vol. 54, pp. 101472, 2020. [Article \(CrossRef Link\)](#)
- [21] Hiltunen, Elina, "The Future Sign and Its Three Dimensions," *Futures*, vol. 40, no. 3, pp. 247–260, 2008. [Article \(CrossRef Link\)](#)
- [22] Lee, Young Joo, and Ji Young Park, "Identification of Future Signal Based on the Quantitative and Qualitative Text Mining: A Case Study on Ethical Issues in Artificial Intelligence," *Quality and Quantity*, vol. 52, no. 2, pp. 653–667, 2018. [Article \(CrossRef Link\)](#)
- [23] H. G. Kim, "On the Leisure Policy of Local Autonomous Entities by the Operation of Five-day Workweek Legislation," *Journal of Leisure and Recreation Studies*, vol. 25, pp. 105–127, 2003.
- [24] Ketokivi, Mikko, and Saku Mantere, "Two Strategies for Inductive Reasoning in Organizational Research," *Academy of Management Review*, vol. 35, no. 2, pp. 315–333, 2010. [Article \(CrossRef Link\)](#)
- [25] Y. J. Lee, "Adolescent's activity needs and policy related Five-Day school week," *Journal of Digital Convergence*, vol. 10, no. 8, pp. 335–340, 2012. [Article \(CrossRef Link\)](#)
- [26] Eisenhardt, Kathleen M., and Melissa E. Graebner, "Theory Building from Cases: Opportunities and Challenges," *Academy of Management Journal*, vol. 50, no. 1, pp. 25–32, 2007. [Article \(CrossRef Link\)](#)

- [27] S. Lee, Y. B. Moh, M. Tabibzadeh, and N. Meshkati, "Applying the AcciMap methodology to investigate the tragic Sewol Ferry accident in South Korea," *Applied ergonomics*, vol. 59, pp. 517-525, 2017. [Article \(CrossRef Link\)](#)
- [28] H. Y. Park, E. J. Lee, Y. W. Ryu, Y. Kim, H. Kim, H. Lee, and S, J, "Epidemiological investigation of MERS-CoV spread in a single hospital in South Korea, May to June 2015," *Eurosurveillance*, vol. 20, no. 25, pp. 1-5, 2015. [Article \(CrossRef Link\)](#)
- [29] Jeong, Seung Ryul, and Imran Ghani, "Semantic Computing for Big Data: Approaches, Tools, and Emerging Directions (2011-2014)," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 6, pp. 2022-2042, 2014. [Article \(CrossRef Link\)](#)
- [30] C. L. P. Chen, and C. Y. Zhang, "Data-intensive applications challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314-347, Aug. 2014. [Article \(CrossRef Link\)](#)



**Dr. Younghan Lee** received his B.S. degree in Marketing from the University of Southern California, Marshall School of Business. Dr. Lee received his Master's and Doctoral degrees in Sport Management from Seoul National University. He is currently an associate professor at Mississippi State University, Department of Kinesiology, Sport Administration. Prior to joining Mississippi State University, Dr. Lee was an assistant professor at the University of Massachusetts, Department of Sport Management. He is currently the director of the Research Institute for Sport Convergence. His research interests include sport marketing, relationship marketing, sport sponsorship, and data analytics in sport.



**Dr. Mi-Lyang Kim** received her B.S. degree from Soonchunhyang University, Department of Sport, Leisure, and Recreation. She received her Master's degree in Statistics and Research Methodology from Sungkyunkwan University, and Ph.D. from Seoul National University majoring in Leisure Studies. Dr. Kim is currently an associate professor in the Department of Sports, Leisure, and Recreation at Soonchunhyang University. Her research interests include serious leisure, leisure career, well-being, and recreation sport.



**Dr. Seoyoun Hong** received her B.S degree (double major) in Dance and Kinesiology from Soonchunhyang University. She received her Master's degree in Physical Education from Ewha Womans University, Graduate School of Education, and Ph.D. from Ewha Womans University, Department of Kinesiology and Sport Studies. Dr. Hong is currently a lecturer at Soonchunhyang University.