International Journal of Institutional Research and Management
International Institute of Applied Informatics
2024, Vol. 8, No. 1, IJIRM836

A Case Study of Using the Structural Topic Model to Grasp Actual Learning through Free-Writing Reports

Tatsuya Tsumagari [*], Yoko Nakazato [†], Takashi Tsumagari [‡]

Abstract

This paper reports the results of an analysis using structural topic model (STM) of career awareness to examine learning in first-year career education courses. The subject of this study was a first-year career education course offered by a required course at public University A in Japan. The free-writing reports of 1,780 first-year students were included in the analysis.

We examined students' career awareness from two perspectives using STM. First, we examined stable career awareness that does not fluctuate from year to year. We identified whether stable career awareness was created by the same syllabus through changes in external factors. Second, we examined the impact of differences in student types on learning.

Fifteen topics were extracted as career awareness. Among those topics, six topics were found to be stable career awareness topics regardless of year. In addition, when students were classified into five types according to their interest in the lectures, eight topics were found to apply to constant career awareness, regardless of their interest. The three topics were independent of both year and student type.

The analysis of students' free-writing reports using STM is useful as an assessment method for grasping the reality of student learning.

Keywords: First-year education, Career education, Learning assessment, Actual learning, Structural topic model

1 Introduction

In developed countries, more than half the population of the same generation has the opportunity to receive higher education. M. Trow called this stage “the universalization of higher education” [1]. As a result of the massification of universities, which has made it possible for anyone who wishes to attend university to have the opportunity to do so, higher education requires a qualitative shift. One such shift is the enhancement of career education.

[*] Takasaki City University of Economics, Gunma, Japan
[†] Kagoshima University, Kagoshima, Japan
[‡] Prefectural University of Kumamoto, Kumamoto, Japan

In Japan, career education at higher education institutions was introduced in 2005. The standards for the university's "system for cultivating the abilities necessary for social and vocational independence" were enforced in April 2011, requiring career education at all higher education institutions [3].

Career education in Japan has been modeled on career education in the United States. In particular, the basic concept was influenced by Super's theory [2,3], who considered a career as a combination of one's age and age-appropriate roles. He also stated that careers are formed by experiencing various roles in the family and society and that they develop throughout life. Careers change dynamically throughout a person's life and are formed in accordance with their own way of life. Being aware of one's career is an important aspect of learning, especially for university students. Previous studies have shown that awareness of how to think about one's own career has a strong influence on student growth [4], and that this awareness is strongly related to student learning [5,6]. From the perspective of student growth, university students' career awareness is crucial.

In addition, we have never been more aware than we are today that we are living in an unpredictable time. In 2015, the Japanese Ministry of Education, Culture, Sports, Science, and Technology used the term "unpredictable era" when announcing its direction for education through 2030 [7]. This educational policy emphasizes living powerfully, even when encountering unpredictable and sudden situations. Under these circumstances, students are expected to shape their lives independently, and career education is becoming even more important.

Based on the abovementioned perspectives, it is desirable for universities to provide career education to students who will eventually join the workforce. Career education during the first year of university is becoming increasingly popular. This is because an increasing number of students are enrolling in university without purpose. Career education at the first-year level should help students understand the concept of a career that they will develop throughout their lives.

What kind of career awareness do university students who take first-year career education courses develop as a result of their learning? Many studies measuring course effectiveness are based on pre- and post-survey comparisons [7]. However, this does not always reveal the reality of student learning. To solve this problem, we attempted a textual analysis of students' free writing to evaluate learning [8].

One text analysis method is topic modeling [9]. A topic model explores multiple subjects (topics) that appear in a document and the words that are closely related to those topics to understand the document. Topic models do not allow for correlations between topics, nor do they allow variables related to topic prevalence or topic content to be incorporated into the model. To solve this problem, structural topic model (STM) was developed.

STM allows for the presence of covariates related to topic prevalence and content, and is an extension of the topic model [10]. It also allows for correlations between topics. STM has been reported as superior to topic models in analyzing the relationships between covariates and topics [11].

As for text analysis using STM, one study analyzed the abstracts of related articles to clarify

the direction of telepsychological research [12]. Our study applied the STM to assess learning among university students. Specifically, we aimed to understand the stable and unique career consciousness of students taking first-year career education courses. We demonstrated that text analysis using STM of students' free texts is a useful method for evaluating their learning.

In the text analysis using STM, we examined the career consciousness generated as stable and unique learning from two perspectives. Career awareness refers to topics extracted by the STM. First, we identified topics that were generated, regardless of the year. Differences among lecturers and other factors may also affect learning. We identified career awareness that students consistently learn in the presence of changes in these external factors. Second, we explored the influence of students' interest in lectures. The results of their learning differed depending on their interest in the lecture, even if they took the same course. However, learning independent of interest should also be generated by attending the same course, and we aim to identify this.

In this study, we used two covariates in the STM—year and student type—categorized based on interests. The purpose of this study was to clarify the career awareness that students consistently learn throughout their first-year career education courses, regardless of the year the course are offered or students' interests, through the STM with these covariates.

2 Sample

2.1 Surveyed University Students and Their Characteristics

This study surveyed first-year students who took a first-year career education course offered as a required course at a public university (University A) in Japan from 2017 to 2022. The target population is students who belong to the Faculty of Social Sciences. A total of 1,780 students took the course from 2017 to 2022.

In this university, more than 80% of the students who enroll each year come from high schools in the same prefecture. Many of them are from the same high school; for this reason, it is expected that many students will enter university with the same attitudes they had when they were in high school. Table 1 shows the results of a survey conducted two months after the entrance of students enrolled in the 2022 academic year. Students were asked about the degree to which they had thought about their post-high school destinations while in high school, with four options provided. A total of 30.5% of students answered, "I thought about it well." However, 51.9% of the students answered "I thought so-so." The content of their thinking was expected to be vague, and it is assumed that many students entered university without thinking deeply about it.

Table 1: Survey on the degree to which university students thought about their post-high school destinations while in high school (Students entering university in 2022, n=285)

Choices	Number of students	Ratio (%)
Thought deeply	87	30.5
Thought so-so	148	51.9
Did not think much	43	15.1
Did not think at all	7	2.5

Table 2 shows the results of a four-choice question asking students if they had a clear image of the kind of work they wanted to do in high school after graduating from university. Only 9.8% of the students entered university with clear careers. The total number of students who answered that they "had a clear careers" or "had a fair clear career" of their career was 44.2%, indicating that approximately half of the students entered university without any particular idea about their careers after graduation. Table 3 presents another survey conducted at approximately the same time among the same students. Students were asked, "Do you have an idea of what kind of job you would like to have after university?" The total number of students who answered that they "had a clear careers" or "had a fair clear career" of their career was 42.2%, indicating that the state of their career image during high school continued after entering university.

Table 2: Survey on whether students had a clear image in high school of the kind of work they wanted to do after university (Students entering university in 2022, n=285)

Choice	Number of students	Ratio (%)
Clear	28	9.8
Fairly clear	98	34.3
Not very clear	117	41.1
Not clear at all	42	14.7

Table 3: Survey on whether students currently had an idea of what kind of job they would like to have after university (Students entering university in 2022, n=256)

Choice	Number of students	Ratio(%)
Clear	23	9.0
Fairly clear	85	33.2
Not very clear	137	53.5
Not clear at all	11	4.3

Table 4: Survey on whether students have an idea of what kind of job they would like to have after university (Social science undergraduate students at a private university located in a prefecture neighboring University A)

Choice	First-year students (n=47)		Second-year students (n=71)		Third-year students (n=36)	
	N	R (%)	N	R (%)	N	R (%)
Clear	1	2.1	4	5.6	4	11.1
Fairly clear	3	6.4	11	15.5	6	16.7
Not very clear	36	76.6	52	73.2	24	66.7
Not clear at all	7	14.9	4	5.6	2	5.6

N: Number of students, R: Ratio

This trend is not unique to the departments surveyed in this study but can be considered a general trend among social science departments in Japan. Table 4 shows the results of the same question asked in January 2023 of undergraduate social science students at a private university located in a prefecture neighboring the surveyed university. Students at this university tended to

have fewer visions of the future. Although students' occupational image solidified as their academic years progressed, many of them still lacked vision even in their third year of study. This phenomenon could be reduced if students were in departments directly related to their career image. However, in a universalized university, the number of students who do not have a career image will inevitably increase in social science departments, which are relatively weakly linked to career image. Therefore, the importance of career education in universities has been emphasized.

2.2 The Course and Free-Writing Reports Surveyed in This Study

The first-year career education courses surveyed in this report were offered to students from April to August in the first semester of their first year of study consisting of eight lectures (90 minutes per lecture), with the aim of cultivating the ability to continuously design their own careers. Table 5 presents the contents of the lectures offered from 2017 to 2022.

Table 5: Lecture Content of First-Year Career Education Courses

year	Lecture content							
	1st	2nd	3rd	4th	5th	6th	7th	8th
2017	President, Dean of faculty	GS evaluation	Lecturer within university	Senior students	Lecturer outside university	Chairman	GS explanation	Graduates
2018	President, Dean of faculty	GS evaluation	Lecturer outside university (1)	Lecturer outside university (2)	Lecturer outside university (3)	Senior students	GS explanation	Graduates
2019	President, Dean of faculty	GS evaluation	Lecturer within university	Lecturer outside university (1)	Lecturer outside university (2)	Senior students	GS explanation	Graduates
2020	President, Dean of faculty	GS evaluation	Lecturer outside university (1)	Senior students	Lecturer within university	Graduate	GS explanation	Lecturer within university
2021	President, Dean of faculty	GS evaluation	Lecturer within university	Lecturer outside university	Senior students	Lecturer within university	Graduate	GS explanation
2022	President, Dean of faculty	GS evaluation	Lecturer within university	Lecturer outside university (1)	Lecturer outside university (2)	Senior students	GS explanation	Graduates

GS: Generic Skill

Lectures by the president, the dean of faculty, and special lecturers, such as lecturers within and outside the university, explained the meaning of university studies and career development for the future. These lectures had strong theoretical aspects. In this study, these lectures are classified as theoretical lectures (Type A). In this course, the measurement of generic skills (hereinafter referred to as "GS") is incorporated into the lectures to understand the students' current abilities. The students were assessed using a commercially available measurement tool (the PROG Test) at the beginning of their enrollment (second lecture). In the seventh lecture, students received the results of the GS measurement, understood their current abilities, and reflected on what they should do in the future. In this study, this seventh lecture is referred to as the Self-

Understanding Lecture (Type B). Finally, lectures by senior students and graduates on their experiences provided students with familiar role models. During the lectures by senior students, six students ranging from sophomores to juniors took the stage and talked about their specific experiences. Three graduates spoke during graduate lectures. These three graduates worked in different occupations from a representative list of jobs available at the university department. These lectures by senior students and graduates provided first-year students with role models for their university life and are classified in this paper as role-model lectures (Type C).

After attending eight lectures consisting of Types A, B, and C, students were required to submit two types of assignments. One assignment was to select two lectures that "caught their interest," and the other was an open-ended report on their understanding of careers, in which the students wrote approximately 600 words (in Japanese) about how they "understood careers" throughout the lectures.

Of the 1,780 students in the course, 1,748 agreed to participate in this survey. Excluding 16 students whose data were missing, 1,732 open-ended reports were used for the analysis in this study. Data were anonymized prior to analysis. The average number of Japanese characters in the open-ended reports of the analyzed students was 622 (σ=67).

3 Analysis Method

3.1 Student Types in Terms of Their Interest in Lectures

Students who took the course were asked to select two lectures of interest from among seven lectures, excluding the second lecture on GS measurements. Based on the combination of choices, students were classified into one of five types. Table 6 lists the number and composition of each type. If all lectures were of equal interest to the students, the expected values were as follows: Type I, 28.6%; II, 19.0%; III, 38.1%; IV, 9.5%; and V, 4.8%. The results are summarized in Table 6. Comparing the actual composition to the expected values, the number of students classified as Type I was very small, whereas the number of students in Types IV and V was very large. One of the characteristics of the students surveyed in this study was that they were strongly interested in self-understanding and role models but not in theoretical lectures on careers.

Table 6: Student types (n=1,732)

Type of Student	Combination of lectures that the students were interested in	Number of students	Ratio (%)	Expectation value(%)
I	A: Theory, A: Theory	97	5.6%	28.6%
II	A: Theory, B: Self-understanding	226	13.0%	19.0%
III	A: Theory, C: Role model	477	27.5%	38.1%
IV	B: Self-understanding, C: Role model	612	35.3%	9.5%
V	C: Role model, C: Role model	320	18.5%	4.8%

3.2 Estimating Structural Topic Models

3.2.1 Setup covariates and tools used

This study used STM to extract topics from open-ended reports on the career understanding of first-year university students after taking a career education course, and to clarify the actual state

of university students' learning. STM enables the estimation of topic prevalence and topic content in students' free-writing reports. STM can specify covariates associated with topic prevalence and covariates associated with topic content and allows us to estimate models that are more robust than conventional topic models.

In the first-year career education courses, although the syllabus is the same from year to year, the lecturers differ from year to year, and it is expected that the difference in lecturers will affect topic prevalence. In addition, differences in the types of students with different interests are expected to affect topic prevalence. Based on this, we set year and student type as covariates for topic prevalence and examined their influence on the topic. Furthermore, the same topic may be discussed differently depending on the students' interests. Based on this, student type was set as a covariate for Topic Content and analyzed.

The stm package [13] for R was used as the analysis tool, and Mecab [14] was used for the morphological analysis of free-writing reports.

3.2.2 Preparation of the data set

This study used 1,732 free-writing reports as data for analysis. First, a morphological analysis was performed on the 1,732 data items using Mecab to extract nouns (general and proper), verbs (independent), adjectives, adverbs, and unknown words. Based on the results and free-writing reports, dictionary registration, stop words, and notational quirks (synonyms) were used to preprocess the text data.

For dictionary registration, a list of 299 words was created and registered as a user dictionary to handle words (e.g., proper expressions) that cannot be processed by Mecab. For stop words, we set words that contained numerical values and words with a high frequency of occurrence that were difficult to interpret as topic content (12 words), as shown in Table 7.

Table 7: Stop words

Words that are difficult to interpret			
do	to be able to	be	a way
myself	become	I think	feel
career	consider	university	certain

The shaking of full/half-width characters, upper/lower case, shaking of notations due to kanji conversion, and shaking of notations due to synonyms were allocated to one specific word. In addition, some speaker-related notational errors were assigned to unique words after carefully examining all the data and identifying the target words. For the speakers, the unique terms were "special lecturers (president, dean of faculty, and lecturers within and outside university)," "senior students," and "graduates." In some cases, the words "special lecturers," "senior students," and "graduates" appeared simultaneously in one sentence in the free-writing reports. These descriptions were judged as references to all speakers. To facilitate interpretation of the analysis results, when the words "special lecturer," "senior student," and "graduate" appeared in a single sentence, they were combined into the single word "speaker."

From the preprocessed data, 1510 words were obtained after extracting nouns (general and proper), verbs (independent), adjectives, adverbs, and unknown words that occurred in five or more students. For these 1,510 words, text analysis using STM was performed on the dataset, which included data on the number of occurrences of each word in the reports, metadata on the year, and student type regarding free-writing reports. All analyses were performed in Japanese

and the final notations were translated into English.

3.2.3 Considering of the number of topics

The number of topics must be determined before estimating a STM. Currently, there is no established method for determining the number of topics. This study examined the number of topics using the held-out likelihood and residual dispersion as evaluation indices, with reference to Muranaka et al. [12]. We examined the behaviors of these two evaluation indices when the number of topics varied from 3 to 40. Figure 1 shows the results for the held-out likelihood. When the number of topics was varied from 3 to 40, the held-out likelihood repeatedly increased and decreased between -6.16 and -6.05, indicating that it was not a useful indicator in this study. Figure 2 shows the results of residual dispersion. The residual dispersion decreases monotonically with the number of topics and starts to slow when the number of topics exceeds 25, finally approaching a value of approximately 1.75, as indicated by the dashed line in Figure 2. Based on this result, we determined the number of topics using residual dispersion as an index. Determining the number of topics such that the residual dispersion is as small as possible is desirable. A smaller number of topics makes them easier to interpret. In this study, we defined the number of topics as the point at which the residual dispersion reached approximately 90% of the asymptote level. This corresponds to the point at which the residual dispersion reached approximately 1.93. This study used fifteen topics that were approximately in this neighborhood.

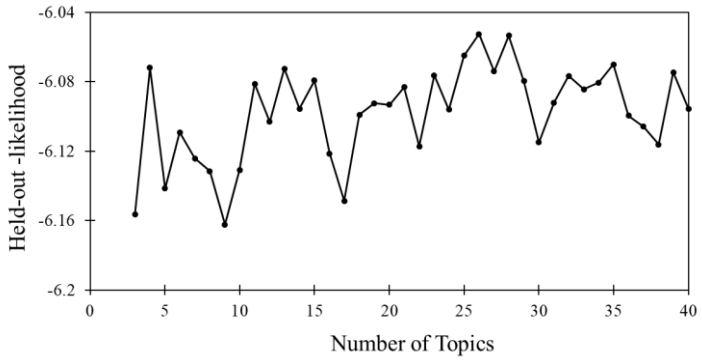

Figure 1: Evaluation index for considering the number of topics: Held-out likelihood

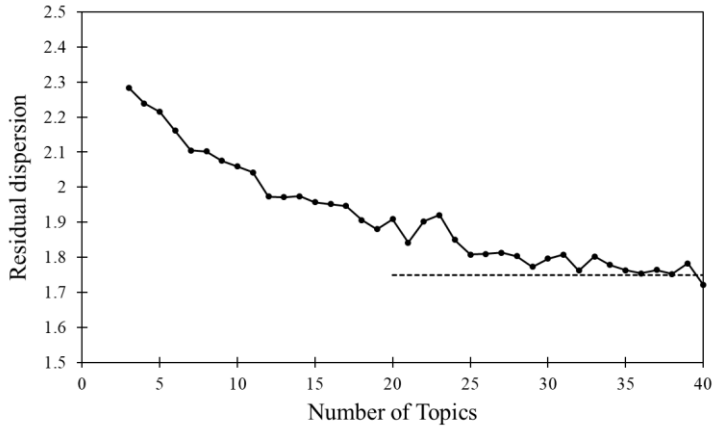

Figure 2: Evaluation index for considering the number of topics: Residual dispersion
(The dashed line is the level considered to be the asymptotic value)

3.2.4 Topic labeling

The labeling of topics, which is the meaning of the topics, was examined with careful reference to four indicators: Highest Prob, FREX, Lift, and Score for each topic. We referred to the content of the reports that contained the highest percentage of the top words appearing most frequently in each indicator and the corresponding topic. According to Ishida, "Intuitively, Highest Prob is the group of words estimated to have the highest each topic content. FREX is the group of words that characterize the topic, Lift is the group of words that are particularly likely to appear on the topic. Score is a measure that is similar to the TF-IDF for frequency information and is the top ranked word group for which all topic distributions are taken into account." [15]. According to Bischof et al. [16], exclusive words are suitable topic summary words. Based on these points, the labeling of topics in this study was conducted mainly with reference to the FREX word groups.

4 Results and Discussion

4.1 Extraction of Topics

Table 8 shows the proportion of each of the fifteen topics extracted by the STM from the free-writing reports submitted by students after taking the course, as well as the top words frequently appearing in the reports for the indicator FREX. Table 8 also lists the labels of the topics considered by the authors. Table 8 presents the topic proportions in descending order.

Table 8: Proportion of topics extracted, top FREX frequency words and labels

Topic	Topic Proportion	FREX: Frequently occurring words	Label
12	0.116	graduates, listen, senior students, very, job hunting, yet, dream, especially, good, PROG	Learning about career development from role models
3	0.087	progress, speaker, role, life, way of life, self, achieve, academic background, reflect, describe	Career as a progress in life
11	0.086	COVID virus, university student, volunteer, internship, summer vacation, English, qualification, club activity, engage in, overseas	Setting goals at university
5	0.082	brush up, knowledge, process, self, build, learn, background, way of life, skill, process	Attitude toward work
1	0.078	competency, task, ability, low, literacy, basic, PROG, information, develop, learn	Challenges of self-abilities
8	0.077	life, spend, choice, live, progress, how, not, university student, career direction, high school student	For a fulfilling life
7	0.073	leave, impression, language, build-up, speak, special lecturer, say, change, make, very	Career as a buildup
4	0.072	do, find, broaden, field of vision, fit, interest, begin to work, be found, good, really	Broadening fields of vision and finding what interests me
2	0.068	person, communication, goo, encounter, exist, bad, idea, different, surrounding, others	Meeting others with different values
14	0.058	advantage, disadvantage, know, learn, develop, characteristic, PROG, life design, how, leader	Self-strengths and weaknesses

10	0.049	company, society, Japan, leave, student, work, old age, child, broad	Working at a Japanese company
9	0.047	self, make up, make, field of vision, broaden, look, people, touch, can, life design	Self-growth through relationships with others
13	0.044	part, dream, view point, goal, civic worker, see, build, look, self-analysis, achieve	Achieving my dreams and goals
6	0.041	way, discipline, survive, intelligence, future, learning, walk, maximum, heart, progress	University as a learning place
15	0.023	enjoy, confidence, send, senior students, people, group, graduates, executive committee, work, university festival	Enjoying university life and gaining self-confidence by modeling the seniors

The free-writing reports had a large topic proportion related to the perspective of what a career is (i.e., "12. Learning about career development from role models" and "3. Career as a progress in life," "8. For a fulfilling life," "7. Career as a buildup."). This was followed by a large topic proportion of behavioral awareness for future career development (i.e., "11. Setting goals at university; 5. Attitude toward work," "1. Challenges of self-abilities," and "4. Broadening fields of vision and finding what interests me," "2. Meeting others with different values," "14. Self-strengths and weaknesses") The students described their careers through these topics after completing the course. This result was satisfactory as an outcome of the first-year career education program.

4.2 Year Dependency of Topics

The first-year career education courses had the same syllabus for all the years targeted. In other words, there is almost no year dependence in course content. However, the lecturers who appeared in the course differed each year, although they were selected based on the syllabus. In addition, external factors influence students who take the course. For example, the surveyed university area suffered significant earthquake damage in 2016, and the disaster experience caused significant changes in students' career awareness [17]. The impact of these factors outside the university may have continued to a lesser extent in the subsequent years. Furthermore, in 2020-2022, the university was affected by the COVID-19 pandemic; for this reason, all lectures in the subject course also changed online. Although the syllabi were identical, the topic proportions may have changed from year to year owing to these external factors.

The results of determining the annual changes in topics are shown in Figure 3. Six topics–Topics 1, 8, 12, 13, 14, and 15–were not found to be year-dependent at a significance level of 5%. These six topics, "1. Challenges of self-abilities," "8. For a fulfilling life," "12. Learning career development from role models," "13. Achieving my dreams and goals." "14. Self-strengths and weaknesses," and "15. Enjoying university life and gaining self-confidence by modeling seniors" were not influenced by external factors in the first-year career education course surveyed. It is highly likely that these six topics are career awareness topics that are always generated as stable learning in any year and represent the reality of learning unique to this course. In particular, "12. Learning about career development from role models," "1. Challenges of self-abilities," and "8. For a fulfilling life" had high topic proportions, and most students taking this course described these topics. Thus, these are important topics for learning in this course.

The STM was able to extract the educational effects unique to this course.

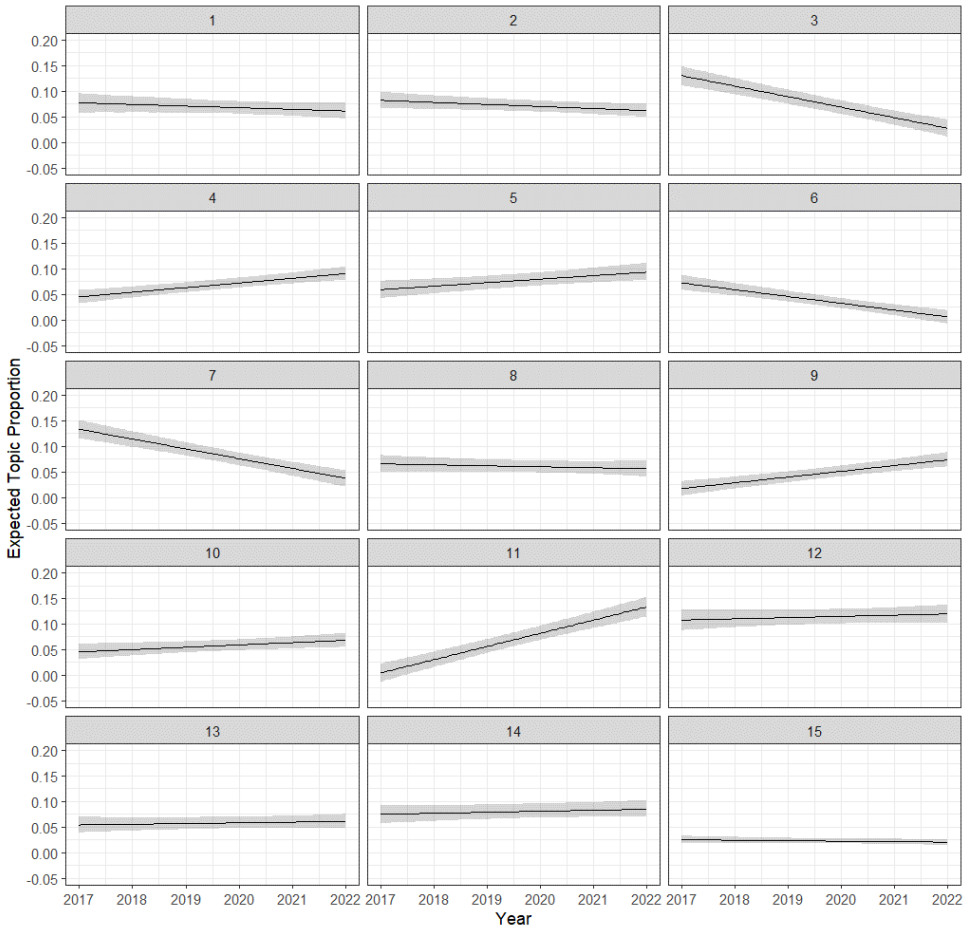

Figure 3: Year change for each topic (shading indicates 95% confidence interval)

4.3 Classification Including the Dependence of Topics on Student Type

To further understand the characteristics of the topics, we examined their dependence on student types according to their interests. A linear regression model was used to analyze each topic as a dependent variable, with year and student type as explanatory variables, which were entered as covariates for topic prevalence. The reference category for the student type was Type V role-model learning. Table 9 presents the results.

Considering the influence of the difference in student type, the fifteen topics can be categorized, as shown in Table 10, according to the presence or absence of significant differences in year and student type. Topics with significant differences included those for which a significant trend was observed.

As mentioned earlier, Topics 1, 8, 12, 13, 14, and 15 were stable regardless of the year. Eight of the fifteen topics (Topics 1, 2, 4, 6, 7, 9, 13, and 15) were not affected by student interest. Of these eight topics, we focused on Topic 1 (Challenges of self-abilities), Topic 13 (Achieving my dreams and goals), and Topic 15 (Enjoying university life and gaining self-confidence by modeling seniors). These topics were not significantly different by year and were not affected by differences in students' interests. These indicate very stable learning, as career awareness is generated by the surveyed course, influenced by neither external factors nor students' interests.

Table 9: The results of multiple regression analysis (reference category: student type V)
(n=1,732, * p<0.1, ** p<0.05)

Topic 1	estimate	std.error	
(Intercept)	6.586	4.743	
Year	-0.003	0.002	
TypeⅣ	-0.006	0.011	
TypeⅢ	-0.006	0.011	
TypeⅡ	0.020	0.014	
TypeⅠ	-0.005	0.019	

Topic 2	estimate	std.error	
(Intercept)	8.189	3.919	**
Year	-0.004	0.002	**
TypeⅣ	0.000	0.010	
TypeⅢ	-0.015	0.011	
TypeⅡ	-0.009	0.011	
TypeⅠ	-0.011	0.015	

Topic 3	estimate	std.error	
(Intercept)	41.622	4.888	**
Year	-0.021	0.002	**
TypeⅣ	0.011	0.012	
TypeⅢ	0.026	0.013	**
TypeⅡ	0.016	0.015	
TypeⅠ	0.038	0.022	*

Topic 4	estimate	std.error	
(Intercept)	-18.373	3.509	**
Year	0.009	0.002	**
TypeⅣ	0.007	0.008	
TypeⅢ	0.012	0.008	
TypeⅡ	-0.003	0.010	
TypeⅠ	-0.014	0.013	

Topic 5	estimate	std.error	
(Intercept)	-13.747	4.479	**
Year	0.007	0.002	**
TypeⅣ	0.012	0.011	
TypeⅢ	0.005	0.011	
TypeⅡ	0.038	0.015	**
TypeⅠ	0.041	0.018	**

Topic 6	estimate	std.error	
(Intercept)	26.829	3.910	**
Year	-0.013	0.002	**
TypeⅣ	-0.003	0.009	
TypeⅢ	0.013	0.010	
TypeⅡ	0.002	0.011	
TypeⅠ	0.027	0.020	

Topic 7	estimate	std.error	
(Intercept)	38.575	4.900	**
Year	-0.019	0.002	**
TypeⅣ	0.016	0.011	
TypeⅢ	0.004	0.013	
TypeⅡ	-0.014	0.014	
TypeⅠ	-0.013	0.020	

Topic 8	estimate	std.error	
(Intercept)	4.015	4.637	
Year	-0.002	0.002	
TypeⅣ	-0.012	0.010	
TypeⅢ	0.013	0.011	
TypeⅡ	0.023	0.014	*
TypeⅠ	-0.003	0.018	

Topic 9	estimate	std.error	
(Intercept)	-22.432	3.803	**
Year	0.011	0.002	**
TypeⅣ	-0.008	0.009	
TypeⅢ	0.009	0.010	
TypeⅡ	0.007	0.013	
TypeⅠ	-0.022	0.015	

Topic 10	estimate	std.error	
(Intercept)	-9.105	4.016	**
Year	0.005	0.002	**
TypeⅣ	0.011	0.009	
TypeⅢ	0.008	0.010	
TypeⅡ	0.011	0.012	
TypeⅠ	0.042	0.017	**

Topic 11	estimate	std.error	
(Intercept)	-51.910	5.186	**
Year	0.026	0.003	**
TypeⅣ	-0.036	0.012	**
TypeⅢ	-0.010	0.013	
TypeⅡ	-0.028	0.015	*
TypeⅠ	-0.010	0.019	

Topic 12	estimate	std.error	
(Intercept)	-4.616	5.161	
Year	0.002	0.003	
TypeⅣ	-0.026	0.012	**
TypeⅢ	-0.048	0.013	**
TypeⅡ	-0.071	0.015	**
TypeⅠ	-0.069	0.020	**

Topic 13	estimate	std.error	
(Intercept)	-2.772	3.606	
Year	0.001	0.002	
TypeⅣ	0.011	0.009	
TypeⅢ	-0.012	0.009	
TypeⅡ	0.011	0.011	
TypeⅠ	0.006	0.015	

Topic 14	estimate	std.error	
(Intercept)	-3.719	4.397	
Year	0.002	0.002	
TypeⅣ	0.023	0.011	**
TypeⅢ	-0.007	0.011	
TypeⅡ	-0.003	0.014	
TypeⅠ	-0.002	0.016	

Topic 15	estimate	std.error	
(Intercept)	1.916	1.777	
Year	-0.001	0.001	
TypeⅣ	0.000	0.004	
TypeⅢ	0.008	0.005	
TypeⅡ	0.000	0.005	
TypeⅠ	-0.005	0.007	

Table 10: Classification of topics by presence of significant differences regarding year and student type

		Student type	
		Not significant	Significant
Year	Not significant	Topics 1, 13, 15	Topics 8, 12, 14
	Significant	Topics 2, 4, 6, 7, 9	Topics 3, 5, 10, 11

5 Conclusion

This study analyzed the career awareness of 1,732 university students who took a first-year career education course between 2017 and 2022 using STM to examine free-writing reports written by the students on their understanding of careers. Fifteen topics were extracted for students' career awareness. These topics were largely related to the perspective of what a career is (i.e., "Learning about career development from role models," "Career as a progress in life," "For a

fulfilling life," "Career as a buildup"), followed by the awareness of future career development (i.e., "Setting goals at university," "Attitude toward work," "Challenges of self-abilities," "Broadening fields of vision and finding what interests me," "Meeting others with different values," "Self-strengths and weakness"). The students who took the survey course learned about these topics in relation to their careers.

The results of the topic dependence analysis showed that the six topics, "Challenges of self-abilities," "For a fulfilling life," "Learning about career development from role models," "Achieving my dreams and goals," "Self-strengths and weaknesses," and "Enjoying university life and gaining self-confidence by modeling the seniors," were stable in proportion in all years. The course syllabus was the same each year. However, different lecturers were on stage in different years, and teaching methods changed significantly because of the COVID-19 pandemic. The results suggest that the six topics were not affected by external factors other than the syllabus, and that they generated stable career awareness among the students who took the course.

The learning content depended on the students' interest in the lectures. We classified the students into five groups according to their interest in the lectures and examined the impact of each group on the content of the lectures. Eight topics were independent of the student type. They were not affected by students' interests and generated constant learning for the students. In addition, some topics were independent of the year and student type. These were the topics "Challenges of self-abilities," "Achieving my dreams and goals," and "Enjoying university life and gaining self-confidence by modeling the seniors." These topics were not influenced by factors outside the syllabus or students' interests and were generated as very stable learning about career awareness. The topic proportion was particularly high for "Challenges of self-abilities," and the students described it frequently. This is a stable career awareness generated by the first-year career education course that was the subject of the survey and is probably the most noteworthy topic when considering the lecture contents of the course.

This study used STM to analyze career awareness acquired from taking a first-year career education course and its correlation with the variables of year and student type. Analyzing students' freewriting reports using STM is a valuable method for evaluating the actual extent of student learning.

This study is based on a paper presented at IIAI AAI 2023 [18]

Acknowledgement

This work was supported by JSPS KAKENHI Grant No. 21K02634.

References

[1] M. A. Trow, Twentieth-Century Higher Education: Elite to Mass to Universal, Johns Hopkins Univ Pr, 2010.

[2] D. E. Super, The psychology of careers, New York: Harper, 1957.

[3] D. E. Super, "A Life-Span, Life-Space Approach to Career Development." J. of Vocational Behavior, Vol.16. 1980, pp. 282-298.

[4] O. Umezaki and M. Tazawa, Studies and Career in Undergraduates. Hosei University Press [in Japanese], 2013.

[5] S. Mizokami, "Student learning and development from a college life perspective: Well-balanced curricular and extra-curricular activities show high development" [in Japanese]. Kyoto Univ. Res. High. Educ., vol. 15, 2009, pp. 107-118.

[6] K. Hatano, "The examination of autonomous academic motivation in university students: From the relationship of learning and career" [in Japanese], Jpn. J. Adolesc. Psychol., vol. 24, 2013, pp. 137-148.

[7] M. Yamamoto and N. Matsuzaka, Class design and management of career education for university students - Focusing on a program under the academic-industrial alliance -, J. Career Design Studies, Vol. 16, 2020, pp. 61-74.

[8] M. Kikuchi, T. Suda, Y. Tange and K. Murakami, "Students' Learning and Thought-Inducing Factors Analyzed form Their Comments on a Career Course," [in Japanese], Jpn. Assoc. for College and University Education, Vo. 41, No. 1, 2019, pp. 117-156.

[9] T. L. Griffiths and M. Steyvers, "Finding Scientific Topics", Proc. of the National Academy of Sciences of the United States of America, 101(Suppl_1), 2004, pp. 5228-5235.

[10] M. E. Roberts, B. M. Stewart, D. Tingley, and E. Airoldi, "The structural topic model and applied social science." Neural Information Processing Society, 2013; http://www.tinyurl.com/y67skpul.

[11] M. E. Roberts, B. M. Stewart, and E. Airoldi, "A Model of Text for Experimentation in the Social Sciences", J. of the American Statistical Association, 111, 2016, pp. 988-1003.

[12] S. Muranaka, Y. Takebayashi, "Isuues in Telepsychology: Abstract Analysis Using the Structural Topic Model," Jpn. J. Behavioral and Cognitive Therapies, Vol. 47, No. 2, 2021, pp. 127-138.

[13] M. E. Roberts, B. M. Stewart, and D. Tingley, "stm: An R Package for Structural Topic Models", J. Stat. Soft., vol. 91, No. 2, 2019, pp. 1-40.

[14] https://taku910.github.io/mecab/

[15] M. Ishida, "Text Mining by R," [in Japanese], Morikita-Shupan, 2020, pp. 23-55.

[16] J. M. Bischof and E. M. Airoldi, "Summarizing topical content with word frequency and exclusivity", In Proc. of the 29th Int'l Conf. on Machine Learning(ICML-12), 2012.

[17] Y. Nakazato, T. Tsumagari, T. Tsumagari, "Analyzing the characteristics of first-year university students' career awareness through free-writing reports," Proc. 9th Int'l Cong. on Advanced Applied Informatics, IIAI-AAI 2020, 2020, pp. 347-350.

[18] T.Tsumagari, Y.Nakazato, and T.Tsumagari, "Analysis of College Students' Career Awareness after Taking First-Year Career Courses Using the Structural Topic Model, " IIAI Letters on Institutional Research, vol.3, 2023, pp.1-8.