Classification of Educational Skills for University Students in Computer Programming Classes

Tadanari Taniguchi

Abstract—This paper analyzes the relationship of educational skills that students should achieve for each computer programming class using a student self-assessment questionnaire. The questionnaire survey, containing 25 educational skills, was conducted in computer programming classes in my university using a computer-assisted web-interviewing technique. The questionnaire results are analyzed using an agglomerative hierarchical clustering based on Ward's method and a self-organizing map, which is a machine learning method. This study shows that the students can be classified into four clusters: highly skilled students, students with high learning and thinking skills but low executing skills, students with high leaning and executing skills but low thinking skills, and students with lower skills.

Index Terms—Educational skills, machine learning, questionnaire survey, self-organizing map.

I. INTRODUCTION

The skills and attributes that should be learned in school have been changing. The United States Department of Education announced the 21st century skills and the Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo) project of the Organisation for Economic Co-operation and Development (OECD) proposed the key competencies. The National Institute for Educational Policy Research in Japan organized a research on curriculum, which fostered attributes and skills.

Tokai University, to which the author belongs, has set four key abilities as a specific evaluation indicator since 2009: thinking, communication, challenging, and accomplishment. The faculty has set the appropriate abilities and evaluation indicators in the syllabus as a skill to be taught in class. However, the skill setting method is neither a theoretical nor a systematic approach. This study aims to set effective educational skills and educational performance indicators analytically per class.

Meta-analyses in [1]-[3], data envelopment analyses in [4], [5], and mediational analysis in [6] were used in a research for the evaluation of indexes and skills. The data envelopment analyses [4] examined the relative efficiency of Australian universities. Furthermore, the technical and scale efficiency [5] were estimated for the population of the Australian universities. The achievement goals, motivational study strategies, and exam performance were examined using mediational analysis in [6]. The author of this paper has examined the relationship between students' educational skills using a multidimensional scaling in [7] and a self-organizing map (SOM) in [8], [9].

In order to investigate the relationships between educational skills and students' consciousness, the aim of this paper is to conduct a questionnaire survey about educational skills before and after students attend computer programming classes. This study used a self-assessment questionnaire to analyze the relationship among students' educational skills. Students who took the introduction and the advanced computer programming classes were the participants of the survey. The questionnaire survey was conducted to analyze 25 educational skills in [10], [11]. The SOM in [12] and an agglomerative hierarchical clustering based on the Ward's method were used to analyze the questionnaire data. The SOM, which is a machine learning method, is an unsupervised neural network method and an efficient tool for visualizing the relationship of complicated data. This study classified the students into four clusters: highly skilled students, students with high learning and thinking skills but low executing skills, students with high leaning and executing skills but low thinking skills, and students with lower skills.

II. METHODS

A. Participants

The questionnaire survey was conducted before and after computer programming classes in the autumn semester of 2018. The questionnaire for the students in the introduction and advanced courses and 11 faculties was surveyed in Tokai University. The introduction course comprises "introduction to computer programming" and "basic computer programming." "Applied computer programming," "computer algorithm," and "computer graphics" are in the advanced course. Students, regardless of their year level in school, or faculties can take these classes. Table I shows the students' year level in study and gender, and faculties of survey respondents. All the participants have volunteered to participate in the study, with a cumulative total of 443 student participants.

B. Procedure

A computer-assisted web-interviewing technique was used to collect questionnaire data. The participants were asked to complete the online questionnaire, containing educational skills. The questionnaire and the study purpose information were provided for the participants by web pages. All the participants joined voluntarily and have read the informed consent terms on the questionnaire web pages.

C. *Questionnaire*

As presented in Table II, a questionnaire survey was conducted to analyze the 25 educational skills in [10], [11] acquired by the students through their classes. The students responded to the questionnaire independently. All items are rated on a nine-point Likert-type scale from 1 (not at all) to 9 (very high degree).

TABLE I: SURVEY RESPONDENTS

		Introduction		Advanced	
		Before	After	Before	After
Total		184	165	58	36
Year	First	28	21	1	0
	Second	67	61	14	10
	Third	59	56	30	19
	Fourth	30	27	13	7
Gender	female	28	22	3	1
	male	156	143	55	35
Faculty	Letters	19	13	1	0
	Cultural and Social Studies	2	1	0	0
	Political Science and Economics	13	14	3	3
	Law	9	9	0	0
	Humanities and Culture	7	3	0	0
	Physical Education	4	3	1	0
	Science	30	30	7	4
	Information Science and Technology	18	15	23	14
	Engineering	80	75	22	14
	Tourism	2	2	0	0
	Health Science	0	0	1	1

TABLE II: EDUCATIONAL SKILLS AND THE ABBREVIATIONS

Educational skill	Abbr.	Educational skill	Abbr.
Learning skills	LRN	Implementing skills	IMP
Thinking skills	THK	Continuity skills	CNT
Inquiring skills	INQ	Accomplishment skills	ACP
Communication skills	COM	Analytical skills	ANL
Collaboration skills	CLB	Modifying and improvement skills	MAI
Relationship building skills	RB	Collecting information skills	CI
Self-assessment skills	SA	Decision making skills	DM
Evaluating other skills	EO	Logical thinking skills	LT
Problem finding skills	FND	Problem solving skills	SLV
Problem setting skills	SET	Artistic skills	ART
Creative thinking skills	CT	Setting goals skills	SG
Planning skills	PLN	Constructing knowledge skills	CK
Executing as planed	EXE		

D. *Data Analysis*

The Mann Whitney U test was used to evaluate how the students acquire their educational skills, before and after the classes. Tables III and IV present the results from the introduction and advanced courses, respectively, where the bold numbers indicate that the means increase after classes compared with before classes.

An agglomerative hierarchical clustering based on Ward's method was applied to the questionnaire results to classify the participants into some educational skills groups. The dendrogram illustrates a tree diagram representing the clustering result in Figs. 1, 3, 5, and 7.

The SOM is an efficient tool visualizing the multidimensional data, which are the relationship of the 21 educational skills based on the student's consciousness using the questionnaire results. The remaining four educational skills (i.e., communication, collaboration, relationship, and artistic skills) were not covered in the classes; therefore they were excluded. This study used the SOM-Toolbox of MATLAB in [13] to create and visualize the SOM for the datasets. The questionnaire results for the introduction and advanced classes were used as the datasets. These data were normalized such that each variable had a unit variance. The SOM results in this study were obtained, regardless of the initial values because the SOMs were initialized and trained through principal component analysis.

III. RESULTS

Firstly, the data were analyzed for statistical significance using the Mann Whitney U test. Table III presents that almost all the means of educational skills increased after the classes compared with before, except for inquiring skills. Table IV shows that the mean values for approximately half of the educational skills increased. However, no significant differences (< 0.05) of the U tests in the educational skills were found in both the results of the introduction and advanced courses. Therefore, there is no statistical significant difference before and after class.

IT education center in Tokai University has been conducting an educational skill survey for students for several years. We examined the relationship among educational skills using a multidimensional scaling in [7] and analyzed the educational skills were classified into different skill groups according to the difficulty of the ICT courses using a SOM analysis in [8]. We showed the students could be classified into several groups based on their attributes, respective academic faculties, and academic years in [9].

This paper analyzed the relationships between the questionnaire results before and after classes. The students attending classes were classified into several groups using an agglomerative hierarchical clustering based on Ward's method for the questionnaire results. The dendrograms presented in Figs. 1 and 3 illustrate the clustering results for before and after the introduction classes, respectively. Figs. 5 and 7 illustrate the before and after results of the advanced classes. The right graph in Fig. 1 illustrates that the gradient is steep at a point less than 4. Therefore, the number of clusters before the introduction classes is set to 4. Similarly, the number of clusters after the introduction classes is set to 4. The numbers of clusters before and after the advanced classes are 4 and 3, respectively. There is no direct relationship between the results of the dendrograms and the following analysis of the SOM. This study considered theses cluster results of the dendrograms because deciding on the number of clusters is difficult for a SOM analysis.

The SOM method was applied to the questionnaire results to investigate the relationship among the educational skills. The unified distance matrices (U-matrices) are in the upper

left of Figs. 2, 4, 6, and 8; the other matrices are the component planes. In the U-matrices, the difference of reference vectors (yellow dot sequence) represents the cluster boundary. The component planes illustrate the 21 educational skills as SOM variables. In the component planes, yellow portion represents a higher skill rating, and blue represents a lower skill rating.

TABLE III: MANN WHITNEY U TEST RESULTS FOR EDUCATIONAL SKILLS BEFORE AND AFTER THE INTRODUCTION CLASSES

ES	Mean before/after		SD before/after		Mann-Whitney	Z	Sig.
LRN	4.67	**4.93**	1.69	1.76	14143	-1.122	0.262
THK	5.10	**5.32**	1.82	1.70	14060.5	-1.209	0.227
INQ	5.32	5.31	1.80	1.91	15047	-0.143	0.886
SA	4.45	**4.59**	1.95	1.98	14583	-0.643	0.520
EO	5.26	**5.55**	2.01	2.00	14006	-1.263	0.207
FND	4.78	**4.99**	1.76	1.67	14303.5	-0.947	0.344
SET	4.47	**4.79**	1.81	1.74	13471	-1.846	0.065
CT	5.13	**5.15**	2.15	2.07	15096.5	-0.090	0.928
PLN	4.71	**4.76**	1.81	1.87	14971.5	-0.225	0.822
EXE	4.26	**4.58**	1.89	2.01	13717.5	-1.571	0.116
IMP	5.03	**5.11**	1.97	1.70	14585	-0.642	0.521
CNT	4.60	**4.87**	2.04	1.97	14001	-1.268	0.205
ACP	5.03	**5.05**	1.74	1.84	15069	-0.120	0.905
ANL	5.13	**5.22**	1.78	1.85	14688	-0.530	0.596
MAI	5.10	**5.13**	1.74	1.70	15093.5	-0.094	0.925
CI	5.22	**5.41**	2.02	1.93	14626.5	-0.595	0.552
DM	5.03	**5.32**	1.99	1.95	14019	-1.249	0.212
LT	4.71	**5.12**	1.91	1.85	13400	-1.918	0.055
SLV	4.96	**5.16**	1.75	1.80	14128.5	-1.137	0.256
SG	4.71	**4.90**	1.81	1.89	14165	-1.097	0.273
CK	4.67	**5.00**	1.85	1.71	13441.5	-1.879	0.060

TABLE IV: MANN WHITNEY U TEST RESULTS FOR EDUCATIONAL SKILLS BEFORE AND AFTER THE ADVANCED CLASSES

ES	Mean before/after		SD before/after		Mann-Whitney	Z	Sig.
LRN	5.47	5.33	2.02	1.85	1018.5	-0.201	0.841
THK	5.59	**5.69**	1.89	1.56	988	-0.447	0.655
INQ	5.91	5.56	2.09	2.03	925.5	-0.933	0.351
SA	4.91	**5.17**	2.11	1.86	982	-0.489	0.625
EO	5.19	**5.53**	1.89	1.92	934	-0.868	0.385
FND	5.38	**5.47**	1.75	1.98	987	-0.452	0.652
SET	4.60	**4.81**	1.93	2.16	997	-0.370	0.711
CT	5.07	**5.14**	2.21	2.23	1009	-0.276	0.783
PLN	4.81	4.72	2.14	2.35	1025.5	-0.145	0.884
EXE	4.84	**4.89**	2.30	2.09	1019.5	-0.193	0.847
IMP	5.69	5.19	1.92	1.89	886.5	-1.248	0.212
CNT	5.07	4.47	2.13	2.21	885.5	-1.246	0.213
ACP	5.28	5.17	1.78	1.99	1009	-0.276	0.782
ANL	5.66	**5.53**	2.00	1.80	1003	-0.324	0.746
MAI	5.24	**5.69**	1.85	1.60	936	-0.857	0.392
CI	5.91	5.33	1.97	1.76	838	-1.624	0.104
DM	5.52	**5.78**	2.07	1.61	981.5	-0.493	0.622
LT	5.84	5.47	2.19	1.90	902.5	-1.115	0.265
SLV	5.40	**5.44**	1.76	1.86	1029	-0.119	0.906
SG	4.95	4.81	2.09	2.20	994	-0.394	0.694
CK	5.55	5.31	1.93	1.56	940.5	-0.818	0.413

IV. DISCUSSION

The U-matrix in Fig. 2 shows two cluster boundaries for

before the introduction classes. The figure also shows that the top-left boundary represents the student group with lower educational skills and the bottom-right one presents the higher skills group. The U-matrix in Fig. 4 shows two cluster boundaries for the after introduction classes. The top boundary represents the lower educational skills, whereas the bottom one presents the higher skills in the same manner as before. Both the upper and lower areas in Fig. 4 are wider after the classes than before. These U-matrices in Figs. 2 and 4 show that the number of students with both lower and higher skills increased after classes. Hence, the problem that should be addressed is the increasing number of students with lower skills.

In the component planes in Figs. 2 and 4, yellow and blue portions present a higher skill rating and a lower skill rating, respectively. These component planes can be classified into three types according to the color patterns. Table V shows the three types of educational skill classifications according to the yellow shapes, and the square area presents the learning skills. The bottom-right corner and bottom-left corner triangle areas present executing skills and thinking skills, respectively. From the cluster boundaries in the U-matrices and these color patterns, the students are classified into four clusters: the bottom boundary students with higher skills, the middle-left boundary students with high learning and thinking skills but low executing skills, the middle-right boundary students with high learning and executing skills but lower thinking skills, and the top boundary students all with lower skills. Interestingly, the middle-left boundary students have the opposite nature of the middle-right boundary ones.
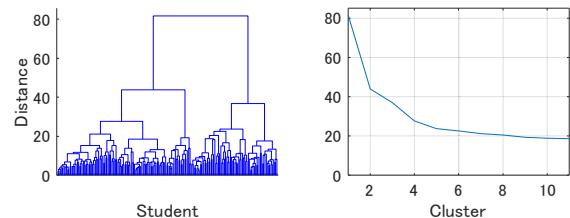

Fig. 1. Dendrograms for before the introduction classes.

Figs. 6 and 8 show the SOM analysis for the advanced classes. The U-matrices of these figures show one clear cluster and one thin boundary of before and after the advanced classes, respectively. The top-left boundary presents students with lower educational skills, and the bottom one presents the higher skills similar to that in the introduction courses. In addition, these U-matrices show that the number of students with lower skills decreased after classes because the top-left boundaries after the classes were smaller than before. These component planes in Figs. 6 and 8 can be classified into three types according to the color pattern similar to that in the introduction classes. Table VI shows the educational skill classifications according to the yellow shapes. From the cluster boundaries in the U-matrices and these color patterns, the students are classified into four clusters similar to the introduction classes.

Therefore, the computer programming students can be classified into four groups regardless of the class difficulty:
1) Students with higher educational skills.
2) Students with high learning and thinking skills but low executing skills

3) Students with high learning and executing skills but low thinking skills
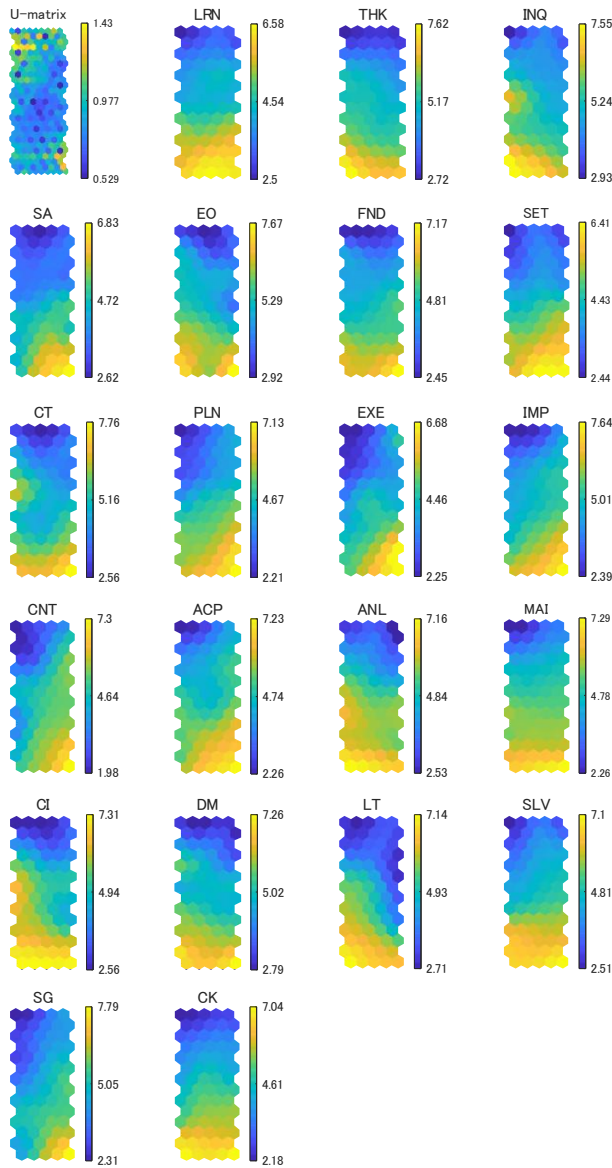4) Students with lower educational skills.

Fig. 2. U-matrix and the component planes for before the introduction classes.
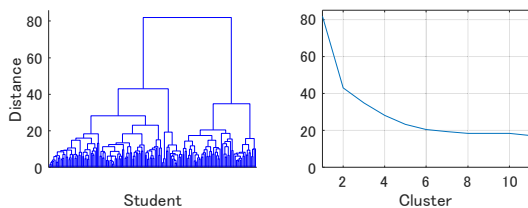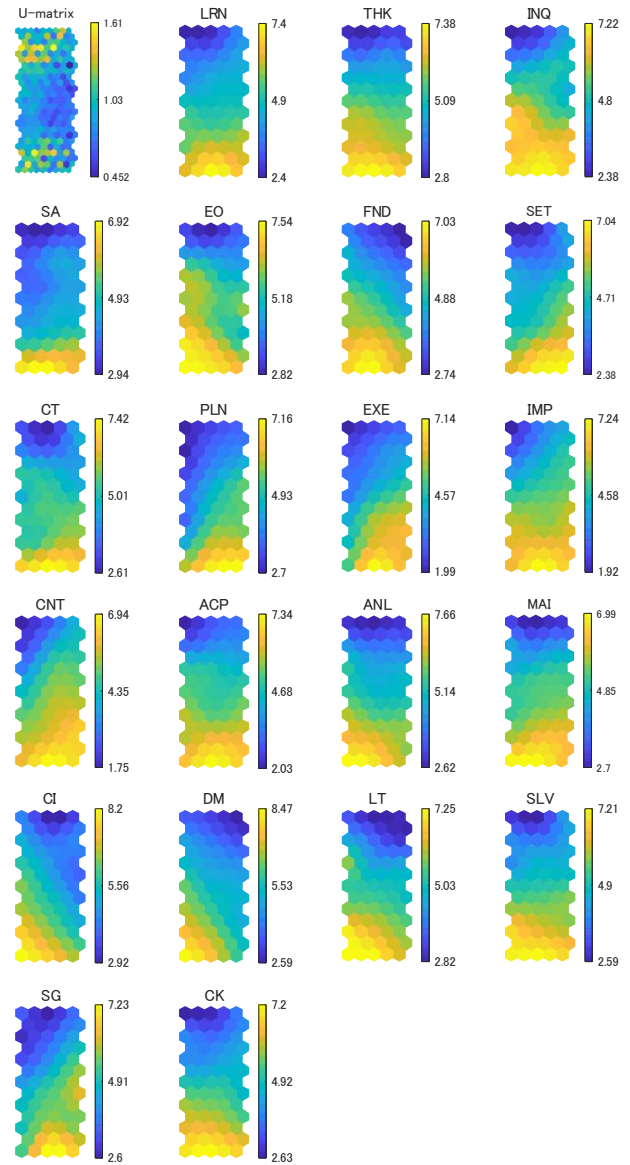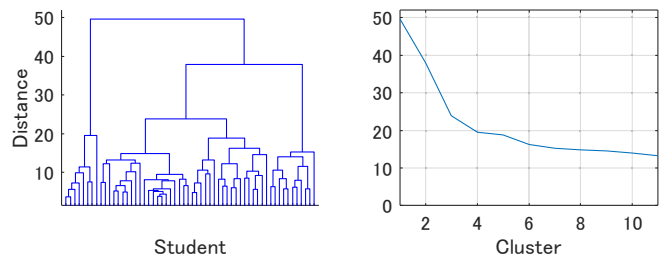

Fig. 3. Dendrograms for after the introduction classes.

TABLE V: EDUCATIONAL SKILLS ARE CLASSIFIED INTO THREE TYPES ACCORDING TO THE COLOR PATTERN FOR THE INTRODUCTION CLASSES

Yellow area	Before classes (Fig. 2)	After classes (Fig. 4)
Square: *learning skills*	**LRN, MAI, SLV**, CK, FND	**LRN, MAI, SLV**, CK, SA, ACP
Bottom-right corner triangle: *executing skills*	**SET, PLN, EXE, CNT, SG**, CT, IMP, SA, ACP	**SET, PLN, EXE, CNT, SG**, CT, IMP
Bottom-left corner triangle: *thinking skills*	**THK, INQ, CI, DM, LT**, EO, ANL	**THK, INQ, CI, DM, LT**, ANL, FND, EO

Bold skills are common in the introduction and advanced classes

Fig. 4. U-matrix and the component planes for after the introduction classes.

Fig. 5. Dendrograms for before the advanced classes.

TABLE VI: EDUCATIONAL SKILLS ARE CLASSIFIED INTO THREE TYPES ACCORDING TO THE COLOR PATTERN FOR THE ADVANCED CLASSES

Yellow area	Before classes (Fig. 6)	After classes (Fig. 8)
Square: *learning skills*	**LRN, MAI, SLV**, CT, IMP, ANL, EO	**LRN, MAI, SLV**, CT, IMP, ANL,
Bottom-right corner triangle: *executing skills*	**SET, PLN, EXE, CNT, SG**, ACP	**SET, PLN, EXE, CNT, SG**, SA, EO
Bottom-left corner triangle: *thinking skills*	**THK, INQ, CI, DM, LT**, CK, SA, FND	**THK, INQ, CI, DM, LT**, CK, ACP, FND

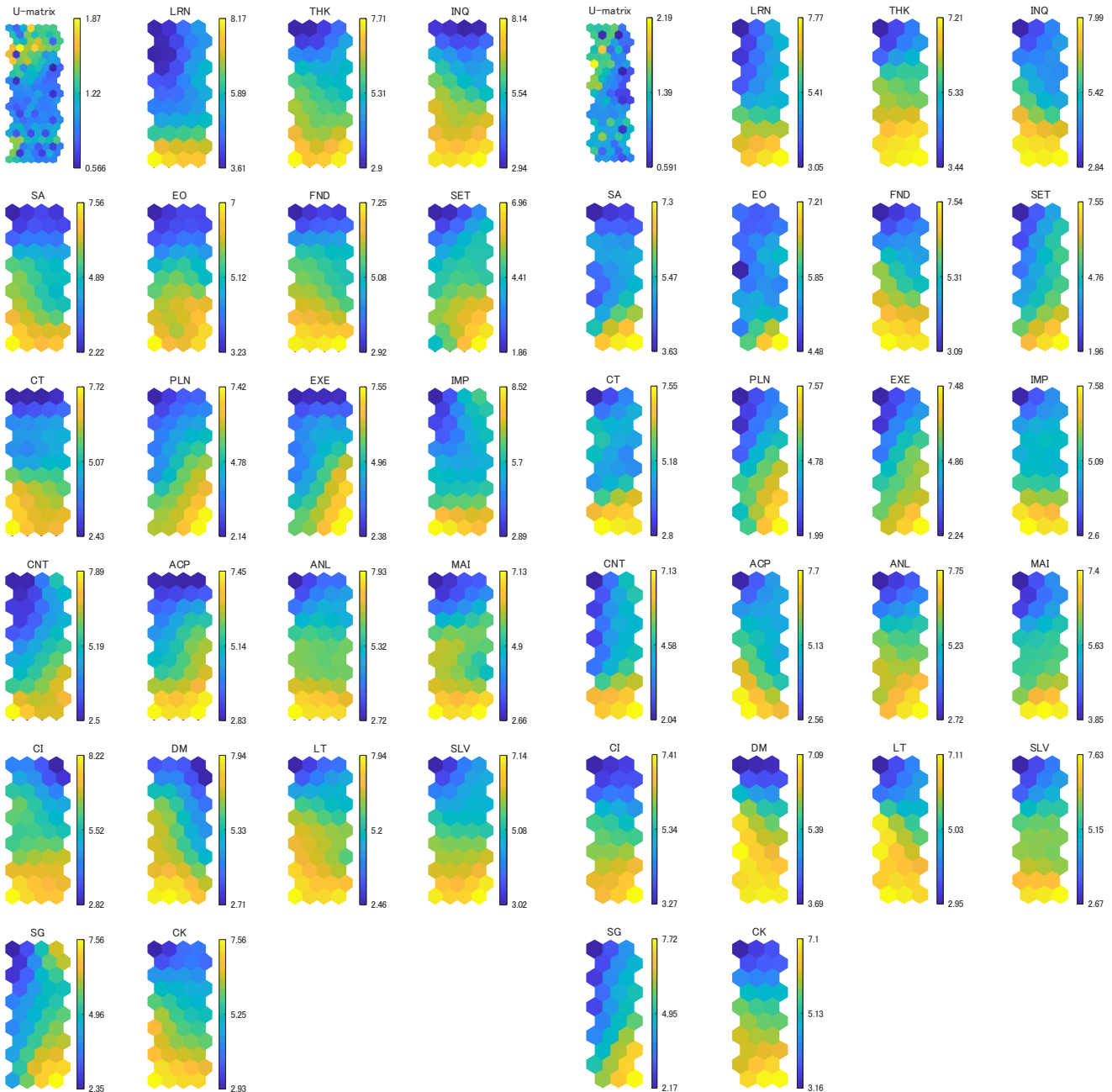**Bold skills** are common in the introduction and advanced classes

Fig. 6. U-matrix and the component planes for before the advanced classes.

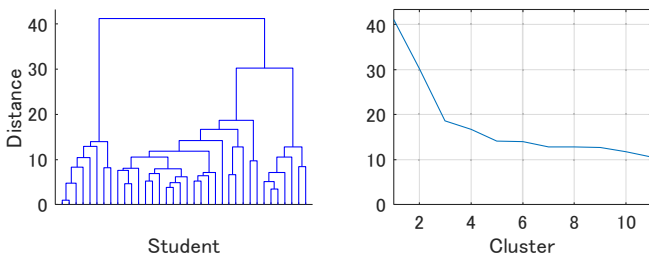Fig. 8. U-matrix and the component planes for after the advanced classes.

Fig. 7. Dendrograms for after the advanced classes.

From these results, the following four measures for the computer programming classes can be considered.

1) Offering more difficult assignments and advanced subjects
2) Offering more assignments and extending computer programming time
3) Compelling the students to think a programming algorithm deeply
4) Compelling the students to ask questions frequently

V. CONCLUSION

This paper has analyzed the relationship of educational skills that students should achieve for each computer programming class through the student self-assessment questionnaire. The questionnaire survey with 25 educational skills was taken to computer programming classes in Tokai University using a computer-assisted web-interviewing technique. The questionnaire results were analyzed using an agglomerative hierarchical clustering based on Ward's method and a self-organizing map, which is a machine learning method. This study has shown that the students can be classified into four clusters: highly skilled students, students with high learning and thinking skills but low executing skills, students with high leaning and executing skills but low thinking skills, and students with low skills. Using the results, present and future works will focus on improving class syllabus and contents.

REFERENCES

[1] V. Robinson *et al*., "The impact of leadership on student outcomes: An analysis of the differential effects of leadership types," *Educational Administration Quarterly*, vol. 44, no. 5, pp. 635-674, 2008.

[2] A. Poropat, "A meta-analysis of the five-factor model of personality and academic performance," *Psychological Bulletin*, pp. 1-68, 2009.

[3] A. Stajkovic and F. Luthans, "Self-efficacy and work-related performance: A meta-analysis," *Psychological Bulletin*, vol. 124, no. 2, pp. 240-261, 1998.

[4] N. Avkiran, "Investigating technical and scale efficiencies of Australian universities through data envelopment analysis," *Socio-Economic Planning Sciences*, vol. 35, pp. 57-80, 2001.

[5] M. Abbott and C. Doucouliagos, "The efficiency of Australian universities: A data envelopment analysis," *Economics of Education Review*, vol. 22, pp. 89-97, 2003.

[6] A. Elliot *et al*., "Achievement goals, study strategies, and exam performance: A mediational analysis," *Journal of Educational Psychology*, vol. 91, no. 3, pp. 549-563, 1999.

[7] T. Taniguchi *et al*., "Analysis and classification of educational skills using questionnaire to university students," in *Proc. KES 2018*, pp. 2021-2029, 2018.

[8] T. Taniguchi *et al.*, "Self-organizing map analysis of educational skills using questionnaire to university students in computing classes," in *Proc. CELDA 2018*, pp.103-110, 2018.

[9] T. Taniguchi *et al*., "Classification of university students attending computing classes using a self-assessment questionnaire," *Smart Education and e-Learning*, pp.29-38, 2019.

[10] G. Ota *et al*., "The comparative survey of computer science and programming education for primary and secondary schools in the UK, Australia and USA," *Japan Journal of Educational Technology*, vol. 40, no. 3, pp. 197-208, 2016.

[11] I. Miyaji, "Difference in Effects of Creating Digital Story Telling by the Difference of Theme", IEICE Technical Report, 3, pp. 83-88, (in Japanese), 2011.

[12] T. Kohonen, "Self-Organizing Maps, Springer Series in Information Sciences," Vol. 30, Springer, 1995.

[13] J. Vesanto *et al*., "Self-organizing map in Matlab: the SOM-Toolbox," in *Proc. the Matlab DSP Conference 1999*, Espoo, Finland, pp. 35-40, 2000.

Tadanari Taniguchi received the B.S. degree in mechanical systems engineering from Kanazawa University, Ishikawa, Japan in 1996, the M.S. degree in mechanical systems engineering from Kanazawa University in 1998, the Ph.D. degree in mechanical and control engineering from University of Electro-Communications, Tokyo, Japan in 2001.

From 2001 to 2005, he was a research scientist, the Brain Science Institute, RIKEN, Japan. Then he has served IT Education Center, Tokai University, Japan. He was a visiting scientist with Universite de Valenciennes, France, in 2015. He is currently an associate professor of Tokai University. His research interests include intelligent control, nonlinear control, statistical analysis, machine learning, and smart education.

He is a member of the Institute of Electrical and Electronics Engineers (IEEE), Japan Society for Fuzzy Theory and Intelligence Informatics (SOFT), and the Society of Instrument and Control Engineers (SICE).