Nakata et al., 2018

Volume 4 Issue 1, pp.631-641

Date of Publication: 22nd May, 2018

DOI-https://dx.doi.org/10.20319/pijss.2018.41.631641

This paper can be cited as: Nakata, Y., Kozaki, Y., Kunisaki, T., Gozu, T., Bannaka, K., & Takamatsu, K. (2018). Assessment Of Rubric-Based Evaluation By Nonparametric Multiple Comparisons In First-Year Education In A Japanese University. PEOPLE: International Journal of Social Sciences, 4(1), 631-641.

ASSESSMENT OF RUBRIC-BASED EVALUATION BY NONPARAMETRIC MULTIPLE COMPARISONS IN FIRST-YEAR EDUCATION IN A JAPANESE UNIVERSITY

Yasuo Nakata
Faculty of Health Sciences, Kobe Tokiwa University, Kobe, Japan
y-nakata@kobe-tokiwa.ac.jp

Yasuhiro Kozaki
Faculty of Education, Osaka Kyoiku University, Osaka, Japan
The Center for Early Childhood Development, Education, and Policy Research, The University of Tokyo, Tokyo, Japan
kozaki@cc.osaka-kyoiku.ac.jp

Taion Kunisaki
Faculty of Education, Kobe Tokiwa University, Kobe, Japan
t-kunisaki@kobe-tokiwa.ac.jp

Tetsuhiro Gozu
Faculty of Education, Kobe Tokiwa University, Kobe, Japan
t-gozu@kobe-tokiwa.ac.jp

Kenya Bannaka
Department of Oral Health, Kobe Tokiwa College, Kobe, Japan
k-bannaka@kobe-tokiwa.ac.jp

Kunihiko Takamatsu
Faculty of Health Sciences, Kobe Tokiwa University, Kobe, Japan
Center for the Promotion of Excellence in Research and Development of Higher Education, Kobe Tokiwa University, Kobe, Japan
Life Science Center, Kobe Tokiwa University, Kobe, Japan
ktakamatu@gmail.com

Abstract

The rubrics have become a widely referenced and utilized form of assessment on campuses across internationally. But rubric can be an asset in any classroom and at any education level but it needs to be implemented correctly. Our research question in this study is whether students were evaluated consistently and equally from teacher to teacher using rubric. To answer this research question, we performed statistical estimation using nonparametric multiple comparisons. This article reports on a normalizing rubric evaluation by nonparametric multiple comparisons in a first-year course called "Manaburu I" offered at Kobe Tokiwa University. "Manaburu" is a word coined by us: "manabu" 'learn' in Japanese + English able. Thus, "Manaburu" refers to Self-Directed Learning I. In the course, about 20 teachers teach about 350 students (16–17 students per teacher). Students are organized into groups of about 6. It is of course difficult for 20 teachers to evaluate their students consistently among them, making this course an appropriate site for the evaluation. We constructed a rubric for the course, under which teachers were meant to evaluate students, and presented it to both teachers and students. Our research question was whether teachers evaluated students consistently and equally according to the Steel–Dwass estimation method, a strict statistical estimation method for nonparametric multiple comparisons. The results show that teachers do not evaluate students equally. Suggestions for future research, more attention to validity and reliability, a closer focus on learning and research on rubric use in higher education.

Keywords

Normalizing Rubric Evaluation, First-Year Education, Nonparametric Multiple Comparison, Steel-Dwass Estimation

1. Introduction

At present, Kobe Tokiwa University is undergoing reform (Kirimura, Takamatsu, Bannaka, et al., 2017). In 2017, Kobe Tokiwa University implemented a new course, "Manaburu I," for first-year students (Kirimura, Takamatsu, Bannaka, et al., 2018). *Manaburu* is a word coined by us: Japanese *manabu* "learn" plus English *able*, thus implying "students are able to learn by themselves." In this course, students learn writing, reading, and thinking with active learning. In the course, about 20 teachers teach about 350 students (Mitsunari, Kirimura, Kunisaki, et al., 2018), that is, 16-17 students each. Students are organized by teachers into groups of 6. We will report what students learned from the course using a free description

questionnaire administered to students and text analysis such as text mining (Kirimura, Mitsunari, Kunisaki, et al., 2018).

Recently, Kobe Tokiwa University proposed a new student support policy to integrate its admissions, curricular, and diploma-award policies. To evaluate and connect these with assessment policy, we developed common evaluation indicators called Tokiwa competencies, or competencies students at Kobe Tokiwa are meant to acquire through regular, quasi-regular (for example, remedial), and extra-curricular (for example, club) activities. The specific competencies are 19 in number: Culture, Common Sense, Professionalism/Expertise, Media Literacy, Logical Thinking, Critical Thinking, Intellectual Curiosity, Exploration, Continuity, Self-Management, Reflection, Design Thinking, Presentation, Judgment, Implementation, Responsibility, Contribution, Communication, and Cooperation & Collaboration. In the syllabus we set for Manaburu I, students will obtain the competencies of Exploration, Reflection, Self-Management, Design Thinking, Presentation, Cooperation & Collaboration (Table 1) (Takamatsu, Murakami, Kirimura, et al., 2017). In our study, we define a competency as a functionally linked complex of knowledge, skills, and attitudes that enable successful task performance and problem-solving, following Spady (Spady, 1994).

In the "Manaburu I" class, teachers form pairs, and the paired teachers teach about 36 students together in the same room. Students are organized into groups of about 6. It is of course difficult for 20 teachers to evaluate their students consistently among them, making this course an appropriate site for the evaluation in this course. So, we needed a fair and appropriate evaluation method. The rubrics have become a widely referenced and utilized form of assessment on campuses across internationally (Association of American Colleges & Universities). To evaluate the students, we created a rubric for the course using the Tokiwa competencies; we defined a rubric as a coherent set of criteria for students' work that includes descriptions of criteria differentiating levels of performance quality (Brookhart, 2013). "Manaburu I" was designed based on a combination of competencies no. 8, 10, 11, 12, 13, and 19. No. 8 (accounting for 20% of students' grade) is "Exploration," meaning thinking deeply about the matters and methods at hand. No. 10 (10%) is "Self-Management," meaning appropriately managing one's physical and mental health. No. 11 (25%) is "Reflection," or looking back on one's thinking and behavior and seeking ways of improvement. No. 12 (10%) is "Design Thinking," or the ability to design solutions and develop a variety of problem-solving ideas and knowledge. No. 13 (15%) is "Presentation," or expressing feelings and thoughts in a way that conveys them clearly to others. No. 19 (20%) is "Cooperation & Collaboration, which

means that beyond narrow individual interests, it is possible to work on things cooperatively. The rubric for "Manaburu I" is given in Table 2.

As it is difficult for as many as 20 teachers to evaluate students consistently and equally across teachers, we constructed an evaluation method using this rubric to help teachers apply it consistently. Also, we explained the rubric in detail to both students and teachers. Table 3 presents the rubric evaluation sheet.

Table 1: *The 19 Tokiwa Competencies (Takamatsu, Murakami, Kirimura, et al., 2017)*

Abbreviated name of competency	Competency
1. Culture	Establishing the liberal arts as the foundation of human nature, which can involve a variety of people
2. Common Sense	Establishing that members of a society should acquire knowledge and behave in certain ways
3. Professionalism/Expertise	Having the necessary knowledge and skills to perform the duties of each profession
4. Media Literacy	Collect, organize, and analyze the necessary information for proper thinking and judgment
5. Logical Thinking	Based on evidence, a situation can be considered logically
6. Critical Thinking	A multilateral, critical perspective captures ideas and can be considered
7. Intellectual Curiosity	To know something, to learn, and remember it with fun and joy
8. Exploration	By thinking deeply about things and methods
9. Continuity	By learning and thinking, it is possible to maintain one's stance and make efforts to act
10. Self-Management	It is possible to handle one's physical and mental health appropriately
11. Reflection	By reflecting on one's thinking and behavior, it is possible to continually seek ways to improve
12. Design Thinking	It is possible to design a solution and develop a comprehensive variety of thoughts and knowledge
13. Presentation	It is possible to convey one's feelings and thoughts to others
14. Judgment	Based on information and thinking, it is possible to make an appropriate decision given the circumstances
15. Implementation	Without fearing failure, it is possible to take a specific action based on one's feelings and thoughts
16. Responsibility	It is possible to face things and take responsibility as a member of society
17. Contribution	Feel joy for someone when something is useful for them, and it is possible to take a specific action
18. Communication	Listen to others' opinions, which can result in creative dialogue

19. Cooperation & Collaboration	Looking beyond one's own interests and those of others, it is possible to work together

Table 2: *Rubric of "Manaburu I"*

Grade / Competency	0	1	2	3	4
I Cooperation & Collaboration	Cannot fulfill a role for oneself.	Cannot fulfill role given to oneself.	Can find role of oneself in the group and fulfill it.	While listening to the opinions of others, including critical opinions, can find role of oneself in the group and fulfill it, and explain the need for the role to others.	Listening to other opinions and critical opinions, can find role of oneself in the group, explain its importance to others, and fulfill it so that group members feel their group performance has improved.
II Exploration	Satisfied with the solutions given by others to the assignments.	Satisfied by giving one idea (opinion) on the assignment.	Can give one idea (opinion) on the assignment with multidirectional thought, and can explain one's reason on one's own.	Can give some ideas (opinions) on the assignment with multidirectional thought, and can logically explain which idea (opinion) is most effective to solve the problem.	Can give one idea (opinion) on the assignment with multidirectional thought, and can logically explain which idea (opinion) most reasonably solves the problem and predict the result.
III Presentation	Cannot describe one's own ideas or initiatives to others.	Can describe one's own ideas or initiatives to others.	Can express one's ideas and initiatives so that others can clearly understand them.	Can show how one's ideas and initiatives differ from others and explain them in an objective and easy-to-understand way.	Can show how one's ideas and initiatives differ from others and explain them in an objective and easy-to-understand way, including what they mean to others.

IV	Reflection	Cannot explain what has been learned.	Can explain what has been learned.	Can reflect collectively on what has been learned and explain on what it meant to ones.	Can explain the results of learning together with own issues and future growth. (Can reflect linking learning with your own growth)	Can explain the results of learning together with own issues and future growth, and can show concrete guidelines on overcoming issues and growth from the results of learning.
V	Self-Management	Cannot prepare the foundation of learning habits and learning environment, for example, cannot put out submissions by due date and/or has nothing to do with group activities.	Can prepare the foundation of learning habits and learning environment, for example putting out submissions by due date and/or actively engaging in group activities.	Can prepare learning environment according to one's own learning style or target content, for example tackling issues on a planned basis, adjusting to an environment suitable for given activities.		
VI	Design Thinking	Cannot provide an idea for one's own assignment.	Can give a general idea for one's own assignment.	Can generate a general idea for the assignment with added ingenuity.	Can generate an idea that is original to the assignment and cannot be seen elsewhere.	Can generate an original idea that can be objectively evaluated on a social scale for the assignment.

Table 3: *Matrix table for evaluation based on rubric of "Manaburu I"*

		I	II	III	IV	V	VI	Total Score
		Cooperation & Collaboration	Exploration	Presentation	Reflection	Self-Management	Design Thinking	
Assignment Report	Midterm		／4		／4			8
	Final		／8		／8			16
Portfolio		／4	／8		／12			24
Other (Group Activities, etc.)		／20		／16		／8	／8	52
Total Score		24	20	16	24	8	8	100
※Please convert "× 2," "× 3," "× 4," "× 5," as raw points on the rubric, to total points "8," "12," "16," "20."								

Using this matrix and based on the rubric using Tokiwa competencies, 19 teachers attempted to evaluate their students. But rubric can be an asset in any classroom and at any education level but it needs to be implemented correctly. A rubric is only as good as its design, support and explanation in its use and conversely the expectations from the use of the rubric should enhance the learning outcomes for the students (Cox, Morrison, Brathwaite, 2015). Without this, a rubric can lead to promotion of shallow learning whilst producing conformity and standardization (Mansilla, Duraisingh, Wofle, & Haynes, 2009).

Our research question in this study is whether students were evaluated consistently and equally from teacher to teacher. To answer this research question, we performed statistical estimation using nonparametric multiple comparisons.

2. Methods

2.1 Confirm whether each item depends on Gaussian distribution for normality estimation

Generally, to confirm whether data for each group depend on a Gaussian distribution or not when the size of the sample is less than two thousand (as here), Shapiro–Wilk estimation is used. The null hypothesis H_0 in Shapiro-Wilk estimation is that the data do depend on a Gaussian distribution; the alternative hypothesis H_1 is that they do not. In this study, we define significance level as 0.05.

2. 2 Nonparametric multiple comparison estimation

After Shapiro–Wilk estimation, which as seen below showed that data did not depend on Gaussian distribution, we performed Steel–Dwass estimation to compare the medians and means

of all pairs of groups using a nonparametric pairwise-ranking method. The null hypothesis H_0 of Steel–Dwass estimation is that medians/means of pairs of groups are different, and the alternative hypothesis H_1 that they are the same.

2.3 Software of statistics analysis

We used JMP, Version 13.0.0, SAS Institute Inc., Cary, NC, 1989–2018.

3. Results and Discussion

The number of students in "Manaburu I" was 263, and the number of teachers, 19. The statistical estimation whose results are presented below was meant to show whether the teachers evaluated the students equally. To answer this question, we first compared means of final grades by the 19 teachers. Shapiro–Wilk estimation confirmed the alternative hypothesis H_1 that the data do not depend on a normal distribution for teachers C (p-value=0.0007), E (p-value=0.0117), L (p-value=0.008), M (p-value=0.0196), and P (p-value=0.0229), and so we used nonparametric methods in the following multiple comparisons.

Next, we performed Steel–Dwass estimation as a multiple comparison method to compare the medians/means of all pairs of groups using nonparametric pairwise ranking. Pairs with ps less than 0.05 are shown in Table 4.

Table 4: *Results of Steel–Dwass estimation*

Target 1	Target 1	p-value
E	B	0.0428
E	G	0.0164
E	M	0.0475
E	O	0.0274
E	P	0.0182

Target 1	Target 1	p-value
F	B	0.0289
F	G	0.0054
F	H	0.0292
F	M	0.0143
F	O	0.0078
F	P	0.0038

These results suggest that the means of data for pairs E-B, E-G, E-M, E-O, E-P, F-B, F-G, F-H, F-M, F-O, and F-P are different. To understand this result in detail, we performed visualization using a matrix (Table 5).

Table 5: *Visualization of Steel–Dwass estimation*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A	–																		
B		–			*	*													
C			–																
D				–															
E					–														
F						–													
G					*	*	–												
H						*		–											
I									–										
J										–									
K											–								
L												–							
M					*	*							–						
N														–					
O					*	*									–				
P					*	*										–			
Q																	–		
R																		–	
S																			–

* p-value less than 0.05

These results suggest that teachers E and F, who were paired, have meaningfully different marking records from the other teachers, which suggests in turn that even using the carefully designed rubric, consistent evaluation is difficult.

4. Conclusion

The results show that teachers do not evaluate students equally. Suggestions for future research, more attention to validity and reliability, a closer focus on learning and research on rubric use in higher education. In the future we will perform additional detailed analysis to investigate which rubric sections and items differed between teachers E and F and the other teachers, and how to avoid such a gap.

References

Association of American Colleges & Universities. (2018, April 20). Retrieved from https://www.aacu.org/value

Brookhart, S. M. (2013). How to create and use rubrics for formative assessment and grading. Alexandria, VA: Association for Supervision & Curriculum Development.

Cox, G. C., Morrison, J., & Brathwaite, B. H. (2015). The rubric: An assessment tool to guide students and Markers. *1st International Conference on Higher Education Advances*, HEAd'15, 26-32. http://dx.doi.org/10.4995/HEAd15.2015.414

Kirimura, T., Takamatsu, K., Bannaka, K., Noda, I., Mitsunari, K., & Nakata, Y. (2018). Design the basic education courses as part of the innovation of management of learning and teaching at our own university through collaboration between academic faculty and administrative staff. *Bulletin of Kobe Tokiwa University*, *11*, 181-192.

Kirimura, T., Mitsunari, K., Kunisaki, T., Gozu, T., Takamatsu, K., Bannaka, K., & Nakata, Y. (2018). Effectiveness of first year experience's course "Manaburu" at Kobe Tokiwa University for university students by using textual analysis. *Bulletin of Kobe Tokiwa University*, *11*, 193-208.

Kirimura, T., Takamatsu, K., Bannaka, K., Noda, I., Mitsunari, K., & Nakata, Y. (2017). Innovation: the management of teaching and learning at our own university through collaboration between academic faculty and administrative staff. *Bulletin of Kobe Tokiwa University*, *10*, 23-32.

Mansilla, V. B., Duraisingh, E. D., Wolfe, C. R., & Haynes, C. (2009). Targeted assessment rubric: An empirically grounded rubric for interdisciplinary writing. *The Journal of Higher Education. The Journal of Higher Education*, *80* (3), 334-353. https://doi.org/10.1080/00221546.2009.11779016

Mitsunari, K., Kirimura, T., Kunisaki, T., Gozu, T., Takamatsu, K., Bannaka, K., & Nakata, Y. (2018). Paradigm shift in education from teaching to learning: Focus on the implementation of academic skills and deep learning I. *Bulletin of Kobe Tokiwa University*, *11*, 7-16.

Spady, W. G. (1994). Outcome-based education: critical issues and answers. Alexandria, VA: American Association of School Administrators.

Takamatsu, K., Murakami, K., Kirimura, T., Bannaka, K., Noda, I., Yamasaki, M., Lim. R.-J. W., Mitsunari, K., Nakamura, T., & Nakata, Y. (2017). A new way of visualizing curricula using competencies: Cosine similarity, multidimensional scaling methods, and

scatter plotting. *Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress On. IEEE,* http://doi.ieeecomputersociety.org/10.1109/IIAI-AAI.2017.29.