

Do Imagined Writing Rubrics Used in University Entrance Examinations Affect Scoring in Classroom?

Takahiro KOWATA
Kogakuin University

Abstract

High-stakes tests significantly impact teaching and learning. This study focuses on writing rubrics used in university entrance examinations in Japan. It investigated how high school teachers imagine the writing rubrics used in university entrance examinations and whether their imaginations affect their own scoring. A total of 129 high school teachers participated in the questionnaire survey. Exploratory factor analysis revealed that the teachers believed university candidates' writing products in entrance examinations are rated using four criteria: Consistency of Text, Accuracy, Effective Expression, and Mechanics. They rated their students' writing products based on the same factors in their classes. Multiple regression analyses showed that their ratings were affected by their beliefs about university entrance examinations. This implied that the writing rubrics used in university entrance examinations may have an important role in the context of Japan.

Keywords

Writing, Rubric, University Entrance Examination

1. Background

Tests impact teaching and learning, which is known as the washback effect. MEXT (2002) stated that university entrance examinations in Japan significantly impact the teaching and motivation of learners and have highlighted the necessity of reforming selection methods. As a result, an English listening section was introduced to the National Center Examination in March, 2007. In this particular exam, the writing and speaking skills of the learners are not measured directly. Only approximately 20% of entrance exams included a writing composition section (Kowata, 2009).

The Central Council for Education (2014) discussed the introduction of a new national examination that would measure all four language skills, including speaking and writing. This change also affects teaching. After the introduction of the new university entrance examinations in Chinese universities, the National Matriculation English Test (NMET), for example, reported that teachers had revised their usage of time for the four skills (Li, 1990). Studies have also

reported the effects of the new Sri Lankan O-level test (Wall & Alderson, 1993) and the new school-leaving examination in Hong Kong, the Hong Kong Certificate of Education Examination (HKCEE) (Cheng, 2005).

Wall and Horák (2006, 2008, 2011) showed the effects of introducing a new writing section in TOEFL iBT on teaching. These studies showed that teachers used a new rubric to rate and give feedback to their students' writing. Since washback effects are complex phenomena and highly dependent on the context (Wall & Alderson, 1993), the Japanese context may not show the same effects. Therefore, we should investigate to determine whether the same effect can be observed in the Japanese context.

One major difference from the contexts of the above-mentioned studies that has been observed in the Japanese context is that most Japanese universities do not publish their scoring rubrics and samples (Kanatani, 2009; Negishi, Matsuzawa, Sato, Toyoda, & Nakano, 2010). Consequently, high school teachers must themselves determine what writing rubrics may be used in university entrance examinations to use them when scoring and teaching students in their classrooms. Therefore, clarifying how high school teachers may create their own ideas of the writing rubrics used in university entrance examinations is essential.

2. Study

2.1 Research Question

This study was undertaken to seek answers to the following two research questions.

RQ1: How do high school teachers imagine the writing rubrics used in Japanese university entrance examinations?

RQ2: Do high school teachers' imaginations of the writing rubrics used in Japanese university entrance examinations affect their scoring of students' writing products in class?

2.2 Participants

A total of 129 teachers (74 males and 55 females aged between 23 and 64) from 33 high schools participated in the study. The selected schools include 2 national schools, 18 public schools, and 13 private schools; all schools are located in 16 prefectures.

2.3 Materials

A questionnaire survey was conducted. It consisted of two parts.

The first part of the questionnaire comprised questions that asked the

participants to rate the extent to which they believed 15 scoring criteria (A01-A15) affected scoring in the university entrance examinations (Table 1). The scoring criteria were selected from the descriptors in the writing scoring rubrics of the ESL Composition Profile (Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey, 1981). This profile adopted an analytic scale, and the writing products are rated based on five differentially weighted aspects: Content, Organization, Vocabulary, Language Use, and Mechanics. This scale was chosen because it has been widely used in ESL studies and in rubrics of such worldwide large-scale standardized tests as Cambridge English (FCE), IELTS, and TOEFL iBT that use a holistic scale including the five aspects.

Table 1 Writing Scoring Perspectives

A01	Relevance to assigned topic
A02	Logical structure
A03	Coherence
A04	Cohesion
A05	Complexity of sentence structure
A06	Wide range of sentence structure
A07	Accuracy of grammatical structure
A08	Accuracy of vocabulary
A09	Appropriate choice of vocabulary
A10	Wide range of vocabulary
A11	Spelling
A12	Punctuation
A13	Capitalization
A14	Paragraphing
A15	Legibility of handwriting

The second part of the questionnaire asked the participants what weight they gave to the same 15 criteria when scoring students' writing products in their classes. These items were labeled B01-B15 correspondingly.

2.4 Analysis

First, items in the participants' imagination of the rubric used in university entrance examinations (hereafter the rubric image) were analyzed (Analysis 1). A maximum-likelihood exploratory factor analysis was conducted with promax rotation to assess the underlying structure for the 15 items. Then, factor scores were calculated. In this analysis, data from 10 participants were excluded

due to their lack of response, and data from the other 119 participants were analyzed. This sample size was too small to conduct reliable factor analysis, so the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was used to check the reliability. According to Field (2005), KMO values between .5 and .7 are acceptable, values between .7 and .8 are good, values between .8 and .9 are great, and values above .9 are superb.

Second, the items on the participants' own scoring in their classes (hereafter scoring in class) were analyzed (Analysis 2). This was followed by factor analysis and the calculation of factor scores. In these items, data for 56 teachers lacked proper response from the participants and thus were excluded; only data from the remaining 73 were analyzed. The main reason for this lack of answers was that some participants did not teach writing composition in their classes. The KMO was used to check the reliability of this factor analysis.

Finally, a multiple regression analysis was conducted to see if the factor scores of the scoring rubric image predicted those for scoring in class (Analysis 3). In this analysis, 71 data were analyzed.

Statistical analyses in this study were performed using SPSS 11.5.

2.5 Results

2.5.1 Teachers' University Entrance Examinations Rubric Images (Analysis 1)

Table 2 shows the results of the descriptive analysis and factor analysis on rubric images. "Relevance to assigned topic" (A01) was given the highest mean value, 4.56, and "Logical structure" (A02) followed it with 4.48. "Complexity of sentence structure" (A05) received the lowest mean value, 2.79, and "Wide range of sentence structure" (A06) followed it with 2.88.

After calculating descriptive statistics, a maximum-likelihood factor analysis was conducted for 15 items on scoring rubric image. Initial eigen values indicated 5.44%, 2.15%, 1.98%, 1.29%, .87%, and .71% of the variance. The four-factor solution was preferred due to the leveling off of eigen values on the scree plot after four factors and the insufficient number of primary loadings.

Next, a maximum-likelihood factor analysis with promax rotation was conducted with four factors. The final factor-loading matrix is presented in Table 2. The four factors before rotation constituted 72.40% of the variance.

The first factor (FA1) was given the name Consistency of Text, the second factor (FA2) was Effective Expression, the third factor (FA3) was Mechanics, and the fourth factor (FA4) was Accuracy. The internal consistency of each of the scales was examined using Cronbach's alpha. The alphas were moderate: .85 for Consistency of Text (4 items), .85 for Effective Expression (3 items), .82 for Mechanics (4 items), and .87 for Accuracy (3 items).

The KMO value was .765, and this indicated that the analysis maintained good reliability.

Table 2 Factor Loadings Based on a Maximum-Likelihood Factor Analysis With Promax Rotation for 15 Items on the Scoring Rubric Images (N=119)

Item		<i>M</i>	<i>SD</i>	Factor Loading			
				FA1	FA2	FA3	FA4
A02	Logical structure	4.48	.67	.895	.129	.295	.294
A01	Relevance to assigned topic	4.56	.65	.820	.076	.260	.265
A03	Coherence	4.32	.85	.710	.284	.279	.331
A04	Cohesion	4.08	.90	.683	.409	.334	.374
A06	Wide range of sentence structure	2.88	.97	.212	.956	.149	.213
A05	Complexity of sentence structure	2.79	.97	.245	.843	.191	.169
A10	Wide range of vocabulary	3.20	1.00	.130	.645	.264	.375
A13	Capitalization	3.71	1.05	.291	.112	.929	.430
A12	Punctuation	3.40	1.00	.253	.306	.784	.462
A11	Spelling	4.19	.84	.378	.166	.770	.640
A14	Paragraphing	3.51	1.01	.376	.382	.455	.336
A08	Accuracy of vocabulary	4.13	.76	.235	.223	.453	.858
A07	Accuracy of grammatical structure	4.25	.77	.452	.104	.655	.837
A09	Appropriate choice of vocabulary	3.92	.81	.381	.317	.399	.829
A15	Legibility of handwriting	3.32	1.07	.159	.311	.296	.161

Note. FA1 = Consistency of Text; FA2 = Effective Expression; FA3 = Mechanics; FA4 = Accuracy

2.5.2 Teachers' Scoring in Class (Analysis 2)

Table 3 shows the results of descriptive analysis and factor analysis on the scoring rubric images. "Relevance to assigned topic" (B01) was given the highest mean value, 4.42, and "Logical structure" (B02) followed it with 4.18. "Complexity of sentence structure" (B05) received the lowest mean value, 2.47, and "Legibility of handwriting" (B15) followed it with 2.49.

After calculating descriptive statistics, a maximum-likelihood factor analysis was conducted for 15 items related to scoring in class. Initial eigen values indicated 5.91%, 2.48%, 1.99%, 1.03%, .69%, and .62% of the variance. The four-factor solution was preferred due to the leveling off of eigen values on the scree plot after four factors and the insufficient number of primary loadings.

Next, a maximum-likelihood factor analysis with promax rotation was

conducted with four factors. The final factor-loading matrix is presented in Table 3. The four factors before rotation accounted for 76.10% of the variance.

The four factors comprised the same items as Analysis 1. The first factor (FB1) was called Consistency of Text, the second factor (FB2) was Accuracy, the third factor (FB3) was named Mechanics, and the fourth factor (FB4) was Effective Expression. The internal consistency of each of the scales was examined using Cronbach's alpha. The alphas were moderate: .86 for Consistency of Text (4 items), .89 for Effective Expression (3 items), .84 for Mechanics (4 items), and .84 for Accuracy (3 items).

The KMO value was .769, and this indicated that the analysis maintained good reliability.

Table 3 Factor Loadings Based on a Maximum-Likelihood Factor Analysis With Promax Rotation for 15 Items About the Scoring Rubric in Classes (N=73)

	Item	<i>M</i>	<i>SD</i>	Factor Loading			
				FB1	FB2	FB3	FB4
B02	Logical structure	4.18	1.00	.990	.326	.163	.293
B03	Coherence	4.03	1.07	.834	.327	.233	.260
B01	Relevance to assigned topic	4.42	.83	.831	.286	.185	.150
B04	Cohesion	3.75	1.04	.614	.384	.232	.219
B08	Accuracy of vocabulary	3.77	1.01	.326	.970	.602	.217
B09	Appropriate choice of vocabulary	3.58	.97	.416	.825	.557	.357
B07	Accuracy of grammatical structure	3.96	.98	.454	.791	.627	.289
B12	Punctuation	3.15	.98	.210	.635	.919	.312
B13	Capitalization	3.32	1.18	.051	.511	.814	.251
B11	Spelling	3.84	1.05	.187	.715	.749	.249
B14	Paragraphing	3.42	1.21	.521	.440	.639	.185
B06	Wide range of sentence structure	2.59	1.12	.156	.282	.376	.901
B10	Wide range of vocabulary	2.67	1.03	.282	.317	.266	.800
B05	Complexity of sentence structure	2.47	1.04	.207	.136	.136	.728
B15	Legibility of handwriting	2.49	1.04	.304	.207	.356	.401

Note. FB1 = Consistency of Text; FB2 = Accuracy; FB3 = Mechanics; FB4 = Effective Expression

2.5.3 Predicting Teachers' Scoring in Classes From Their Scoring Rubric Images of University Entrance Examinations (Analysis 3)

Using the enter method, four multiple regression analyses were conducted to predict the factor scores of scoring in class. In the analyses, the independent

variables were the factor scores of scoring rubric image (Consistency of Text, Accuracy, Mechanics, and Effective Expression), and dependent variables were the factor scores of scoring in class (Consistency of Text, Accuracy, Mechanics, and Effective Expression). The variance inflation factors (VIF) of independent variables were 1.18-2.01, indicating that covariance problems were not found.

Table 4 shows the results of the multiple regression analyses. In the four analyses, only the same named factor scores indicated significant standardized partial regression coefficient (β), for example, .699 for Consistency of Text of scoring rubric image predicting Consistency of Text of scoring in class. Similarly, the significant index indicated .322 for Accuracy, .734 for Mechanics, and .533 for Effective Expression.

Table 4 Results of Multiple Regression Analysis (N=71)

		Scoring in Class (dependent variables) β			
		Consistency of Text	Accuracy	Mechanics	Effective Expression
Scoring Rubric Image (independent variables)	Consistency of Text	.699***	.108	.024	.009
	Accuracy	.049	.322*	-.041	-.006
	Mechanics	-.071	.218	.734***	.048
	Effective Expression	-.024	-.126	-.007	.533***
R^2		.477	.285	.511	.303

Note. * $p < .05$, *** $p < .001$

3. Discussion

Using factor analysis for participants' rubric images, four factors (Consistency of Text, Effective Expression, Mechanics, and Accuracy) were extracted. These categories do not exactly correspond to those of Jacobs et al.'s (1981) scoring profile: Content, Organization, Vocabulary, Language Use, and Mechanics.

Items related to Consistency of Text were from the descriptors of Content and Organization of the profile. This result implies that the teachers believed the content of a writing product is highly related to its organization. Teachers might believe good contents support good organization, and/or vice versa.

A huge difference was found in Accuracy and Effective Expression, as shown in Table 5. In the ESL Composition Profile, the characteristics of writing products were divided into Vocabulary and Language Use. The former includes accuracy, appropriateness, and range of vocabulary. The latter includes accuracy, complexity, and range of sentence structure. In contrast, in the rubric image, the

same characteristics were divided into Accuracy and Effective Expression. The former includes the accuracy and appropriateness of the vocabulary and sentence structure. The latter includes the range and complexity of the vocabulary and sentence structure. This implies that the rubric imagined by teachers might differ from those used by universities in their entrance examinations.

Table 5 Difference Between the Scoring Rubric Image and the ESL Composition Profile

		the ESL Composition Profile	
		Vocabulary	Language Use
the scoring rubric image	Accuracy	Accuracy of vocabulary Appropriate choice of vocabulary	Accuracy of grammatical structure
	Effective Expression	Wide range of vocabulary	Wide range of sentence structure Complexity of sentence structure

According to factor analysis of participants' scoring in class, the same four factors as in their scoring rubric image were extracted. Multiple regression analyses were conducted based on the factor scores of the scoring rubric image and scoring in class. One result showed that only the Consistency of Text of the rubric image predicted the Consistency of Text of scoring in class. Only the same structured factor scores predicted their scoring in class for each factor. This implies teachers' university entrance examination rubric images affected their scoring in class. There is some possibility that the scoring in class affects the rubric images; this leads to a chicken-and-egg problem. However, if universities publish their scoring rubrics, teachers will use them to rate their students' writing products.

Universities do not publish their scoring rubrics and teachers must analyze the test specifications from question statements. In such cases, teachers can often determine *what* should be written as the content of written product, but they have difficulty figuring out the criteria of *how* it should be written from the perspective of organization, etc. They must imagine how the rubrics are designed, both from the content and organizational perspectives. Therefore, whether their ideas are correct, they guide lessons and affect teachers' grading.

It is to be expected that when the new type of Japanese university entrance examination with a writing section is introduced, the writing rubric will be published, just as is done with other large-scale tests such as TOEFL iBT and GTEC for STUDENTS; this will affect teaching at high schools. Kowata (2008) interviewed freshmen who had prepared a writing composition for their university

entrance examination. He reported that most of them had never thought the analytic perspectives for good writing as writing rubric descriptors show and had followed teachers' advice when they practiced writing. This result also indicates teachers' understanding of the importance of scoring rubrics in students' learning.

In the summer of 2015, the Tokyo University of Foreign Studies published their basic writing scoring rubric and four high-scoring samples from their entrance examinations, held in March, 2015. To the best of my knowledge, other universities in Japan do not make such detailed information available. This might affect next year's candidates and their teachers.

If this effort spreads to other universities, fewer teachers will have to be concerned about unexpected ratings. Simultaneously, universities will be able to show ideal or adequate writing products to their new students, and high school teachers need the ability to analyze and understand proposed writing products.

4. Conclusion

This study investigated teachers' images of writing rubrics used in university entrance examinations and how teachers rate their students' writing products in class. As for the first research question, which explored teachers' images of writing rubrics, it was found that high school teachers believed candidates' writing products in university entrance examinations are rated on the following four criteria: Consistency of Text, Accuracy, Effective Expression, and Mechanics. As for the second question, which investigated whether teachers' images affect their scoring in class, results indicated that teacher beliefs affect how students' writing products in class are assessed.

5. Further Research

This study showed that the writing rubric used in university entrance examinations would affect teachers' ratings in class and contributed to understanding the role of rubrics. One of the limitations of this study is that teachers' real scoring behaviors were not investigated. Teachers may rate differently to how they self-report. Actual writing products scored by teachers should be examined in the future study.

References

Central Council for Education (2014). *Atarashii jidai ni fusawashii koudai setsuzoku no jitsugen ni muketa koutougakkoukyouiku daigakukyoku daigakunyugakusha senbatsu no ittaiteki kaikaku ni tsuite (toshin)* [An integrated reform of high school education, university education, and selection of university entrants to realize the connection between high schools and universities suitable for a

- new era (report)]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/_icsFiles/afieldfile/2015/01/14/1354191.pdf [2015, August]
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). CA: Sage Publications, Inc.
- Jacobs, H., Zinkgraf, S., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *English composition program*. Rowley, MA: Newbury House Publishers Inc.
- Kanatani, K. (2009). *Kyoukasho dakede daigakunyushi wa toppa dekiru* [It is possible to pass the university entrance examination only with a textbook]. Tokyo: Taishukan Shoten.
- Kowata, T. (2008). Jukensha no writing mondai no toraekata to sono taisaku—Nihon no daigakunyugakushiken kankyou ni okeru hakyuukouka [How writing tests interpreted and prepared for? : Washback effect in the context of Japanese university entrance examinations]. *ARCLE REVIEW*, 3, 122-133.
- Kowata, T. (2009). Nihon no daigakunyugakushiken ni okeru hakyuukouka no kanousei —Writing test bunseki ni shouten wo atete [The possibility of washback effects in the context of Japanese university entrance examinations: Focusing on the analysis of writing tests]. *Language, Area and Culture Studies*, 15, 81-93.
- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11(5), 393-404.
- MEXT (2002). “*Eigo ga tsukaeru nihonjin*” ikusei no tameno senryaku kousou [A strategic plan to cultivate “Japanese with English abilities”]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chousa/shotou/020/sesaku/020702.htm#plan [2015, August]
- Negishi, M., Matsuzawa, S., Sato, R., Toyoda, Y., & Nakano, T. (2010). Daigakunyushi ga kawareba eigo kyouiku mo kawarunoka [If the university entrance examinations change, does English education also change?]. *Eigo Kyouiku*, 59(5), 10-19.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study*. Princeton, NJ: ETS.
- Wall, D., & Horák, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 2, coping with change*. Princeton, NJ: ETS.
- Wall, D., & Horák, T. (2011). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 3, the role of the coursebook, and Phase 4, describing change*. Princeton, NJ: ETS.