

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 20, Number 6, March 2015

ISSN 1531-7714

What is your teacher rubric? Extracting teachers' assessment constructs

Heejeong Jeong, *Hanyang University*

Rubrics not only document the scales and criteria of what is assessed, but can also represent the assessment construct of the developer. Rubrics display the key assessment criteria, and the simplicity or complexity of the rubric can illustrate the meaning associated with the score. For this study, five experienced teachers developed a rubric for an EFL (English as a Foreign Language) descriptive writing task. Results show that even for the same task, teachers developed different formats and styles of rubric with both similar and different criteria. The teacher rubrics were analyzed for assessment criteria, rubric type and scale type. Findings illustrate that in terms of criteria, all teacher rubrics had five areas in common: comprehension, paragraph structure, sentence structure, vocabulary, and grammar. The criteria that varied were mechanics, length, task completion, and self-correction. Rubric style and scales also were different among teachers. Teachers who valued global concerns (i.e., comprehension) in writing designed more general holistic rubrics, while teachers who focused more on sentence-level concerns (i.e., grammar) developed analytic rubrics with more details. The assessment construct of the teacher was shown in the rubric through assessment criteria, rubric style, and scale.

When assessing writing, teachers have various ways of operationalizing the construct of a student's writing ability. Each teacher has a different set of criteria of what qualifies as good writing compared to poor writing (Erdosy, 2003; Lumely, 2002; Lumely, 2006). This calls for rigorous rater training and clearly defined rubrics for teachers to agree upon a similar assessment construct (Alderson, 1991, Lovorn & Rezaei, 2011). However, there are assessment traits that cannot be fully erased with training; therefore, it is important to find what these differences are, as well as their origins. Different perspectives in assessing writing among teachers can be based on various factors, such as a teacher's rating style, personal characteristics, rating experience, and educational background (e.g., Lumley & McNamara 1995 ; Weigle, 1998) . This study investigates variations among teachers on what constitutes effective EFL(English as a Foreign

Language) descriptive writing by looking into the rubrics developed by teachers.

A rubric is not only a document showing the scales and criteria of what is assessed and how scores are given, but it can also function as a piece of evidence to represent the teacher's assessment construct. Rubrics contain teacher's beliefs in the factors that should be included to assess the task and show the scale and criteria that should be covered. Rubrics display what is meaningful to the teacher, and the simplicity or complexity of the rubric illustrates how much detail he or she looks for when grading a student's paper.

In the literature, there have been many studies reporting the increase of rater reliability through rating training (Lumely & McNamara, 1995; McNamara, 1996; Weigle, 1998; Lovorn & Rezaei, 2011). Unfortunately, for classroom teachers it is difficult to

have professional rater training opportunities (Knoch et al., 2007). In this paper, I want to propose a method for classroom teachers to have a better understanding of their own assessment constructs by developing and analyzing rubrics developed by teachers. Visualizing one's assessment construct can help teachers have a better understanding of themselves as an assessor which can result into an increase in the reliability and validity of teachers' assessment practice.

For this study, teachers were asked to design a rubric to assess short (1-3 paragraphs) EFL descriptive writing samples. Teachers were given total freedom in the design of the rubric, but were not allowed to use any references. The reason for this research design was to see how differently or similarly teachers responded to the same writing task when they had the liberty to decide on the assessment criteria. The purpose of the study was not to evaluate the quality of the rubric, but to focus on the variations among teachers. Past studies in rater judgment variations have focused on what raters attended to by analyzing rater think-alouds (Barkaoui, 2010; Cumming, 1990; Cumming, et al. 2002; DeRemer, 1998; Lumley, 2006) and by investigating rating justifications (Gamaroff, 2000). This study will take a step further in researching rating variations by examining teacher-developed rubrics. To minimize different variables in the design, teachers were selected from a similar background; all worked for the same language program, had similar rating experiences, and taught similar classes.

Meaning behind the score

Studies on rater reliability and validity have a long and rich history in assessing productive language skills. Raters are needed for both speaking and writing performance assessment; moreover, the goal of achieving higher rater reliability has long been the mission for many testers and testing organizations. Massive amounts of money are invested in developing rater training programs, and a variety of different statistical analyses (e.g., FACETS, G-studies, D-studies) are used to verify the effect of rater training. One under-researched area in rater reliability and validity studies is the meaning associated with a score. Gamaroff's (2000, p.42) study, which looked at the relationship between rater judgments and scores, states that "similar scores between raters do not necessarily mean similar judgments." Raters in his study were

asked to assign a score for students' essays in the areas of *topic relevance*, *content*, and *grammatical score*. After assigning a score, raters wrote reasons for the criteria they specified in their scores. Gamaroff found that raters gave different reasons for the same scores; for example, "A score of 3 for one rater represented 'meaningless cloudy' and for another rater the same score of 3 represented 'misspelled many words but not too bad'" (p.42). He argues that variance in rater judgments can be a threat to validity. Similar findings were reported in DeRemer's (1998) study, which examined what raters think and attend to when they grade student essays. In this paper, DeRemer (1998) states that the rating process is similar to a problem-solving activity, where raters construct scoring decisions. Rating criteria can be interpreted in different ways, thereby resulting in different meanings for the same scores.

Rubric analysis

While there have been many studies on rater training (Knoch et al., 2007; Lovorn & Rezaei, 2011) and rater reliability (Lumley & McNamara, 1995; McNamara, 1996; Weigle, 1998), there has hardly been a study that looks specifically at a teacher's assessment construct for writing assessment through teacher-developed rubrics. There are many rubric-related sources available for classroom teachers on how to develop and use rubrics, but very few that discuss the construct of a teacher rubric. As Jonsson and Svingby's (2007) review paper shows, the majority of rubric-related articles discuss the development and benefits of using rubrics. Rubric validation studies are difficult to find, and most studies do not mention the assessment construct of classroom teachers. In short, empirical studies concerning rubrics largely focus on rater reliability issues.

Investigations on rating behavior (Rezaei & Lovorn, 2010) and rating processes (Barkaoui, 2010; Barkaoui, 2011; DeRemer, 1998; Lumley, 2002) have lightly touched on how raters interpret and use rubrics; however, none have looked into the existence or analysis of a teacher rubric. General steps on how to design (Mertler, 2001) and analyze rubrics have been proposed by multiple researchers (Arter & McTighe, 2001; Moskal, 2003) and a systematic analysis procedure is stated in Arter and Chappuis's (2006) book. In this book, the authors present steps on evaluating rubrics by looking into factors concerned

with assessment criteria description, criteria coverage and organization. Rubrics are also analyzed by the number of levels, level definition and level consistency. Tierny and Simon's (2004) paper investigates the quality of rubrics by examining consistency of performance criteria across scale levels. The authors in this study stress the importance of referencing the same attributes in the descriptors across levels and the precision of language in rubrics to advance the rubric design.

Research on teacher made rubrics is a new area, given that teachers are not often invited to take part in the rubric development process from school districts or large language institutions. This study attempts to define teachers' assessment construct by analyzing teacher-made rubrics.

Research Questions

1. *What criteria do teachers use in assessing short EFL descriptive writing?*
2. *What information can be inferred through teacher rubrics analysis?*

Methodology

Teachers

The teachers for this study were EFL instructors from a large private university. The five teachers selected for the study were native English speakers with extensive teaching experience in a higher education context (Table 1). The teachers' rating experiences were mostly within the university setting. The courses they taught were graded on the basis of students' performance in activities, and they had experience rating the English placement test conducted every year for freshman students. A few had experience taking part as judges for an essay contest and a presentation contest held on campus.

Data collection process

Data for this study came from teacher-developed rubrics and teacher interviews. Teachers were asked to develop a rubric for a placement test (Appendix A) for EFL university freshmen students in face-to-face meetings with the researcher. The teacher rubric was developed on a given piece of blank paper(s), and

teachers had the freedom to design it in any kind of form or style with the only requirement being that they had to cover a 6-point scale (A+, A, B+, B, C+, C). No

Table 1. Teacher background and experience.

	Teaching Experience	Nation-ality	Rating experience
Susan	5yrs	American	Placement Test, Essay Contest
Eunice	12yrs	American	Placement Test
Matt	10yrs	British	English Placement Test, Essay Contest, Presentation Contest
Logan	12yrs	Australian	Placement Test, Essay Contest
Ben	12yrs	Canadian	Placement Test, Large-scale speaking test

other references, except for the student essay samples and the teacher's essay justifications, were given. Twenty student essays from the placement test were given to the teachers ahead of time, and they were asked to rate the essays on a 6-point scale. For the essay ratings, teachers were not given a rubric; rather, they were asked to assess them based on their own intuition. A short justification for their ratings was required, along with their letter grade ratings. Results of the student essay ratings are beyond the scope of this paper, and reported elsewhere (Jeong, in press). Nevertheless, the essay justifications were analyzed in order to identify the key assessment criteria (Appendix B) that appeared in the justifications. The two sources that allowed teachers to develop their rubrics were the teacher's own rating justifications and the student essay samples.

For the sake of this study, the teacher rubrics were purposefully created at the teacher-researcher meeting, without any external references because the intention of this task was not to develop a solid rubric that could be used to rate the placement test, but rather to extract teachers' assessment construct in rating the essays.

After the teacher rubric was developed, an in-depth interview (Appendix C) was conducted to elicit the reasons behind their choice of criteria, layout, and format of the teacher rubric. The interviews were digitally audio-recorded and were transcribed for themes and patterns. Collected teacher rubrics were analyzed and compared, and were contrasted by form and content.

Data analysis

Teacher rubrics were analyzed first by format and later by content. Rubric analysis, focusing on format and style, followed the rubric examination steps adapted from Arter and Chappuis (2006). Teacher rubrics were compared and contrasted on i) assessment criteria; ii) criteria descriptor language; iii) rubric type; and iv) scale type. Next, the teacher rubric content analysis was conducted by following the nine key assessment criteria (Appendix B) derived from student essay justifications written by the teacher. The rating justifications were summarized and condensed using descriptive coding, and nine key assessment criteria were identified: comprehension (COM), paragraph structure (PS), sentence structure (SS), vocabulary (VOC), grammar (GM), mechanics (MC), length (LNT), task completion (TC), and self-correction (SC). These criteria were the assessment features appearing in the essay justifications. The descriptions in the teacher rubric were condensed and summarized, following the nine assessment criteria.

Findings

Teacher rubric format

Five teacher rubrics were first examined based on structure and format (Table 2). First, the assessment criteria were identified. Eunice's rubric was broken down into four criteria: vocabulary, sentence and paragraph structure, coherence and transitions, and word usage. Ben had five, which were structure, vocabulary, mechanics, sentence variety, and grammar. Susan and Logan did not have separate criteria in their rubric, but instead had an overall description for each level. Matt's rubric was a list of 11 questions, with no separate criteria.

Next, the criteria descriptor language was analyzed to check whether the teachers used qualitative,

quantitative or a combination of both types. Eunice included specific numbers (e.g., 2 parts missing, 1-2 verb tense mistakes) in the descriptors, along with qualitative labels, such as "sophisticated use of language, high-level words." The other four teachers' descriptive language was mostly qualitative, but included some numerical language (e.g., many errors, no errors, few errors, all correct). The descriptor language used in the teacher rubrics focused on the quality of the writing (e.g., excellent, good, poor), the proficiency of the writer (e.g., native like, high level, advanced), and the frequency of a specific criterion (e.g., few grammatical errors, no spelling errors).

The scale of the teacher rubric was developed within a 6-point letter grade (A+, A, B+, B, C+, C). Teachers developed the rubrics following this guideline, but the calculation of the final grade was presented differently. Susan's rubric and Logan's rubric were holistic, with a list of descriptions explaining their expectations next to each letter grade. Eunice had a detailed description for each criterion at every level. For Matt's rubric, points (ranging from 1-6) were assigned for each question but did not have a scoring band to convert the total sum into a letter grade. Ben converted each criterion into points and gave a scoring band at the bottom, which could be used to assign a letter grade (e.g., A: 24-26, B:19-21). Ben's rubric included a scale definition for the top and bottom level. In terms of scale consistency, all teachers except Logan were parallel in content. Thus, if a certain criterion was discussed at one level, it was covered at all levels.

Teacher rubric content

Table 3 shows a summary of the teacher rubrics by criteria and their descriptions, based on the nine assessment criteria mentioned in the essay rating justifications. Even though the teachers did not interact or communicate with one another in the rubric development task, the teacher rubrics shared common features in regard to content. Overall, all teacher rubrics had five areas in common: COM, PS, SS, VOC, and GM. This meant that the rubric developed by each teacher included these assessment areas. The criteria that varied were MC, LNT, TC, and SC.

The specific description used for describing each criterion had similarities and differences. For Eunice, COM mostly concerned flow and coherency, with

Table 2. Teacher rubric format summary

	Eunice	Susan	Ben	Matt	Logan
Criteria	4 criteria	Not separate, overall description	5 criteria	Not separate, 11 questions	Not separate, overall description
Rubric Language	qualitative, quantitative	qualitative	qualitative	qualitative	qualitative
Lg by Level	<i>sophisticated high level</i>	<i>well structured, developed, organized, few/ no mistakes, excellent use, varied</i>	<i>well organized, proper use, advanced, no errors, variety</i>	<i>appropriate, clear, extensive, variety, correctly, effectively</i>	<i>excellent, native like, few errors, complex</i>
A+ / A	<i>complex natural used correctly</i>				
B+ / B	<i>not high level, simple, attempts to, missing, choppy, 3-4 mistakes</i>	<i>simple, some areas, good use of, easy to follow, under-standable, good use, attempts</i>			<i>good, some errors, intelligible, limited, some ability, starting to develop, simple</i>
C+ / C	<i>simple, childlike, no attempt to, 7-8 mistakes</i>	<i>limited use, insufficient, difficult to follow, only simple, unable, many mistakes</i>	<i>poor, limited, incorrect, many errors, simple, no variety, multiple errors</i>	<i>many errors</i>	<i>inconsistent, simple, able to follow, borderline, very basic</i>
Scale Type	6-point letter grade	6-point letter grade	points converted to 6-point letter grade	points converted to 6-point letter grade	6-point letter grade
Rubric Type	analytic	holistic	analytic	analytic	holistic
Scale Consistency	parallel	parallel	parallel	parallel	non-parallel
Scale Definition	all levels	all levels	top and bottom	none	all levels

Table 3. Teacher rubric criteria summary

	Eunice	Susan	Ben	Matt	Logan
COM	coherency, transitions, logical flow, idea, meaning	idea, readability	idea	clarity, fluency	intelligibility, able to follow
PS	understanding of paragraph structure, topic, supporting, concluding paragraphs	organization, use of paragraphs	organization	intro, body, conclusion	organization
SS	sentence variety	sentence variety	sentence variety; simple/complex /compound sentences	sentence variety	sentence style
VOC	academic word list, use, understanding	use	word choice, proper use	expression	English idioms, expression
GM	verb tense, parts of speech	grammar	verb tense, S/V agreement, conjunctions	verb tense, article, preposition	verb tense
MC	punctuation		punctuation, capitalization, spelling	punctuation, capitalization	spelling
LNT		sufficient amount		multi- paragraphs	
TC				under-standing the task	
SC		can/limited/ unable to self-correct			

a minor focus on idea and meaning. In Susan’s rubric, a comprehensible essay was one that was easy to read and had good ideas. Ben also thought of COM as organization of ideas; however, in Matt’s case, clarity and fluency were most significant. COM, which is a broad concept, had similar (fluency, flow) and different (idea, readability) definitions among the teachers. The description of paragraph structure (PS) was quite similar across teachers. Paragraph structure (PS) involved paragraph organization with a clear introduction, body, and conclusion. Sentence structure

(SS) described the variety and complexity of sentences. Explanations for vocabulary (VOC) were similar, which covered the choice, use, and understanding of words. Some teachers were more specific in their expectations for VOC. For instance, Logan went into more detail, stating VOC-covered idioms and expressions, and Eunice included using words from an academic word list. Grammar (GM) included verb tense agreement for all four teachers, except Susan. Other specific areas described under GM were conjunctions, articles, prepositions, and parts of

speech. Detailed descriptions of the particular types of grammar assessed for each teacher varied. Mechanics (MC) covered punctuation, capitalization, and spelling, but the degree of attention given to this criterion was not the same. Similar to MC, length (LNT), task completion (TC), and self-correction (SC) criteria appeared in one or two teacher rubrics, but not in all of them.

Individual Teacher Rubric Description

Eunice

Eunice's rubric was the longest and most detailed out of the five teacher rubrics. Her rubric was developed across two pages and had four criteria: vocabulary, sentence and paragraph structure, coherence and transition, and word usage. Vocabulary appeared first on her rubric and contained detailed descriptions for that criterion. To qualify for the "A+" level, sophisticated words from the academic word list had to be included, and for the "C+" level, she stated simple and somewhat childlike vocabulary as requirements. In evaluating vocabulary, she included detailed descriptors such as "sophisticated, high level, natural, correct, awkward, and simple." The second criterion in Eunice's rubric was sentence and paragraph structure. These two assessment criteria were combined into one section. The sentence and paragraph structure criteria covered sentence structure, sentence variety, and parts of a paragraph. Paragraph structure involved topic, supporting, and concluding sentences. Coherence and transitions were stated as the third category and covered the logical flow of time, space, and ideas through transitions. The idea of the essay was embedded within the coherence and transition category. The last assessment criterion in Eunice's rubric was word usage, which concerned verb tense usage and the parts of speech.

Susan

Susan's rubric resembled a style commonly found in a holistic writing rubric. The left-hand side of the rubric stated the six levels, following a level description. She had 5~8 descriptors for each letter grade, which were not categorized into separate criteria across levels. Even though the descriptors were not classified, they were parallel across levels. Descriptions

for each level first covered essay ideas (e.g., well-structured ideas, difficult-to-follow ideas), organization, grammar, vocabulary, sentence structure, and use of paragraphs. She also included expectations for sufficient essay length and self-correction ability in her rubric. The language used in her rubric was broad and vague. She would state "many grammar mistakes, only use simple sentences," without any specific information on error or sentence type.

Ben

Ben developed a short half-page rubric with five assessment criteria: structure, vocabulary, mechanics, sentence variety, and grammar. These five criteria were written at the top of the page, followed by a simple description in parentheses. Structure was stated first and was described as "organization of ideas/paragraphs." Vocabulary was described as word choice, and mechanics covered "punctuation, capitalization, and spelling." Ben was specific in what he expected for sentence variety, namely complex and compound sentences. For grammar, Ben was looking for correct verb tense, subject-verb agreement, and correct conjunction use. Ben did not define the expectations for each level, but included the requirements for the top and bottom levels for the five assessment criteria.

Matt

Matt's rubric was similar to a checklist, with questions on the left-hand side and points on the right-hand side. Each point was aligned with a letter grade (e.g., A+ = 6), and there was a check box for the points. The rubric had 10 questions (e.g., Did the student understand the question and provide an appropriate answer?) and one statement (overall writing fluency) at the end. Among the teacher rubrics, Matt's had the most assessment points. Matt's rubric consisted of 8 criteria: TC, LNT, PS, SS, VOC, MC, GM, and COM. In his rubric, TC appeared at the top. For Matt, it was important for the students to read the question and provide an answer, following the required features (e.g., writing a multi-paragraph essay). Even though Matt's rubric did not have descriptions for separate grade levels, the questions in his rubric were stated in a clear order with global writing concerns (e.g., Is there a clear introduction, body and

conclusion?) followed by sentence level issues (e.g., Does the student use a variety of sentence types?), and finally editing concerns (e.g., Are there many article and preposition errors?).

Logan

Logan's rubric was holistic in format and contained short descriptions of each level. The style of the rubric was similar to Susan's, but was much less detailed. The descriptors in the rubric began by stating the features related to COM (e.g., demonstrates native-like English ability, demonstrates strong ability in written expression) and moved on to sentence-level writing skills (e.g., a few errors are permissible); however, the descriptors were not consistent across levels. For example, COM was not mentioned at the B+ or C+ levels; thus, the content and the amount of descriptor language varied by level. For the A level, Logan wrote, "Few errors, in spelling, demonstrates a strong ability in written expression, strong vocabulary and English idioms." However, the C+ description is written as follows: "Borderline intelligibility, as in often have to reread and ask what he is trying to say, very basic expression."

What was not included in the teacher rubrics?

During the interview, the teachers were asked to discuss the criteria they wanted to include but did/could not in the teacher rubric; thus, criteria that was part of their writing assessment construct but was not shown on the teacher rubric. Susan and Ben said that they included all intended assessment criteria in their rubrics, but Matt, Eunice and Logan talked about their unstated assessment criteria.

For Matt, the unstated criterion was students' future writing potential. In the interview, Matt commented that teachers should be able to look between the lines and predict how well a student could perform with a little help from the teacher. He thought that teachers should look beyond what is written at the surface level and predict what writing skills students possess, but may not convey in the essay. For example, Matt pointed out a sample essay that had good content, but was written as a list of sentences rather than in paragraph form. He thought that this essay should not be penalized because of its appearance. Matt felt its

future potential was a criterion that shaped his assessment construct, but he decided not to include it in his teacher rubric. When asked why, he stated that such a category was not "objective or quantifiable."

Similar to Matt, Eunice said that a criterion she was aware of but could not include in the teacher rubric was the impact of the student's first language (L1). She noted that, having taught in an EFL context for the past 10 years, she was quite familiar with certain expressions students commonly used, which had been generated from their L1. During the interview, Eunice referred to a sentence in one student's essay: "This gave me a lesson that self-proud make me handsome." Eunice commented that as an English teacher who has strong background knowledge of the student's L1, she easily understood what the student was trying to say. Even though the sentence was clearly ungrammatical, as a teacher, she felt that she should give some credit to the student for trying to transfer his or her L1 expression into English. Along with the L1 impact, Eunice shared that cultural bias was also an area that had an impact on her ratings, but she did not include it in her rubric. She said that topics which she had read multiple times (e.g., students' experience with the college entrance exam) failed to grab her attention when she was doing the rating, which had a negative influence in her final ratings.

For Logan, there was an "X-factor." Logan commented that from time to time, students would surprise him by showing work that was totally unexpected. Whether it was exceptionally good or bad, he commented that there is an "X-factor" not included in his rubric. Logan said that this "X-factor" is something he does not know about ahead of time; therefore, it is impossible to put into a rubric, although he admitted it does have an impact on his ratings.

Discussion

What criteria do teachers use in assessing short EFL descriptive writing?

By analyzing the teacher rubrics, the assessment criteria that teachers used to assess EFL descriptive writing could be detected (Table 4). The assessment criteria that all five teachers had in common were COM, PS, SS, VOC, and GM. Regardless of any individual differences; all teachers included these five

factors as assessment points in evaluating EFL descriptive writing. The criteria that varied were MC, LNT, TC, and SC. Susan did not choose MC as an assessment criterion, and for other teachers, the specific content assessed under the MC category varied. Eunice only looked for punctuation; Logan assessed spelling; Matt checked punctuation and capitalization; and Ben examined all three areas (Table 2). This finding shows that even though teachers assess the same criterion, their focus can be different; thus, meaning embedded within the assessment criterion can be different, depending on the teacher. This not only applies for MC, but the findings were also similar for COM and GM. Comprehension can represent logical flow and fluency to one teacher, while it can signify an essay with good ideas to another. Assessing grammar can be broken down into detailed sections, such as grammar, articles, and conjunctions.

inferred. The criterion that was believed to be most important appeared at the very beginning, or was most frequently mentioned in the teacher rubrics. Susan and Logan’s top criterion was COM, specifically, the strength of ideas and readability. Both teachers stated descriptions concerning COM first in their rubrics and this finding was confirmed through the teacher interviews. When teachers were asked what they considered the most important criterion for this writing task, Logan’s response was “communication ability”; Susan’s response was “ability to construct an idea”; for Eunice, it was “paragraph structure” and “the ability to express specific examples using appropriate vocabulary.” During the interview, Eunice noted that for the purpose of this type of task (short EFL descriptive writing), it was more important to look at the distinguishing features, such as vocabulary, grammar, and sentence structure, rather than the ideas. In Ben’s case, the most important criterion was

Table 4. Teacher rubric analysis findings

		Eunice	Susan	Ben	Matt	Logan
Assessment Criteria	Same	COM, PS SS, VOC, GM	COM, PS SS, VOC, GM	COM, PS SS, VOC, GM	COM, PS SS, VOC, GM	COM, PS SS, VOC, GM
	Different	MC	LNT, SC	MC	MC, LNT, TC	MC
	Unstated	L1, culture			potential	X-factor
Top Criterion		VOC	COM	PS, SS	TC	COM
Rating Style		Analytic	Holistic	Analytic	Analytic	Holistic
Rubric Preference		Detailed	Simple	Detailed	Detailed	Simple
Writing Assessment Focus	Global Level		V		V	V
	Sentence Level	V		V	V	
	Editing Level			V	V	

What information can be inferred through teacher rubrics?

A teacher’s personal assessment construct was visible through teacher rubrics; in addition, the teacher’s top criterion in assessing writing could be

“structure.” Structure was stated at the top of Ben’s rubric, and he gave a broad definition for structure, which included the flow of the essay and the organization of paragraphs. In the interview, Ben said he gave most attention to sentence fragments and paragraph form. In Matt’s case, he felt that all of the assessment criteria, “clarity, flow, grammar, sentence

structure, and vocabulary,” were important, and all features were included in his rubric.

Rubrics developed by the teachers were similar and different in format and content. Susan and Logan produced holistic rubrics that were similar in design, while Matt and Ben both assigned numerical points aligned with the letter grade. Eunice developed an analytic rubric containing detailed descriptions for each criterion. The ways the rubrics were developed may suggest the teachers’ rating styles. Logan and Susan developed holistic rubrics with overall descriptions for each level. Both thought of global concerns relating to COM as the most important assessment criterion in assessing EFL descriptive essays. On the other hand, Eunice, Ben, and Matt valued the importance of sentence-levels issues (SS, PS, VOC, GM). These three teachers designed more detailed rubrics with accurate numeric measures. Thus, teachers who valued global concerns (i.e., ideas) as most important designed more general holistic rubrics, while teachers who focused more on sentence-level concerns developed analytic rubrics with more details. Teacher rubrics show a teacher’s assessment construct and the ensuing implications for a teacher’s assessment style.

Suggestions for using teacher rubrics: Teacher rubrics for classroom assessment

Through an analysis of teacher rubrics, stakeholders can have a better understanding of the meaning assigned to a score. This is especially helpful in classroom contexts, where the teacher usually plays a dual role as teacher and assessor. By examining the rubrics given by a teacher, students and parents will be able to identify what assessment criteria are thought to be important, and what kind of rater the teacher is. Analyzing teacher rubrics is not only limited to the area of writing assessment, but can be applied to other fields. Rubrics are widely used in all subject areas (e.g., language, science, math, social studies) from primary to higher education. Similar applications can be applied to teacher developed rubrics for different fields and grade levels.

According to the teacher rubrics developed in this study, the meaning of an “A” can differ, depending on the teacher. For Susan and Logan, an “A” level essay means an essay with strong ideas and good flow, despite some grammatical and mechanical errors. For

Eunice, Ben, and Matt, an “A” essay will have a variety of sentences with good organization, grammar, and mechanics, but may not have a strong idea. The meaning associated with a grade or score varied among teachers and this was evidenced through teacher rubrics. This difference can be problematic in a classroom context, where a student can receive a different grade for the same performance depending on who did the assessment. If a teacher has low expectations for an ‘A’ level essay, the students in this teacher’s class are likely to have better grades compared to a teacher who has higher expectations for an ‘A’ level essay. To resolve these problems, it is important for teachers who work in the same program or teach similar courses to discuss their expectations of each grade level or task prior to assessing student’ work. Teachers can get together and develop a rubric for the same task. Similar to what was done for this study, each teacher can develop his or her own rubric and later compare and contrast the criteria, scales, and rubric style. Based on the similarities and differences of individual rubrics, a standard rubric that can be used across all teachers can be produced. This method can give an opportunity to discuss and visualize teachers’ assessment constructs and develop common criteria for a grade or score. Instead of presenting a standard rubric for the teachers to use from the beginning, asking teachers to work together and develop a shared rubric seems more effective to ensure quality and coherency in assessment. If a formal rater training is not possible for classroom teachers, a workshop on developing teacher rubrics can give an opportunity to have a better understanding of one’s assessment patterns. Like the language teachers in this study, teachers from other fields can take part in developing, comparing and discussing rubrics to have a clear understanding of their assessment constructs.

Limitations: How well does a teacher rubric represent the assessment construct?

As shown from the findings of this paper, each teacher possessed a unique assessment construct. A teacher’s assessment construct cannot be easily visualized and is difficult to extract; thus, developing and analyzing teacher rubrics may be a method to envision one’s assessment construct. The limitation of this method is that no matter how hard teachers tried, it was difficult to articulate every attribute and rating

criterion in a rubric. Even the teachers in this study stated that there were criteria they used but did not or could not include in the teacher rubric. The most we can achieve from teacher rubrics is to extract the most important and significant factors in assessing the given task. Even in this process, as seen from this study, teachers prioritized and selected criteria that were visible, quantifiable, and widely accepted. However, not all performances can be dissected into measurable components thus; teacher rubrics can only represent part of a teacher's assessment construct, not the whole.

Conclusion

By analyzing teacher rubrics, we can visualize the assessment construct of teachers. The teacher rubric development procedure requires teachers to think of their expectations and characteristics of a good/poor quality performance or product. Teachers must consider the number of levels required to assess the given task and the assessment criteria that can be used to measure student's work. Also, teachers need to carefully pick specific wording that will be used to describe each criterion. Through these steps, the teacher rubric was made which allowed a glimpse of the teacher's assessment construct.

The five teachers in this study shared both common and different assessment criteria in assessing descriptive EFL writing. Teacher rubric analysis also provided implications in identifying the assessment focus and rubric preference of the teachers. The teachers who produced detailed, analytic rubrics focused on sentence-level writing issues. In contrast, teachers who developed holistic rubrics valued global writing concerns, such as idea development and comprehension. The findings of this study show that through teacher rubric analysis, the assessment construct of a teacher can be explicitly visualized, which can better explain the meaning associated with a rating or score and increase validity and reliability in teachers' assessment practices.

References

Alderson, J. C. (1991). Bands and scores. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 71-86). London: Macmillan.

- Arter, J. A., & Chappuis, J. (2006). *Creating and recognizing quality rubrics*. Boston: Pearson.
- Arter, J. A., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press/Sage Publications.
- Barkaoui, K. (2010). Variability in ESL Essay Rating Processes: The Role of the Rating Scale and Teacher Experience. *Language Assessment Quarterly*, 7(1), 54-74.
- Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, 28(1), 51-75.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), pp. 31-51.
- Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal*, 86(1), pp. 67-96.
- DeRemer, M. (1998). Writing assessment: Teachers' elaboration of the rating task. *Assessing Writing*, 5(1), 7-29.
- Erdosy, M. U. (2003). *Exploring variability in judging writing ability in a second language: A study of four experienced teachers of ESL compositions*. TOEFL Research Reports, RR-03-17. Princeton, NJ: Educational Testing Service.
- Gamaroff, R. (2000). Teacher reliability in language assessment: The bug of all bears. *System*, 28, 31-53.
- Jeong, H. (in press). Rubrics in the classroom: Do teachers really follow them? *Language Testing in Asia* 5. doi: 10.1186/s40468-015-0013-5
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26-43.
- Lovorn, M., & Rezaei, A. (2011). Assessing the Assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research & Evaluation*, 16(16), January 13, 2015..
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: what do they really mean to the teachers? *Language Testing*, 19(3), 246-276.
- Lumley, T. (2006). *Assessing second language writing: The teacher's perspective*. Frankfurt and Main: Peter Lang.
- Lumley, T., & McNamara, T. F. (1995). Teacher characteristics and teacher bias: Implications for training. *Language Testing*, 12(1), 54-71.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25), August 1, 2014.

Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment Research & Evaluation*, 8(14), June 14, 2014.

Rezaei, A., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15, 18-39.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment Research and Evaluation*, 9(2), August 12, 2013.

Weigle, S. C. (1998). Using FACETS to model teacher training effects. *Language Testing*, 15, 263-287.

Appendix A: Essay Rating Teacher Instructions

Read the essay prompt and rate the given essays, based on a 6-point-band scale, with *A+* being the highest, and *C* the lowest; *A+*, *A*, *B+*, *B*, *C+*, *C*. Next, write a short description of your rating next to your letter grade. Please **DO NOT** refer to any other sources (e.g., rubrics) while you do this rating.

Essay Task

Time: 50 min

Instructions: Write a multi paragraph essay on one of the given topics. (no minimum or maximum word limit)

Text Type: Descriptive or Narrative

1. Describe a time in your life when you felt extremely proud of yourself. What did you learn from the experience?
2. Describe a disagreement you had with a friend (or family member). How was the disagreement resolved?

Essay #	Letter Grade	Description
1	B+	Ideas are well structured and developed. Good organization. Easy to read, writing is simple, but well organized. Student is able to self-correct. Good, but limited choice of vocabulary. Good sentence structure.
2	C+	It was easy to follow, but used simple sentences. Inconsistent tense use and limited use of vocab, insufficient length.

Appendix B. Rating Justification Code

Criteria	Acronym	Justification Description
Comprehension	COM	flow, readable, understandable, good language ability, clarity, easy to follow, able to tell a story, coherence, good writing, meaning
Paragraph Structure	PS	topic, supporting, concluding sentences, paragraph format, layout, intro, body, conclusion, logic, organization, transitions
Sentence Structure	SS	complex sentence structure, simple sentences, sentence fragments, sentence style, run-on sentences
Vocabulary	VOC	word choice, adjectives
Grammar	GM	tense, article usage, conjunctions, SV agreement, parts of speech, prepositions
Mechanics	MC	spelling, capitalization, punctuation
Length	LNT	multi-paragraph essay, single paragraph, inadequate length
Task Completion	TC	answer the question, respond to the question
Self Correction	SC	correct own errors

Appendix C. Teacher Interview Questions

1. Why did you choose this style or form in developing the rubric?
2. When developing your own rubric, what were the important factors? How are they represented in your rubric?
3. What could be factors that influence you, but are not presented in the rubric? What is an implicit criterion that is not stated in the rubric? What are things you would like to include, but did not?

Acknowledgement:

I would like to thank the five teachers who participated in this study for their help and support.

Citation:

Jeong, Heejeong (2015). What is your teacher rubric? Extracting teachers' assessment constructs. *Practical Assessment, Research & Evaluation*, 20(6). Available online: <http://pareonline.net/getvn.asp?v=20&n=6>

Author:

Heejeong Jeong
College English Education Committee
Hanyang University
222 Wangsimni-ro, Seongdong-gu
Seoul, 133-791, Republic of Korea.

Email: jeongheejeong [at] gmail.com