

## **Reflections on Designing and Using an Instructional Rubric in Support of an Academic Debate Course**

Jamie Lesley  
English for Liberal Arts  
International Christian University

### **Abstract**

This paper reflects on the design and use of an instructional rubric administered in an academic debate course as part of the first-year undergraduate EFL curriculum at a private bilingual university in Tokyo. In the absence of any existing materials, a rubric with descriptors for five criteria was introduced to guide student understanding of performance targets and help students prepare and review their debates in relation to established standards of quality. In this way, the rubric was a planning and rehearsal tool for students, as well as a primer for formative feedback operationalised through self-assessment, peer-reflection, and teacher-fronted input. The paper provides an explanation of those processes by examining the rubric's key features and the model upon which it is based, and reviewing how the rubric was used. It ends with suggestions for improving the design and application of the rubric in future iterations of the same or similar debate courses by using video technology with a revised course and modular structure.

According to Popham (1997), the term *rubric* was historically an ecclesiastical concept appearing in English around 600 years ago for a typographical practice by Christian monks of writing section headings in religious texts in large red letters. In Latin, red is *ruber*, which in turn led to *rubric*. Since then, rubric has also been a referential label for general rules or instructions (Brown, 2012). Today, it is commonly understood as an educational tool (Moskal, 2000) that “articulates the expectations for an assignment by listing the criteria or what counts, and describing levels of quality from excellent to poor” (Reddy & Andrade, 2010, p. 435). Rubrics achieved “mainstream academic popularity” (Leader & Clinton, 2018, p. 86) from the end of the 1970s, although a lack of application in higher education today is noted (Wolf & Stevens, 2007).

This paper begins with a literature review of rubrics to establish features of format, content, type, scope, and purpose. It next outlines reported advantages and disadvantages of rubrics before introducing the current study's teaching context, the author's rubric, and the model it was based on. It then analyses strengths and weaknesses of the rubric's design and application, before ending with suggestions to help improve both aspects.

## Features of Rubrics

### ***Format***

Rubrics are typically construed as measuring instruments against which categories of learner performances, processes or products may be judged along a cline or scale of proficiency for assessment or feedback (Brookhart, 1999). Information is usually presented in a matrix or table with two axes, one for target criteria and the other a ratings scale (Brown, 2017). This allows users to identify incrementally stronger or weaker accomplishments as they move up or down the scale (Green & Hawkey, 2012). It is the *extent* to which criteria are achieved that distinguish rubrics from checklists (Andrade, 2005), which do not differentiate by degree, but determine if criteria are fulfilled (Moskal, 2000).

### ***Type***

There are two basic rubric types. One is *analytic* and parses criteria into categories (in rows or columns) that isolate discrete aspects for independent analysis; the other is *holistic* and describes and evaluates attributes together to produce a single, global rating (Brown, 2012). Analytic rubrics generally collect more information and take longer to use than holistic rubrics, which capture whole measurements as opposed to one for each component part (Mertler, 2001).

### ***Purpose***

Since they are quicker and easier to administer, holistic rubrics are often used for *summative* assessments to gauge what students have learned from a period of instruction (Katz, 2012, 2014). By contrast, analytic rubrics often support *formative* assessment purposes owing to their greater detail (Brown, 2012; Green & Hawkey, 2012; Wolf & Stevens, 2007). Formative assessments also differ from summative ones by seeking active modifications and improvements of learners' future behaviours (Black & William, 1998; Sadler, 1989; Shute, 2008). This is why Andrade (2005) argued that rubrics designed for formative use are more *instructional* by determining current progress in relation to subsequent actions taken.

### ***Scope***

Rubrics can be further categorised according to whether their scope serves *general* skill and knowledge evaluations or whether they are attached to *specific* tasks requiring specific skills and knowledge (Moskal, 2000). General rubrics offer flexibility allowing them to be administered in multiple activities, courses, and programs, which can save time (Tierney & Simon, 2004). However, loss of specificity comes at a potential cost of reduced validity and reliability due to over-generalised performance qualities (Popham, 1997). Such risks must be carefully handled to match rubrics to desired purpose.

### ***Formative Focus***

When formatively applied, instructive rubrics aid learning by generating answers to what Hattie and Timperley (2007) called the three major questions of effective feedback: 1) *Where am I going?* (or What are my goals?) 2) *How am I going?* (or What is my progress?) and 3) *Where to next?* (or What is my next step to close the gap between where I am now and where I want to be?) Rubrics help teachers answer these questions about students, as rubrics can help students answer the same questions for themselves. Consequently, rubrics can promote self-regulated learning since learner autonomy and student-centred instruction are best supported when learners clearly understand what constitutes good performance of

## Reflections on an Instructional Rubric

target criteria. When guided via teachers and peer dialogue towards appreciating the qualities of their own performances, students are then better placed to move closer to desired standards in subsequent attempts (Nicol & Macfarlane-Dick, 2006). Students need to be receptive, of course, for feedback to be acted on (Shute, 2008). Teaching and learning adjustments must also be grounded in evidential findings (Black & William, 1998; Burns, 2014) and a rubric is one source from which such findings can be sought, provided the rubric is of sound construction.

### ***Other Factors***

Research into rubric development reveals a range of design and usage factors. In a consolidation of many studies, Dawson (2017) outlined a framework of 14 elements of rubric format or application: (1) *specificity* (general versus task-specific); (2) *secrecy* (publicly sharing criteria); (3) *exemplars* (to provide benchmarks of quality); (4) *scoring strategy* (to determine how ratings are arrived at); (5) *evaluative criteria* (i.e. the target attributes); (6) *quality levels* (to establish rating range and number); (7) *quality definitions* (to differentiate accomplishment levels); (8) *judgement complexity* (based on required evaluations); (9) *users and uses* (e.g. teachers, students, or institutions plus lesson planning, communicating and comprehending goals, conducting assessment and giving feedback); (10) *creators* (teachers, students, or both combined); (11) *quality process* (to ensure validity and reliability); (12) *accompanying feedback* (to aid score interpretations); (13) *presentation* (i.e., format); and, lastly, (14) *explanation* (for how instructions are conveyed and to what level of detail). Given the breadth of considerations, successful rubric creation is by no means a foregone conclusion and may be determined by any number of influences.

### ***Benefits***

The value of effectively crafted rubrics is well documented. For example, rubrics are reported to help teachers communicate learning targets and expectations with greater clarity (Brown, 2017, 2019) and provide guidance on lesson planning and delivery (Glickman-Bond & Rose, 2006). Rubrics can generate more consistent, transparent, and objective assessments via their measurable standards of quality (Moskal, 2000). Moreover, feedback can become more focused and meaningful because of the framework of reference from which it derives (Andrade, 2005; Mertler, 2001; Tierney & Simon, 2004). Other studies highlight positive effects on motivation (Leader & Clinton, 2018), while opportunities for self-assessment and peer-feedback are also praised (Arter, 2000). Beyond this, rubrics can inform program development and assist minority background learners (Wolf & Stevens, 2007).

### ***Drawbacks***

Potential advantages appear plentiful, yet rubrics are not without caveats. Most importantly, perhaps, evidence of learning and achievement as a *direct* result of rubrics is inconclusive (Reddy & Andrade, 2010). Problems also arise when the criteria are too vague or excessively detailed, thereby hindering clarity (Popham, 1997). Similarly, rubrics may generate confusion if the consistency of attributes or descriptors lapse (Tierney & Simon, 2004). Creating effective rubrics is known to be difficult and time-consuming (Wolf & Stevens, 2007), especially to ensure validity, reliability, objectivity, and practical usage (Andrade, 2005; Reddy & Andrade, 2010). To be administered correctly, raters require training with models of criteria and quality to foster strong intra-rater and inter-rater reliability (Green & Hawkey, 2012; O'Sullivan, 2012). If students are to understand ratings or feedback from rubrics, or if they are to use rubrics to self-assess, they also require training

## Reflections on an Instructional Rubric

(Arter, 2000; Tai, Ajjawi, Boud, Dawson, & Panadero, 2018). Otherwise, rubrics' instructional value is undermined (Reddy & Andrade, 2010).

### ***Adaptations***

Given the challenges of creating good rubrics (Wolf & Stevens, 2007), one option is to locate an existing one and adapt it (Tierney & Simon, 2004), which is what this author did. Brown (1995) suggests five aspects to consider materials from when adopting or adapting them: (1) materials background, (2) fit to curriculum, (3) physical characteristics, (4) logistical characteristics, and (5) teachability. When adapting rubrics, attention must also be duly paid to its criteria, ratings scale, performance descriptions, type, and scope to ensure suitability for the intended context.

This paper addresses a single research question: How can an instructional rubric be used to support an academic debate course in an undergraduate EFL curriculum? All answers, analysis, and interpretations are based primarily on the author's reflections with additional input from students' end-of-course survey data and informal feedback.

In the next section, the paper outlines the author's teaching context and describes the adaptations of a model rubric that were made for the author's academic debate courses. It explains the adapted rubric's defining features before analysing its design and use with suggestions to improve aspects of both.

## **Method**

### ***Participants***

All participants were freshmen undergraduate English learners at a private bilingual university in Tokyo in an intermediate proficiency stream within TOEFL ITP range 450-580 or IELTS bands 4.0-5.5. Academic Debate (AD) was an elective course option taught once a week in 70-minute lessons over 9 or 10 weeks of the 2019 autumn and winter terms. In total, 28 students enrolled in two AD classes taught by the author with a further 41 enrolled in three classes of the equivalent winter term.

### ***Teaching Context***

The AD course has four main goals, to:

- develop critical thinking and argumentation skills;
- build listening and note-taking skills;
- reinforce argumentation skills from the concurrent Academic Reading and Writing (ARW) course; and
- conduct basic research.

The goals are covered across three course modules connected to common term themes (*Culture, Perception, and Communication* and *Issues of Race* in the autumn; and *Ethical Issues* and *Visions of the Future* in the winter). Resolutions related to these themes are researched and debated in small teams. Affirmative teams argue in favour of resolutions and negative teams argue against them. In the first two modules, teachers choose the resolutions, whereas students select them in the third.

### ***Module Flow***

Each module spans a three-lesson cycle: (1) preparation, (2) practice, and (3) performance. In the first lesson, students research the resolution topic by finding, reading and

## Reflections on an Instructional Rubric

analysing articles on both sides of the argument. They then meet their team outside of class to develop arguments and counterarguments. In the second lesson, students work within their teams to rehearse debate delivery under timed conditions that replicate an assessed debate, which follows in the third lesson of the cycle. After the assessed debates, students write short reflection papers reviewing their own performance and summarizing what they learned about the topic. At the start of the next module cycle, they reflect on their last debate before starting to research the next resolution in new teams.

### ***Debate Structure***

Each debate is contested between two teams and lasts 18 minutes. Teams contain two or three members, but due to unforeseen absences, teams may be reduced to one member on the day, heightening the importance of preparation. Teams initially take two alternating turns of 2.5 minutes to present their arguments starting with the affirmative team (see Table 1). After the negative team’s second turn, a 3-minute break exists to plan rebuttals. The speaking order is reversed with the negative team delivering the first rebuttal for another 2.5 minutes. After the affirmative team’s rebuttal, the audience (acting as judges) render a verdict by a show-of-hands to nominate a winner. While watching the debates, judges have copies of the rubric and paper for notetaking. Quiet consultation is permitted in the 3-minute break, but not the final verdict.

Table 1  
*Debate Stages and Timing*

Turn 1	Turn 2	Turn 3	Turn 4		Turn 5	Turn 6	
1st Affirmative	1 <sup>st</sup> Negative	2 <sup>nd</sup> Affirmative	2nd Negative	Break to strategize	Negative Rebuttal	Affirmative Rebuttal	Judges’ Decision
2.5 min.	2.5 min.	2.5 min.	2.5 min.	3 min.	2.5 min.	2.5 min.	

### ***The Model***

No rubric was available prior to the author teaching this course. This left three options: create one from scratch, adopt an existing one, or adapt an existing one. Option three was selected and a rubric was designed based on Shoemaker’s (2002) Debate Grading Rubric. It was chosen for its task-specific format, analytic criteria and ratings scale, and simplicity. Containing one table for each team, its four criteria and key indicators of performance are:

- *Organization and Clarity* - of main arguments and responses
- *Use of Argument* - for reasons given to support or oppose the stated resolution
- *Use of Cross-Examination and Rebuttal* - for identifying weaknesses in the other team’s arguments and defending one’s own arguments from attack
- *Presentation Style* - for tone of voice, clarity of expression, argument precision, keeping the audience’s attention, and persuading them of the team’s case

Ratings are delineated along a 4-point scale from 4 (highest) to 1 (lowest). For each criterion, quality is differentiated via modified descriptions of presence versus absence of desirable and unwanted attributes and their frequency of sustained use. A rating is assigned to each criterion for a combined total of 16 points. A team average is then created by dividing all totals by the number of members.

### ***The Rubric and its Applications***

The Shoemaker model was modified to create a new rubric (see Appendix A). Due to stylistic preference, the ratings scale and criteria were placed on opposite axes to the model with the key attributes across the horizontal axis and the scale of quality along the vertical axis. The 4-point scale was retained but changed to 3-0 rather than 4-1. The criteria expanded from four to five:

- *Organization and Timing* - ideas are organized clearly and with good time management
- *Reasoning* - ideas and support are relevant and logical
- *Evidence* - relevant support, examples and source materials strengthen the reasoning
- *Counterarguments* - opposition team arguments are acknowledged and responded to effectively
- *Style and Delivery* - academic register, voice projection, use of notes, body language, and enthusiasm are convincing to the audience

The rationale for the *Organization and Timing* criterion was to increase focus on time management, which is challenging for students at this level. To place more emphasis on logical, supported arguments and to align the AD course with the logical argumentation and critical thinking skills taught in students' other academic reading and writing classes, separate *Reasoning* and *Evidence* criteria were created. The *Counterargument* criterion replaced *Use of Cross-Examination and Rebuttal* for two reasons. Firstly, cross-examination is not permitted in the debate structure of the AD course. Questions can be posed to the opposition, but answers may only be given in the opponent's turn, and only then if they choose to provide any, i.e., there is no designated period between turns when questions might otherwise be posed. The second motivation for change was to reward effective listening and notetaking. The revised *Counterargument* criterion thus credits reporting opposition team ideas (based on the ability to take notes accurately) and refutation. The *Style and Delivery* criterion continues the model's regard for academic language, but adds importance to voice projection (not just tone), as well as handling of notes, assertive body language, and enthusiasm to help convince the audience.

The revised criteria can accumulate a maximum of 15 points using the 3-point scale. This mirrors the percentage value each debate was worth to students, i.e., 15% of final grades, which simplified calculations. Unlike the model, only the counterargument score was averaged and shared because not all speakers could contribute verbally to counterarguing to the same degree (e.g., by being the first speaker of the team) but all team members shared responsibility for planning refutations and rebuttals between turns and in the 3-minute break. Hence, members were graded individually for their organization and time management, argumentation, and delivery, but together for counterarguments. Boxes for scores were removed from the rubric for simplicity's sake.

The performance qualities are differentiated across the rating scale with deliberate parallelism. The highest rating of 3 points indicates that *all* aspects of a criterion are present, followed by *most* aspects at 2 points, *some* aspects at 1 point, and *few to no* aspects at 0 points. An optimal performance achieving level 3 for all criteria thus means *all* ideas are clearly organized with *no* time management issues; *all* reasoning is relevant and logically supports arguments; *all* sources and examples are relevant and clearly able to strengthen reasoning; *all* counterarguments from the opposition team are acknowledged and responded to effectively; and finally, *all* style features (register, voice, notes, body language, and enthusiasm) are convincing. It was hoped this parallelism would assist students and the author when making judgements.

### ***Procedure***

Students were introduced to the rubric in the first lesson of the first module cycle. Initially, they considered the course goals in relation to what a good debate performance might include. These student-generated criteria centred on argumentation, critical thinking, and public speaking skills. They next listed potential indicators of the attributes they brainstormed for a strong, moderate, and weak performance. This laid the platform for the presentation of the actual rubric, the five criteria, and the key indicators. Students were guided towards identifying how the quality of performance differed along the ratings scale to cement understanding before starting their research on the first resolution.

In the second lesson of each module, i.e. the practice lessons, the rubric facilitated focused rehearsals of debate delivery. Working within their own teams only, students completed two full deliveries with opportunities to review their efforts using the rubric. In the third lesson of each module, when students completed their assessed debates against real opposition, the rubric was used by the audience to judge each debate. When judging, students were encouraged to consider content and delivery in relation to the criteria as they were for writing the 300-400-word reflection papers post-lesson.

At the beginning of the second and third modules, before students began researching the new resolution, they discussed their assessed debate with an opposition team member. This was so that students could reflect on their individual performances and possible ways to improve matters. Discussions centred on a rubric-based question worksheet (see Appendix B). Afterwards, general and group-specific teacher feedback was communicated. Then, after class, debate grades were published online.

### **Results**

Overall, based on the author's experiences and reflections, including a rubric where previously none existed benefited the course for the clarity, equity, and accountability it afforded, as well as its formative focus. Specifically, the rubric helped the author more clearly communicate the main goals, the core skills being practiced and assessed, and the incremental progress students made through the course. It also offered some transparency by identifying what was assessed and how, so that students knew what to prepare and what to demonstrate for ratings of a particular level. Finally, the rubric framed reflective discussions with self-directed learning targets materialising to complement the teacher's feedback.

The rubric was generally (but mostly informally) well-received by the majority of students as a valuable way to review debate performance. It is the author's belief that the rubric was beneficial beyond simple review by providing students with a clear and usable framework of reference when planning, practicing, and reviewing each debate. However, to confirm this belief, more detailed feedback would be required than the single item in the end-of-course survey administered to all students in all classes (not just the author's) who took the course in the 2019 winter term. Regrettably, deeper investigation was not instigated and the equivalent end-of-course survey in the autumn contained no item about the rubric's value. Despite this, of the 76 students who answered the winter survey, 31 respondents strongly agreed and 28 agreed that the rubric was effective for reviewing debates. Only four respondents questioned its effectiveness, although it is unclear if this was due to the instrument's design being problematic or its use.

## Reflections on an Instructional Rubric

The students' data provides little more than basic insight about potential endorsement of the rubric, but indicates a degree of learner support for a rubric's inclusion in future debate courses. What is not known are the details of that endorsement nor the exact way in which the rubric was used in classes outside of the author's. The rubric was available to both instructors who taught AD classes in the autumn and all three instructors who taught them in the winter, but unfortunately no approach was standardised. A more thorough investigation would have ensured standardised understanding and use across all teachers before courses started with a control group of students to more accurately judge the rubric's effects, if any, on learning and learning outcomes.

### Discussion

This paper addresses a single, broad research question: How can an instructional rubric be used to support an academic debate course in an undergraduate EFL curriculum? In short, this reflection asserts that instructional rubrics can provide greater clarity and accountability of goals, more consistent, fairer, and transparent assessments, as well as opportunities for formative feedback. These are beneficial for teachers, students, and institutions. However, aspects of the rubric's design and use can also be improved.

#### *Appraising the Design*

The rubric housed five criteria, which allowed the quality of performance to be analysed for clarity and organization of ideas, time management, strength of argumentation and evidence, as well as stylistic delivery. These criteria were chosen to complement the skills of logical argumentation and critical thinking in the ARW course and to develop them under time-restricted conditions of planned and unplanned output. The criteria could perhaps have been evaluated without a rubric, but clear descriptions with scaled gradations of quality likely made matters less abstract, less obscure, and more understandable than they might otherwise have been.

The criteria and key indicators were selected by the author, and although there was strong agreement between what the teacher and students identified in lesson 1 as necessary for good debate performance, the students could have been more involved in the selection process. Including learner input when deciding rubric criteria may have been motivating for the students and raised the possibility of generating ideas hitherto not considered by the author. That said, the time required to act on students' selections before the second lesson may have been too great to guarantee an appropriate rubric could be created just one week later. It should also be noted that some students were completely unfamiliar with debates and so tasking them with deciding criteria upon which they should be judged may not have been prudent. The option is nonetheless interesting and one to consider for future courses.

The rubric's criteria were not the only ones available. For example, if the debate structure had included a stage when teams cross-examined each other, then it would have been sensible to retain the *Cross-Examination* criterion from Shoemaker's model. In addition, the *Style and Delivery* indicators of *academic register*, *voice projection*, *use of notes*, *body language*, and *enthusiasm to convince the audience* were less than essential. Although considered important by the author and students, these features were not explicitly isolated for controlled and semi-controlled practice in the same way as the other criteria. Hence, an argument could be made to exclude them from assessment so that only attributes that were taught were evaluated to make the assessment itself more valid. In future courses, this is



## Reflections on an Instructional Rubric

something to reconsider, either by removing the *Style and Delivery* features altogether, or retaining them but increasing lesson time to practice them so as to justify their inclusion.

The 4-point ratings scale ranging from a strong 3 points to a weak 0 points was useful for delineating performance, but a wider scale may have helped differentiate performances with more detail. Creating this scale would require careful construction and precision of wording to make the variations clear and practically measurable. A longer scale could be created by categorizing all performance levels by their comparative presence/absence of desirable/undesirable features, as well as their frequency of sustained use. In this way, a more incrementally graded scale of performance quality may be established. The advantage of this would be the scope for more accurate assessments. The disadvantage would be the difficulty in creating such descriptors, as well as an even greater need for something already conspicuous by its absence - exemplars to illustrate the performance levels of the criteria. Such exemplars were unfortunately unavailable when making the rubric. This was a shortcoming, not of the rubric per se, but in how the criteria were first presented and later reviewed, as well as how well the key indicators of performance could be understood.

### *Appraising the Use*

In the first lesson of the first module cycle, although students were guided through the five criteria and how the key indicators determined whether the criteria were present or absent and with varying degrees of quality or control, exemplars from actual debates would have been more effective. Students would likely have been able to appreciate the differences in the descriptors better. Ideally, this would have been done using video. For instance, to help understand the distinction between the 2-point performance of the *Organization and Timing* criterion, where *most ideas are organized*, students could watch a comparison video with a 3-point performance of the same criterion in which *all ideas are clearly organized*. In the future, if students' debates are recorded to create an archive, short extracts of footage might then be used to introduce the criteria and identify standards of quality with strong, moderate, and weak performance samples. An archive would clearly require time, effort, and careful planning, but it would be a valuable learning tool.

Video technology might be beneficial at other stages of the course. In the practice lessons (the second in each module cycle), watching and reviewing video footage post-practice to review or rate the performance using the rubric might have been useful for students. As it was, the rubric was still referenced during these practices, and was perhaps sufficient to judge the relative strengths and weaknesses of content and delivery. However, a recorded copy of the same practice to review during or outside class might have led to stronger performances in the following assessed debate. For this to be effective, students would need to be trained to administer the rubric, rather than merely understanding its component parts. This could be useful, although time limitations might make it impractical. In the assessed debate lessons (i.e. the last in the module cycle), video recordings would also allow the teacher to review or confirm in-class debate evaluations post-facto. Similarly, students could watch their debates at a later date and reflect on them in relation to the rubric and hopefully with greater accuracy. Viewers might also get different impressions watching the videoed debates than in real time, which might produce other interesting points of comparison and discussion.

However, use of video technology to support the rubric brings into question the structure of the three-lesson module cycle. Currently, the review of a module's assessed debate occurs in the first lesson of the next module, i.e., the first assessed debate is reviewed at the beginning of module 2, and so on. If the module cycle were extended from three lessons

## Reflections on an Instructional Rubric

to four and the assessed debates reduced from three to two, then a full lesson could be devoted to reviewing each assessed debate exclusively and exhaustively. This would enable more thorough self-assessment, peer-reflection, and formative feedback follow-up tasks. Although students would complete one fewer assessed debate in total, more time to understand, practice and fine-tune discrete aspects of performance through dedicated practice would be afforded. Presently, this receives only cursory attention before students move on to researching the next module's resolution. In future courses, reviewing and researching could then be addressed in separate lessons so that neither has to share lesson space with or be undermined by the other.

### Conclusion

Overall, the design and use of the instructional rubric in this study offered support by helping the author introduce learning targets and measure progress towards them. It also guided students in practicing their formal debate delivery and reviewing it against standards of quality in formative feedback and reflection tasks. It seems this was appreciated by students, although further investigations are required to understand more about the efficacy of the instrument and its value to learning outcomes when or if supported by video technology. It is hoped that the reflections contained in this paper might also be of benefit to those teaching academic debate in their own EFL/ESL contexts.

Successful rubrics need strong validity, reliability, and practicality of use, which takes time and effort to create. Frequent exposure to the criteria, the descriptors, and the ratings scale in training scenarios are also required to make a rubric a viable instrument of learning and assessment. However, when a rubric is administered effectively, the end results should mean more continuity and consistency of assessment, greater transparency and accountability, and better learning outcomes for students, which, ultimately, is the measurement by which all courses should be judged.

### References

- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30.
- Arter, J. (2000). *Rubrics, scoring guides, and performance criteria: Classroom tools for assessing and improving student learning*. Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.
- Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-144, 146-148.
- Brookhart, S. M. (1999). *The art and science of classroom assessment. The missing part of pedagogy*. Washington, DC: The George Washington University Graduate School of Education and Human Development.
- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston, MA: Heinle, Cengage Learning.
- Brown, J. D. (2012). Introduction to rubric-based assessment. In J. D. Brown (Ed.), *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages* (pp. 1-9). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

## Reflections on an Instructional Rubric

- Brown, J. D. (2017). Developing and using rubrics: Analytic or holistic? *Shiken* 1(2), 20-26.
- Brown, J. D. (2019). Assessment feedback. *The Journal of Asia TEFL*, 16(1), 334-344.
- Burns, A. (2014). Concepts for teaching speaking in the English language classroom. *Language Education and Acquisition Research Network (LEARN) Journal, Special Issue 2014*, 12-22.
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360.
- Glickman-Bond, J., & Rose, K. (2006). *Creating and using rubrics in today's classrooms: A practical guide*. Norwood, MA: Christopher-Gordon Publishers, Inc.
- Green, A. & Hawkey, R. (2012). Marking assessments: Ratings scales and rubrics. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 299-306). Cambridge: Cambridge University Press.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Katz, A. (2012). Linking assessment with instructional aims and learning. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 66-73). Cambridge: Cambridge University Press.
- Katz, A. (2014). Assessment in second language classrooms. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language (4th Edition)* (pp. 320-337). Boston, MA: National Geographic Learning.
- Leader, D. C. & Clinton, M. S. (2018). Student Perceptions of the effectiveness of rubrics. *Journal of Business and Educational Leadership*, 8(1), 86-99.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, 7, 1-8.
- Moskal, B. M. (2000.) Scoring rubrics: What, when and how? *Practical Assessment, Research, and Evaluation*, 7(7), 1-5.
- Nicole, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- O'Sullivan, B. (2012). Assessing speaking. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 234-245). Cambridge: Cambridge University Press.
- Popham, W. J. (1997). What's wrong and what's right with rubrics. *Educational Leadership*, 55(2), 72-75.
- Reddy, Y. M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
- Shoemaker, D. W. (2002). *Debate grading rubric*. California State University. Retrieved from <http://www.csun.edu/~ds56723/phil338/hout338rubric.htm>
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467-481.
- 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(9), 1-7.

## Reflections on an Instructional Rubric

Wolf, K. & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, 7(1), 3-14.

Appendix A

Debate Rubric

Assessed Debate Rubric & Key Indicators of Performance					
Criteria					
Levels	Organization and Timing	Reasoning	Evidence	Counter-Arguments	Style and Delivery
<b>3</b>	All ideas are clearly organized with no time management issues	All reasoning is relevant and logically supports the arguments	All sources and examples are relevant and clearly strengthen the reasoning	All counter-arguments are acknowledged and responded to effectively	All style features are present and delivered convincingly
<b>2</b>	Most ideas are clearly organized and with only minor time management issues	Most reasoning is relevant and logically supports the arguments	Most sources and examples are relevant and strengthen the reasoning	Most counter-arguments are acknowledged and responded to effectively	Most style features are present and delivered convincingly
<b>1</b>	Some ideas, but these may be disorganized and/or timing may be mismanaged	Some reasoning, but it may be irrelevant or illogical and fail to support the arguments	Some sources and examples, but these may fail to strengthen the reasoning	Some acknowledged counter-arguments, but responses may be ineffective	Some style features are present, but these may be unconvincing
<b>0</b>	Few or no clearly organized ideas, timing is largely mismanaged	Little or no relevant or logical reasoning, arguments are largely unsupported	Few or no sources and examples, evidence is largely irrelevant	Few or no acknowledged counter-arguments, responses are largely ineffective	Few or no style features are present, delivery is largely unconvincing

Key Indicators of Performance

- **Organization and Timing:** Ideas are organized clearly and with good time management
- **Reasoning:** Ideas and support are relevant and logical
- **Evidence:** Relevant support, examples, and source materials strengthen the reasoning
- **Counter-Arguments:** Opposition team arguments are acknowledged and responded to effectively
- **Style and Delivery:** Academic register, voice projection, use of notes, body language, and enthusiasm are convincing to the audience

# Reflections on an Instructional Rubric

## Appendix B

### Post-Debate Reflection Worksheet

#### Assessed Debate Reflections

Work with a member of the opposition team from last week's assessed debate. Discuss your performance. What went well? What could be better next time?

##### 1. Organization & Timing:

Ideas are organized clearly and with good time management

- What went well with your preparation before the debate? What could be improved?
- How organized were your team's ideas during the debate? Did they get progressively more persuasive?
- How well did you manage your time during each debate stage (including the break)?

##### 2. Reasoning:

Ideas and support are relevant and logical

- When planning your arguments, did you follow an ARE approach? (Assertion > Reasoning > Evidence)
- How relevant and logical were your reasons? Did you have enough?
- What could have strengthened your reasoning?

##### 3. Evidence:

Relevant support, examples, and source materials strengthen the reasoning

- Where did you source your evidence from?
- How relevant was your evidence? Did you have enough?
- What could have strengthened the evidence?

##### 4. Counter-Arguments:

Opposition team arguments are acknowledged and responded to effectively

- What notes did you take during the debate? What techniques did you use? How effective were they?
- In your rebuttals, did you simply acknowledge the other team's ideas or did you acknowledge and refute?
- How well were your own arguments countered by the other team?

##### 5. Style and Delivery:

Academic register, voice projection, use of notes, body language, and enthusiasm convince the audience

- Was your delivery style appropriately academic in register and language? (vocabulary, sentence formality)
- Did you speak clearly and project your voice to the whole room?
- In what ways did your notes aid your delivery? In what ways did they take away from it?
- Did you make eye contact with all members of the audience?
- Was your posture and body language positive and professional?
- Did you end your turns with an enthusiastic and convincing plea to the judges to side with your team?
- What did you learn from watching other teams' debates?
- How else can you improve your next debate performance?