Grading Groups: Developing a Rubric for Evaluating Discussion

Imogen M. Custance

ディスカッションを評価するためルーブリックの開発

カスタンス　イモジェン M.

Abstract

How learning outcomes are assessed in any course is of key importance. Clear guidelines help both learners and instructors to understand what is expected in a course, and administrators to ensure that standards are being maintained. These assessment procedures must be fair, appropriate to course content, and practical to implement for instructors. This is especially true in situations where multiple instructors teach a course. It is also essential to examine how well current methods of assessment are fulfilling their roles. In this paper, I examine how well the current assessment criteria in the Academic Discourse course evaluate the skills the course aims to develop. I argue that the current system can be improved by incorporating more specific criteria that should promote learning-oriented assessment (Carless, 2007), enhance the development of interactional competence (Galaczi & Taylor, 2018), and facilitate more consistent evaluation between different instructors.

Keywords: interactional competence, learner-oriented assessment

（Received September 20, 2022）

抄　　　録

　学習成果をどのように評価するかは重要なポイントである。明確なコース目標は学習者と講師が何が期待されているかを理解するのに役立ち、管理者は基準が維持されていることを確認することができる。評価手順は、公平であることをはじめ、コースの内容に適切で、講師にとって実践的でなければならない。現在の評価方法がどの程度役割を果たしているのかを検証することも必要である。本稿では、Academic Discourse において、中間・期末の試験を採点するための評価基準が、本コースが育成しようとする能力をどの程度評価しているのかを検討する。私は学習指向の評価や相互作用能力の考えを取り入れることで評価方法を改善することができると論じる。

キーワード：学習指向の評価、相互作用能力

（2022 年 9 月 20 日受理）

Background

The role of assessment in education cannot be underestimated. It shapes and is shaped by the educational environment. "Teaching to the test" and "negative washback" are often blamed for the limited development of language skills in countries like Japan, where the use of pen-and-paper tests continues to restrict the skills that can be evaluated. This is particularly problematic for the development of skills related to spoken interaction which cannot be evaluated in this way. If learners are to develop communicative competence, time must be allocated in the curriculum to achieve this, and appropriate forms of assessment must be used to evaluate this type of development.

Evaluation of learning outcomes is usually determined through a mixture of formative and summative assessment (Brown, 2004). Formative assessment is generally low-stakes and incorporates overall impressions of what learners can understand at different times in a course. It tends to include feedback provided "with an eye toward the future continuation (or formation) of learning" (Brown, 2004, p. 6). Summative assessment, on the other hand, is usually a more significant, formal evaluation of how well learning goals have been achieved. It does not necessarily indicate "the way to future progress" (Brown, 2004, p. 6), especially if it is used at the end of a unit or course. Summative assessments are usually taken more seriously than formative assessment as they are typically, and sometimes required to be, a larger proportion of a final grade. They can therefore have a powerful influence on learners' behaviour in a course, making the design and grading criteria used a key consideration when determining assessment procedures.

Carless (2007) suggests that one way to harness the focus that learners (and instructors) have on more summative assessments is through "strengthening the learning aspects of assessment" (p. 59) to create what he terms learning-oriented assessment (LOA). He suggests that LOA makes assessment more efficient as the two main purposes of assessment, to evaluate achievement and to promote learning, are well-aligned. He lists three principles that can be used to achieve LOA: Designing assessment tasks so they "stimulate sounds learning practices amongst students" (p. 59); involving students in deeply considering the quality of their own and peers' performance; and providing feedback to support continued learning. In particular, he highlights the need for assessment tasks to "promote the kind of learning dispositions required of [learners] and should mirror real-world applications" (p. 59).

It is also essential that instructors do not lose sight of the need for assessment to evaluate achievement. Backwards design (Wiggins & McTighe, 2005) is therefore a useful concept in maintaining the balance between the dual aims of assessment. Backwards design involves first identifying what the aims of a course are, then working backwards to

determine how those aims can be achieved. Careful consideration of how achievement of the aims will be assessed is also necessary. Approaching course design in this way allows for a deeper awareness of how evaluation can be planned not only to assess learning outcomes, but also be oriented to further learning opportunities, i.e., to make it learning-oriented.

The focus on identifying ways in which to effectively evaluate the learning goals of a course is of particular use in achieving other essential features of a test or assessment. According to Brown (2004), the effectiveness of a test is determined by how practical, reliable, valid, and authentic it is, along with the nature of the washback it provides. Practicality requires that a test or assessment can be implemented, while reliability relates to how consistent and/or dependable the scoring is. Validity, which concerns whether the test or assessment actually evaluates the construct or learning goals that it is intended to evaluate, is "arguably the most important principle" (Brown, 2004, p. 22) in determining the effectiveness of assessment. For language learning assessment, authenticity relates to the how "authentic" the type of language used in the assessment is when compared to real-world use. Finally, washback is a how the teaching and learning practices are affected by the nature of the assessments used for evaluation. Both authenticity and validity can be increased if backwards design is used, while the implementation of learning-oriented assessment improves effectiveness in terms of direct impacts on learning after a test through increased understanding of one's weaknesses, and prior to a test as a result of teaching and learning decisions made to help prepare for the test.

Research Questions

As a student population and/or the learning environment change, it is necessary to examine assessment procedures to ensure that they are effective and fair. It is also important to incorporate ideas from recent research that can expand on and increase the efficacy of both learning and assessment. This leads to the following research questions:

1. How effective are the current assessment criteria for evaluating the learning outcomes of Academic Discourse and promoting behaviours that result in skill development?
2. In what ways could the assessment criteria be improved to make them more effective?

Current Course Goals and Assessment Criteria

Academic Discourse is a second year required course for university students. The stated aim of the course is to continue developing speaking skills in academic settings.

The specific learning goals are that by the end of the course, students should be able to actively participate in discussions, take notes during discussions, ask questions to clarify information and improve understanding, ask follow-up questions to further discussions, and adapt their speaking style to the demands of interlocuters and occasion. To achieve these learning goals, students engage in multiple discussions with varied group members on a regular basis. The first part of the course involves discussions based on specific skills related to sharing opinions, making suggestions, and making decisions, with a particular skill covered over a two-week period. The latter part of the course involves synthesis discussions with a single topic, for example "Create a program for improving English education," receiving focus each week. Students also take on a specific discussion role (leader, summarizer, time keeper, or language monitor) each week. Students are expected to learn and use specific key expressions for the different roles and types of discussion as the course progresses.

For the course, 35% of students' grades are determined by a midterm (15%) and final (20%) group discussion test. These both follow the same format, with students completing the assessment in groups of four, or three if necessary. Students decide their roles before the test and are given 10 minutes to prepare before discussing the topic. Tests are eight minutes for groups of four, and six minutes for groups of three. The grading criteria for both assessments are the same, with the midterm taking place relatively earlier in the semester (usually in the sixth class). This is to encourage students to focus on and use the learning opportunities available in the latter half of the course more effectively. In this respect, while it is part of the summative assessment for the course, it is treated in such a way that it orients students towards further learning. The weightings for each are to ensure that students take both tests seriously, with the slightly higher final assessment weighting reflecting the fact that students should be more capable of successfully engaging in discussion by the end of the course.

Student scores are a combination of a group and individual score. All students in the group receive the same group score, which is based on the number of sentences said by the student who says the least number of sentences. This is to encourage all students to participate and to highlight the importance of involving everyone in a discussion. If all students in a group say at least 10 sentences, they receive full points for the group score, which is five. Reactions to other students' utterances are also included within the sentence count. The individual score is worth 10 points and intended to be assigned based on the general descriptors listed in Figure 1 below. It is a largely holistic rubric with instructors free to interpret how important individual elements of level descriptors are when choosing which grade to assign.

Figure 1 *Grading Information for Individual Scores*

0-2 points:
People scoring 2 points or less contribute very little to a conversation. If a member scores in this range, the team score is usually low. A grade of 2 or less may also be used when speakers do not allow others to participate in the conversation.

☐ Minimal participation in conversation
☐ Answers only to direct questions
☐ Very slow response
☐ **Does not try to use conversation phrases**

3-4 points:
People scoring 4 points or less have may have attempted short sentences, but have not actively participated in the conversation. A grade of 4 or less may also be used when speakers do not allow others to participate in the conversation.

☐ Usually uses short turns (lots of one-word answers)
☐ Doesn't pay attention much attention to other speakers
☐ Has long silences with slow responses
☐ Often speaks off-topic
☐ Responds slowly
☐ **Tries only a few conversation phrases (or often repeats same phrase)**

5-6 points:
A score of 6 indicates a basic level of participation, with a small range of skills used, but some success in communication. Typically, a conversation at this level sees every member doing well, but experiences occasional breakdowns or remains limited in depth.

☐ May move off-topic
☐ Limited response or follow-up and short focus on topic
☐ Might focus on some speakers and exclude others
☐ Some long silence and slow responses
☐ Only a few long sentences
☐ **Makes a good attempt to use some conversation phrases**

7-8 points:
A score of 7 or 8 indicates a reasonable variety of strategies, and a good attempt to pursue the topic. Breakdowns may occur, but the speaker makes a good attempt to sustain the conversation with all speakers.

☐ Is an active listener, paying attention to others
☐ Asks follow-up questions
☐ Shows a positive focus on using English
☐ Pays close attention to the other group members and the conversation
☐ **Uses a good variety of conversation phrases**

9-10 points:
A speaker scoring 9 or 10 has made a significant contribution to the group. Although language may not be perfect, the conversation is positive and enjoyable. This usually involves:

☐ Staying focused on the main question or related topics
☐ Including all other speakers and Helping others to speak
☐ Giving extra information and detail
☐ Using "English" thinking and conversation sounds
☐ Responds quickly (e.g. echo key words, repeat question)
☐ Mixing long and short turns
☐ **Uses many conversation phrases for self and encouraging others**

Strengths

The current assessment criteria are relatively easy to use and understand for course instructors. The types of behaviour described are readily identifiable and can be used to separate stronger students from weaker. In addition, using the criteria to grade groups of three or four is not burdensome making it a practical option for assessment. The criteria also emphasise the use of various discussion phrases and that students should support each other in the discussion. This has a positive washback effect with students actively practicing the use of phrases and supportive behaviour during lessons. It is also possible for students to receive full points, even if language is not perfect. This is reflective of real-

world language use where all language users, including native speakers, reformulate and adjust their speech regularly (Rost, 2011). By removing the focus on 'perfect' language use, students are more likely to prioritize participating in the discussion as opposed to overly worrying about the accuracy of their language. For many students, this can be a pivotal change in behaviour which can lead to increased willingness to communicate using English and the seeking out of opportunities to use English outside of the classroom.

It should be stated that the criteria were not originally designed for this course, and were introduced to provide a form of guidance for instructors when evaluating students. To this extent, the criteria are appropriate as they allow for the separation of stronger from weaker students. The focus on participation that they encourage also makes it easier to give every student a score based on their output. However, the extent to which the criteria effectively evaluate the learning outcomes of the course needs improvement. The following section lists the weaknesses of the current assessment criteria when considering how to more effectively encourage the development of interactional competence and making assessment more learning oriented.

Weaknesses

There are clear weaknesses with the group score, particularly in terms of how fair and appropriate it is. The aim of the group score is to motivate all students to contribute to the discussion and to encourage more interactive behaviour. All students receive the same group score to highlight the shared responsibility they have to each other in the discussion. While scoring in this way can emphasise that everyone needs to contribute, in terms of actual assessment, it is uninformative. Students who interact well and try to involve everyone in a discussion can receive a low score if a single group member is uncooperative, shy, suffering from test anxiety, or simply having a bad day. Conversely, a group who barely interact, instead effectively giving speeches one after another, can achieve full points if everyone's speeches are long enough. The group score represents a third of students' overall score for the midterm and final assessments. It is not appropriate that such a large proportion of the evaluation is determined by something that is not a reliable measure of an individual's development and an imprecise measure of interaction.

The way in which the group score is decided is similarly problematic. While easy to understand on paper for both students and instructors, in practice there are many ways in which discrepancies can arise when actually assigning a score. Of fundamental issue is determining what constitutes a "sentence," a concept derived from written output, within a spoken setting. Interactional speaking is, by its nature, messy (Galaczi & Taylor, 2018). If a student starts to reformulate what they were saying before completing an utterance, should the incomplete utterance be counted or not? How do we consistently determine

whether a coordinating conjunction is being used in the middle or at the start of what is being said? How do we deal with interrupted speech? The list goes on. This is not only a problem for instructors who need to maintain consistency in their decisions; fairness also requires consistency across instructors and that students are aware of what is, or is not, considered a sentence while being assessed.

The individual score assessment criteria are challenging too. One issue is that the general descriptions of speakers achieving different scores are confusing and inconsistent. Despite the description aiming to help elucidate which individual score should be assigned, there are frequent mentions of group behaviour and the conversation as a whole. Thus, a comment such as "a conversation at this level···remains limited in depth" is difficult to make use of when determining an individual score. The mixture between a holistic description of achievement and specific behaviours for the scores is not only confusing, but also erratic. There is a resulting lack of clarity as to whether a certain behaviour is central to receiving a particular score or not. Given students receive a single score for individual performance, the impact that the absence of a behaviour might have is large. This issue with scoring can harm the consistency of grading if instructors feel a student has done well enough in the rest of their performance to achieve a higher grade. In this way, it is easy for grading to become amorphous, which makes interpreting and using grades to inform study or teaching challenging, achieving consistency across different sections of a class impractical, and assessment overall less than optimum.

A more specific issue with the individual score is how use of discussion phrases is evaluated. There are two problems connected to this. The first issue relates to an instructor's ability to recognize whether, for example, a student has "[Used] a good variety of discussion phrases." Many of the phrases are short and very functional. Just as it is argued that English language learners find it difficult to acquire accurate use of articles or third person-s because they are not salient (Goldschneider & DeKeyser, 2001) and can largely be ignored without loss of understanding (VanPatten, 2003), so too can a well-used set phrase go unnoticed as other aspects of evaluation are attended to. Keeping track of which students have used which phrases and whether this represents sufficient "variety" is no easy task. The second problem is that there is nothing in the assessment criteria that indicates that there is an evaluation of the appropriacy of the phrases used. While instructors are likely to include this factor when grading a discussion, a student who used many phrases inappropriately might not realise this has impacted their grade. Using set phrases appropriately helps to reduce the cognitive burden on both speakers and listeners in an interaction, so appropriacy should receive more attention in the assessment criteria.

A further, and major, weakness with the current assessment criteria is the absence of anything related to the comprehensibility of the language used. A fundamental benefit of

using discussion in language learning is that it provides ample opportunities for learning through interaction. The Interaction Hypothesis (Long, 1996) states that breakdowns in communication that require negotiation for meaning lead to language acquisition. With no explicit focus on ensuring comprehensibility in the assessment criteria, breakdowns in communication might be ignored rather than addressed, and these important opportunities for learning lost. Japanese is also considered to be a listener-responsible language, increasing the likelihood that a confusing or unclear utterance will not be questioned and negotiation for meaning will not occur. By not assessing comprehensibility, students are less likely to encounter opportunities that require them to adjust their output to suit their interlocutors, a skill that successful communicators need to develop.

Finally, another factor missing from the current criteria is any assessment of how well students are performing their roles within the discussion. The focus on practicing different roles and using role-associated phrases is a strongly emphasised element within the course. Given the focus that performing these roles well receives, how well students fulfil the role requirements within the midterm and final assessments should be included in evaluation.

Potential Improvements

In this section I will first step outside of the current course to consider how the skills developed in the course are used outside the classroom. I will then suggest changes to the assessment criteria that should enhance learning. I conclude the section by arguing that developing an analytic rubric is necessary to introduce standardization and fairness to the assessment procedures for the Academic Discourse course.

For Academic Discourse, the goal of the course is to "continue developing speaking skills in academic settings." Nevertheless, the majority of the students who take the course do not continue in academia beyond university. While they need these skills at university, after graduation more general interactive skills are likely to serve them better. However, the more specific learning goals (see *Current Course Goals and Assessment Criteria* above) are clearly of relevance in a wide range of interactions, especially if "discussions" is substituted for "conversations." As such, the current course goals seem appropriate for the target group of learners.

With the exception of notetaking, the learning goals are all related to skills that underlie interactive competence. Interactive competence is a complex skill that describes the ability of an individual to communicate with others in a variety of settings (Galaczi & Taylor, 2018). This includes both linguistic and pragmatic skills, and acknowledges that interactions are co-constructed by the individuals involved. Considering how the discrete skills described in the learning goals are effectively used in tandem can highlight the

overarching skill that students need to develop. In addition, this approach encourages a focus on the interaction as a whole and how the different subskills each perform different roles within it. This provides an opportunity to develop a deeper understanding of what leads an interaction to be successful or unsuccessful. This knowledge provides a strong foundation for lifelong learning and the continued development of interactional competence beyond the classroom.

Interactional competence is a composite of a wide range of underlying subskills involving both linguistic and behavioural elements. This division can be utilized to help evaluate students' overall success in interactive communication without the kind of overt focus on language subskills that is often found in proficiency ratings descriptors (e.g., Council of Europe, 2020; IELTS, n.d.), and which would not be appropriate given the learning goals of Academic Discourse. In Crosthwaite and Raquel's (2019) criteria for group oral assessment, they separate "ability to interact with others" and "ability to communicate comprehensibly" (p. 60). In this way, it is possible to have a more language-focused criteria which is connected to how well an individual can manipulate language to achieve communicative goals, and a behaviour-focused criteria connected to awareness of and acknowledgement of others in the interaction. These two components of interactional competence do not develop synchronously for many second language learners and can therefore make assessing growth in a single "interaction" criterion more difficult. Orienting students to these two subskills should help them to recognize which element(s) of interaction they struggle the most with and focus their efforts accordingly. It is thus a useful division for orienting learning.

The course is not, however, only focused on interactive skills. It is based on the development of these skills within the context of discussion. Students engage in discussion on a weekly basis and are expected to achieve the aims of the particular discussion each week. They are also expected to develop an understanding of and perform specific discussion roles. This helps learners to develop a different skill to the interactional skills described above. It is also a skill connected to the overall learning goal of developing speaking skills in academic settings and should therefore be included in course evaluation. Including elements of contribution and fulfilling role requirements should also have a positive washback effect on non-assessed discussions and lead to stronger learning overall. Performing one's role well is also a relatively easy skill to acquire which can boost motivation and feelings of accomplishment.

Finally, the course would benefit from introducing an analytic rubric for evaluating performance within the midterm and final group assessments. An analytic rubric provides a more standardized form of assessment with clear descriptions of the different aspects of performance that should be evaluated. The scales used on rubrics must be monotonic in

nature, that is, as the scale increases, it is indicative of a higher level of performance on the criteria being measured. For the grading criteria to be useful for evaluating development between the midterm and final assessments, this necessitates the development of such a scale. Clear descriptions are beneficial for instructors as they should make rating more reliable, and for students as they receive better feedback on their performance (McDonald, 2018). A more defined grading rubric should increase the reliability of scores in the course overall, which should positively impact the fairness of the scores received and help to improve the standards of the course in the long term.

Results

The first research question asked how effective the current assessment criteria for evaluating the learning outcomes of Academic Discourse are, and whether they help to promote behaviours that result in skill development. The critical evaluation above identified many weaknesses that indicate a failing to fulfil either of these roles effectively. The main failings in this regard are that they provide an essentially holistic score which is unlikely to offer students the information they need to focus on overcoming their weak points; criteria that should be evaluated in the course are missing; and the inter- and intra-rater consistency for the current assessment criteria is likely to be low, which is not appropriate for a course taught by multiple instructors where standards should be consistent. Essentially, the inability of the current criteria to evaluate the achievement of the course learning goals in a meaningful or consistent way means they cannot be effective in this context. Although there are some contexts in which the criteria might be appropriate, this is not the case for Academic Discourse.

The second research question asked how the assessment criteria could be improved to make them more effective. There are three main ways in which the criteria could be improved. First, the introduction of an analytic rubric would be beneficial as it should increase the reliability of scoring and opportunities for learning through feedback. Second, framing assessment through the lens of interactional competence is likely to be advantageous as it reflects the skills that should be developed as part of the course. Finally, it is necessary to ensure that the assessment reflects and encourages learners to contribute during the test, but in a way that does not result in negative impacts on others' scores for lack of contribution.

Discussion

In this section, I present a new grading rubric developed for evaluating group discussion in Academic Discourse (Table 1). This is followed by an explanation of each criterion and how it incorporates the potential improvements listed previously.

Table 1 *Midterm and Final Assessment Grading Material for Academic Discourse*

Score	Comprehensibility & Language Use	Interaction	Contribution
5	• Very easy to comprehend. • Any mistakes do not interfere with understanding. • Vocabulary, grammar, and phrases are very appropriate and varied. • Uses language with ease.	• Does not dominate the conversation. • No individual turn is too long; allows others to add comments/confirm understanding. • Contributions are naturally linked to what was said before. • Active listening skills (nodding, eye contact, etc.) used as appropriate. • Excellent ability to respond to/question the contributions of others.	• Made a significant contribution to the discussion. • Fulfilled role requirements very well. • A very positive focus on using English and the topic.
4	• There might be one or two times when it is not easy to comprehend what was said. • Any mistakes do not interfere with understanding. • Vocabulary, grammar, and phrases are appropriate but might not always be varied. • Generally uses language with ease.	• Does not dominate the conversation. • One or two individual turns might be too long, but some effort made to allow others to add comments/confirm understanding. • Contributions are often well-linked to what was said before. • Active listening skills (nodding, eye contact, etc.) used as appropriate. • Some ability to respond to/question the contributions of others.	
3	• There might be many times when it is not easy to comprehend what was said. • Some language mistakes interfere with understanding, e.g., using an incorrect word. • Vocabulary, grammar, and phrases are generally appropriate; they might lack variation. • Struggles with language at times.	• Might dominate the discussion at times with long turns and little opportunity for others to comment/confirm understanding. • Contributions are sometimes well-linked to what was said before. • Active listening skills are usually used as appropriate. • Some attempts to respond to/question the contributions of others, though they might not be successful.	• Some good contributions to the discussion but might not have performed discussion role well. *OR* • Weak/inappropriate contributions but discussion role performed well.
2	• Frequently incomprehensible. • Mistakes are frequent and might interfere with understanding often. • Vocabulary, grammar, and phrases are often inappropriate and/or lack variation. • Often struggles with language.	• Might dominate the discussion frequently. • Contributions might not be well-linked to what was said before. • Active listening skills are only sometimes used as appropriate. • Might not try to respond to/question the contributions of others.	
1	• Few attempts to produce language that others can understand/respond to. • Contributions cause repeated and sustained strain on the listener.	• Only answers direct questions. • Little evidence of listening/paying attention to others.	• Very little or no contribution to the discussion. • Discussion role performed very poorly.

Note. "Contributions cause repeated and sustained strain on the listener" from Crosthwaite & Raquel (2019, p. 60).

Comprehensibility and Language Use

The *Comprehensibility and Language Use* criterion is designed to assess how well students can communicate what they want to convey in a manner appropriate for their interlocuters and the discussion setting. Modern technologies have made finding an unknown word or translating an entire phrase much easier. While use of these technologies can be very beneficial for producing language quickly, it can make comprehending what was said more difficult for interlocuters who are also unlikely to know the word that the speaker looked up. The focus on ease of comprehension in addition to appropriacy of vocabulary and grammar aims to help students become more aware of how their language choices can make it easier or more challenging to communicate successfully. It encourages students to make greater use of their own linguistic resources instead of relying on dictionaries or online translators. Additionally, if students are more focused on understanding each other, they should engage in more negotiation for meaning to ensure that understanding is achieved. This should encourage learners to engage in the type of behaviour that results in language acquisition (Mackey, 1999).

Unlike most assessments of oral production, there are no specific measures relating to complexity or fluency in the rubric, and minimal focus on accuracy. This is a deliberate choice as development in these areas, while hoped for, is not part of the listed learning outcomes for the course. However, there are elements in the evaluation that still encourage a focus in these elements, but only to the extent that they encourage the development of interactional competence. Higher evaluation for use of varied language encourages students to experiment with language, as does penalizing mistakes only if they interfere with understanding. Students are also rewarded for using language with ease. Variation in language use can make interaction more interesting as the lack thereof can make output seem robotic, while overly disfluent production can increase comprehension difficulties. In addition, it is easy to interpret fluency as meaning "faster is better" and complexity as "use more difficult vocabulary and make sentences longer." These interpretations are far from appropriate in many situations outside the classroom, let alone when interlocuters are also second language learners. Aiming to speak faster or use more complex syntax could even be detrimental to the development of students' ability to learn how to adapt their speaking style depending on their interlocuters, one of the learning outcomes of the course. Focusing on comprehensibility and appropriacy of language use is therefore suitable for assessment in this course.

Interaction

The *Interaction* criterion aims to assess the elements of interactional competence that focus on behaviours within an interaction that make it successful. The explicit focus

on allowing others to comment and confirm understanding is a direct response to the commonly observed problem of discussions becoming more a series of speeches than an actual discussion. Shorter turns should also make it easier for students to ask follow-up questions or add their own comments on a similar point. The requirement to link the ideas through the discussion naturally should also help encourage students to become more active listeners and take more notes to help recall and navigate the discussion. Increased awareness of how easy it is for others to take notes can also help learners to recognize that adjusting speed and using intonation to highlight key points benefit comprehension. In this way, the link between how the two elements of interactional competence (linguistic and behavioural) is also highlighted. By actively rewarding higher levels of interaction in the formal assessments, students should focus on this type of behaviour during lessons and develop the skills they need to speak and interact appropriately in a range of situations.

Contribution

The *Contribution* criterion incorporates both the content of a test-taker's input to the discussion and how well they have fulfilled the necessary functions of their assigned discussion role. A successful discussion requires participants to understand the responsibilities that different people have within the discussion, what types of outcomes are expected, and the restrictions placed on the discussion as a result of the time available. How explicitly roles are defined or assigned outside of the classroom will vary, as will the degree to which a discussion participant will take on different roles as the discussion proceeds, but the underlying skills inherent in the Academic Discourse discussion roles are still present. For example, if there is no clear leadership in a discussion, it can easily lose focus and be unproductive; if a decision must be made before the end of a meeting, time must be monitored carefully. While the role of language monitor is more specific to the Academic Discourse course, the idea of checking and finding information within a discussion is clearly similar. The phrases associated with the different roles are also applicable whether a specific role is assigned or not. During lessons, students practice different discussion roles and associated phrases. Through this practice, they learn the phrases necessary for different functions within a discussion and what types of actions are necessary to achieve a successful discussion. Emphasizing and rewarding students who perform their roles well in the formal assessments encourages greater awareness of the process through which a successful discussion is achieved, thereby enhancing opportunities to acquire the associated skills.

Conclusion

In this paper, I examined the assessment criteria used for evaluating group oral discussion tests in Academic Discourse. I argued that the current criteria are unlikely to fulfil the requirements of effective assessment, and should be changed to increase their efficacy. I then developed a new analytic rubric for use in these evaluations using ideas related to learning-oriented assessment and focusing on how assessment can be used to promote development of interactional competence. However, work remains to determine whether the new analytic rubric is effective for the evaluation of group oral discussion tests in this context. It is important to gain both a qualitative understanding of how the rubric works for instructors, and to obtain feedback from students about how they use the information in the rubric to adjust their behaviour in lessons. Further, analysis of scores assigned using the rubric, for example using many-facet Rasch analysis, is necessary to provide validity evidence supporting use of the new scale in Academic Discourse.

References

Brown, H. D. (2004). *Language assessment: Principles and classroom practices.* Longman.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44* (1), 57-66. https://doi.org/10.1080/14703290601081332

Council of Europe (Ed.). (2020). *Common European framework of reference for languages: Learning, teaching, assessment - companion volume.* Council of Europe Publishing.

Crosthwaite, P. R., & Raquel, M. (2019). Validating an L2 academic group oral assessment: Insights from a spoken learner corpus. *Language Assessment Quarterly, 16* (1), 39-63. https://doi.org/10.1080/15434303.2019.1572149

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly, 15* (3), 219-236. https://doi.org/10.1080/15434303.2018.1453816

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51* (1), 1-50. https://doi.org/10.1111/1467-9922.00147

IELTS. (n.d.). *Speaking: Band descriptors (public version).* https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21* (4), 557-587. https://doi.org/10.1017/S0272263199004027

McDonald, K. (2018). Post hoc evaluation of analytic rating scales for improved functioning in the assessment of interactive L2 speaking ability. *Language Testing in Asia*, *8* (1), 19. https://doi.org/10.1186/s40468-018-0074-3

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Taylor & Francis Group.

VanPatten, B. (2003). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Taylor & Francis Group.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Pearson Education.