

Kwansai Gakuin University
Humanities Review
Vol.20, 2015
Nishinomiya, Japan

Learner and Instructor Perspectives of Group Oral Discussion Task Performance

Robert STROUD*

Abstract

In the last few decades, Japanese high schools and universities have been pushed towards creating a more 'communicative' generation of English speaking graduates. The Japanese Government's Ministry of Education, Culture, Sports, Science, and Technology has focused upon creating learners who are more internationally-minded and able to take part in everyday conversations with speakers of English (MEXT, 2002, 2003). Frameworks adopted by educational institutions to assess the oral communicative competence of learners often include well known international language tests, such as TOEIC, TOEFL and IELTS. One common classroom approach to preparing learners to undertake such tests is oral group discussions. Learners are given a set time in class to share their opinions with each other on a given topic and may be asked to agree upon and report a group decision after negotiation in English. However, with a great deal of variety in testing rubrics available, rater viewpoints and learner perspectives of 'good' discussion performance, issues with performance and assessment can arise. An examination of how both learners and instructors (also the raters of tests) consider the importance of a variety of common performance measures in a discussion can help expose any such issues. An experiment was performed involving ninety-four students studying in their first-year of university and seven language teachers/raters of discussion tests in the same university department in Japan. A survey was administered at the end of fifteen-week communication courses which focused on the teaching and testing of oral group discussions. Findings showed some similarities between teacher and student perspectives of the importance of different performance measures, but also some significant differences. Additionally, perspectives amongst the seven teachers were shown to largely vary for some survey items and indicated inconsistencies in rater viewpoints for

evaluating discussion performance. This further demonstrated that the understanding of what it means to display 'good communication' during a discussion is challenging and may not always be clear or consistent amongst teachers/raters of the same tests or for students undertaking such tests. This paper concludes with recommendations for teachers of university communication course for creating a clearer and more consistent learning environment for oral group discussion tests.

I. Communicative competence focuses in Japanese universities

In Japan, the need for graduates who can confidently communicate in English through speech has become a clear focus of the government. A generation of young people who can orally interact with English speakers from Western countries is viewed as a high priority and has been set out as a long-term goal for the Ministry of Education, Culture, Sports, Science, and Technology (MEXT, 2002, 2003). However, there needs to be more consideration of what 'communicatively competent' speakers of English might actually mean for Japanese university students (Iwai, 2009).

What a good 'communicator' of English actually is has been put under question worldwide in the last few decades. In addition to having a general understanding of and ability to use the grammatical elements of English, it has become clear that other competences when using language as a tool to negotiate meaning are essential for learners to be considered capable of doing so. Non-linguistic factors, such as how speakers interact appropriately to the social context at hand, as well as use discourse strategies to negotiate meaning, are examples which have been considered (Canale & Swain, 1980; Hymes, 1972). After revisiting several years of research into what elements can be said to constitute 'discourse competence' for learners during interaction, Celce-Murcia (2007) concluded it to be made up of five important factors. They are (1) *linguistic* (the use of grammar and vocabulary for example), (2) *strategic* (discourse strategies adopted during interactions), (3) *socio-cultural* (communicating appropriately within the social and cultural context), (4) *interactional* (turn-taking and expressing opinions well in a discussion for example) and (5) *formulaic* (the use of prefabricated and useful chunks of everyday language) competences. By carefully considering how well a student can demonstrate these five competences within a discussion with others, a teacher can get closer to determining the communicate competence (or discourse competence for a discussion) of learners, rather than

basing assessment on just linguistic elements such as vocabulary or grammar use for example.

II. Communicative competence testing in Japan

Although the previous section lays out an overview of what factors may demonstrate communicate competence for a learner, effectively and reliably assessing such skills within a classroom or test setting is not so simple. In order for the Japanese government to gain adequate feedback from educational institutes on the progression of student language levels from year to year, standardized tests are required. In order to do this, tests such as the Test of English for International Communication (TOEIC), the Test of English as a Foreign Language (TOEFL) and the International English Language Test System (IELTS) are commonly used. The speaking sections of each of these involve interaction on various topics in an interview-style test with a native English speaker (often the teacher). The grading rubrics are different for each test, but all three of them are focused upon assessing the same fundamental performance measures in the interviews. They are the delivery of the language (fluency of speech, pronunciation, intonation and stress mainly), the actual language used (grammar, phrases and vocabulary) and the development of topics by the speaker (completeness, relevance and development of speaking turns). Therefore, in order to succeed within any of these tests, a student must be able to take part in a discussion fluently, using appropriate grammar and vocabulary, whilst developing their ideas clearly across time. This is no easy challenge and involves a large amount of different skills that need to be demonstrated within a short amount of time (see the previous section again). Teachers must consider how to prepare students well for such tests with oral communication courses which involve the development of all of these skills, whilst being realistic in terms of resources and time available. One such option commonly used to do this is classroom group discussions, which will now be discussed.

III. The classroom oral group discussion approach

One approach widely used within Japanese universities to improve the communicative competence of students is the oral group discussion. Although this approach does not perfectly mirror the interview-style testing used in Japan (see the previous section), there are many reasons why it is so popular amongst

teachers. Firstly, group discussions for classwork (and often also testing) are more time-efficient and cost-effective when compared to pair or individual speaking tasks (Hilsdon, 1995). For example, a teacher is able to watch and give feedback to a larger number of students at the same time when they are working in groups. This may be especially important and perhaps the only practical option for teachers of very large classes. Also, it has been argued that using group discussions for feedback and testing is more consistent, as the teacher does not participate within the discussion, as they would need to do in a two-person interview setting (Ockey, 2001). Thus, by letting students demonstrate their communicative abilities to create discourse together, rather than the teacher playing a key role in the direction and perhaps repairing of breakdown in a discussion, students can be judged more consistently across a class.

Additionally, there are many benefits for the raters of classroom tasks or tests when group discussions are used compared to interviews. It is easier for teachers to grade students when acting as an observer, rather than a participant in a discussion test. In a similar fashion, it would also be easier to train teachers to become raters of discussion tests rather than interviews (Bonk & Ockey, 2003). Training to be a rater of discussions would only involve watching and assessing students, rather than simultaneously being an active participant in an interview. It can also be said that for large classes, teachers/raters will not become as tired when assessing group discussions compared to taking part in many interviews one after the other. By simply watching students to assess them a teacher can maintain their energy levels, as well as alertness, and thus continue to focus better on assessing students when compared to testing the same group of students through a high number of individual interviews (although there is no actual data to prove this available).

A final point which supports the use of group discussions in class to teach and test communicative competence is the reported validity of such a methodology. Ockey (2011) states that many teachers and researchers have reported that group discussion learning and testing is a valid way to teach and assess the oral ability of language learners. Additionally, Ockey et al. (2014) recently found significantly high correlation between group discussion test scores (for three speaker groups) and scores on the TOEFL interview-style speaking test (a very commonly used measure of discussion performance across Japan and in the world). This is not definite proof, but may suggest that students who can improve their scores at group discussion tests may also be able to improve at the common language tests adopted within Japan such as TOEFL. If this is so, then the use of group discussions in university classrooms seems to be a valid and effective way of learning English communication skills for students.

IV. Student issues for group discussion tasks

Despite the reasons given above for group discussions being used as an effective teaching methodology for communicative competence, some issues may lie with such tasks. First of all, the fact that students will be asked to demonstrate their language abilities with a group of classmates, rather than with a single teacher in an interview, will give rise to new considerations. Factors such as the size of the group (Cao & Philp, 2006), fears about making an English mistake in front of several other class members (Williams & Andrade, 2008) and power-struggle issues amongst group members (Fushino, 2010) can alter the behavior of students and even influence what they view as appropriate communication within such a set-up (which may not perhaps match up with what a teacher is looking for).

Also, students who are inexperienced with group discussions (as Japanese university students can sometimes be) may struggle to understand what it is that their teacher expects them to do or say within a discussion. The importance of feedback on performance and progress across time from a teacher is thus a crucial element of a communication course in order for students to direct their efforts in the way which will help them improve as language learners and also score as highly as possible on tests. However, students may not receive clear feedback on what 'good' performance in a discussion consists of or how to improve their performance across time. Teachers may not always have or take the time to provide students with such direct feedback on performance or the details of testing rubrics they will use to assess them.

It may be assumed by some teachers that by allowing students to discuss topics in English that they will automatically improve at doing it through practice of output of their speech (Swain, 1993). This may be true to some degree, but if students practice improving areas of their language performance which do not match up with their teacher's test rubric, then they may become disappointed at the outcome of their efforts in terms of their test scores. This again highlights the importance for a teacher to clarify what communicative competence entails in a discussion (perhaps speaking as much as possible and with good supporting reasons for arguments for example) when compared to other work that the students may undertake such as written essays (focused more on linguistic accuracy and complexity) or speeches (often graded more for accuracy of rehearsed language than a spontaneous discussion).

Considering the issues discussed above, it is understandable that a student can feel confused as to how they should 'perform' in a group oral discussion. They may even try to use more Japanese with other group members (something they

would probably avoid doing in an interview test with an English speaker) or even just remain silent and let other group members speak (also something they cannot really do in a one-to-one interview). In discussion tests where students are not comfortable speaking with their group members, do not understand how to succeed in the test (rather than a more straight forward interview where a teacher can almost lead the discussion) or do not really know what skills they should try and demonstrate to pass the test, issues with performance can occur. Therefore, it is important for us to gather feedback from both students and teachers in such a situation. This was undertaken for the experiment in this paper and will be discussed later on.

V. Teacher issues with group discussion assessment

As mentioned earlier, using discussion tests to evaluate a class of students can be considered a more consistent method compared to an interview-style approach (where the teacher must be an active and perhaps leading participant). Although this may create a degree of consistency in testing for a single teacher, it does not ensure consistency from teacher to teacher in the same educational institute. Different teachers may value certain aspects of discussion skills more than other teachers, due to differences in their own background for example (Winke, Gass, & Myford, 2012). Even with a shared testing rubric, different teachers will almost certainly score students at least slightly differently for the same test performance. Xi (2007) showed that different raters of the TOEFL speaking test can give a variety of scores for the same test performances by students. In a similar fashion, a teacher of one class may score groups differently from how a different teacher may score the same groups. This is understandable as even assessing performance in a group discussion after having unlimited time to analyze recordings of the data still results in controversy. Researchers of task performance for discussions (who have unlimited time to analyze recordings of spoken English) still disagree on the classification of oral accuracy, complexity and fluency in language use during the negotiation of meaning between speakers (Larsen-Freeman, 2009). Therefore, expecting teachers to be able to do it for large numbers of groups on perhaps a regular basis will obviously create difficulty for them, and inconsistency between raters.

It is clear from the above discussion that if we are to reasonably expect students to become better at doing group oral discussions, that we must ensure that the understanding of communicative competence during such discussions matches up for both the students and teachers. A starting point to do this is to

investigate the views of both students and teachers as to what they believe are the most important factors for doing well on discussion tests, as so many variables exist (discussed above) which may create differences in opinions and resultant issues for test scores later on. The experiment in this paper approached this issue and will now be discussed.

VI. Method

1. Research questions

The focus of the study undertaken in this paper is summarized by the following two research questions:

- (i) *How similarly do students and teachers in the same university English department rank the importance of different communicative competence measures for group oral discussion test performance?*
- (ii) *How consistently do the teachers rank the same measures of group oral discussion test performance?*

2. Participants

Ninety-four non-English major first year Japanese university students from six different classes undertaking weekly orally communication classes participated in the study. All of the students took part in English group discussions with classmates each week and were tested on their communicative competence within discussion tests (with the same group members, timing and similar topics) at the end of their fifteen-week courses. Additionally, data from seven native-speaker English teachers, who were teaching these courses in the same department as the students, was included. The teachers all had at least one year of experience of teaching and testing group discussion skills, and were preparing students within the department to undertake group discussion tests at the end of their communication courses. Each teacher was free to assess students as they felt necessary, with their own individual choice of group discussion test rubric.

3. Procedure

The same survey (see the appendix) was administered to the students (in their first language, Japanese) and to the teachers (in English) two weeks before the end of their fifteen-week communication courses. The survey asked the participant to rank different items (from the most important to tenth most important from a choice of thirteen) in terms of how important they were

perceived to be for a student to demonstrate good ‘communicative competence’ within an eight-minute group discussion test (which all of the students were about to undertake at the end of their courses). As all of the survey items could be said to be highly important to demonstrate communicative competence, the participants were asked to rank the items in terms of importance, rather than state how important each item was on a scale for example. That would have potentially resulted in some very unclear data, as the participants may have rated all of the items as highly important and shown no distinguishable differences between scores for each item.

The survey items adopted were a collection of regularly occurring measures used within the tests discussed at the start of this paper (TOEIC, TOEFL and IELTS), such as how fluently a student can deliver speech or how varied their spoken grammar is. In addition to these items, factors related to the participation students (how many words they say or turns they take for example) were added to examine any differences between how the teachers and students might view the importance speaking as much as possible within a discussion (something that students can avoid doing more easily in a group than when they are being interviewed alone). A final item added to the list was how much English (as opposed to a student’s second language, Japanese) a student uses in a discussion. This is also something that discussions need consideration for, as Japanese students may speak to each other in Japanese if they choose too, but might not be able to do so in an interview-style test with a native speaker of English.

VII. Results

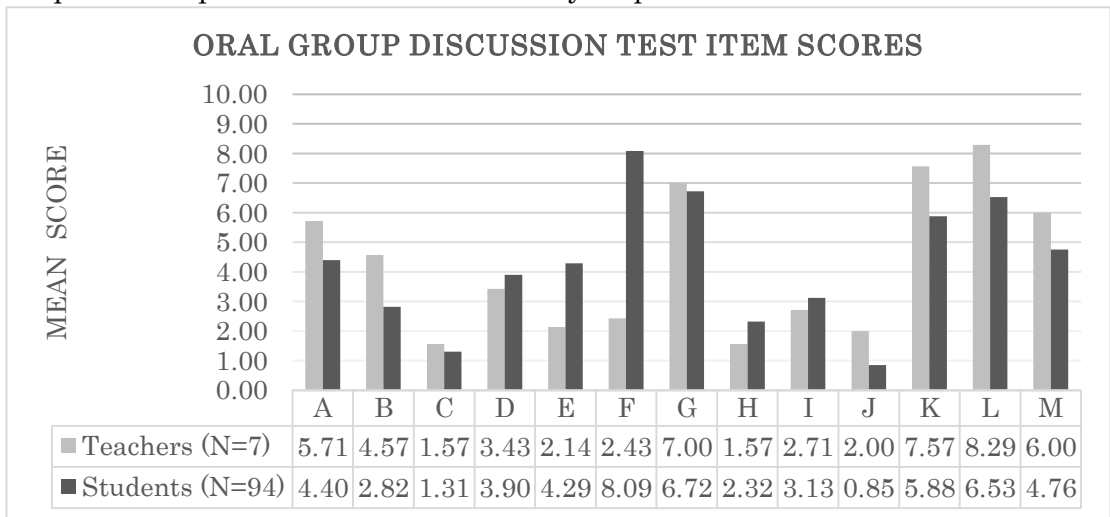
Discussion test items were awarded scores according to the ranking they received from the participants. When an item was ranked as the most important it received ten points. The second most important received nine points, and so on. Hence, the more often an item was ranked highly by participants, the higher the average score it received.

Table 1. Group oral discussion test survey responses.

Oral group discussion test items	MEAN SCORES			
	Teachers (N=7)	Standard Deviation	Students (N=94)	Standard Deviation
A. Saying as many total words as they can in the discussion	5.71	4.07	4.40	3.17
B. Saying as many words as they can in each sentence / speaking turn	4.57	3.69	2.82	2.67
C. Speaking as quickly as they can	1.57	5.13	1.31	1.70
D. Pronouncing words well	3.43	1.13	3.90	2.56
E. Having good intonation	2.14	2.34	4.29	2.63
F. Speaking without pausing much	2.43	2.57	8.09	2.63
G. Using only English	7.00	2.83	6.72	3.42
H. Speaking without repeating / correcting my English	1.57	2.70	2.32	2.84
I. Using accurate vocabulary / grammar	2.71	3.02	3.13	2.93
J. Trying to use complex / varied grammar	2.00	1.63	0.85	1.24
K. Giving as many reasons as they can to support ideas	7.57	1.62	5.88	3.04
L. Asking / answering questions well	8.29	2.21	6.53	2.88
M. Active-listening to others well	6.00	3.00	4.76	2.98

The overall averages of these scores awarded to each of the survey items by the students and teachers can be seen in Table 1 above and a more visual comparison of the student and teacher responses can be seen in Graph 1 below:

Graph 1. Group oral discussion test survey responses



VIII. Discussion and conclusion

Several key points can be made from the data above to answer the first research question in this paper. By looking at Graph 1, is it reasonable to say that, on the whole, the teachers and students surveyed in the experiment had similar views about the ranking of the discussion performance measures. Two of the most highly rated items by both the students and teachers were 'using only English' and 'asking and answering questions well', with the most similarly ranked item by both the students and teachers being 'using only English' (as opposed to using Japanese) during discussion tests. Because of this similar high rating of only using English, it could be said that the understanding between teachers and students about the importance of doing it was not a major issue for the participants in the study. It seems that the view of most participants was that a discussion done only in English and which involves many questions and answers is the best way for students to demonstrate communicative competence.

Although similar ranking of the items occurred between the students and teachers, there were some factors which they appeared to value differently for discussion test performance. Generally, it seems that the students valued 'fluency'-related items a little more than the teacher, whilst the teachers valued items related to 'participation' and 'topic development' a little more than the students did. One example of this is how the students rated 'speaking without pausing much' and 'having good intonation' significantly higher than the teachers. The students scored these measures of fluency of spoken speech at least twice as highly as the teachers did. On the other hand, the teachers scored 'saying as many total words as they can' and 'saying as many words as they can in each sentence/speaking turn' significantly higher than the students, suggesting that they saw speaking up as much as possible in a discussion as more important on the list of factors than the students did. Additionally, the teachers scored 'giving as many reasons as they can', 'asking/answer questions' and 'active listening' all slightly higher than the students did. This may suggest that the teachers rated 'topic development' issues such as these more than the students. However, the difference between the teacher and student scores for this area of performance was not very large and perhaps not hugely significant.

The second research question in this paper addressed how consistently the teachers might rate the same performance measures of discussion test performance. From the data in Table 1, this can be answered by looking at the large standard deviation values for rankings of the items amongst the teachers. For instance, although the seven teachers gave an average scoring of only 1.57 for students 'speaking as quickly as they can' (and hence not valuing it very highly as

a demonstration of communicative competence), the standard deviation of that scoring was as high as 5.13. This suggests some major differences in opinions between the teachers about how important that item is for discussion performance. Standard deviation figures as large as this (as can be seen for several of the items in Table 1) indicate that there is an issue with consistency amongst the teachers in the same department, as to what items will demonstrate communicative competence by a student in a discussion test, which can lead to unfairness in grading of tests from class-to-class within a department. It may be necessary for the teachers to review such data and discuss their grading thoughts and approaches with each other to narrow this gap in opinions and create more consistency for students taking group discussion tests. In a similar fashion, perhaps teacher feedback for students as to how important each of the items in Table 1 would be for performing well on a test would help bring the student and teacher ratings of the items closer, as well as lower the inconsistencies also shown between the student responses in the survey (demonstrated by their high standard deviations for many of the item scores by the students). By doing so, students will have a clearer view of what their teacher expects of them in order to gain higher test scores and can thus focus their efforts on the most appropriate performance factors to do so during their classroom learning.

Exact reasons as to why differences may have occurred between teacher and student rankings of performance items cannot be deduced from the data available in the experiment. However, the fact that differences do exist highlights the potential for misunderstanding between examiner and examinee for group discussion tests in Japanese universities. It is important for teachers of communication courses to ensure that their students understand the value of different performance factors (such as those shown in Table 1 above) when they are preparing for or undertaking a group discussion test. Taking time to ensure students understand what will be scored highly as communicative competence within a discussion can help them focus well on improving significant skills to do well on future tests. Without such feedback and support from a teacher, it is not realistic for students to be expected to improve in the ways that their teachers will be assessing them.

One suggestion to achieve higher levels of clarity for students about group oral discussion performance would be to survey them (in the way the students in this paper were) and have the teacher take time to give feedback on any differences between their own and their students' views of how to do well in a test. This would focus students early on in a course to improve the most relevant areas of language use (perhaps on speaking more and worrying a little less about pausing during speech for the students in this paper's experiment for example)

which would increase their chances of gaining higher group discussion test scores. If teachers within Japanese universities are to use group discussions as an effective way of improving student oral language test scores, then it is important that teachers take the time to collect and discuss data, such as that collected and analyzed in this paper.

References

- Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20(1), 89–110.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to Second language teaching and testing. *Applied Linguistics* 1(1), 1–48.
- Cao, Y., & Philp, J. (2006). Interactional Context and Willingness to Communicate: A Comparison of Behavior in Whole Class, Group and Dyadic Interaction. *System*, 34, 480-493.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In Soler, E. & Jorda, M. (eds.) *Intercultural language use and language learning* (pp. 41-57). Netherlands: Springer.
- Forsyth, D. (2006). *Group Dynamics*. Belmont, USA: Thomson Learning Inc.
- Fushino, K. (2010). Casual Relationships Between Communication Confidence, Beliefs about Group Work, and Willingness to Communicate in Foreign Language Group Work. *TESOL Quarterly*, 44(4), 700-724.
- Hilsdon, J. (1995). The group oral exam: advantages and limitations. In Alderson, J. and North, B., editors, *Language testing in the 1990s: the communicative legacy*. Hertfordshire: Prentice Hall International, 189–97.
- Hymes, D. (1972). On communicative competence. In: Pride, J.B., & Holmes, J. (eds). *Sociolinguistics: Selected Readings*. Penguin, Harmondsworth, pp 269–293.
- Iwai, Y. (2009). Toward Communicative Competence in the Japanese English Teachers, 6(2), 77–99.
- Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 579–589.
- Ministry of Education, Culture, Sports, Science, and Technology. (2002). Developing a strategic plan to cultivate “Japanese with English abilities”: Plan to improve English and Japanese abilities. Retrieved July 1st, 2015, from <http://www.mext.go.jp/english/news/2002/07/020901.htm>.

- Ministry of Education, Culture, Sports, Science, and Technology. (2003). The course of study for foreign languages. Retrieved July 1st, 2015, from <http://www.mext.go.jp/english/shoutou/030301.htm>.
- Ockey, G. (2001). The group oral and the oral interview. Paper presented at the Teachers of English to Speakers of Other Languages (TESOL) Conference, St Louis, Missouri, USA.
- Ockey, G. J. (2011). Assertiveness and self-consciousness as explanatory variables of L2 oral ability: A latent variable approach. *Language Learning*, 61(3), 968–989.
- Ockey, G. J., Koyama, D., Setoguchi, E., & Sun, A. (2014). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing*, 32(1), 39-62.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, 50(1), 158-164.
- Van Moere, A. (2006). Validity evidence in a group oral test. *Language Testing*, 23(4), 411–440.
- Williams, K., & Andrade, M. (2008). Foreign Language Learning Anxiety in Japanese EFL University Classes: Causes, Coping, and Locus of Control. *Electronic Journal of Foreign Language Teaching*, 5(2), 181-191.
- Winke, P., Gass, S., & Myford, C. M. (2012). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30, 231–252.
- Xiaoming Xi. (2007). Evaluating analytic scoring for the TOEFL(R) Academic Speaking Test (TAST) for operational use. *Language Testing*, 24(2), 251–286.

Appendix

Group Oral Discussion Test Survey

Please rank the following in order of how important you think they are to demonstrate good 'communicative competence' in an eight-minute group discussion test. There are thirteen items, but you only need to choose ten.

(1st choice = most important, 10th choice = tenth most important)

- Saying as many total words as they can in the discussion
- Saying as many words as they can in each sentence/speaking turn
- Speaking as quickly as they can

- Pronouncing words well
- Having good intonation
- Speaking without pausing much
- Using only English
- Speaking without repeating/correcting my English
- Using accurate vocabulary/grammar
- Trying to use complex/varied grammar
- Giving as many reasons as they can to support ideas
- Asking/answering questions well
- Active-listening to others well