Reflections in the mirror: The contribution of self and peer assessment in the teaching of speaking skills

Violetta Aeginitou, Eleftheria Nteliou and Nectaria Vlahoyanni

Charokopio State University of Athens

Abstract

Student familiarisation with self and peer assessment techniques (Magin and Helmore 2001, Boud and Holmes 1995), presents many benefits, provided that there is adequate preparation and gradual introduction of these concepts in the syllabus design. Apart from the benefits, students' awareness raising regarding the skills they are taught may prove to be necessary. This paper will present a study, in which undergraduate students of intermediate and advanced levels of proficiency were trained in the self and peer assessment of oral presentation skills. These skills had been taught to them, as part of their ESP course syllabus. The research tools that have been used were assessment checklists and questionnaires designed for the specific students' needs. A quantitative and qualitative analysis of results followed, revealing the relationship between students' involvement in oral presentation tasks, their familiarisation with the assessment of oral skills and the motivating factor behind the afore-mentioned process.

Keywords: peer and self assessment, ESP, feedback, motivation

1. Introduction

Students' practice in peer and self-assessment highlights the vocational role of ESP (Sivan 2000) and teaches students the importance of providing feedback, which is normally left to later stages of their training. Such training also helps tutors move away from a traditional teacher-centered mode of classroom discourse and adopt teaching strategies that "reflect the diversity of current classroom practices" (Fairclough 1992: 14). The purpose of the present paper is to discuss the benefits of peer and self-assessment in ESP and then quantitatively analyse the results of our research on examining the agreement between the tutors' and students' evaluation of presentation assessment criteria and the students' response towards peer assessment. Moreover, a qualitative analysis of the students' comments concerning the self-assessment process follows. Finally, concluding remarks and implications for further research are given.

2. Assessment benefits

2.1 Peer Assessment

In our effort to initiate our students to evaluating their peers' oral presentation skills, we considered the following benefits of peer assessment: First and foremost, it develops such professional skills as responsibility, judgment and autonomy (Boud and Holmes 1995, Hughes and Large 1993, Lejk, Wyvill and Farrow 1999, Magin and Helmore 2001, Stefani 1998). Second, it will enable students to see the learning benefits of such practices since peer assessment ultimately leads to self-criticism that has proven to help students improve their presentation skills (Falchikov 1986, 1995, Magin and Churches 1989, Mockford 1994). As concerns the speaker-listener relationship in public speaking, peer assessment contributes to a better speaker rapport with the audience, since the former's performance will be marked by the latter (Lynch 1988). Last, summative peer assessment increases the objectivity of the results since "… the reliability of the averaged scores will increase as the number of raters increases" (Magin and Helmore 2001: 288).

2.2 Self assessment

It is essential that teachers provide opportunities to their learners to self-reflect and evaluate their performance, because in this way the latter monitor their learning and progress and set goals for the future, therefore encouraging responsibility for learning, promoting critical thinking and helping them construct and reconstruct their knowledge. Self-assessment enhances the gradual development of understanding, because it helps learners to form links between what they already know and new experience. Weak students become more aware of the entire learning process and so the gap between high and low achievers is reduced.

3. Our study

The current study was carried out at Charokopio University and the participants involved were students from the departments of Home Economics and Ecology and Dietetics and Nutrition. Students were divided according to their levels of proficiency in English in two groups: Intermediate and Advanced. The purpose of the current study was to help students to familiarise themselves with the self and peer assessment of their oral presentation skills. For that reason, we carefully organised and implemented a prior training session.

3.1 Methodology

The assessment checklists used by the tutors in our study were based on Rignall and Furneaux's (1997) checklists for assessing oral presentations and contained 18 assessment criteria divided into four categories (content, language, organisation, presentation techniques and visual aids). The assessment criteria on the students' checklists were fewer but they were also divided into the same four categories. That is, the students' assessment checklists were simplified and the criteria were rephrased avoiding any metalanguage so that the students' training in peer assessment and familiarisation with the assessment criteria for oral presentation could be more effective. Some of the assessment criteria which are common in both the tutors' and students' checklists can be seen on Tables 1 and 2 under the 'Variables column'.

The first stage of our study involved a revision of the theoretical principles of giving presentations as well as a discussion on the students' own strengths and weaknesses, according to the presentations they had given in the previous semester. Then the students listened to two senior students' audio-taped presentations and applied the criteria on the assessment checklists we had already distributed and explained to them (Aeginitou, Nteliou and Vlahoyanni 2006), in order to assign marks. Group discussions followed.

The students gave their presentations a week later and they also completed a questionnaire (ibid.). A week after the presentations, students had to attend group tutorials, where they were asked to comment on their presentations, their strengths and weaknesses, in order to trigger their self-assessment skills. Feedback was then provided.

3.2 Analysis of data on peer assessment

Based on the assessment checklists (ibid.), there are two rater categories for statistical analysis, i.e., tutors and students. Our main purpose is to examine the agreement between the tutors' evaluation of the assessment criteria and the students' respective ones. More specifically, in order to assess the inter-rater reliability on the assessment criteria that our checklist contains, we applied the Cohen's Kappa statistics (Cohen 1968), which is most suitable when examining categorical variables.

The statistical analysis revealed different results in the two levels of proficiency. Regarding the Intermediate course, there were seven variables ('Topic support',

'Cohesion', 'Technical vocabulary', 'Clarity of visual aids', 'Speed', 'Loudness' and 'Eye contact') in which K and K-weighted coefficients (applied only for ordinal variables) are different from zero in 5% level of significance[1] (see Table 1). On the other hand, in the Advanced level course, there are just three variables ('Topic support', 'Clarity of visual aids' and 'Speed') in which K and K-weighted coefficients are different from zero in 5% level of significance and, therefore, the null hypothesis of zero Kappa coefficient is rejected (see Table 2). That is, the Intermediate students seem to be consistent in their judgments on many more criteria than the Advanced level students.

Table 1: Checklist variables - Intermediate level

Variables	*Kappa*	*Kappa p-value*	*Kappa-weighted*	*Kappa-weighted p-value*
Topic support	0.483	0.001	0.503	0.014
Cohesion	0.637	0.000	0.637	0.000
Clarity of visual aids	0.543	0.000	0.550	0.017
Speed	0.787	0.000		
Loudness	0.653	0.033		
Eye contact	0.386	0.001	0.479	0.004
Content relevant to topic	-0.031	0.517	-0.031	0.512
Sources	-0.021	0.591	0.105	0.265
Technical vocabulary	0.108	0.182	0.164	0.1921

Table 2: Checklist variables - Advanced level

Variables	*Kappa*	*kappa p-value*	*Kappa-weighted*	*Kappa-weighted p-value*
Topic support	0.380	0.006	0.353	0.049
Clarity of visual aids	0.540	0.000	0.580	0.001
Speed	0.382	0.004		
Sources	0.146	0.131	0.293	0.063
Cohesion	-0.019	0.546	0.031	0.454
Technical vocabulary	0.160	0.092	0.133	0.227
Loudness	0.079	0.408		
Eye contact	0.192	0.072	0.289	0.066

[1] Since p-values <0.050, and the null hypotheses of zero Kappa coefficients are rejected

Regarding the variables 'Content relevant to topic', 'Sources', 'Technical vocabulary' (Intermediate level) and 'Sources', 'Cohesion', 'Technical vocabulary', 'Loudness', and 'Eye contact' (Advanced level) the null hypothesis that Kappa coefficient equals zero cannot be rejected in level of significance 5% (see Tables 1 and 2), which means that agreement is only because of chance. We wanted to further investigate what was wrong with these variables and there was no agreement. The analysis was made with SPSS, where Spearman correlation was computed. Kappa (unweighted) coefficient and McNemar test of symmetry have also been computed with SPSS. Matlab has been used in order to compute Kappa-weighted and K unweighted for the cases where SPSS was not able to compute them.

3.3 Discussion of results

3.3.1 Discussion of assessment checklists results

The analysis of data showed that there are two common problematic variables ('Sources' and 'Technical vocabulary') in both levels. Regarding 'Sources', students of both levels tend to be more lenient than tutors (see Tables 3-8). A characteristic example was the case of the advanced level, in which there was a great asymmetry in the results, since there were ten cases that students valued as 'Yes' and tutors as 'Quite', while the opposite happened only three times (see Table 7). Also, there were five cases that students rated as 'Quite' and professors as 'No', while the opposite never happened (see Table 7). A possible explanation for the undergraduates being so lenient could be that they are not fully aware of the importance of providing adequate bibliographical data in their speech, although this need was emphasised throughout the course as well as by tutors in the training sessions.

Table 3: Sources (Intermediate level): Symmetric measures

		Value	*Asymp. Std. Error*[2]	*Approx T*[3]	*Appr. Sig.*[4]
Ordinal by Ordinal	Kendall's tau-b	.456	.128	3.349	.001
	Spearman Correlation	.488	.137	3.112	.004[3]
Interval by interval	Pearson's R	.436	.127	2.700	.011[3]
N of valid Cases		33			

Table 4: Cross tabulation of tutors' and intermediate students' responses on the 'Sources' criterion

Count

		Sources (tutors)			*Total*
		No	*Quite*	*Yes*	
Sources (students)	No	3	0	0	3
	Quite	13	2	1	16
	Yes	5	7	2	14
Total		21	9	3	33

Table 5: Sources: Chi-Square tests

	Value	*df*	*Asymp. Sig. (2-sided)*
McNemar-Bowker Test	22.500	3	.000
N of valid Cases	33		

Table 6: Sources (Advanced level): Symmetric measures

		Value	*Asymp. Std. Error*[1]	*Approx. T*[2]	*Approx. Sig.*
Ordinal by Ordinal	Kendall's tau-b	.475	.127	3.534	.000
Interval by interval	Spearman Correlation	.530	.137	3.477	.002[5]
	Pearson's R	.534	.116	3.520	.001
N of valid Cases		33			

[2] Not assuming the null hypothesis
[3] Using the asymptotic standard error assuming the null hypothesis
[4] Based on normal approximation
[5] Using the asymptotic standard error assuming the null hypothesis.

Table 7: Cross tabulation of tutors' and advanced students' responses on the 'Sources' criterion

Frequency distribution

		Sources (tutors)			*Total*
		No	*Quite*	*Yes*	
Sources (students)	No	5	0	0	5
	Quite	5	5	3	13
	Yes	1	10	4	15
Total		11	15	7	33

Table 8: Sources: Chi-Square tests

			Asymp.Sig. (2-sided)
	Value	*df*	
McNemar-Bowker Test	9.769	3	.021
N of valid Cases	33		

Concerning the 'Technical vocabulary', the intermediate level students still tend to be more lenient than their tutors (see Tables 9 and 10), whereas advanced level students do not seem to have a clear idea of how to assess their peers (see Table 11). A characteristic example is the case in which tutors chose 'No' for 14 out of the 33 students, while all students chose 'Yes' (see Table 10). Higher marks on the part of the students could possibly be explained by the fact that students feel that the distribution of a technical vocabulary list at the beginning of their presentation is sufficient. Tutors, however, consider further explanation of these terms necessary, from an academic point of view. Moreover some students have difficulty in drawing the line between general English vocabulary and sub-technical or even technical vocabulary.

Table 9: Technical vocabulary (Intermediate level): Symmetric measures

		Value	*Asymp. Std. Error*	*Approx. T*	*Appr. Sig.*
Ordinal by Ordinal	Kendall's tau-b	.326	.152	2.135	.033
	Spearman Correlation	.345	.161	2.045	.049
Interval by interval	Pearson's R	.347	.161	2.062	.048
N of valid Cases		33			

Table 10: Cross tabulation of tutors' and Intermediate students' responses on the 'Technical vocabulary' criterion

Frequency distribution

		Technical Vocabulary (tutors)			*Total*
		No	*Quite*	*Yes*	
Technical Vocabulary (students)	Quite	10	5	3	18
	Yes	4	4	7	15
Total		14	9	10	33

Table 11: Technical vocabulary (Advanced level): Symmetric measures

		Value	*Asymp. Std. Error*	*Appr. T*	*Approx. Sig.*
Ordinal by Ordinal	Kendall's tau-b	.145	.167	.863	.388
	Spearman Correlation	.155	.179	.873	.389
Interval by interval	Pearson's R	.152	.186	.859	.397
N of valid Cases		33			

Another problematic variable that was further examined was 'Content relevant to title'. The Intermediate level students appeared totally inconsistent when assessing their peers on that point (see Table 12), in contrast to the advanced level students, who showed total agreement on that variable (see Table 13). This impressive difference among the students of the two levels could be explained by the fact that the Intermediate level students have not been exposed to subject specific texts as much as advanced level students have and for that reason they may not feel very confident when they judge a presentation on its content.

Table 12: Content relevant to title (Intermediate level): Symmetric measures

		Value	*Asymp. Std. Error*	*Approx. T*	*Approx. Sig.*
Ordinal by	Kendall's tau-b	-.031	.022	-.730	.466
Ordinal	Spearman Correlation	-.031	.022	-.174	.863
Interval by interval	Pearson's R	-.031	.022	-.174	.863
Measure of Agreement	Kappa	-.031	.022	-.180	.858
N of valid Cases		33			

Table 13: Cross tabulation of tutors' and Advanced students' responses on the 'Content relevant to title' criterion

Frequency distribution

		Content relevant to title (tutors)	*Total*
		Yes	
Content relevant to topic (students)	Yes	33	33
Total		33	33

Regarding the Advanced level students, there are some more problematic variables that have been further investigated, such as the use of cohesion and the presenters' loudness of voice and eye contact.

As far as cohesion is concerned, the null hypothesis of zero Spearman correlation could not be rejected (see Table 14), which means that the students appeared totally inconsistent when judging their peers on these variables. It was quite surprising that students were not consistent on judging that point, since they were extensively taught cohesive devices and linking words throughout the course. However, they seem to feel quite satisfied even when they listen to some common cohesive devices at the beginning of a given presentation, without paying attention to that variable till the presentation ends. Therefore, by having this attitude, they are not able to identify any weaknesses regarding the use of cohesive devices in the presentations as a whole.

Table 14: Cohesion (Advanced level): Symmetric measures

		Value	*Asymp. Std. Error*	*Approx T*	*Approx Sig.*
Ordinal by Ordinal	Kendall's tau-b	.085	.167	.501	.617
	Spearman Correlation	.089	.176	.496	.623
Interval by interval	Pearson's R	.119	.161	.670	.508
N of valid Cases		33			

Regarding loudness, a reasonable explanation for the inconsistency on that variable (see Table 15) cannot be found. What remains to be done is to investigate this point anew in future research to find out whether the same results will be repeated.

Table 15: Loudness (Advanced level): Symmetric measures

		Value	*Asymp. Std. Error*	*Appr. T*	*Approx. Sig.*
Ordinal by Ordinal	Kendall's tau-b	.194	.234	.761	.447
	Spearman Correlation	.196	.236	1.110	.275
Interval by interval	Pearson's R	.117	.193	.656	.517
N of valid Cases		33			

Finally, when judging eye contact, advanced level students tended to be much more lenient than their tutors (see Tables 16-18). Although the importance of maintaining eye contact with the audience has been stressed a lot during the teaching of oral presentation skills, students often state that it is very difficult for them to keep eye contact, because this entails memorising their speech. This may be due to inadequate rehearsal of their speech at home and the presentation of too much information on the transparencies. Reading aloud is therefore unavoidable, a fact which is unacceptable only by tutors.

Table 16: Eye contact (Advanced level): Symmetric measures

		Value	*Asymp. Std. Error*	*Approx. T*	*Approx. Sig.*
Ordinal by Ordinal	Kendall's tau-b	.458	.124	3.477	.001
	Spearman Correlation	.496	.134	3.184	.003
Interval by interval	Pearson's R	.496	.123	3.176	.003
N of valid Cases		33			

Table 17: Cross tabulation of tutors' and advanced students' responses on the 'Eye contact' criterion

Frequency distribution

		Eye contact (tutors)			*Total*
		No	*Quite*	*Yes*	
Eye contact (students)	No	4	0	0	4
	Quite	7	7	2	16
	Yes	2	7	4	13
Total		13	14	6	33

Table 18: Eye Contact: Chi-Square Tests

			Asymp. Sig.(2-sided)
	Value	*df*	
McNemar-Bowker Test	11.778	3	.008
N of valid Cases	33		

3.3.2 Discussion of questionnaire results

Regarding the students' response towards peer assessment, the analysis of the questionnaire results revealed a positive attitude, which is clearly shown in the pie charts (see Appendix). In short, the prevailing values in the charts are 'quite' and 'a lot'. It is remarkable that in Question 2 almost half of the students agreed on the importance of organising a presentation. In Question 3, however, they seem indecisive, concerning the notion of subject-specific vocabulary, a fact which can be clearly shown by their choices of 'little' and 'a lot'.

3.4 The self assessment process

In order to investigate the connection between peer and self-assessment practices and raise our students' awareness of the latter, we organised a number of tutorials a week after the presentations and before giving feedback to each student. Dickinson (1987) and Oscarson (1997) have further suggested that participating in self-assessment procedures can help students become skilled judges of their performance and set realistic goals for themselves as far as their future presentations are concerned. Researchers such as Carr (2002) recognise the importance of tutorials as a self-evaluation tool. The tutorials took the form of 'guided interviews', which can be used as an assessment tool in qualitative studies. Some of the questions we asked students were the following:

1. Were you satisfied with your presentation?
2. In which aspects of your presentation you feel you need more practice? Why?
3. In which aspects of your presentation you feel you performed well and thus you wouldn't change?
4. Is it easy for you to assess yourself?

Some students seemed to be quite aware of their performance levels, whereas others did not manage to make any substantial comments on their performance,

probably because of their absence from the training sessions or because they have not been familiarised with reflecting on their performance.

According to relevant research (Bachman and Palmer 1989, Ready-Morfitt 1991 cited in Carter and Nunan 2001), learners find it easier to identify their weaknesses rather than their strengths, shown by their more elaborate answers in question 2 rather than question 3.

4. Concluding remarks

As concerns peer assessment practices, the limited number of subjects prevents us from generalising results. It would be interesting if the specific study could be realised with a larger number of participants. Another limitation is that the absence of some students from the training course might have affected the way they assessed their peers. Regarding the design of the self-assessment practice, a more structured tool in the form of a questionnaire may yield more conclusive results in the future. In fact, future research could also involve factors such as learning styles and personality traits, in order to examine their role in the effectiveness of self-assessment.

All in all, prior training and the students' active involvement in the marking process indicated an improved performance in oral presentation skills and modified the results of our pilot study (Aeginitou *et al.* 2006), despite some discrepancies in the findings. The results have also shown that the use of technical vocabulary and the reference to bibliographical sources are two major problematic areas for students of both levels of language proficiency. In addition, the high motivation rates shown by the questionnaire results as well as the contradictory but encouraging statements expressed during the tutorials, should be seriously taken into consideration when designing tasks for peer and self-assessment of presentation skills. Finally, the findings analysed above suggest that the preparation and active involvement of students in self-reflective practices from the very early stages of learning enhance oral presentation skills.

References

Aeginitou V., E. Nteliou and N. Vlahoyanni (2006). Familiarising ESP undergraduate students with public speech principles. *Proceedings of the 5th International Conference of ESP/EAP*. Thessaloniki: University of Macedonia, Greece.

Bachman L.F. and A.S. Palmer (1989). The construct validation of self-ratings of communicative language ability. *Language Testing* 6/2: 14-25.

Boud D. and H. Holmes (1995). Peer and self marking in a large technical subject. In D. Boud (ed.), *Enhancing learning through self assessment*. London: Kogan, 63-78.

Carr S.C. (2002). Self-evaluation: Involving students in their own learning. *Reading and Writing Quarterly* 18: 195-199.

Carter R. and D. Nunan (2001). *Teaching English to speakers of other languages*. Cambridge: CUP.

Cohen J.A. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70: 213-220.

Dickinson L. (1987). *Self-instruction in language learning.* Cambridge: Cambridge University Press.

Fairclough N. (1992). *Discourse and social change*. Cambridge: Polity Press.

Falchikov N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education* 11: 146-166.

Falchikov N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International* 32: 175-187.

Hughes I.E. and B.J. Large (1993). Staff and peer-group assessment of oral communication skills. In *Studies in Higher Education* 18/3: 379-385.

Lejk M., M. Wyvill and S. Farrow (1999). Group assessment in systems analysis and design: A comparison of the performance of streamed and mixed-ability groups. *Assessment and Evaluation in Higher Education* 24: 5-14.

Lynch T. (1988). Peer evaluation in practice. In A. Brookes and P. Grundy (eds), *Individualisation and autonomy in language learning*. In *ELT Documents* 131: 23-27.

Magin D. and A. Churches (1989). Using self and peer assessment in teaching design. *Proceedings, world conference on engineering education for advancing technology*. Australia: Institution of Engineers 89/1: 640-644.

Magin D. and P. Helmore (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education* 26/3: 287-298.

Mockford C. (1994). The use of peer group review in the assessment of project work in higher education. *Mentoring and Tutoring* 2: 45-52.

Oscarson M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham and D. Corson (eds), *Encyclopedia of language and education, Vol. 7: Language testing and assessment*. Dordrecht: Kluwer Academic, 175-187.

Rignall M. and C. Furneaux (1997). *Speaking.* London: Prentice Hall.

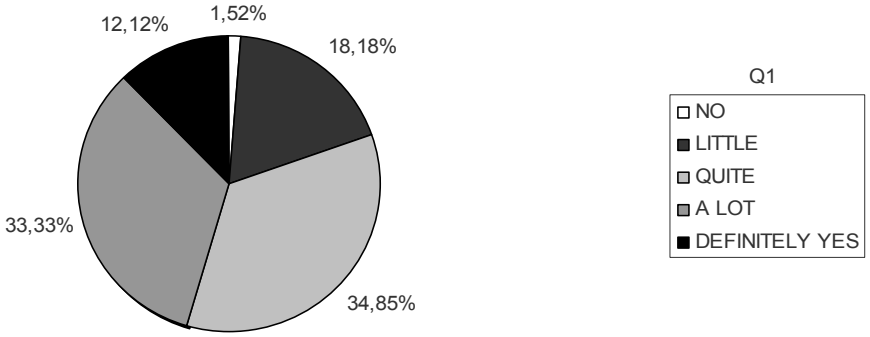
Sivan A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education* 7(2): 193-213.

Stefani L. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education* 23: 339-350.

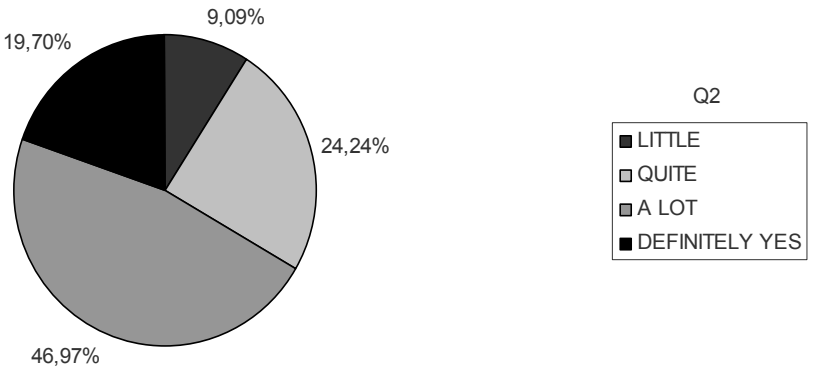
Appendix

Questionnaire results on the students' responses towards peer assessment

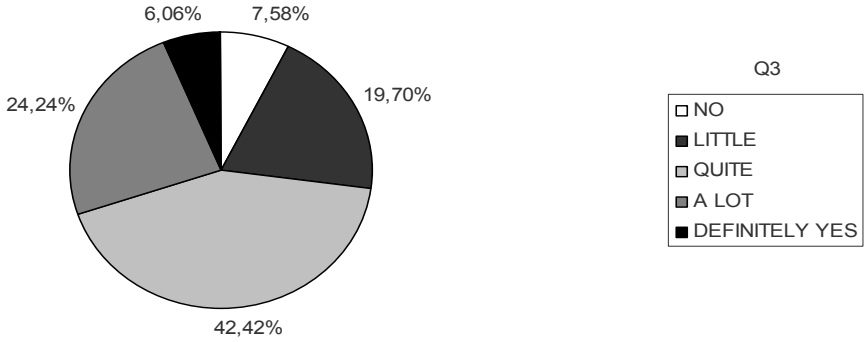
1. While listening to the presentations of your classmates, have you learned anything new on the topics under discussion?



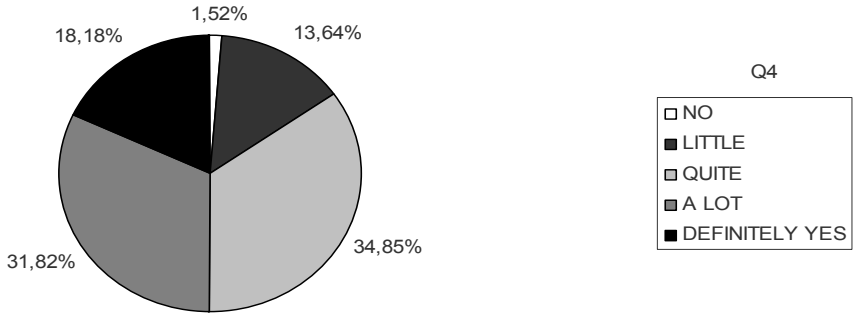
2. Has the organisation of the presentations helped you in the way you will organise your future presentations?



3. Have you learned useful words / expressions in your subject area?



4. Were the visual aids helpful in your future selection of relevant graphics?



5. Did you find evaluating your classmates interesting?