

## Peer Assessment of Oral Presentations: What These Two Cases Tell Us

---

**Anil Pathak and Chan Soon Keng**

School of Computer Engineering

Nanyang Technological University, Singapore

### **Abstract**

This paper explores the use of peer assessments of oral presentations in teaching and evaluation. The experiment was conducted on undergraduate and graduate university students of communication courses. While the undergraduates were asked to evaluate their peers' oral presentations using a rather structured assessment sheet, the graduate students used a relatively open feedback sheet. A comparison was made between the peer ratings and teacher ratings. It was found that the graduate students differed significantly from the teachers' ratings and other peer ratings. The undergraduates, however, displayed a tendency towards uniformity and towards a broad agreement with the teacher ratings.

The paper examines the value of using assessment sheets, and comments on the use of peer assessment for teaching as well as evaluation purposes. It is concluded that the more structured assessment sheets provide more reliable assessment measures, while the less structured sheets provide a better scope for critical thinking, feedback, and instruction.

Peer Assessment is a well-researched area. On one hand, researchers are concerned with achieving greater learner control through this instrument (Nunan, 1994), while they struggle with the concepts of Test Reliability and Test Validity (Carlson &

Smith-Howell, 1995). It is now generally agreed that the system of peer evaluation allows for multiplicity, plurality and independent choice. Some educators have also considered its use in the development of critical thinking skills. Macpherson (1999)

has recently reported on the efficacy of a module designed to develop students' critical thinking skills in the context of peer evaluation in a second-year supervisory management course unit. The results of her experiment indicate that tuition in critical thinking strategies and skills may influence peer/tutor-grading agreement. Researchers have also concentrated on the area of student perceptions of peer assessment exercise. Cheng and Warren (1997) made an attempt to gauge the students' attitudes towards peer assessment. They used a two-fold process: the students' attitudes were canvassed both prior to the peer assessment and at the end of it. Their research focuses on those students who had second thoughts about peer assessment and the reasons given for these shifts in attitude. Roskams (1999) has recently explored the issue of a popular teacher belief that active participation and accurate, appropriate and meaningful feedback in Asian cultures is constrained by fear of mistakes, politeness norms, and the belief that peer assessment lacks credibility. His research on 217 Chinese students suggests that peer feedback is generally perceived as useful, and occurred often. Although the overall student response was positive, students were unsure about its fairness, and felt less conformable about it as an assessment exercise than as a learning exercise.

In this paper our focus is on peer assessment of oral presentations. Even this is a well-trodden area. Weaver and Cotrell (1990) discuss how and how well peer speech evaluation worked in a hybrid

communication course. Carlson and Smith-Howell (1995) have discussed how reliable speech evaluation forms are when filled out by reasonably trained people (students). In the context of oral presentation the system of peer assessment is particularly useful to keep the student audience engaged in a learning experience. Secondly, peer feedback is usually given in terms that the presenter is in a position to appreciate and more closely parallels the student's own level of development. The multiplicity and plurality are relevant here since the students gain a greater sense of audience. More fundamentally, peer evaluation adds variety to the instructional process. The student presenters do not then think that the presentation is meant only for the teacher-evaluator and this affects the way they develop their presentation style.

### ***Premises***

In this paper we describe two related experiments carried out on students of Computer Engineering in Singapore. The first experiment was carried out with an undergraduate class. The class was studying an elective course *Developing Professional Image*, and oral presentations were an important part of the course. The second experiment was performed with graduate students undertaking a specific-purpose course in Research Communication.

Prior to the experiment both groups of students received preparatory instructions on how to prepare for and conduct oral presentations effectively. During these sessions the structure of a presentation and

elements that contribute towards the quality were discussed. Students were advised to prepare an outline to guide them and the audience in the course of the presentation. They were also instructed on the importance of good organisation of ideas, selection of topic and content as well as audience awareness when preparing for a presentation. At the same time elements like effective use of non-verbal behaviours, appropriate and effective language use including accurate pronunciation, and the judicious use of visual aids were dealt with. In addition to instruction, the students watched videotapes which showed them how to make good speeches and samples of good public speaking.

Assessment formed an integral part of both the courses and carried considerable weight in the overall assessment scheme. Students were encouraged to use the presentation software and marks were specifically allocated for the use of this media. Students were given some liberty to choose a topic although they were required to make a presentation related to either the corporate world (undergraduates) or academic communication (graduates). Topics getting closer to the arena of popular public speaking were disallowed. The presentation time given ranged from 15 to 19 minutes and the presentations were made in a small theatre equipped with a video projector.

### ***Preparation***

We assumed that careful preparation (or the lack of it) would significantly affect the reliability, validity and backwash effect of

peer assessment. In both experiments, the following two factors were primarily explored in the preparation phase.

**1. Assessment forms:** We used two different assessment forms in the two experiments. The intention was to study intra-rater consistency as well as the quality of student assessors' feedback. We, therefore, made crucial changes in the two forms related to these two variables.

Both experiments used assessment forms that listed the factors leading to a successful presentation. These factors were derived from the course content and had been explored in detail in lectures and tutorials. Although this was a common factor in both experiments, there was some difference in the way these checklists were structured. The undergraduate experiment used a rather narrow 40-point scale to mark the overall presentation. Also, a content analysis of the presentation task was done and a list of impact factors was prepared. This was later used to prepare the assessment checklist. (See Appendix 1.) Equally importantly, the impact factors were further broken down into sub-factors. There was relatively little space for the assessors to write their comments. This checklist was prepared to address the fear regarding lack of agreement between teacher ratings and peer ratings. It gave the student assessors an elaborate idea of the marking scheme. Since a narrow 40-point scale was used, unnecessary deviations could be eliminated. (Note that 'good' raters of oral presentations usually do not use the full scale of 100

points.) Since the impact factors were broken into sub-factors, the list supposedly gave the student assessors a more exact idea of what they were looking for. Also, this extensive list helped to ensure test validity in the sense of a valid test coverage. The undergraduate checklist, thus, was largely geared towards ensuring reliability and validity of the test instruments. It did not have much emphasis on the aspects of feedback and critical thinking, which are equally crucial in the context of peer assessment.

Our experiment with the graduate students, on the other hand, encouraged participants to give their comments freely by pointing out to their peers both strengths and weaknesses of their performance. At the end of each presentation the evaluators were given about three minutes to write comments on their forms. They also awarded marks to each aspect of the presentation. The assessment forms used in the graduate experiment were less structured (See Appendix 2), and thus gave a better scope for personal expression. At the end of the experiment, the investigator collected the assessment forms and collated the feedback comments (in verbatim) and marks for each of the participants. These comments were collected from each peer assessment form with the instructor's assessment heading the list and typed out into sheets, one for each speaker. These feedback sheets were returned to the speakers to serve as their reference and guide to future oral presentation assignments.

**2. Briefing of assessors:** An important part of the assessment was appointment and briefing of student assessors. They were asked to form panels of four or five assessors and were specifically briefed on the following four aspects:

**What to look for in the presentations:**

They were advised to avoid an overt emphasis on the correctness of language. At the same time they were advised to specifically look for the relation between the meaning intended and the effect produced.

**Use of meta-language and terminology:**

The terms used in the checklists were discussed, and it was ensured that there was a consensus regarding their meaning. This factor was emphasised more in the undergraduate experiment.

**Comments and feedback:** They were advised to write their comments on the feedback sheet. It was explained that the whole exercise should be more of an instructional activity. This factor was emphasised more in the graduate experiment.

**Consistency:** They were advised to check their notes frequently and achieve intra-rater consistency.

***Reliability and Validity***

A usual concern in the area of peer assessment is related to the fact that many times peer assessment comes nowhere near the assessment given by the teacher. However, in our experiment with the undergraduates the average score of the peers matched well with the instructor's

assessment in many cases. From our experience we know that intra-rater consistency is difficult to achieve in the assessment of presentations. Also, there is no significant variation among marks awarded by different raters to a presentation. In most cases the differences range from 0 to 10 marks.

However, in the case of the graduate students, we observed a distinct tendency of peers to be very reluctant in awarding low marks despite claims of using categories of measurement values similar to the instructor's. For example, the instructor failed a particular speaker with a 46 mark because he presented a talk on an uninteresting topic, maintained no eye contact with the audience, read entirely from the paper, made lots of pronunciation errors and used transparencies ineffectively. In contrast, his peers awarded him an average of 68, which was in the category of, "*Fairly good with some major faults*," despite the fact that their comments were similar (although somewhat more restrained) to those of the instructor. The peers too made comments like: "*font size too small, too much content on transparencies, no idea what's going on, pronunciation not good and weak voice*." Obviously, in terms of awarding marks, the peers showed more empathy with their peer and were more tolerant of major presentation weaknesses. This could be more positive than the instructor's 'bluntness,' but the marks may be misleading if taken out of context.

We noted that the deviation from instructor mark is smaller if the speaker is at

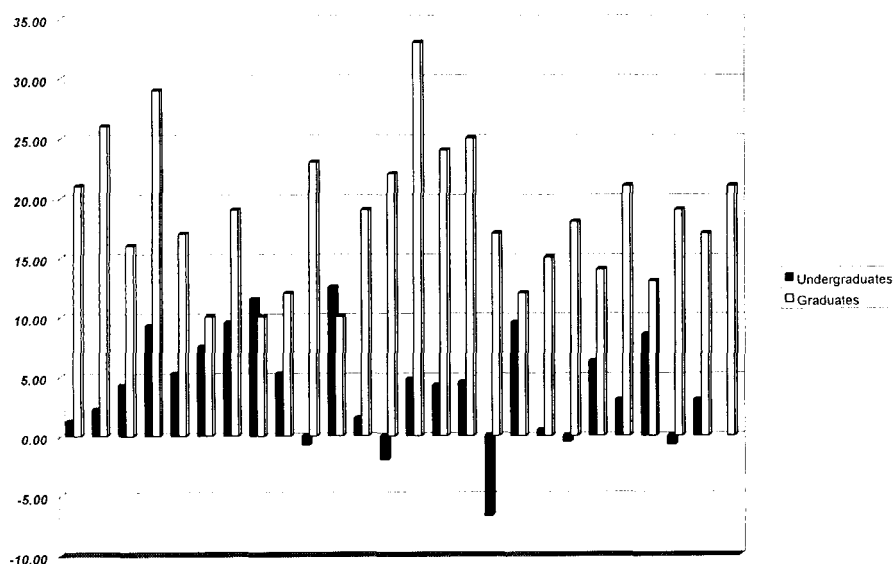
high performance level. For example, when the instructor awarded a good mark of 70 to one of the better speakers, his peers awarded marks between 76 and 86, and when instructor mark was 72, the peers gave between 79 and 89. To conclude, it appears that peers tend to be more consistent when assessing good speakers and more tolerant and divisive when assessing weaker students.

The actual assessment marks awarded by the graduate peers were varied and spread over larger ranges. In some instances the mark range (difference between the highest and lowest marks awarded by a set of peer evaluators) was smaller like 7 – 10 mark differences, but in others the variation spread across more than 25 points. In the most extreme case, a particular speaker was awarded 88 marks by a peer and 53 by another. This reflects the different perceptions of peers with regards to a particular speech or speaker, a phenomenon not uncommon of a group of real-life audience. It appears that each presentation appeals to different listeners differently and it would be formative for students of public speaking to realise this. Perhaps the most significant observation in this comparison is the pattern of assessment by peers and its marked contrast with that by the instructor. While the instructor's marks ranged between 42 and 72, the peer assessments hovered between 71 and 86, although in a few instances an individual student did coincide more with instructor assessment. It is apparent that peer assessment is consistently more generous than instructor assessment.

Besides peer mark range, the results also showed up very significant differences between peer averages and instructor assessments. The differences range between 10 and 33 marks. These results indicate very different perceptions between peer and instructor of a particular presentation despite their earlier shared measurement values used

for assessment. An explanation could be that at point of assessment peers were influenced by other factors other than the selected categories of measurement. Evidence of this is seen in their responses to the question on what they were thinking and feeling during the assessment.

**Figure 1: Student-instructor marking differences (The Y-axis shows the difference between the marks awarded by teacher assessor and marks awarded by student assessors.)**



In Figure 1, we can see that in most cases students have awarded higher marks than the teacher-assessor. However, in four cases, the graduate students have awarded less marks than the teacher-assessor. (See the numbers with a minus sign above.)

As the above figure shows, in a large number of undergraduate cases (13 out of 25) the difference amounted to less than 5%, while another large number of cases (10) showed a difference in marks ranging from 5 to 10. On the other hand, in the case of

graduate students there are rather significant differences between peer averages and instructor assessments.

From the above discussion, it is obvious that the way we structure our assessment forms significantly affects the intra-rater consistency. The undergraduates (who used a more structured assessment form) achieved greater intra-rater consistency.

### **Quality of feedback**

A more crucial area was peer feedback. Feedback was important, since the assessment was intended to be a learning experience for the presenters. We are glad to observe that the experiment generated many useful and specific comments by peers, which the speakers could find encouraging and formative. Receiving positive comments boost the speaker's confidence for future assignments. As discussed earlier, the graduate students were encouraged to write more elaborate comments. The structure of assessment forms also allowed them to write detailed comments such as the following:

*Had attention; clear main ideas, interesting content, supported by details; well structured; appropriate eye-contact, suitable hand gestures and good posture; used effective expressions, good grammar; good font size, clear and no crowded visuals; good overall performance.*

The exercise also appears to have provided the audience cum evaluators an open forum to give negative feedback constructively.

*No topic sentence given, clear main ideas, interest of content so so; no summary but impactful concluding remarks; well sequenced; little non-verbal, good posture; correct grammar, no good pronunciation and effective expressions; used transparencies and tables for comparison, highlighted important sentences on transparencies, but better not to print all*

*words you want to say, too crowded; title of speech not related to rest of content, but organisation is good.*

On the other hand, the undergraduate students offered largely cryptic comments (probably due to the nature of assessment form) of less critical nature. Here are some examples of actual feedback. We have included our perception of the feedback value in brackets.

*You should not show whole transparency. Distracting. You used laser pointer. But did not use it effectively.* (Very constructive, specific, and useful)

*Excellent effort.* (Not at all specific, not useful)

*Quite well-presented. Relevant physical models to show when presenting her speech.* (Less specific, not very useful)

*Hands inside pocket; reading from notes. Health effect and safety problem do not seem to be related to your topic.* (Specific, useful, but no reference to presentation media)

*Presentation rather fast. Too fast transition while flashing the slides.* (Very specific)

### **Student Perceptions**

In general, students felt that peer feedback was useful in gaining a conscious awareness that they were making the

presentation in a virtual setting. That affected how they organised and delivered their presentations. It also helped them to provide a clear goal for their work and organise the details in relation to the goal.

We made a more systematic attempt to measure perceptions of the graduate students. After the assessment session the graduate students were requested to answer a questionnaire on their experience in evaluating their peers. The objectives of this questionnaire were to determine some major affective features of the experiment, observe the measurement values perceived and used by the students, and find out their opinions on the concept of peer evaluation.

Questions in the follow-up questionnaire aimed to establish the affective aspects influencing the assessors during the experiment. They checked for maturity and reliability of the assessors (See Table 1). It is interesting to note that participants were

of the opinion that they were objective (56%), confident and constructive during their assessments. They based their evaluation on their knowledge of making effective oral presentation (64%) and on their perception of what good presentations should be (76%) and on how a particular presentation appealed to them (56%). Also interesting to note is that, 36% of the participants would probably have awarded high marks even if they thought the presentations were not of high quality in order to avoid embarrassing their friends. Apart from these considerations the participants appeared to have been rather mature and objective during their task and the results of experiment should be regarded as relatively reliable with high evaluative value. A more thorough discussion of the implications of these observations will be dealt with in the following section.

**Table 1: Affective influences during assessment**

During my assessment of each peer I was influenced by	Responses (Total:25)	
	N0.	%
1. Objectivity	14	56
2. My knowledge on making effective oral presentations	16	64
3. Constructiveness in my comments	10	40
4. Feelings of my evaluatees	5	20
5. My interest in maintaining good relationship in class	1	4
6. What I think a good presentation should be	19	76
7. How a particular presentation appeals to me at that time	14	56
8. Not wanting to embarrass my peer	3	12
9. My confidence in my evaluation	12	48
10. My doubts about my ability to assess accurately	2	8



Another major area of interest is major opinions held by participants on the value of peer assessment. Participants were asked if peer evaluation had a place in their training and if there were any disadvantages they had encountered. The next two paragraphs highlight some of their opinions.

The first significant benefit is the objectivity of the assessment. By that the students probably felt that evaluation by more than one person was definitely more fair and reliable. Some said, "It's a comprehensive evaluation of my presentation." They were glad that the audience was involved because more people's opinions were involved and when the average score was calculated the final mark evened out more fairly, especially considering the large mark ranges we observed earlier. Participants were happy to be both assessor and assessee because through this practice they could learn from each other. As assessors, they believed they could help to improve a peer's presentation by pointing out his weaknesses and providing constructive feedback for improvement. As the assessor, they were obliged to concentrate on their peers' presentation and this helped to focus their attention better on the elements of good presentation. They also obtained more information from their peers and got more involved in others' interests because during assessment they had to pay very close attention. They also felt that peer assessment supported self-evaluation. Listening to a good presentation helped them pick up useful tips for their own presentations later.

Bad presentations also helped them identify what to avoid when they became the evaluated. Being the evaluators also made them reflect on their own problems and awakened in them a greater awareness about how the audience perceived a presentation. As evaluatees, they felt that they could benefit a lot from their peers too. Their peers would help to pick out their mistakes, which they would be otherwise unaware of and offer them good advice for improvement. Also positive comments from peers would certainly boost their confidence in oral presentation. Overall the students believed that peer assessment has an important place in evaluation of presentations.

However, some disadvantages were identified. The main objection to peer assessment seems to be the fear of its subjectivity since evaluators are so varied themselves. A substantial number thought that peer evaluators did not have the expert knowledge about evaluation and may then be inaccurate and even biased in their judgement. Also, evaluators may be biased towards some of their peers or could have personal preferences. Therefore, how could peers, unless they were the experts, ensure fairness in awarding marks when there were no standard measures to follow? Others thought that evaluators being different, with different interests, different knowledge background, different evaluation standards, should not be entrusted to assess a presentation accurately. Furthermore, some participants felt that, "*As an oriental person it's not easy to evaluate other guys (sic).*" Finally, some participants were of the

opinion that it was very time consuming and tedious, and if too much “*attention is put on assessing peer then the purpose of presentation is sidestepped (sic).*”

### ***More Reflections***

The whole cycle of preparation-implementation-review was rewarding to all those involved in the process. However, we felt there were some weak areas which needed much better treatment. First, many undergraduate students were still hesitant to criticise or write any constructive comments at all on the feedback sheets. This could be an effect of the rather structured checklists given to them. The lists had only little space at the bottom to write their comments. (See Appendix 1). The structure might have put a little pressure in terms of an open expression of feedback. There were of course many useful comments, but the overall feedback experience could have been much improved by (a) an emphasis on this aspect during the briefing, and (b) a feedback sheet that encourages to provide constructive feedback.

Towards the end of the experiment, we also realized the importance of grouping of

assessors. We had given them complete liberty to form into groups. This may have been a cause of the rather weak evaluation and feedback coming from some groups. We could have perhaps achieved a more balanced result, if we had placed the assessors in appropriate groups to match their abilities with other members.

The undergraduates were specifically encouraged to make use of presentation software during their presentation. However, some tried to rely on conventional media such as overhead projectors and charts. A statistical analysis showed that the assessors generally seemed to be impressed with the presentations that used graphics software rather than the conventional media. This analysis probably reflects the fact that, other factors being equal, software-based presentations will always have an edge over conventional media-based presentations.

### ***Review and Conclusion***

The following table provides a quick summary of the similarities and differences in the two experiments.

**Table 2: Comparison of the two experiments**

	<b>Graduates</b>	<b>Undergraduates</b>
<b>Checklists</b>	Less structured	Structured
<b>Assessors' briefing</b>	Extensive	
<b>Topics</b>	Academic (Broader range)	Professional (Narrower range)

	<b>Graduates</b>	<b>Undergraduates</b>
<b>Test Reliability (Intra-rater consistency)</b>	Low	High
<b>Test Validity</b>	High	
<b>Beneficial Backwash</b>	High	Low
<b>Practicality</b>	High	
<b>Quality of feedback</b>	Extensive comments by student assessors	Cryptic comments by students assessors
<b>Student perceptions</b>	Varied, but supporting peer assessment in general	Not measured
<b>Affective influences</b>	In general, mature and affective	Not measured

From the comparison of the two experiments, we can make the following concluding remarks.

1. Structure of assessment forms and briefing of assessors seem to emerge as the strongest factors affecting the peer assessment process.
2. There seems to be a tension between achieving intra-rater consistency and *Beneficial Backwash* (a term used to refer to the effect of assessment on teaching). In other words, if we wish to achieve quality instruction through
3. Peer assessment seems to be a fruitful instructional exercise. It can be seen from the experiment with the graduates that in the case of oral presentations, this exercise provided a multi-dimensional feedback. Such multiple perspectives (often clashing with each other) are a unique and essential part of a quality instructional process.

assessment of oral presentations, we have to sacrifice part of intra-rater consistency. Evidently, this is what has happened in the experiment with graduates.

### **The Authors**

Anil Pathak is an Assistant Professor at the Nanyang Technological University in Singapore. He has taught communication skills in industrial, conventional, as well as distance education systems. His current interests include teaching multimedia-based presentations and use of web for communicative teaching.

Chan Soon Keng is a lecturer at the Nanyang Technological University. She is working towards a doctorate in academic and professional writing. Her research interests include analyzing different academic and professional text types. She teaches Communication Skills and English Proficiency to NTU undergraduates, Research Writing to international scholars, and Negotiating Skills to graduating students. Her special interests are developing interpersonal communication skills as well as analyzing, planning and writing business documents.

### **References**

Carlson, R. E. & Smith-Howell, D. (1995). Classroom public speaking assessment: Reliability and validity of selected evaluation instruments. Communication Education, 44, 87-97.

Cheng, W. & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. Studies in Higher Education, 22 (2), 233-239.

Nunan, D. (1994). Self-assessment and reflection as tools for learning. Assessment colloquium, international language in education. Hong Kong: University of Hong Kong.

Macpherson, K. J. (1999). The development of critical thinking skills in undergraduate supervisory management units: Efficacy of student peer assessment. Assessment & Evaluation in Higher Education, 24, 273-284.

Roskans, T. (1999). Chinese EFL students' attitude to peer feedback and peer assessment in an extended pairwork setting. RELJ Journal 30(1), 79-123.

Weaver, R. L. & Cotrell, H. W. (1990). Peer evaluation in the basic speech-communication course: A student supported process. World Communication, 19(1), 67-76.

**Appendix 1: Checklist for assessing oral presentations (Undergraduates)**

<b>Subject Matter</b>	<b>Prescribed Range</b>	<b>Delivery</b>	<b>Prescribed Range</b>
<b><i>Exposition (10)</i></b> ➤ Attention getting ➤ Establishing credibility ➤ Preview of presentation	<b>3-7</b>	<b><i>Gestures (10)</i></b>	<b>3-7</b>
<b><i>Relevance (5)</i></b> ➤ Choice of topic ➤ Choice of points ➤ Choice of matter ➤ Choice of examples	<b>2-4</b>	<b><i>Eye contact (10)</i></b> ➤ Establishing Rapport ➤ Covering the entire audience	<b>3-7</b>
<b><i>Credibility (5)</i></b> ➤ Evidence ➤ Support/Substantiation	<b>2-4</b>	<b><i>Voice and pronunciation (5)</i></b> ➤ Use of voice energy ➤ Proper volume ➤ Tone variations	<b>2-4</b>
<b><i>Logical order (10)</i></b>	<b>3-7</b>	<b><i>Language (5)</i></b> ➤ Accuracy ➤ Appropriateness	<b>2-4</b>
<b><i>Smooth progression (10)</i></b> ➤ Links ➤ Unity	<b>3-7</b>	➤ <b><i>Presentation media (15)</i></b> ➤ Choice/Suitability ➤ Effective use	<b>4-12</b>
<b><i>Summary/Conclusion (10)</i></b> ➤ Review ➤ Ending	<b>3-7</b>	<b><i>Appearance (5)</i></b>	<b>2-4</b>

**Appendix 2: Checklist for assessing oral presentations (Graduates)**

<b>Tick applicable details to describe speaker's performance</b>	<b>Comment on strength and weaknesses of speaker's performance</b>	<b>Marks</b>
<p><u>Content:</u> Had outline:</p> <p><i>Introduction:</i> Had attention-getter, developed attention-getter, gave topic sentence.</p> <p><i>Body:</i> clear main ideas, accurate information, interesting content, supported by details.</p> <p><i>Conclusion:</i> summary, impactful concluding remarks.</p>		<p>Max: 20</p>  <p>Your mark</p> <p>_____</p>
<p><u>Organisation:</u></p> <p>Well-structured with introduction, body and conclusion, well sequenced, cohesive, sign-posted with logical connectors.</p>	<p>Introduction:</p>  <p>Body:</p>  <p>Conclusion:</p>	<p>Max: 20</p>  <p>Your mark</p> <p>_____</p>
<p><u>Presentation style:</u></p> <p>Appropriate non-verbals, e.g., eye contact, suitable hand gestures, good posture; good pacing of speech, good voice, enthusiastic.</p>		<p>Max: 20</p>  <p>Your mark</p> <p>_____</p>
<p><u>Language use:</u></p> <p>Correct grammar, effective expressions, clear, simple expressions, good pronunciation, accurate vocabulary.</p>		<p>Max: 10</p>  <p>Your mark</p> <p>_____</p>
<p><u>Use of visual aids:</u></p> <p>Suitable supporting materials, enhance verbal content, motivating to listeners, aid verbal presentation, e.g., good font size, clear, not crowded, well organised, well related to speech.</p>		<p>Max: 10</p>  <p>Your mark</p> <p>_____</p>
<p><u>Overall Performance:</u></p> <p>How the speaker and speech impressed you on the whole.</p>		<p>Max: 20</p>  <p>_____</p>