ISTANBUL GELISIM UNIVERSITY
SCHOOL OF FOREIGN LANGUAGES

Standardisation Manual

on the Assessment of Writing and Speaking Examinations

Quality Assurance

2019-2020

NOTE

This document is designed solely for enabling coordination in the assessment of students who are enrolled in the preparatory programme –especially for Progress (PT), Proficiency and (YET) tests.

It is based on the established Writing and Speaking evaluation scales for Preparatory Programme, and should not be used for any other examinations.

References

- Weigle, S.C. *Assessing Writing*. Cambridge University Press
- Luoma, S. *Assessing Speaking*. Cambridge University Press

Training

A. Writing

Objectives

The objectives of the Standardization Workshop are:

- to present the writing criteria for assessment clearly to all graders
- to maintain a shared understanding of writing grading throughout the school
- to prevent any errors of fact regarding writing assessment
- to offer a conflict resolution method in a case of any errors occurring

Standardization Procedure

1. The Unit reads through scripts to find **anchor/benchmark scripts** that exemplify different points on the scale. (three to ten anchors, depending on the number of graders, complexity of the rubric, and the experience of the graders.)
2. The first set of scripts is given to graders in order (from highest to lowest) with appropriate scores indicated, and should be as unambiguous a set as possible. This set is used to familiarize graders with the scale and to instantiate certain features of the rubric. The Unit can use these scripts to describe for the graders what is meant by phrases used in the rubric.
3. At this point questions are raised by the graders and discussed with the whole group.
4. Once the graders feel comfortable with the scale as defined by the Unit and instantiated in the first set of anchor scripts, another set is given that includes one script at each level in random order.
5. Graders are told that there should be one recording at each level and given a chance to rate the recording themselves.
6. At this point questions are raised by the graders and discussed with the whole group.
7. Once graders are able to handle this task, more problematic sets can be given out.

It should be noted that it is virtually impossible to get a large group of graders to agree on exact scores and that some disagreement among graders is inevitable. It is more important to communicate to graders the amount of variability that is acceptable and let them know that they are not required to be perfectly accurate at all times.

Notes to Consider

- Checks on the grading in progress by the Unit help to ensure that individual graders are maintaining the agreed-upon standards for grading.
- Evaluation (of the graders) and record keeping are essential for an ongoing assessment program so that reliable graders are kept on and unreliable graders are retrained or dropped if necessary.
- In a large grading, additional workshops may be required at certain points. For example, if the grading takes place over more than one day, or one or two sets of anchor scripts can be used to recalibrate graders each day.

B. Speaking

Objectives

The objectives of the Standardization Workshop are:

- to present the speaking criteria for assessment clearly to all graders
- to maintain a shared understanding of speaking grading throughout the school
- to prevent any errors of fact regarding writing assessment
- to offer a conflict resolution method in a case of any errors occurring

Standardization Procedure

1. Training session should begin with an introduction to the exam and the criteria. Different levels on the rubric are then illustrated, through taped performances that have been rated by the Unit before the workshop.
2. The Unit goes through recordings to find **anchor/benchmark recordings** that exemplify different points on the scale. (three to ten anchors, depending on the number of graders, complexity of the rubric, and the experience of the graders.) If this is not possible, sample recordings of exam sessions should serve as anchor as long as they are matched with the rubric.
3. The first set of recordings is shown to graders in order (from highest to lowest) with appropriate scores indicated, and should be as unambiguous a set as possible. This set is used to familiarize graders with the scale and to instantiate certain features of the rubric. The Unit can use these recordings to describe for the graders what is meant by phrases used in the rubric.
4. At this point questions are raised by the graders and discussed with the whole group.
5. Once the graders feel comfortable with the scale as defined by the Unit and instantiated in the first set of anchor recordings, another set is given that includes one recording at each level in random order.
6. Graders are told that there should be one recording at each level and given a chance to rate the recording sessions themselves.
7. At this point questions are raised by the graders and discussed with the whole group.
8. Once graders are able to handle this task, more problematic sets can be given out.

It should be noted that it is virtually impossible to get a large group of graders to agree on exact scores and that some disagreement among graders is inevitable. It is more important to communicate to graders the amount of variability that is acceptable and let them know that they are not required to be perfectly accurate at all times.

Test Administration

All test administration procedure is carried out as described in the School of Foreign Languages Quality Manual. There are also additional considerations that need to be taken into account.

Writing

Unless specified otherwise, no use of physical or digital dictionaries are allowed.

Speaking

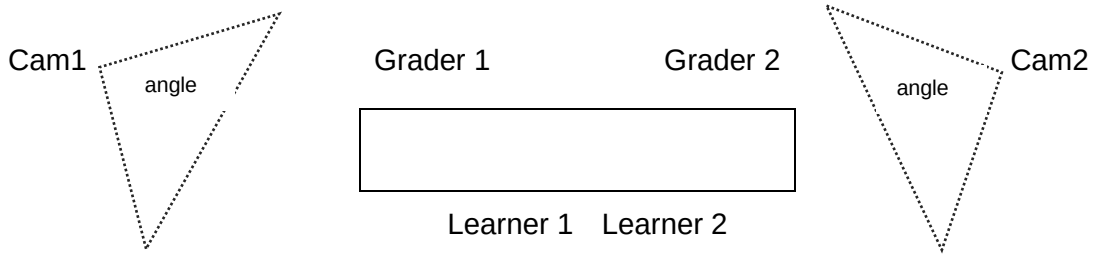
Unless specified otherwise, speaking examinations are conducted as interviews in pairs, during which two students take the test simultaneously.

Two classrooms are allocated for each group for test administration. One classroom is assigned as “exam hall” in which the interviews are held, and the other is assigned as the “waiting room”, in which students wait for their turn.

Students are interviewed in the order their names appear in the register.

The two graders are not allowed to interact with each other during the exam session.

A possible seating plan for a speaking session is described below:



*Cam1 & Cam2 are possible camera locations. The camera should clearly be directed at both learners whose facial expressions should be visible when recorded.

Grading

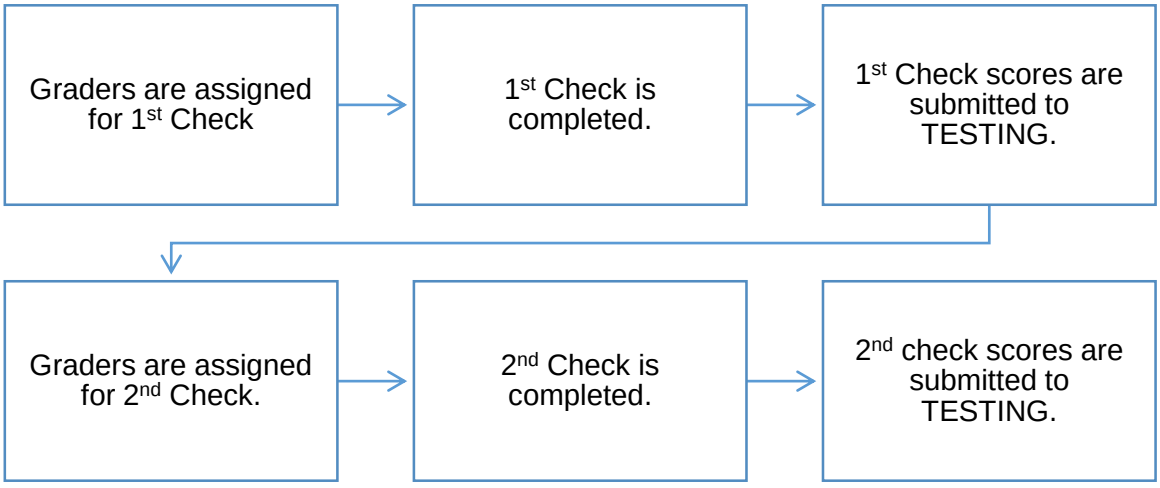
Writing and Speaking tests, both being types of subjective assessment, require independent grading by two graders.

Because of the nature of the tests, each grading procedure achieves this in a different way.

Writing

Analytic scoring is used for the assessment of Writing. This allows separate reporting of performance in each criterion to provide statistical data on monitoring progress. The Writing Evaluation Scale developed by TESTING is used.

The process is as follows:

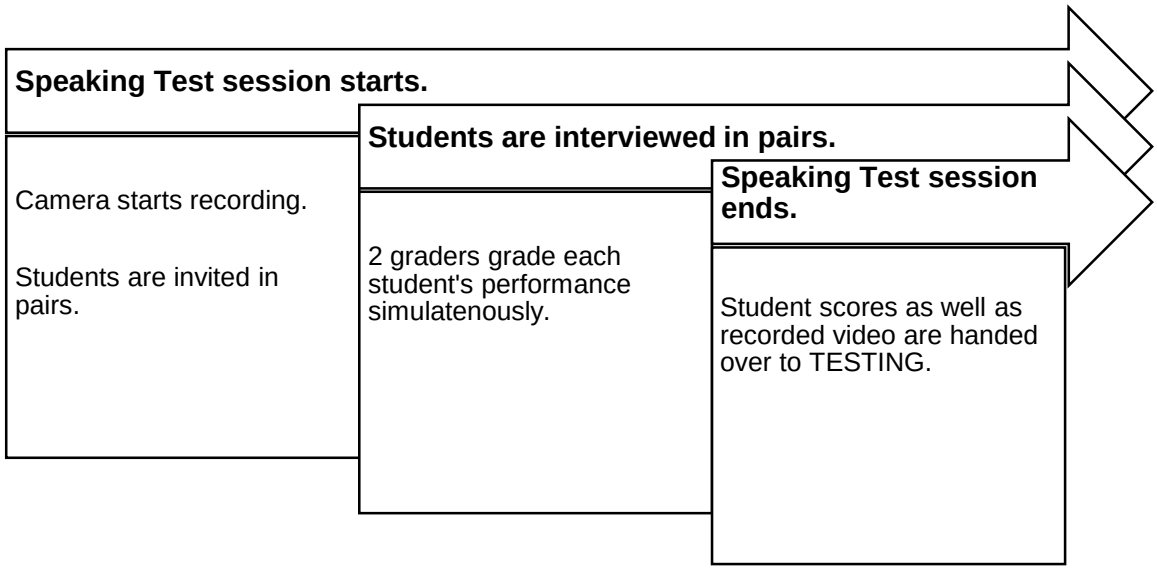


Once live grading is under way, it is important to ensure that grading is independent –that is, that graders do not see and therefore cannot be influenced by scores given by other graders. It is essential for the integrity of the scoring process that graders arrive at their scores independently, without reference to scores given by other graders. For this reason, it is also important that graders do not write comments or underline errors when scoring scripts to avoid influencing the scores given by other graders.

Speaking

Analytic scoring is used for the assessment of Speaking. This allows separate reporting of performance in each criterion to provide statistical data on monitoring progress. The Speaking Evaluation Scale developed by TESTING is used.

The grading of Speaking tests takes place synchronously with the test, at the moment of interviewing by two separate graders.



Notes to Consider

- Live grading should commence *after* the Standardization Workshop is completed.
- For writing tests, grading should be done in a controlled grading (if possible). A controlled grading is a group of graders meeting together to grade scripts at the same place and time. (to eliminate unnecessary sources of error variance and a positive social environment to maintain grading standards.)
- The tone set by the Unit has a tremendous influence on the success of the grading. If it is led with sensitivity and respect, it can be enjoyable and professionally valuable experience for graders, on the other hand, poorly run gradings in which graders feel exploited or coerced, can turn graders against the grading process which in turn can have negative effects on the scoring itself.

Discrepancy

Once the exam scores are handed over to the Assessment and Evaluation Unit, the unit members compare the 1st and 2nd Check scores for discrepancy.

The maximum limit for discrepancy between first and second checks is twenty percent. Anything beyond that requires a 3rd check.

Method of Resolve

If the discrepancy between 1st grader and 2nd grader scores is above 20% (21 and above out of 100 marks), then a consensus must be reached. This is called "3rd check."

Discrepancy Ratio	Action required
20% and below	The average of the sum of 1st and 2nd check scores gives the final score for exam. **Example:** *Alice takes the Writing test. The 1st grader scores her paper as 50 (out of 100 marks.), and thr 2nd grader scores her paper as 60 (out of 100 marks.) Since the discrepancy between two scores is within the accepted discrepancy limit, no 3rd check is required.* *So;* *50+ 60 = 110* *110 / 2 = 55* *So Alice's Writing test score is 55.*
21% and above	A consensus between two graders are reached at a meeting in which two graders review the script (if Writing) or the video recording (if Speaking) and decide upon a final score for the student. **Example:** *Alice takes the Speaking test. The 1st grader scores her performance in the interview as 40 (out of 100 marks), and the 2nd grader scores her performance in the interview as 65 (out*

	of 100 marks.) Since the discrepancy between the two scores is beyond the accepted limits, a 3rd check is required. *TESTING invites the 1st and 2nd graders on a schedule time and the two graders together watch the video recording of Alice's performance, giving reasons for their choice on the evaluation rubric. Two graders discuss the criteria one by one and find a common ground.* *After their meeting, the two graders jointly score Alice's performance as 55. Then they fill in another Speaker Evaluation Scale and mark it as 3rd check.* *This serves as Alice's speaking test score.*