



Guidelines for Making and Scoring of **Free-Response Questions**



Co-created by
**CBSE - Centre for Excellence in Assessment
and
Educational Initiatives**

November 2023



GUIDELINES FOR MAKING AND SCORING OF FREE-RESPONSE QUESTIONS

First Edition: November, 2023
Unpriced e-Publication, not for sale

**Co-created by
CBSE Centre for Excellence in Assessment
and
Educational Initiatives**

Executive Summary

The National Education Policy, 2020 emphasizes the importance of transforming assessments and adapting the education system to the changes in the global landscape. To do this, high-quality questions that test skills like conceptual understanding, critical thinking, problem-solving, creativity, and collaboration can enable students to learn and acquire relevant skills for the future. A variety of question types that range from Multiple-Choice Questions (MCQ) to Free-Response Questions (FRQ) enable a holistic assessment of student learning outcomes.

This document offers guidelines and principles for creating Free-Response Questions (FRQs) to help assess and evaluate students' understanding of important concepts. It addresses areas like question development, grading of student responses, and helpful evaluation strategies while also pointing out potential mistakes to avoid. It highlights the importance of developing well-crafted FRQs that provide clarity and context to the student. The document suggests using real-life examples that connect with what students are supposed to learn and outlining clear expectations from the students' responses.

Simultaneously, the document delves into the creation of rubrics for free-response questions, emphasizing the importance of fairness and consistency in evaluating student responses. It underscores the need for a clear and well-defined marking scheme, taking into account both response quality and the underlying reasoning.

Further, the guidelines within this document outline best practices for assessing student answers. These practices encompass measures such as limiting the number of responses per evaluator, utilizing exemplar responses from experienced evaluators as reference points, and enhancing evaluator skills to ensure consistent grading. The document also enumerates specific guidelines for evaluators to adhere to and common pitfalls to avoid during the evaluation process.

Adhering to these principles would enable educators to ensure that assessments are fair and effectively measure students' knowledge and understanding. Ultimately, these practices facilitate student growth and readiness for success in an ever-evolving and dynamic world.

Table of Contents

A. Role of Good Questions in an Assessment	3
B. Key Principles of Question Creation.....	4
C. Making Free-Response Questions	11
D. Key Principles of Free-Response Questions and Marking Scheme Creation.....	12
E. Best Practices for Evaluation of Student Responses on Free-Response Questions	21
F. Do's and Don'ts for the Answer Script Evaluators.....	30

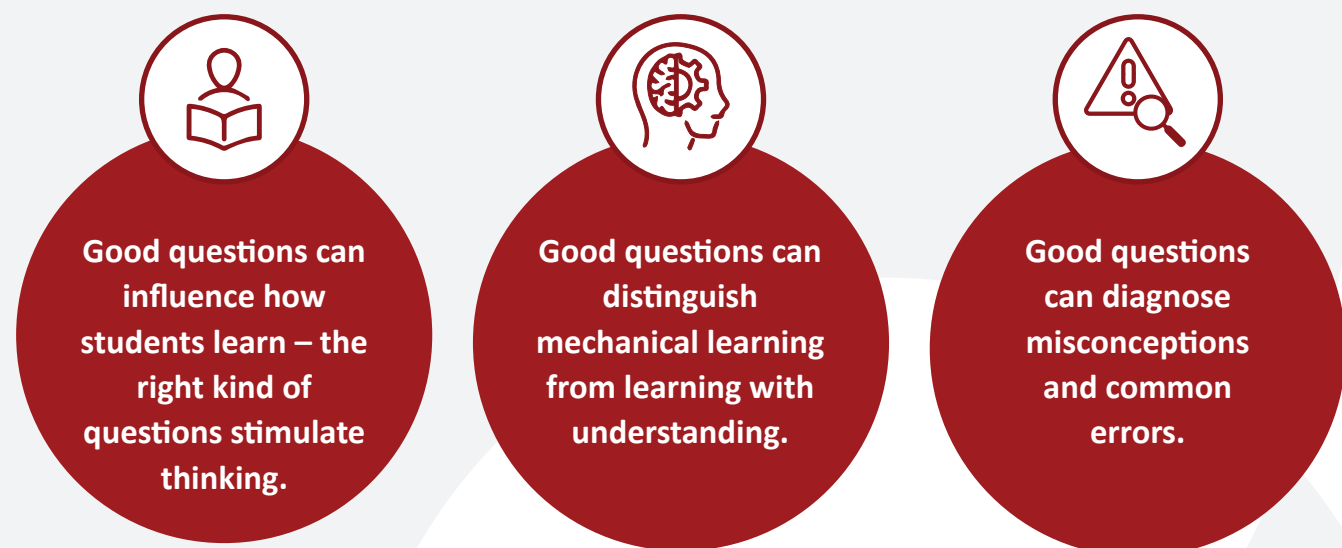
***“The one real goal of education
is to leave a person asking questions.”***

- Max Beerhohm

A. Role of Good Questions in an Assessment

Questions are central to any good assessment and the quality of the questions determines the quality of the insights that one can derive based on the data on those questions. A good question is one that challenges and stimulates a learner to think deeply and apply concepts. The ability to ask questions that make students think both at the time of instruction and assessment is the hallmark of a good teacher. A good question, correctly framed, can help a teacher understand how a student thinks and how well a student has internalized a concept or mastered a skill.

Questions are important because of the following reasons:



Tests are an important tool to help gauge how much has been learnt, and ‘how well’ children are doing. They are yardsticks that also provide valuable feedback about the effectiveness of instructional methods. Assessment with good questions should serve a dual purpose of measuring what students have understood as well as providing actionable insights. If we are able to empower teachers with the data they need to take the necessary action in order to overcome the gaps in learning, we should be able to see a shift in learning levels in the positive direction.



B. Key Principles of Question Creation

This section highlights some of the key principles of question creation. We have used examples to illustrate the points and clarify how to identify good questions from average or poor ones.

1. Test understanding of concepts and not just recall of facts.

While recollection of knowledge is important and there should be questions testing that, it is equally important to have a good balance of questions and there should be a substantial number of questions testing understanding of concepts and application of concepts in unfamiliar contexts. Evaluating students' conceptual understanding is useful as it can pinpoint areas that need further action in order to develop the concept or build future concepts. For example, while there can be a question asking 'Define a peninsula.', a question like the one illustrated below can check whether students can demonstrate an understanding of the concepts of the tilt of the Earth's axis in the given context.

Sample question 1: Question testing understanding of a concept

Identify the characteristic that is common to all the regions shaded on the following map.



The question requires students to use the understanding that peninsulas are landforms that are surrounded by water on almost all sides except one where it is connected to land. If students understand that, they should be able to see that the common characteristic for all the regions highlighted is that they are peninsulas. Even if students are able to express the meaning in words, without mentioning the term 'peninsula', it should be okay as it demonstrates that they have understood the concept.

2. Test core learning and not peripheral facts.

The question should test an important concept and not a trivial fact. While trivial facts should be avoided in any assessment, prioritisation of key ideas becomes important in a summative assessment as there is a limit to the number of questions that can be asked in a fixed time. Key ideas to be included in an assessment should be wisely selected to gather maximum information about students' learning levels and their understanding of concepts. Here are examples of two questions, one testing recollection of a trivial fact and one testing recollection of an important fact. Questions like the first one should be completely avoided in an assessment.

Sample question 2: Recall of a trivial fact

What is the temperature of the Sun?

Sample question 3: Recall of an important fact

At what temperature does water boil?

While both sample questions 2 and 3 are testing the skill of recollection of knowledge, what question 2 tests is recall of a trivial fact which doesn't have any implication on understanding the concept or learning other concepts whereas question 3 tests recall of an important aspect, of the boiling point of water. Also, if the different answers in sample question 3 are analysed, it can help in understanding specific wrong notions students may have about the boiling point of water.

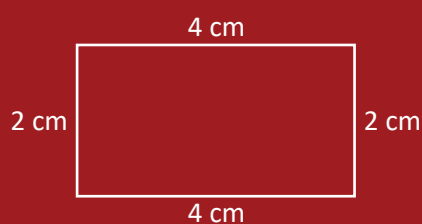


3. Use unfamiliar format of questions to differentiate real learning from apparent learning.

In order to prevent students from answering questions based on a question pattern or style that matches the ones from the textbook or past papers, it is important to ask the question in a way that makes them think deeply before they answer. So, while the concepts may be from the curriculum, the contexts used, or the framing of the questions should not be something that they are familiar with. This will help in differentiating between apparent learning and real learning.

Sample question 4: Textbookish question

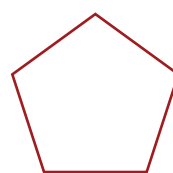
What is the perimeter of this rectangle?



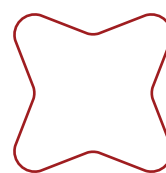
Answer: _____

Sample question 5: Question with an unfamiliar context

Which of these shape(s) has/have a perimeter



shape 1



shape 2



shape 3

Answer: _____

Most of the textbooks cover the procedural aspect of the calculation of the perimeter of a given shape. So, when a question like sample question 4 is asked, students tend to instantly apply the formula or rush to add the four numbers to calculate the perimeter. But does that indicate whether students understand what perimeter means? A question like sample question 5 can help in uncovering whether students really understand the concept or not. The choices selected by the students can also help in understanding their misconceptions related to the concept of perimeter.

4. Use real-life, authentic contexts that students can relate to.

The use of real-life contexts and data can make questions more engaging, and help students understand the practical importance of their education. Therefore, in addition to testing concepts, these questions become teaching tools in themselves. Their use in examinations will also encourage teachers to structure classroom instruction accordingly.

This principle can apply to questions across subjects. In Mathematics and Physics, for example, data and examples used in questions should be framed such that they relate to students' households, commutes, chores, etc. In language examinations, comprehension passages can be from real texts across domains such as history, science, or economics. While difficulty levels should be moderated based on grade levels, the principle should be followed as far as possible.

Sample question 6: Good real-life context

Infrared light shows us the heat radiated by the world around us. By viewing objects with a thermal infrared camera, we can actually "see" the differences between warm and cold objects. For example, we can use the temperature scale on the right side of the infrared image to find out the temperature range of different objects in the image. (Note that the temperature range on the right is Fahrenheit scale.)



Source: http://coolcosmos.ipac.caltech.edu/image_galleries/ir_zoo/

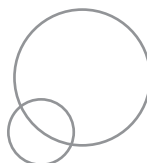
Shown here is an infrared image of a pencil. Check the temperature of the writing tip of the pencil.



Source: http://coolcosmos.ipac.caltech.edu/cosmic_classroom/light_lessons/thermal/images/

Give one reason why the tip is showing this temperature compared to the rest of the pencil.

The question uses an authentic real-life context of infrared cameras and a question that is based on the description given. It is possible that not all students are aware of what infrared photos mean and how infrared photos are to be interpreted. That's the reason a stimulus material explaining the working of infrared photos is given and the question expects students to use their understanding of other concepts and their application in this context.



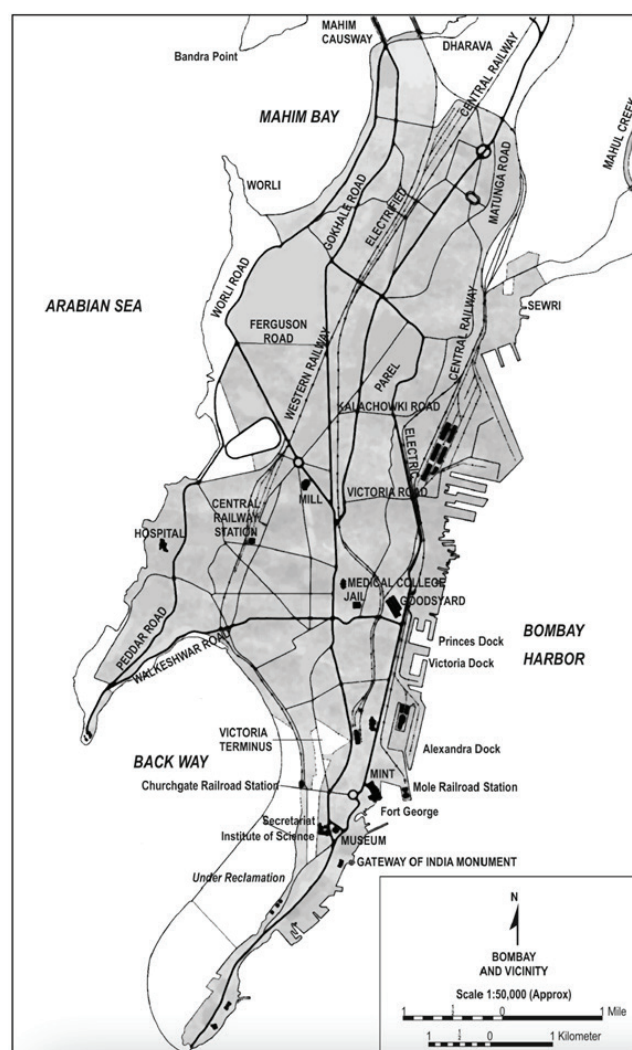
5. Use stimulus material as needed and test required skills.

Many effective assessments make good use of stimulus material by using them as a passage or a piece of information and asking questions based on it. It can be a good way to present a real-life context to the student and check for the application of a concept. What is more important in such cases is to ensure that the questions are linked to the stimulus material; students should not be able to answer the questions based on the material independently as it would defeat the purpose of using the material.

Here is a sample that illustrates how an excerpt from a book can be used as a good stimulus material.

Sample question 7: Question using a stimulus material

1. State 2 things about which the map does not provide any information. Justify your answer.
2. A port, metro railway line, railway stations and catholic churches are some of the features found in the present-day Mumbai. Which of them are NOT likely to be represented in maps created in the early 20th century? Justify your answer.



The set of questions uses a map as a stimulus material. Students are expected to study the map and answer questions that require interpretation of the map or are related to the concept in the context of the given map.

6. Use authentic data, maps, graphs or any other information.

As far as possible, data, maps, graphs or other information used in questions should be from an authentic source. This can allow students to relate to the context better and read about real-life applications of what they are learning.

Sample question 8: Question using data from an authentic report

Read 'Can Africa Feed the World?' and answer the following questions.

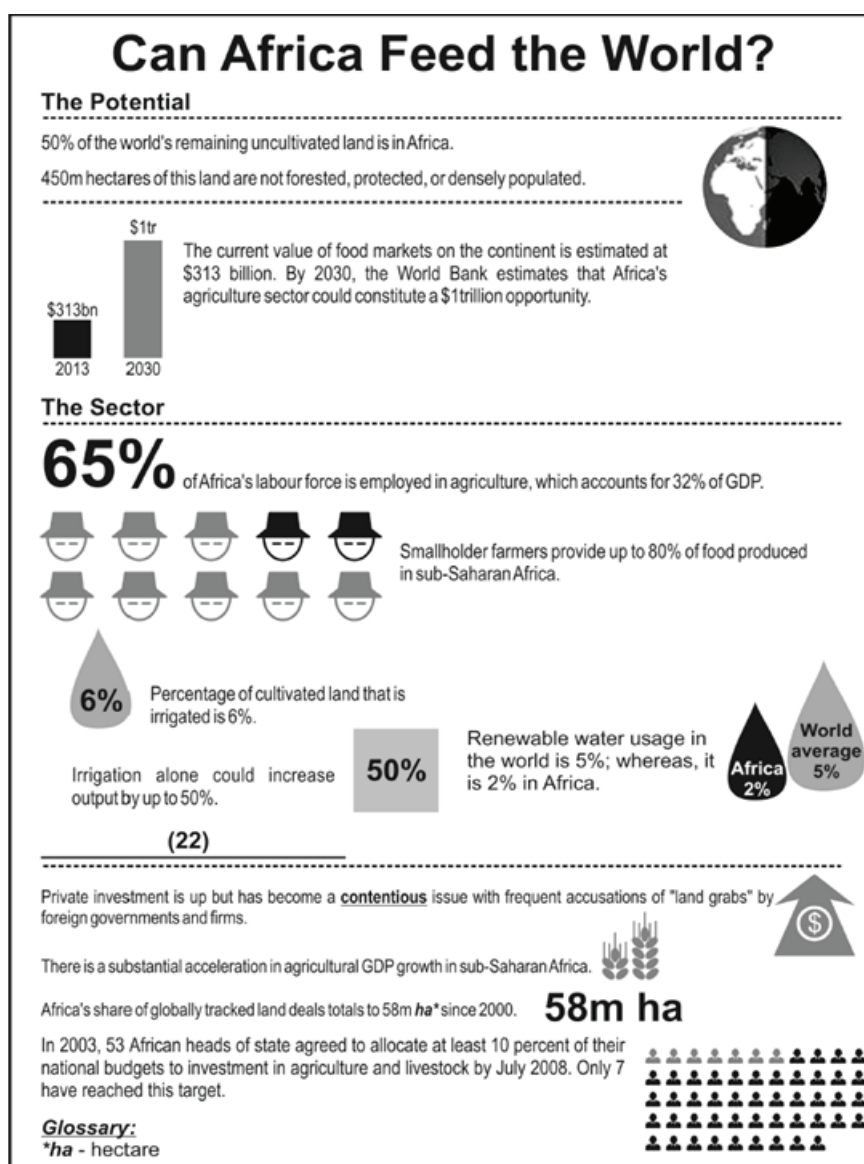
1. What aspect/aspects of the infographic makes it factual?

(i) not based on personal opinions

(ii) graphical representation of information

2. What is the main purpose of the infographic? Explain your answer in up to 200 words.

3. Give a suitable title for the infographic and explain your choice.



Source: 'Can Africa Feed the World?' - <http://www.legacyhotelsnigeria.co.za/wp-content/uploads/can-africa-feed-the-world.jpg>

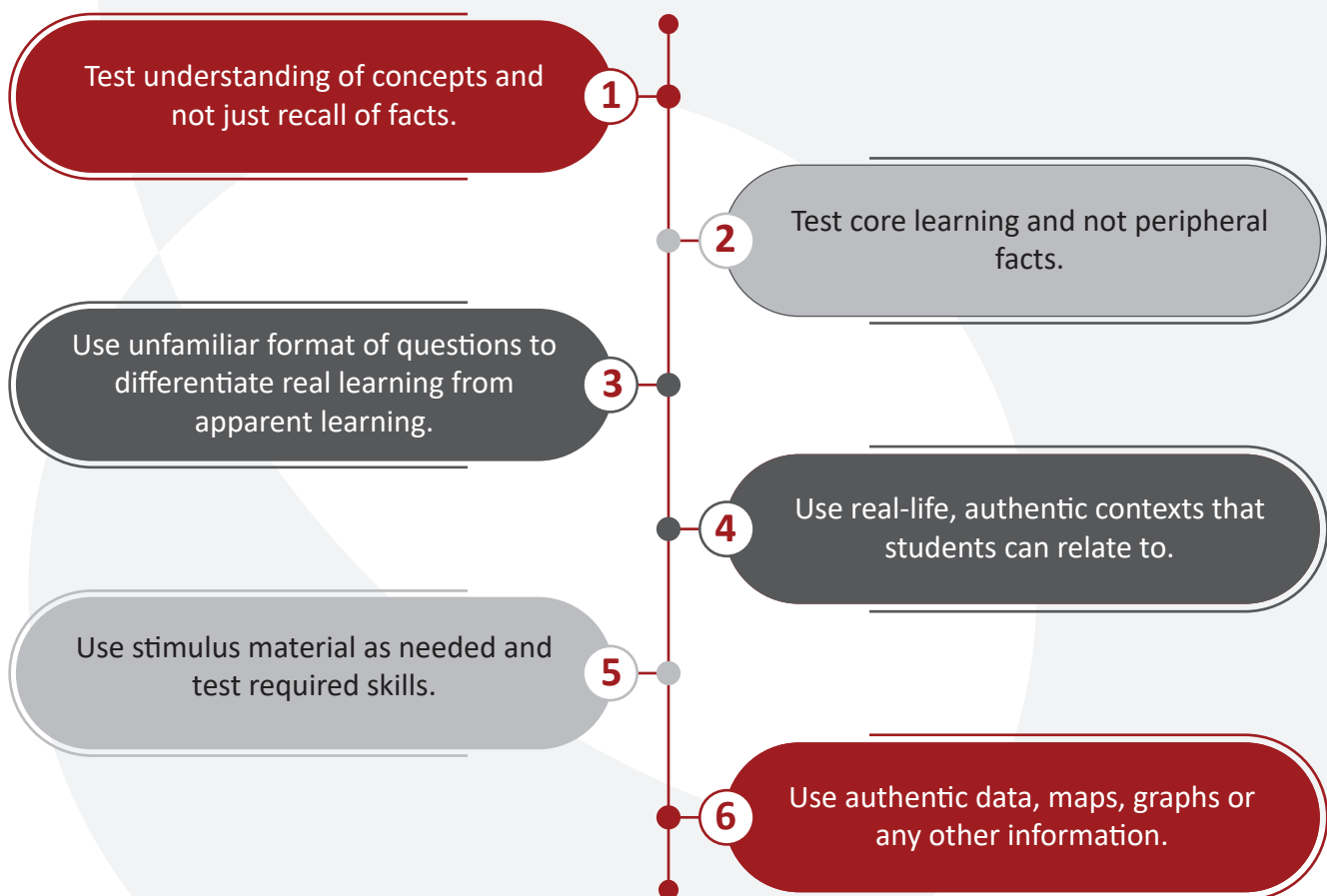
The question uses an infographic, acknowledges the source from where it has been taken asks a question that expects students to answer questions based on its interpretation.

7. Use age-appropriate content, context as well as language.

The question should be age-appropriate (in terms of content as well as language). This is another criterion that is important as it determines the validity of the question. If students cannot answer the question because of specific terms that they have not encountered or specific concepts that they are yet to learn/be exposed to, it affects students' performance, and the question doesn't help in revealing the learning level of the students.

Key Takeaways:

In summary, these are the main points to keep in mind when framing a good question. A good question should:



(a/b)



C. Making Free-Response Questions

A Free-Response Question is one in which a student is expected to create an answer as opposed to simply choosing from a list of choices. The response could be a single word or an essay or an opinion or an experiment to be designed. A Free-Response Question consists of a:

- **Stem** – the text of the question (with/without images, tables, graphs, etc.)
- **Scoring rubric/Marking scheme** – the scoring guide for the question. It helps the assessor score a question by describing the attributes of the response and also indicates the value to be awarded for that particular response

In the image given here, a man is moving a table in the indicated direction on a rough surface.

- a) *What are the two horizontal contact forces acting on the table?*
- b) *What can be said about the direction of these forces with respect to the indicated direction?*

Stem

Response	Marks
(a) Award 0.5 mark for each: <ul style="list-style-type: none">• muscular force / force applied by the man• friction	1
(b) The direction of muscular force is along the indicated direction, while the direction of friction is opposite to that.	1

**Scoring rubric/
Marking scheme**

The stem is the beginning part of the question that presents the question to be solved, a question asked of the respondent, or an instruction directing what the student needs to do, and contains any other relevant information. Depending on the testing objective, it may even include an image or a graph or a table or any other source of information.

The scoring rubric/marketing scheme describes the different responses that one may look for and the marks that are to be allotted for a given response.

For certain questions, such as an applied knowledge question, the stem can consist of multiple parts. The stem can include authentic material such as a graph, a table, or a detailed description which has multiple elements to it. In such cases, the stem ends with a lead-in part explaining what the respondent is expected to do.

D. Key Principles of Free-Response Questions and Marking Scheme Creation

This section of the document discusses critical aspects that a Question writer needs to consider while creating Free-Response Questions. This is followed by some do's and don'ts that need to be kept in mind.

1. The question stem should give rise to an intended response that can be scored.

Question writers should ensure that the students should be able to understand what is expected in terms of an expected response from the question. A clear indication should be provided, wherever possible, of the details needed and the extent to which the student should write the response. For example, if the intention of the question is to test the uses of satellites in a particular domain such as weather forecasting, the question should clearly state the domain as well as how many uses the student is expected to list.

Sample question 9:

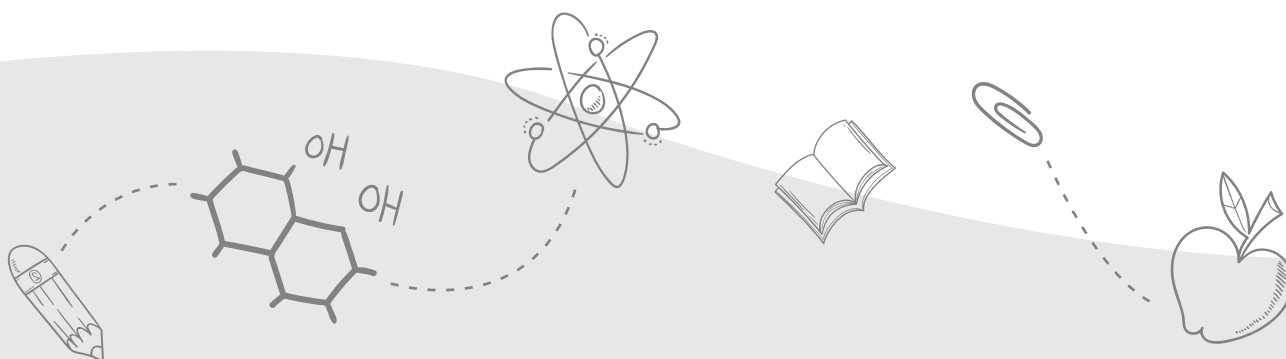
What are the uses of satellites?

Sample question 10:

State two ways in which satellites can be useful in weather forecasting.

Sample question 9 is a question that can invite different kinds of correct answers from the students. Responses may include the use of satellites in agriculture, weather, military, communication, and many others. Sample question 10 clearly asks for responses on their use in weather forecasting, with details around how many uses the student is expected to list. Here, scoring the question becomes easier and more objective.

To be able to do this, it is crucial for question writers to consider answering a question created. It helps with the construction of the rubric, allotting appropriate marks, eliminating ambiguity, and also ensuring the testing objective is met.



2. It is usually a good practice to indicate the length of the answer expected.

In assessments that are time-bound, it becomes vital to ensure that the student should be able to answer the questions within the time frame set for the assessment. So, it is important to frame the question such that it clearly states the length of the answer expected. If this is not mentioned, students may end up writing more for a question that will fetch them lesser marks, thereby losing out on time.

Sample question 11:

In the poem, 'The Elf and the Dormouse', when and why did the dormouse cry, "Goodness Gracious me"?

Sample question 12:

Describe two signs that told Ray that the two men who had entered his shop were not shoppers. Answer in 40–50 words with reference to 'The Old Clock Shop'.

Sample question 13:

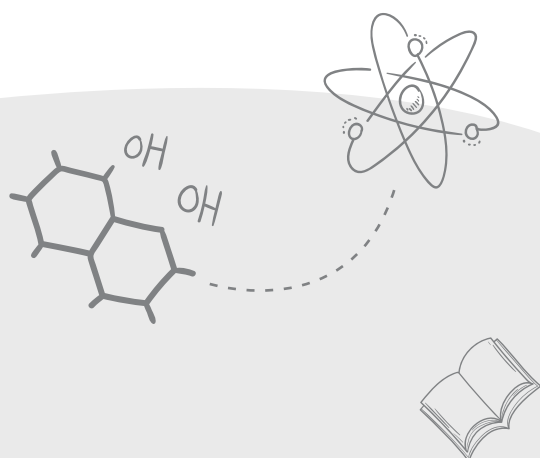
How are photosynthesis and respiration different?

Sample question 14:

State TWO points of difference between photosynthesis and respiration.

In sample questions 11 and 13 above, it is not clear whether one needs to write multiple lines or even a paragraph or simply a few words that are commensurate with the marks that will be awarded. Whereas in sample questions 12 and 14, the instructions are very clear – whether it is number of words that are expected or number of specific points that need to be stated.

Although students gauge the length of the answer expected from the number of marks allotted to the question, the intention of the assessor may be different. Therefore, it is important to specify this detail in the stem of the question itself.



3. Other good practices.

There are some more good practices that need to be followed in the process of creating good question stems.

Sr. No.	Practice	Example of a poor Free-Response Questions	Example of a good Free-Response Questions
1	<p>Unless critically required for a question, the stem should be complete with a closed question line. Cognitively, closed question lines are clearer in setting expectations with the student than open question lines. With the first example, the student is unclear what answer is expected and can write any use of plants making multiple correct answers possible whereas the second example clearly defines the expectations making it a better question.</p> <p>In fill in the blanks, the length of the blank should not give an unnecessary hint as to what the correct answer could be.</p>	<p><i>Plants are useful to humans because _____.</i></p>	<p><i>State one use of plants to humans in medicine.</i></p>
2	<p>Avoid double negatives in the stem. This adds to the cognitive load in an undesirable manner and may lead to unnecessary confusion which is not the aim of an assessment question.</p>	<p><i>Identify a process that will not take place in plants if light is not present.</i></p>	<p><i>Identify a process in plants that occurs only in the presence of light.</i></p>
3	<p>The language of the question should be clear, concise as well as grade-appropriate.</p>	<p><i>For grade 4: Nikita is relocating to a city in Maharashtra from her native village. What kind of features can she expect in her new city house?</i></p>	<p><i>For grade 4: Nikita is shifting to a city in Maharashtra from her native village. What kind of features can she expect in her new city house?</i></p>
4	<p>While the question itself can be long and be text-heavy, using long sentences in the question should be avoided which may</p>	<p><i>250 mL of a gas at 27 °C in a vessel fitted with a movable piston</i></p>	<p><i>A gas at 27 °C occupies a volume of 250 mL. The temperature</i></p>

Sr. No.	Practice	Example of a poor Free-Response Questions	Example of a good Free-Response Questions
4	unnecessarily add to the reading load for the student.	<i>is heated to 87 °C at constant pressure. What is the volume of the gas at 87 °C?</i>	<i>of the gas is raised to 87 °C keeping the pressure constant. Find the volume of the gas at 87 °C.</i>
5	The words/terminology used should not give unfair advantage to any specific groups of students. For example, a question on idlis (a savoury rice cake that is a popular breakfast food in southern India) and the way they are cooked may give a benefit to students who are used to eating it regularly.	<i>Idlis are a popular breakfast which is made using a cooker. Which technique of cooking is used to prepare idlis?</i>	<i>Rice is cooked using a cooker in many households. What is the technique used to cook rice?</i>
6	No sensitive or controversial info to be used in the question. It should be very neutral and should not use the names of any country, minister, government. For example – Avoid the use of specific countries, especially in a certain stereotyped connotation. Instead, write Country X and Y (except when absolutely necessary like in Geography)	<i>In February 2022, [REDACTED] invaded [REDACTED] to take over the country by force.</i> <i>State TWO likely impacts on [REDACTED] due to this act by [REDACTED].</i>	<i>Country X has recently passed an open-door policy for refugees from conflict-stricken neighbouring countries.</i> <i>What will be TWO likely impacts of this policy on country X?</i>

The second part of a Free-Response Question that requires equal attention while framing is the rubric or the scoring guide. This helps the assessor understand the marks allotted to each aspect of the expected response. So, it becomes important to clearly state all instructions required for scoring in the rubric for smooth evaluation of responses.

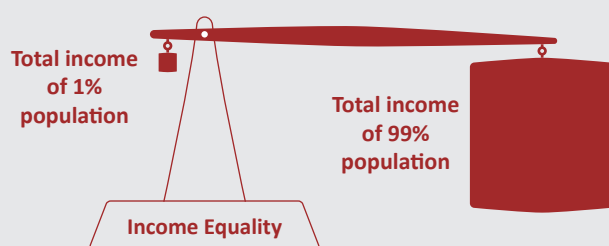
4. Instructions in the marking scheme should be clear so that there is no subjectivity involved in its interpretation.

The purpose of the marking scheme is to support evaluation in an objective manner. For it to achieve this objective, it should be clear and objectively defined to as much extent as possible.

Here are a couple of examples of good clear marking schemes informing what to look for in the answers and how many marks to give.

Sample question 15

Look at the image given below about the income distribution of a country X, and answer the questions that follow.



- In two points, describe what the image is depicting.
- The calculated per capita income of this country is high. Does this accurately reflect the income of an average individual in the country? Give reasons to support your answer.
- What is the main demerit of using average income as an indicator of development?

[2+2+1 marks]

Marking scheme:

- Award 1 mark to each of the following:
 - Income inequality
 - The total income of the richest 1% is equal to the total income of the rest of the country
- Award 1 mark to each of the following:
 - No
 - Any one of the following explanations, or any other relevant explanation:
 - The calculated per capita income will be higher than that of an average individual.
 - The average is influenced by the outliers (the rich 1%)
- Hides disparities/inequalities [1 mark]

Sample question 16

State true or false and justify your answer.

- It is not possible to find the largest negative integer.
- 0 is a positive integer.

[1.5 marks]

Marking scheme:

- False [0.5 marks]
 - (-1) is the largest negative integer [1 mark]
- False [0.5 marks]
 - 0 is neither positive nor negative [1 mark]

As shown above, the marking scheme of sample question 15 describes answers for each of the sub-question and also indicates marks to give if the expected answer is found in the student responses. As seen in the 15 (b) sub-question, the marking scheme provides possible reasons as well as provision for other possible reasons in case student has thought independently of another reason. This ensures that students are not penalised for thinking afresh and are credited for any possible correct answer. Marking scheme of sample question 16 indicates that even for a straight-forward question as true/false with justification, mark distribution between each of the sub-questions is clearly indicated. This ensures that different evaluators don't give different marks for stating true/false and its justification.

5. The marks allotted should justify the cognitive process involved in answering the question.

In assessments, especially those that are high stakes, it is important for the Question-writer to ensure that the student spends as much time in thinking and writing a response to a question as is proportionate to the marks that will be awarded for that question. It is unfair to expect a student to write an answer that involves thinking and integrating multiple steps or a long answer for a lesser number of marks. Similarly, it is also not appropriate to award high marks for a very short answer. So, student reasoning and the cognitive process used in formulating the response to a question should be proportional to the marks awarded for that question.

In certain questions where a higher-order cognitive process is involved, due to the open-ended nature of questions asked, it becomes even more critical for the Question writer to consider different possible answers and award marks accordingly. Given below are two examples of such questions.

Sample question 17

She was one of those pretty, young ladies, born as if through an error of destiny, into a family of clerks.

Examine the above line from 'The Necklace' and state what it reveals about the relationship between appearances and class. Give two examples from the story to support your answer. Answer in 70–90 words.

Marking scheme:

- *States what the given line reveals about the relationship between appearances and class [1 mark]*
- *Gives two examples from the story to support the answer [1 mark]*

[Accept any other valid response supported by the text]

Sample question 18

“You don’t understand who or what I am!” he shouted. “Very well — I’ll show you.”

Why did the scientist say the above line in the story ‘Footprints without Feet’? Substantiate your answer in 70–90 words.

Marking scheme:

Content

- *States the reasons for the scientist’s outburst [2 marks]*

Organisation

- *Presents a logical connection between the ideas [1 mark]*

[Award the full allotted marks if the criteria have been met.

Award half of the allotted marks if the criteria need improvement.

Award 0 marks if the criteria have not been met]

Language Mechanics

- *Uses correct grammar, spelling, and other language mechanics [1 mark]*

[Award the full allotted marks if there are two minor errors or one major error.]

Here both questions require students to make inferences and substantiate those inferences with appropriate examples. As per the rubric of sample question 17, marks are awarded only for making the inference and providing two examples. However, when the response is being constructed by the students from scratch, in English, it becomes imperative to pay attention to the organisation and language of the response and award separate marks for it, as is done in sample question 18.

Based on the word count and the cognitive process required to answer these questions, they should ideally be awarded 2 marks simply for making the correct inference and substantiating it. Separate marks could have been awarded for organising the answer appropriately and for grammar and sentence structure (if that is equally important) in sample question 18. Awarding only 2 marks for all of these may not correspond to the time a student will be required to spend in answering such a question.

6. The marking scheme/rubric should state the possible complete answer as well as break-up of marks for the expected parts in the answer

The marking scheme/rubric should state the possible complete answer along with a breakup of marks for each aspect of the complete answer. Alternatively, it should at least in points list out what to expect in the students' answers.

Sample question 19

Jolita wants to find the density of a brine solution that she has prepared. She uses a density bottle and makes the following measurements.

Mass of empty density bottle = 45 g

Mass of density bottle when completely filled with water = 105 g

Mass of density bottle when completely filled with brine solution = 117 g

Help Jolita calculate the density of the brine solution. [3 marks]

Sample marking scheme 1:

Mass of empty density bottle = 45 g

Mass of brine solution = $117 - 45 = 72$ g

Mass of water = $105 - 45 = 60$ g

Sample question 20

How would you describe Toto's relationship with the other animals in the narrator's house? Support your answer with two examples from the story 'The Adventures of Toto'. Answer in 70–90 words.

Sample rubric 1:

Toto did not have friendly relations with the other animals in the narrator's house. He was never at peace with his animal companions. [1 mark]

When Toto was put in a cage and housed in the servants' quarters, he did not let the other pets sleep peacefully. Moreover, when Toto is finally accepted by the narrator's grandmother and housed in a stable with Nana, the family donkey, he does not let Nana live in peace but bit her ears with his teeth. [1 mark]

Volume of density bottle = Mass of water/density of water = $60/1 = 60 \text{ cm}^3$

Density of brine solution = mass of brine solution/volume of brine solution = $72/60 = 1.2 \text{ g/cm}^3$

Sample marking scheme 2:

Possible complete answer:

Mass of empty density bottle = 45 g

Mass of brine solution = $117 - 45 = 72 \text{ g}$

Mass of water = $105 - 45 = 60 \text{ g}$

Volume of density bottle = mass of water/density of water = $60/1 = 60 \text{ cm}^3$

Density of brine solution = mass of brine solution/volume of brine solution = $72/60 = 1.2 \text{ g/cm}^3$

- 0.5 marks each for calculating mass of water and brine solution
- 1 mark for calculating volume
- 1 mark for calculating density

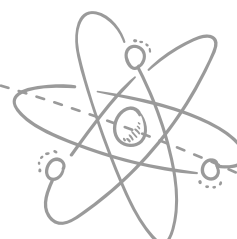
Sample marking scheme 2:

Sample Answer: Toto did not have friendly relations with the other animals in the narrator's house. He was never at peace with his animal companions. When Toto was put in a cage and housed in the servants' quarters, he did not let the other pets sleep peacefully. Moreover, when Toto was finally accepted by the narrator's grandmother and housed in a stable with Nana, the family donkey, he did not let Nana live in peace but bit her ears with his teeth.

Marks break-up:

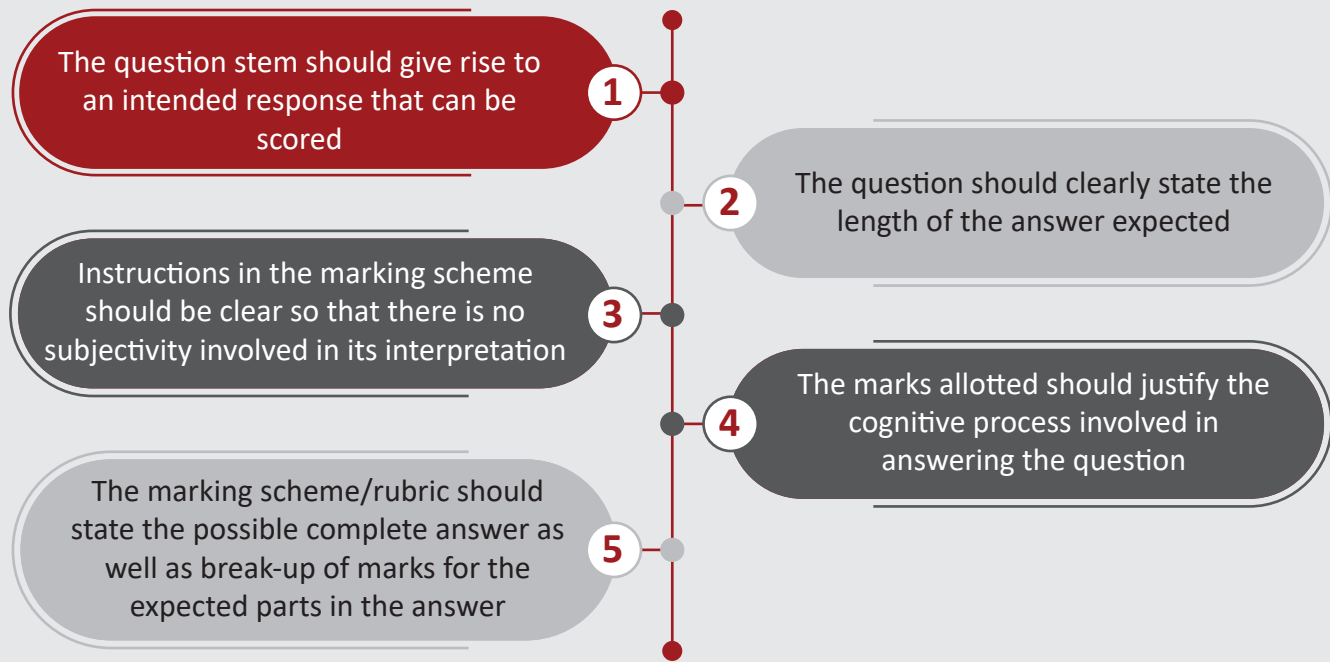
<p>Content</p> <p>♦ Describes Toto's relationship with other animals [1 mark]</p> <p>(did not have friendly relations with the other animals in the narrator's house/was never at peace with his animal companions)</p> <p>♦ States any two suitable examples [1 mark]</p> <p>(When Toto was put in a cage and housed in the servants' quarters, he did not let the other pets sleep peacefully/when Toto was finally accepted by the narrator's grandmother and housed in a stable with Nana, the family donkey, he did not let Nana live in peace but bit her ears with his teeth)</p> <p>Accept any other valid response supported by the text.</p>	2
<p>Organisation</p> <p>♦ Presents a logical connection between the ideas</p> <p>Award the full allotted marks if the criteria has been met. Award half of the allotted marks if the criteria needs improvement. Award 0 marks if the criteria has not been met.</p>	1
<p>Language Mechanics</p> <p>♦ Uses correct grammar, spelling and other language mechanics Award the full allotted marks if there are two minor errors or one major error. Award half of the allotted marks if there are three to four minor errors or two major errors. Award 0 marks if there are more than four minor errors or more than two major errors.</p>	1

As shown in both the sample questions 19 and 20 above, sample marking scheme 1, the possible correct answer is mentioned but the marks distribution is not. So, it is unclear to the assessor what marks should be deducted if a step has been missed. However, sample marking scheme 2 shows how the incorrect rubric can be made better by specifying the breakup of marks to be awarded. Also, try mentioning all possible acceptable alternate answers or forms of answers. One can also mention '[Accept any other valid answer]' so that the assessor can accept any correct answer that is not mentioned in the rubric.



Key Takeaways:

In summary, these are the main points to keep in mind when framing a good Subjective question and its Marking scheme:



E. Best Practices for Evaluation of Student Responses on Free-Response Questions

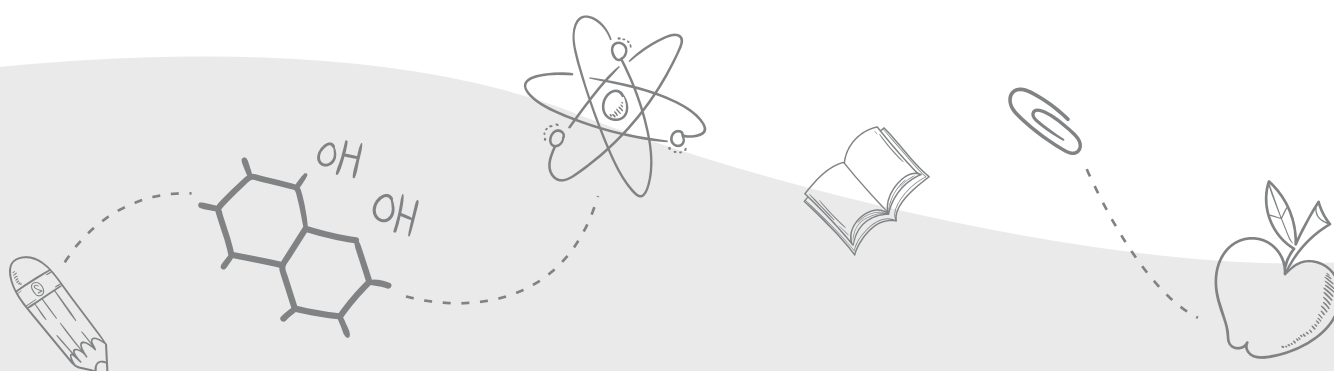
In any educational system, the assessment of students' knowledge and understanding is an essential component of the learning process. While objective questions such as multiple choice or true/false questions are easier to grade, subjective questions that require students to think critically and express their thoughts in their own words are also crucial components of evaluation. Many subjective questions may require students to demonstrate a deeper level of understanding and comprehension of the material being covered, and as such, they are an integral part of a well-rounded educational assessment.

However, grading subjective questions can be challenging as it is often difficult to establish a clear and objective standard for evaluating responses. Unlike objective questions, which have a single correct answer, subjective questions may have multiple correct responses, or students may provide responses that are partially correct. As a result, having an effective marking scheme for subjective questions and a robust process to ensure that the evaluation of students' responses using those marking schemes is of high rigour is critical to ensure that evaluations are consistent and fair.

Here are some recommended best practices around evaluation of student responses that can help in increasing the quality of evaluations and the reliability of the scores on subjective questions.

1. Using a variety of pre-selected responses for effectively training and monitoring the evaluators during the evaluation process.

In order to achieve standardisation across different evaluations, before the evaluators (scorers/examiners) start marking the answer scripts, experienced senior examiners are required to get together to mark a sample set of answer scripts. In the process, they are expected to curate a range of student responses, which can be used to further train the evaluators. The training helps the evaluators know how different responses are to be coded and it also helps the senior examiners in monitoring the variation in the scoring done by the different evaluators.



Shown below are two such processes followed by other international examinations.

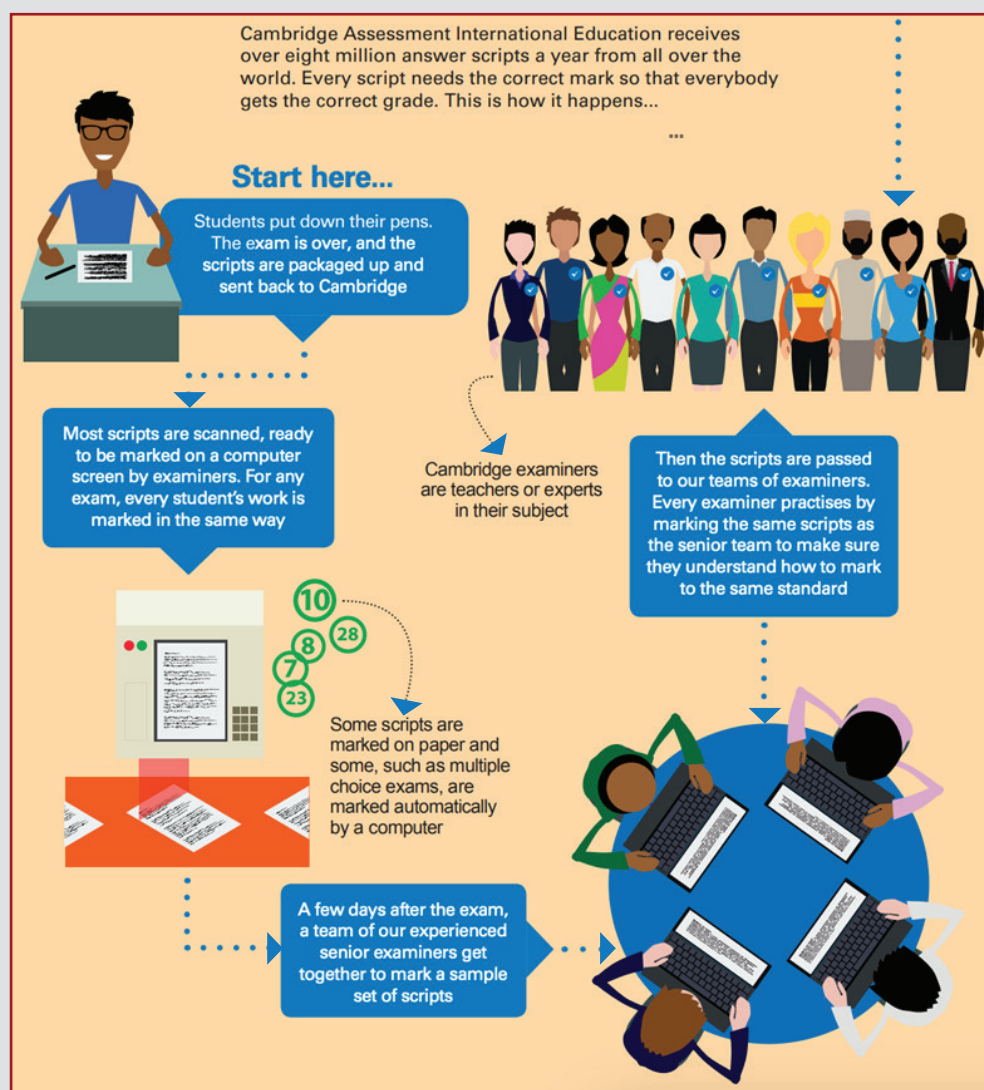


Exhibit: Marking of Sample Scripts by Senior Examiners at Cambridge

(Source: <https://www.cambridgeinternational.org/Images/321435-marking-and-grading.pdf>)

Standardization is not a single meeting but a period of time that has several aims:

- for the principal examiner to set the standard for the assessment
- to test and refine the markscheme
- to produce definitively marked responses (practice, qualification, seeds)
- to share understanding with all examiners to confirm examiners' understanding of the standard.

It is important to understand the different purposes of the three types of definitively marked responses. The purpose of practice responses is to support examiners in learning the marking standard, while qualification responses are intended to demonstrate/prove examiners can mark to the correct standard. Seed responses are used to demonstrate/prove examiners are continuing to mark to the correct standard.

The responses that are selected should be of a clear standard as their purpose is to check whether examiners are applying the markscheme/criteria appropriately and marking to the principal examiner's standard, not to "catch them out". It is important that the responses selected contain a range of responses and a range of marks in order to check that examiners have a wide and deep understanding of the markscheme/criteria, which is essential in order to recreate the principal examiner's marking standard.

The standard-setting stage has been successful if examiners have a good understanding of the standard and the IB has a set of definitively marked responses (practice, qualification, seeds) to support examiners in learning and demonstrating they have mastered it.

(Source:

<https://www.ibo.org/contentassets/8ecbee9f96af4db98293b97c133516f5/chief-examiners-and-principal-examiners-amended.pdf>)

2. Limiting the number of questions that an evaluator has to evaluate in any given answer script to increase the effectiveness of marking on any given question/set of questions and also reduce the bias arising due to evaluator-specific errors.

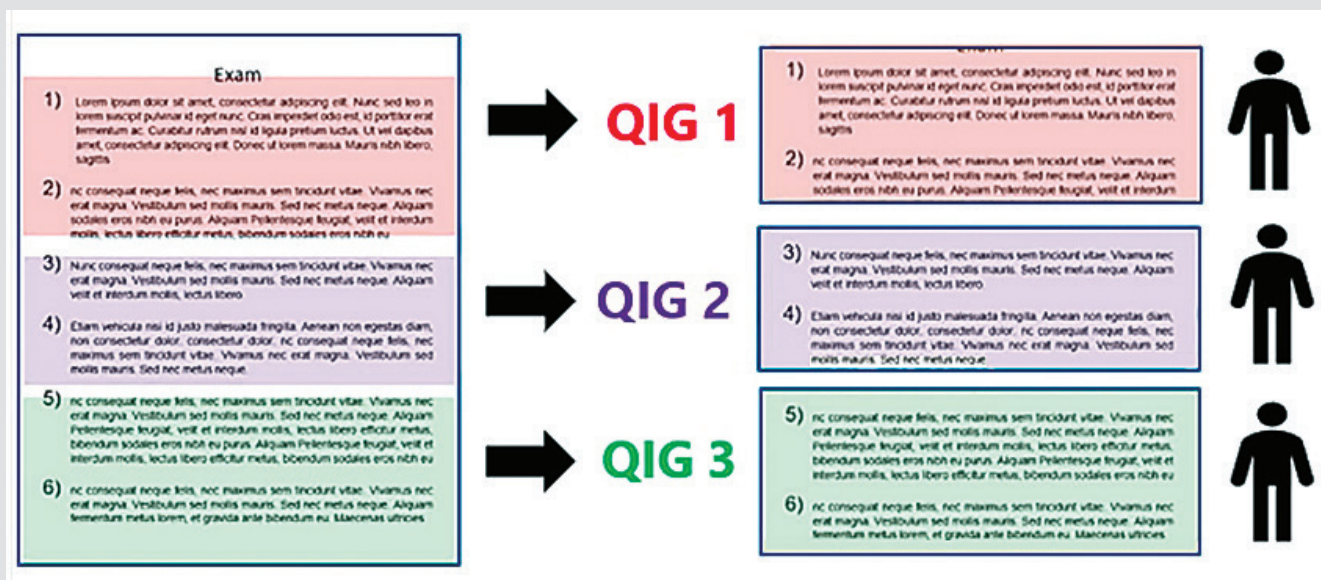
Scoring multiple Free-Response Questions back to back would mean switching from one question to another and in some cases, from one subject domain to another. Especially when multiple answer scripts are to be evaluated back-to-back, it could affect the effectiveness of the evaluator and impact the scores given. In some cases, it may also lead to the evaluator getting biased either positively or negatively by responses on a few questions initially in the paper which may impact his/her judgement on the following responses in the paper. One of the approaches to mitigate this could be to limit the number of questions that an evaluator has to evaluate in any given answer script. This can ensure that the evaluators are able to focus on fewer questions and their expected responses, thereby improving the consistency in the scores for that question. The limited set of questions that the evaluator has to evaluate can be determined either by the ease of separation of the questions in an answer script or even by topic in some subjects as there could be evaluators with specific domain expertise.

Shown below are two such processes followed by other international examinations.

When there are multiple subjective tasks on the test, the scoring system should include specifications for how the responses from different tasks are assigned. The contribution of any single rater to a test taker's final scores should be considered when designing the scoring process. For example, if there are many subjective tasks on a given assessment, limits might be placed on how many responses a rater can be assigned from a given test taker. This will prevent a test taker's overall test score from being affected by any systematic errors common to the ratings specific to a rater. Moreover, this will also prevent halo effects (i.e., a rater's decision on a test taker's response being affected by that rater's score on other responses from that same test taker). This may yield highly correlated ratings that possibly reduce the ability to distinguish the strengths and weaknesses of a particular test taker (Thorndike, 1920). Online scoring platforms, which are widely used in large-scale assessments, generally are designed to follow such rules when assigning responses to raters.

(Source: Best Practices for Constructed-Response Scoring – ETS Research Report Series |

<https://onlinelibrary.wiley.com/doi/full/10.1002/ets2.12358>)



International Baccalaureate (IB) forms Question item Groups (QIGs) during scoring. A QIG can be a question, section or topic. Examiners mark these smaller sections as they do with whole exam responses (that is, completing practice and qualification before going on to live responses with seeds). By splitting exam responses into QIGs examiners can mark those topics in which they are specialists and focus on marking one question at a time, thereby increasing the reliability of marking.

Exhibit: Question Item Groups (QIGs)

(Source: <https://blogs.ibo.org/2020/10/28/the-pros-and-cons-of-marking-exams-one-question-at-a-time/>)

3. Assigning a sample of answer scripts to multiple evaluators to check the variation in scores given, thereby checking and taking steps to ensure that the reliability of scoring is high.

A common practice to establish that the scores given are reliable is to check how much the scores would change if a different evaluator were to evaluate the responses. This may not be needed for all the answer scripts. If done systematically, a selected representative sample of the larger number of answer scripts should suffice.

Sharing below two quotes from other international assessments around their practices.

Hong Kong: Answer scripts can be readily allocated to different markers who can be assigned to mark the answers to a particular question. They can also be sent simultaneously to two markers for double marking, allowing for greater objectivity and avoiding the need to circulate scripts among markers, check-markers and HKEAA staff.

TIMSS: The method for establishing the reliability of the scoring within each country was for two independent scorers to score a random sample of 200 responses for each Free-Response Question. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. In collecting the within-country reliability data, it was vital that the scorers independently scored the questions assigned to them, and each scorer did not have prior knowledge of the scores assigned by the other scorer. The within-country reliability scoring was integrated within the main scoring procedure and ongoing throughout the scoring process.

(Source: *Methods & Procedures: TIMSS 2019 Technical Report, Chapter 6: Survey Operations Procedures* / https://timssandpirls.bc.edu/timss2019/methods/pdf/T19_MP_Ch6-survey-operations-procedures.pdf)

4. Using seed answer scripts to monitor and increase the quality of live evaluations.

While the evaluators would have been screened for their ability to evaluate student responses and also trained on the specific set of questions and their marking schemes, it is important to monitor their quality of evaluation and give them feedback around their quality of evaluations. One of the approaches for this could be to use seed answer scripts. These are answer scripts which have been pre-evaluated by the senior examiner, with scores available for each Question. The evaluator gets these answer scripts just like any other answer script within their lot of answer scripts. Their scores on the seed answer script can be compared with that of the senior examiner and variations if any can be pointed out.

Sharing below a relevant practice from International Baccalaureate (IB).

Once exams are taken by candidates, their responses are either sent to scanning centres or uploaded. They are then made accessible to examiners through an online marking tool that randomly allocates responses. The responses are anonymous, which prevents examiner bias.

The process for examiners has three stages: practice, qualification and live marking. Live marking is monitored using seeds.

The practice stage allows examiners to look at responses that were already marked by the principal examiner (definitively marked) to learn how to apply the markscheme. They receive automatic feedback to clarify their understanding of how marking should be carried out.

The qualification process requires examiners to demonstrate they can apply the markscheme appropriately by testing them with a selection of responses selected by the principal examiner. An examiner who cannot show that he or she can apply the mark scheme appropriately will not be allowed to start marking candidate work.

Once an examiner has qualified to start marking, his or her marking is monitored throughout by seeds being introduced randomly into their allocation. A seed is a response that has already been definitively marked by the principal examiner. Examiners are unaware that they are marking a seed, which ensures the seeds are marked in the same manner as the other responses in their allocation.

To help determine that the examiners are marking within the standard set by the principal and the standardization team, each subject is assigned a tolerance value. An examiner's marking should be within this tolerance. A tolerance reflects the legitimate differences in the marks awarded by different examiners to the same piece of work. Think about two teachers in your school marking a piece of work; both agree it is good, but one could award 29 and the other 30.

(Source: IB assessment principles and practices: A guide to assessment for teachers and coordinators | <https://www.ibo.org/contentassets/4d92e48d38a4415a87e11555e143a39f/assessment-guide-for-teachers-and-coordinators-en.pdf>)

5. Monitoring the evaluators' work in real-time where senior examiners and assistant examiners can give live feedback on the quality of evaluations .

Evaluators should be monitored in real or near-real-time during operational scoring sessions so that timely feedback can be provided to them to ensure continued scoring accuracy. This can be done using different approaches:

- (i) Using Seeds and Tolerance Levels: Comparing raters' scores to embedded pre-scored sample responses (seeds) to estimate scoring accuracy, also known as validity response monitoring and measure the deviation from senior examiner's scores.
- (ii) Backrater: Using expert raters, such as scoring leaders, who score responses and compare their score to those assigned by a rater, known as backrating.
- (iii) Compiling individual evaluator performance data: Individual evaluator performance data should be compiled and evaluated over time to better understand his/her overall performance level.

Cumulative performance data should be used to determine if a rater's performance levels do not meet minimum standards. Such raters can then be appropriately remediated or removed from scoring or can be offered opportunities for additional training that targets their deficiencies.

Here are some of these approaches from international assessments:

IB: *To help determine that the examiners are marking within the standard set by the principal and the standardization team, each subject is assigned a tolerance value. An examiner's marking should be within this tolerance. A tolerance reflects the legitimate differences in the marks awarded by different examiners to the same piece of work. Think about two teachers in your school marking a piece of work; both agree it is good, but one could award 29 and the other 30.*

(Source: IB assessment principles and practices: A guide to assessment for teachers and coordinators |

<https://www.ibo.org/contentassets/4d92e48d38a4415a87e11555e143a39f/assessment-guide-for-teachers-and-coordinators-en.pdf>)

Hong Kong - HKDSE: *Control rooms are set up at the assessment centres where the Chief Examiners and the Assistant Examiners monitor the marking progress, using such statistics as the marking speed of the different panels and the number of scripts marked by individual markers. Markers who stray off standard can immediately be advised where they are being too lenient, or too strict. We can tell them if they have missed something or if they have made a mistake. The markers themselves can then correct and adjust their work.*

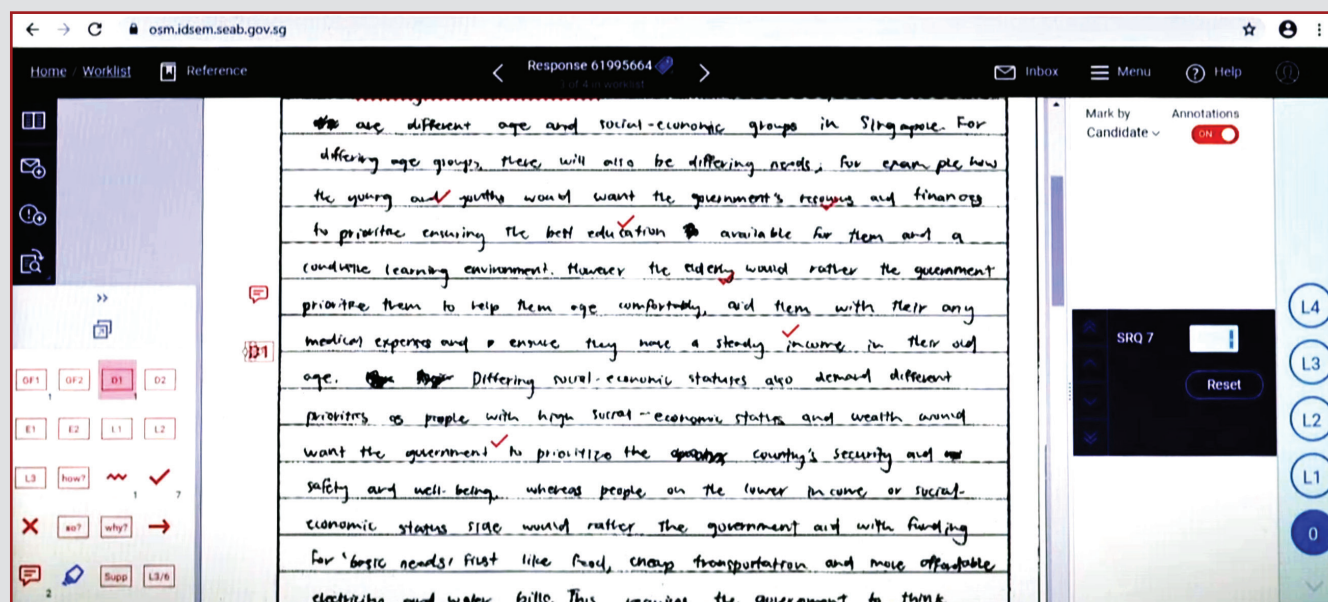
Cambridge Assessment International Education: *Marks on scripts and portfolios of evidence: Senior examiners check the marking of every examiner to make sure that their marking is consistent across all scripts and in line with the mark scheme.*

Digital storing and evaluation of student answers offer multiple advantages over the traditional scoring on paper. On the logistical front it helps in avoiding physical transfer of answer sheets between centres/evaluators. It also enables secure, easy access from any location. It also enables real-time monitoring of the progress of evaluation and catch any delays/inaccuracies. On the quality of the evaluation, it can help in easy, random allotment to independent evaluators thereby allowing for inter-evaluator reliability measurement. Any errors arising due to manual entry of marks can also be avoided. On the data availability front, it can allow easy processing of the data and post-examination analysis of data. The Hong Kong Examinations and Assessment Authority uses such an ‘online marking system’ to evaluate student answers. A similar system is also used by Finland for its school-leaving examination.

Finland Student Examination Board:

Guidelines for Making and Scoring of Free-Response Questions

Ministry of Education, Singapore: The onscreen marking initiative by the Singapore Examinations and Assessment Board (SEAB) has helped to increase marking quality and efficiency. Unlike traditional pen-based marking, candidates' hardcopy answer scripts are now digitised after the examinations and marked on a computer screen. Scripts no longer have to be moved and counted physically, enhancing security on script handling and reducing manual labour.



7. Using statistical tools to calculate consistency between evaluators and having internal records of such metrics to measure and track year-on-year improvement in the overall evaluation process

Naturally, the goal is to have 100% or perfect agreement among evaluators. An agreement above 85% is considered good and agreement above 70% is considered acceptable. Percentages of agreement below 70% are a cause for concern. (Source: IEA)

There are multiple ways to calculate consistency between evaluators (e.g., percent agreement, kappa, intraclass correlation), but all methods rely on a sample of multiple (i.e., at least two) ratings per response. Trend reliability scoring technique measures to ensure reliability from one assessment cycle to the next.

If internally such methods are deployed and metrics are recorded and made available, it can help measure evaluation practices year-on-year and provide feedback around where and how much to improve.

Key Takeaways:

In summary, these are some of the best practices for evaluation of student responses on subjective questions:

- 1 Using a variety of pre-selected responses for effectively training and monitoring the evaluators during the evaluation process
- 2 Limiting the number of responses from any answer script that can be evaluated by an evaluator to increase the effectiveness of marking on any given Question/set of questions and also reduce the bias arising due to evaluator-specific errors
- 3 Assigning a sample of answer scripts to multiple evaluators to check the variation in scores given, thereby checking and taking steps to ensure that the reliability of scoring is high
- 4 Using seed answer scripts to monitor and increase the quality of live evaluations
- 5 Monitoring the evaluators' work in real-time where senior examiners and assistant examiners can give live feedback on the quality of evaluations
- 6 Moving to a seamless onscreen marking process and implementing it through the duration of the evaluation process where all key steps like scanning answer scripts, randomised question allotment to evaluators, scoring, monitoring and scrutiny are executed, documented and analysed
- 7 Using statistical tools to calculate consistency between evaluators and having internal records of such metrics to measure and track year-on-year improvement in the overall evaluation process

F. Do's and Don'ts for the Answer Script Evaluators

Subjective questions are an essential part of assessments and evaluations. Unlike objective questions, which typically have a single correct answer, subjective questions require students to provide a written response, often requiring critical thinking and analysis. However, grading subjective questions can be a challenging task, as it requires examiners to evaluate a wide range of responses and perspectives. Here are some dos and don'ts for examiners evaluating subjective questions.

Do's

1

Solve the question paper independently and review the marking scheme for each question:

Before grading a subjective question, it is important to be familiar with the question and what it is expecting. It is also important to review the marking scheme provided. This will help understand the expected answer format, the distribution of marks and the key points to consider while grading the answer. It is essential to understand the marking scheme as it sets a standard for grading each response, ensuring that grading is consistent across all examiners.

2

Read each answer carefully:

It is essential to read each answer carefully to ensure that one understands the student's perspective and the points they are trying to make. Sometimes, students may not present their answers in a structured manner, which can make it difficult to understand their responses. Therefore, it is necessary to read the answer multiple times to comprehend the essence of the answer.

3

Give equal attention to all parts of the answer and not just the highlighted or underlined keywords:

Subjective questions often have multiple parts, and it is necessary to give equal attention to each part of the answer while grading. Students may provide accurate responses to some parts of the question while missing out on others. Or in some cases may just highlight parts of their answers in some way, like underlining keywords. It is possible that the description provided along with the highlighted parts isn't accurate. Therefore, it is essential to read the entire response and award marks for each part of the answer based on the marking scheme.

4

Give partial marks for different aspects of the answer as indicated in the marking scheme:

Subjective questions may not always have a single correct answer, and students may provide partially correct responses. Also, the marking scheme may indicate how many marks to give for each of the sub-answers or steps in the answer. In such cases, it is essential to award partial marks for the parts of the answer that are correct as indicated in the marking scheme. This encourages students to continue attempting the question even if they are unsure of the complete answer. For this it is assumed that the marking scheme has provision for partial marking.

5

Ensure credit is given for innovative or insightful responses not always stated in the manner as given in the marking scheme:

Subjective questions can provide students with the opportunity to showcase their creativity and originality. Students may often write the answers in their own words as well. If students present such responses, examiners should award marks as far as it is along the lines of the expected responses or demonstrates the understanding of the required concept. This also motivates students to think out of the box and provides them with the incentive to present unique and original responses.

6

Consider the context of the question and the knowledge level of the student:

It is necessary to consider the context of the question and the level of knowledge expected of the student while grading. Students may come from different backgrounds and may have varying levels of knowledge. Therefore, it is essential to take into account their context and level of knowledge while grading their answers.

7

Provide constructive feedback:

Examiners should provide constructive feedback on the areas where the student can improve. This feedback should focus on the weak areas of the answer, along with suggestions on how the student can improve their response. This feedback will not only help the student improve their performance in the future but also help them understand the reasoning behind the grading.

When it comes to evaluating subjective questions, examiners must be careful not to fall into common pitfalls that can lead to inaccurate or biased grading. Here are some important "don'ts" for examiners to keep in mind.

Don'ts

1

Don't grade based on handwriting or presentation or length of the answer:

It can be tempting to be swayed by neat handwriting or a well-organized answer sheet, with elaborate, long answers, but these factors should not affect the marks awarded. Rather, the content of the answer should be the sole basis for evaluation.

2

. Don't give excessive marks for unsupported or incorrect answers:

It can be tempting to give marks for an answer that seems plausible, or that is based on a student's personal experience, but this can lead to inflated grades. Each answer must be supported by evidence, and incorrect assumptions or assertions should be penalized.

3

Don't be biased:

It is crucial that examiners remain impartial when evaluating answers. This means not being swayed by factors such as the answers on previous questions, or in cases where it is known - student's gender, ethnicity, or academic record. Each answer should be evaluated solely on its merits.

4

Don't make assumptions:

Examiners should avoid making assumptions about what a student meant to say, or what they might have written if they had more time. Each answer should be evaluated based solely on what is actually written.

5

Don't ignore the context of the question:

When evaluating an answer, it's important to keep in mind the context of the question. For example, if a question asks for a critical analysis of a specific text, then a student's personal opinions on the topic may not be relevant. Examiners should evaluate each answer in light of the question asked.

6

Don't penalize for unconventional or innovative answers:

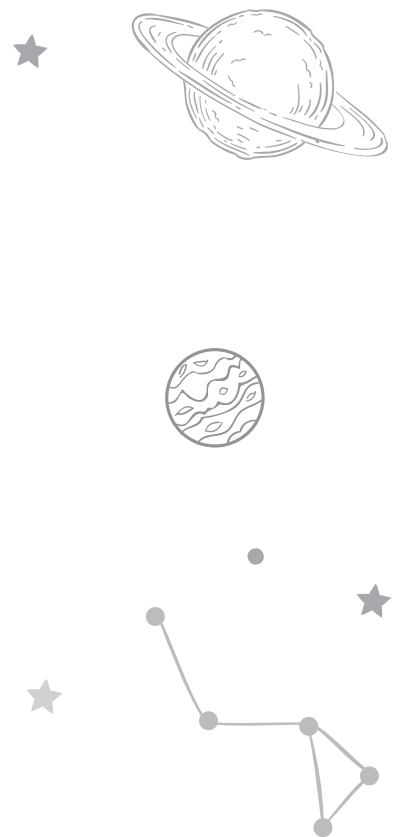
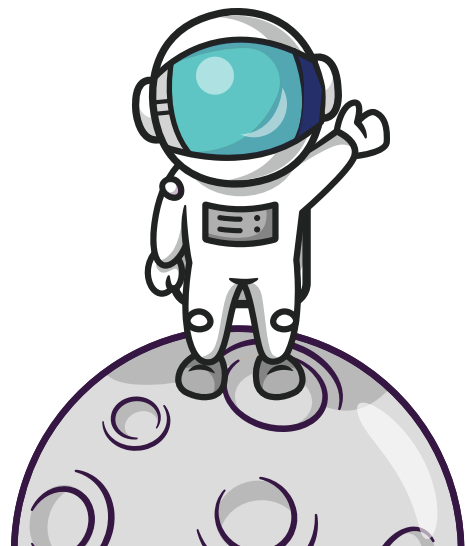
While examiners should avoid giving marks for unsupported or incorrect answers, they should also be open to innovative or unconventional responses. If a student provides an answer that is not the expected response but is still accurate and insightful, they should be given credit for it.

7

Don't hesitate to consult with other examiners or the marking scheme:

If an answer is particularly complex or difficult to evaluate, or if an examiner is unsure about how to grade a particular response, they should not hesitate to consult with other examiners or the marking scheme. It's better to take the time to get the evaluation right, rather than rushing to a decision that may be inaccurate or unfair.

In conclusion, evaluating subjective questions is a complex and challenging task, and it's important for examiners to avoid common pitfalls that can lead to inaccurate or biased grading. By keeping these "don'ts" in mind, examiners can ensure that their evaluations are fair, accurate, and helpful for students.



Acknowledgements

ADVISORY

- Smt. Nidhi Chhibber, IAS, Chairperson, CBSE

GUIDANCE AND SUPPORT

- Dr. Joseph Emmanuel, Director (Academics), CBSE
- Dr. Praggya M. Singh, Director, (Academics - Assessment), CBSE
- Dr. Sweta Singh, Joint Secretary (Academics - Assessment), CBSE
- Mr. Sridhar Rajagopalan, Chief Learning Officer, Ei
- Mr. Nishchal Shukla, Vice President – Content and Pedagogy, Ei

PLANNING AND EXECUTION

- Mr. Ritesh Agarwal, Associate Vice President, Ei
- Mr. Varun Kappal, Program Director, Ei
- Ms. Manisha Upreti, Manager, Ei
- Mr. H.M Shahnawaz Khan, Associate Manager, Ei
- Mr. Muzaffar Ahmad, Education Specialist, Ei



Central Board of Secondary Education
Shiksha Sadan, 17, Rouse Avenue,
New Delhi-110002