Ministry of Education
Government of India

सत्यमेव जयते

विद्यया ऽमृतमश्नुते

एनसीईआरटी
NCERT

PARAKH

Guidelines for Setting Balanced Question Papers to Promote Equivalence in Assessment Practices of School Boards

Guidelines

for

SETTING BALANCED QUESTION PAPERS

to

PROMOTE EQUIVALENCE IN ASSESSMENT PRACTICES

of

SCHOOL BOARDS

Foreword

The *Guidelines for Setting Balanced Question Papers to Promote Equivalence in Assessment Practices of School Boards* reflects a commitment to establishing fairness and equity in assessments across school boards in India. These guidelines aim to address longstanding disparities in question paper design by providing a structured framework that promotes consistency, inclusivity, and competency-based assessment practices.

Aligned with the vision of the National Education Policy 2020, the document emphasizes a shift from rote learning to methods that encourage critical thinking, problem-solving, and conceptual clarity. It highlights the importance of well-balanced question papers that are not only fair but also capable of evaluating students' abilities holistically. By standardizing the processes involved in question paper creation, these guidelines aim to ensure that assessments become meaningful tools for learning and growth.

This document is designed as a practical resource for educators, administrators, and policymakers, offering detailed steps and recommendations for setting balanced and effective question papers. It is hoped that the adoption of these guidelines will lead to a more robust and equitable education system, fostering transformative learning experiences for all students.

Prof. Dinesh Prasad Saklani
Director, NCERT

Preface

The *Guidelines for Setting Balanced Question Papers to Promote Equivalence in Assessment Practices of School Boards* is a comprehensive resource designed to address the pressing need for standardized and equitable assessment practices across school boards in India. Rooted in the principles of the National Education Policy (NEP) 2020, this document serves as a pivotal step in transitioning from traditional assessment models to competency-based systems that prioritize critical thinking, problem-solving, and conceptual clarity over rote learning.

This guideline document outlines a meticulously designed seven-step process for developing balanced and effective question papers. Each step is integral to creating assessments that are fair, reliable, and aligned with curricular goals. The process begins with the preparation of an assessment design, which focuses on defining the cognitive objectives, content areas, types of questions, and difficulty levels. This ensures that assessments are aligned with the curriculum and cater to diverse learning needs. The next step involves the development of an assessment blueprint, which acts as a detailed map, specifying the placement of questions according to cognitive domains, content areas, and marks distribution. This step ensures coherence and consistency in the question paper.

The third step emphasizes question writing, which requires crafting questions that are clear, fair, and aligned with learning objectives. The guidelines stress the importance of clarity, unity, and economy in question design, alongside considerations for inclusivity and accessibility. This is followed by the assembly of the question paper, where questions are organized in a logical sequence to ensure a balance in content, difficulty, and format. The inclusion of clear instructions and section-wise structuring further enhances ease of administration.

Another critical aspect of the process is the development of a scoring key or marking scheme. To ensure objective and transparent assessment, the document provides detailed recommendations, including aligning marks with cognitive domains, recognizing valid alternative responses, and minimizing subjectivity in scoring. The sixth step, question paper analysis, involves a thorough review to identify any misalignments with the blueprint and design. Comprehensive checklists help ensure alignment with cognitive levels, content, and question quality. Finally, the process concludes with moderation, which ensures that the

question paper adheres to prescribed standards of fairness, reliability, and validity by addressing biases, accessibility, and competency-based assessment principles.

One of the standout features of this document is its emphasis on fairness and inclusivity. The guidelines provide practical recommendations for eliminating cognitive and affective barriers, ensuring gender neutrality, and creating assessments that cater to diverse learning needs. The use of diverse question types, such as multiple-choice questions, short answers, and essays, allows for comprehensive assessment while accommodating different learning styles. Stimulus-based questions, marking schemes, and question paper analysis are also highlighted to enhance the overall quality of assessments.

This document serves as a valuable resource for educators, administrators, and policymakers, providing actionable insights and tools to transform assessment practices. By aligning with the broader goals of NEP 2020, this document paves the way for a future-ready education system that prioritizes fairness, equity, and competency-based learning.

I extend my sincere gratitude to all the school boards for their collaboration with PARAKH in finalizing this document.

Prof. Indrani Bhaduri
CEO and Head, PARAKH, NCERT

Contents

1 Background

In July 2024, PARAKH released its first report on[1] the equivalence of school education boards. The report outlined a series of recommendations to bring greater parity among all recognized Indian educational boards across five broad categories: 1.) Administration, 2.) Curriculum, 3.) Assessment, 4.) Infrastructure, and 5.) Inclusiveness. These recommendations were informed by a detailed study of recent question papers set by various boards.

The section on Assessment examined Grade X and XII question papers from the previous academic year and identified several inconsistencies in assessment design, particularly in the following areas:

- **Inconsistent distribution of marks** relative to the cognitive demands of the content;
- **Uneven allocation of marks** according to the nature of the items; and
- **Lack of a systematic approach** to distributing marks across different difficulty levels.

Such disparities in assessment design present significant obstacles to achieving parity across boards and undermine equity in educational opportunities. A high-quality and robust assessment system is essential for ensuring educational equity. Recognizing this, NEP 2020 emphasized the need to establish equivalence across school education boards to raise assessment standards and foster fairness[2]. Restructuring assessment methodologies is also critical for reducing dependence on rote memorization and promoting critical thinking and the evaluation of core competencies.

Furthermore, establishing a greater level of equivalence among the question papers set by state boards will facilitate the ability of students to transfer between boards and schools will reduce disadvantages due to differing assessment systems and practices.

To address these challenges, PARAKH developed training sessions on designing balanced question papers and piloted the trainings with teachers and board members of the Himachal Pradesh Board of School Education (HPBoSE) during a two-day workshop at the HPBoSE campus in Dharamshala, Himachal Pradesh. Subsequently, a series of workshops on balanced question paper design were conducted at the PARAKH-NCERT campus in New Delhi. These sessions emphasized standardizing question paper development practices based on the principles of competency-based assessment.

[1] The Equivalence of School Education Boards Report
[2] National Education Policy 2020 [Section 4.41]

This guideline document will serve as a first step towards creating more balanced, consistent and higher quality assessments at the Board level. The recommendations are based on extensive interactions with workshop participants and the collective decade-long experience in education and test design, development, analysis, scaling, and reporting of the PARAKH team. Administrators, question writers, and educators should feel free to adapt the recommendations to their circumstances but should always keep in mind the goals of creating assessments that encourage student learning and growth for the challenges of the future.

2 Steps to Develop a Balanced Question Paper

In India, written exams are widely used and dominate the educational landscape. They are almost synonymous with evaluation, with all Educational Boards conducting annual and supplementary exams to assess students' achievement levels. Schools also hold their own terminal and annual exams. Given the importance of written assessments, it becomes essential to thoroughly examine the processes involved.

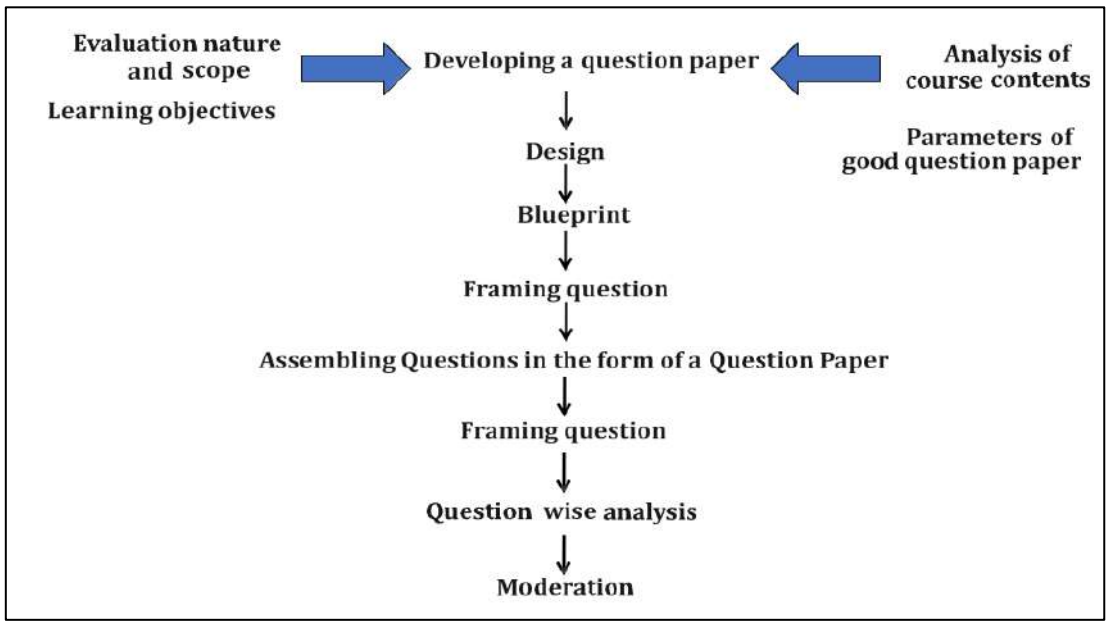
The question paper, a cornerstone of written exams, plays a crucial role in shaping learning outcomes. Its quality directly impacts student achievement and certification. Question papers also have strong influence on classroom teaching and pedagogy. Without an appropriate degree of standardization across time and student groups, variations in student performance across years and subjects will reflect both the quality of the question paper and genuine differences in student abilities, which limits the validity of claims based on question paper scores across time and student groups. For instance, a poorly designed or overly challenging paper may negatively affect overall achievement. Factors like content coverage, difficulty level, number of questions, types of questions, and the options provided must therefore be sufficiently balanced to ensure a fair assessment. Any imbalance in these areas can threaten the validity of claims based on assessment results. Creating a well-balanced question paper requires a systematic approach that incorporates design, blueprint creation, question formulation, marking schemes, and detailed analysis.

Once an examiner understands the characteristics of a good assessment, they can begin constructing it—whether it is a classroom unit test or a comprehensive question paper covering the entire syllabus. Regardless of the assessment type, key steps of assessment construction should include at minimum the following[3]:

1. Preparation of an Assessment Design
2. Preparation of an Assessment Blueprint.
3. Writing of Questions.
4. Assembling Questions in the form of a Question Paper.
5. Preparing Scoring Key/Marking Scheme.
6. Carrying out Question-wise Analysis.
7. Moderation of a Question Paper

[3] Please note, depending on the assessment purpose, the resources available, and the timelines for assessment development, inclusion of additional steps, such as small-scale pre-testing or large-scale piloting, may be advisable.

The chart below illustrates the process flow of a balanced question paper setting:



Note that different individuals or groups may be responsible for different steps in the process. Thus, clear documentation is essential.

2.1 Preparation of an Assessment Design (Step 1)

The first step in preparing a question paper is creating its design. A question paper is not merely a collection of random questions. It must be thoughtfully planned as an assessment, keeping in mind the objectives or cognitive domains, the course content, and the forms of questions required to assess these. This involves assigning weightage to various objectives/cognitive domains, content areas, and forms of questions while estimating the time required to answer each question, along with the scheme of sections and pattern of options. During the assessment design phase, the designers need to answer why they are planning to test, whom they are testing, what is being tested, what methods the test should include, what methods of quality control can be used during test design, and how much time, money, and other resources are available to design the test.

Test Design: Test design is the process of deciding how to meet the requirements established for a test within the unavoidable constraints that exist for the test.

The design serves as a framework of the question paper, much like an architectural map of a building. It determines how the completed question paper will appear and function. The design outlines essential components, such as weightage to different elements, to ensure alignment with the objectives of the test.

The figure below illustrates a design template for a question paper.

Format of Design (Subject other than language)

Question Paper/Test

Subject:	
Unit/Paper:	
Class:	
Time:	
Marks:	

Weightage to Objective/Cognitive Domains

Objective	Knowledge	Understand	Apply	Analyze	Evaluate	Create	Total
Percentage of Marks							
Marks							

Weightage to Major Content Areas:

S. No.	Unit/Sub-Units	Marks
1.		
2.		
3.		
4.		
5.		
6.		
Total		

Weightage to form of question:

Forms of Questions	E/LA	SA	VSA	O (MCQ)	Total
No. Of Questions					
Marks allotted					
Estimated time					

Estimated difficulty level:	Difficult	%marks
	Average	%marks
	Easy	%marks

Schemes of Sections	
Pattern of Options	

Index of Abbreviations:

(E/LA: Essay/Long Answer; SA: Short Answer; VSA: Very Short Answer; O: Objective)

The design document should clearly state the following five major specifications:

1. Weightage to objectives/cognitive domains
2. Weightage to major content areas
3. Weightage to types of questions
4. Weightage to different estimated difficulty levels
5. Scheme of sections and pattern of options

2.1.1 WEIGHTAGE TO OBJECTIVES/COGNITIVE DOMAINS:

To ensure the validity of a test, it is crucial to analyze the objectives or cognitive levels outlined in the course and determine which ones should be assessed and in what proportion. Marks are then assigned to each objective based on its significance, which is typically influenced by factors such as:

- The proportion of time spent covering that objective,
- The complexity and nature of the content in the course,
- The learning levels of the students

While assigning weightages to these objectives or cognitive levels, question paper setters must align their decisions with the vision and guidelines of their respective boards. However, while setting these weightages, a question paper setter must be well informed about these cognitive levels and how they are used in assessing the learning outcomes of the examinees. These cognitive levels most widely used are *Know, Understand, Apply, Analyze, Evaluate, and Create*. In the design, the weightages assigned to these cognitive levels should be specified in both percentages and actual marks.

Below is an example to showcasing, how it could be filled for non-language subjects e.g. mathematics test of 80 marks of a secondary student:

Objective	Know	Understand	Application	Analyse	Evaluation	Creation	Total
Percentage of Marks	17.5%	17.5%	35%	22%	4%	4%	100%
Marks	14	14	28	18	3	3	80

Similarly, in language subjects, the focus is typically on four major objectives: knowledge of language elements, comprehension, expression, and appreciation. For a 50-mark test, the weightages for these objectives can be distributed accordingly in percentages. Below is an example of a possible distribution:

Objective	Knowledge of language elements	Comprehension	Expression	Appreciation	Total
Percentage of Marks	10%	40%	40%	10%	100%
Marks	5	20	20	5	50

2.1.2 WEIGHTAGE TO MAJOR CONTENT AREAS

It is equally important to analyze the syllabus and assign weightages to different areas of content to ensure the test's validity. The weightage for each content area is determined by factors such as the amount of instructional time dedicated to it during the academic year.

To illustrate, consider a scenario where an academic year includes 100 instructional periods to complete 10 units or sub-units of a subject, which will have a final assessment of 80 marks Instead of distributing marks equally (e.g., 8 marks per unit), the weightages should reflect the proportion of time required to cover each unit/sub-unit out of total available time for course completion in an academic year. This ensures a fair representation of the content's relative importance. The weightage to unit/sub-unit is calculated using the following relation:

This implies that if unit/sub-unit 1 takes 5 periods to complete while unit 2 takes 15 periods to complete. Then,

Weightage of unit 1 should be, $= \frac{5}{100} \times 80 = 4$ Marks

Weightage percentage to the unit 1 $= \frac{4}{80} \times 100\% = 5\%$

And weightage of Unit 2 $= \frac{15}{100} \times 80 = 12$ Marks

Weightage percentage to the unit 2 $= \frac{12}{80} \times 100\% = 15\%$

In addition, the following factors should also be considered when deciding the weightage for different content areas:

- Units/sub-units that build foundational concepts for senior classes or competitive exams should be given greater weightage,
- Units or sub-units with more complex learning objectives could receive higher weightage if desirable,
- The density of content elements within each unit or sub-unit should also influence the assigned weightage.

2.1.3 WEIGHTAGE TO TYPES OF QUESTIONS

After analyzing the objectives and content, it is essential to determine the most suitable ways to assess them. Different objectives and content areas are better tested using specific forms of questions. Therefore, a variety of question types should be included in the test, along with a clear allocation of marks and estimated time for each type. This process enhances the reliability of the test.

A question paper setter may consider the following points, while deciding the weightage to different forms of questions:

- Within a particular group of the same type of question, the time and marks distribution can vary. For example, in a question paper, a MCQ of knowledge domain may carry 1 mark for instance, whereas a MCQ targeting higher-order cognitive skills (such as analyze or evaluate) could carry 2 or 3 marks.
- Similarly, questions of differing estimated difficulty levels within the same type may have varying marks. For example, a question judged to be easy may carry fewer marks than a question judged to be difficult even if they are of the same type.
- The time allocated to answer a question depends on its cognitive demand and estimated difficulty level.

2.2 Weightage to Different Estimated Difficulty Levels

When planning a question paper, it is important to include questions of varying difficulty levels to address the needs of all students and to differentiate between high and low achievers. The difficulty level of a question should align with the abilities and achievement levels of the target students.

The question paper setter may estimate the difficulty level of a question based on:

- The complexity of mental processes involved.
- The specific content area being assessed.
- The time required to answer it.
- Observed difficulty of similar questions on previous exams

Because of the difficulty of accurately predicting difficulty levels of test questions, preliminary estimates of difficulty are best used as a rough sorting device. Question writers and reviewers are advised to adopt a simple three-level scheme for estimated difficulty (Easy, Medium, and Hard) and use generous ranges rather than precise numbers for each level. The true difficulty level is determined only after the question is administered and analyzed.

2.2.1 SCHEME OF SECTIONS

The design of a question paper may also indicate the scheme of sections for the paper. For example, a question paper may consist of both Selected response questions and Constructed-response questions. Such a test may have two sections, one consisting of Selection-type questions like MCQs, Alternative choice questions (True-False), and Matching type questions. While the other consists of constructed-response questions like essay or long answer type, short answer and very short answer type questions.

Similarly, for non-language subjects, sections might be organized by content areas, such as Physics, Chemistry, and Biology for Science, or History, Geography, and Civics for Social Studies. In language subjects, sections could be based on skill areas, such as reading comprehension, writing tasks, and grammar.

2.2.2 PATTERN OF OPTIONS (QUESTION CHOICE)

The design should indicate the pattern of options i.e. whether there are any options between the questions in the question paper, if so, how many and what their nature is. From the academic point of view, there is no justification for providing options except to make the test actually or apparently simpler for the examinees. The best comparison of achievement of the examinees can be made only if all of them are put to the same test without any choices. Thus, test-takers should not be allowed to choose a subset of questions choices to attempt. For example, if a question paper allows students to attempt any 25 out of 40 questions, some objectives or content areas may go untested or receive limited responses, thus compromising the test's reliability.

However, if there is a mandate to provide some type of question choice, it is best to provide limited **internal choice** within selected questions (typically constructed response questions). For instance, a constructed response question on a history question paper might give the student a small choice of significant events of the same type and time period to write about.

When providing such a choice, the alternatives must:

- Be comparable in terms of the objectives being tested.
- Maintain consistency in question format and difficulty level.
- Cover similar content areas as far as possible.

The statistical performance of questions featuring internal choice should be monitored. Any significant score discrepancies or difficulties scoring these questions should be caused to reconsider the policy of offering choice.

2.2.3 FORMAT OF DESIGN (LANGUAGE SUBJECTS)

The question paper design is typically created by the Examining Board in consultation with subject experts, teachers, and examiners. Decisions on weightage to content units are made by a committee appointed by the Board administration. Once finalized, the completed design template is provided to paper setters for implementation.

Design templates for language and non-language subjects are provided below for reference:

Question Paper/Test

Class: _____

Subject: Unit/Paper:

Time: Marks:

Weightage to Objectives

Objective	Knowledge of language elements	Comprehension	Expression	Appreciation	Total
Percentage of Marks					
Marks					

Weightage to Major Content Areas:

S.No.	Unit/Sub-Units	Marks
1.		
2.		
3.		
4.		
5.		
6.		
Total		

Weightage to form of question:

Forms of Questions	E/LA	SA	VSA	O (MCQ)	Total
No. Of Questions					
Marks allotted					
Estimated time					

Estimated difficulty level:	Difficult	%Marks
	Average	%Marks
	Easy	%Marks

Schemes of Sections	
Pattern of Options	

Index of Abbreviations:
(E/LA: Essay/Long Answer; SA: Short Answer; VSA: Very Short Answer; O: Objective)

2.2.4 FORMAT OF DESIGN (SUBJECT OTHER THAN LANGUAGE)

Question Paper/Test

Class: _____

Subject: Unit/Paper:

Time: Marks:

Weightage to Objective/Cognitive Domains

Objective	Knowledge	Understand	Apply	Analyze	Evaluate	Create	Total
Percentage of Marks							
Marks							

Weightage to Major Content Areas:

S. No.	Unit/Sub-Units	Marks
1.		
2.		
3.		
4.		
5.		
6.		
Total		

Weightage to form of question:

Forms of Questions	E/LA	SA	VSA	O (MCQ)	Total
No. Of Questions					
Marks allotted					
Estimated time					

Estimated difficulty level:	Difficult	%Marks
	Average	%Marks
	Easy	%Marks

Schemes of Sections	
Pattern of Options	

Index of Abbreviations:

(E/LA: Essay/Long Answer; SA: Short Answer; VSA: Very Short Answer; O: Objective)

2.3 Preparation of an Assessment Blueprint (Step 2)

The preparation of a blueprint is a crucial step after finalizing the test design. A blueprint serves as a detailed plan that clearly maps out the placement of each question based on its cognitive level/objective, content area, form, sequence number, and allocated marks. A well-constructed blueprint helps bring the test design to life in practical terms. It goes beyond merely assigning weightage to objectives, content areas, and question types, as it specifies the exact relationship between these elements. This ensures that the test maker has a clear understanding of which question assesses which cognitive domain/objective, from which content unit, through which form of question, weightage it carries, and at what position it will appear in a question paper.

Creating a blueprint involves technical skill and subject-specific insight. Paper setters must have a clear understanding of the content hierarchy, the learning outcomes associated with each unit, and how these align with assessment objectives. This clarity enables them to decide on the most suitable types of questions for each unit and the cognitive domains/objectives these questions will assess.

A blueprint represents the overall structure of the question paper, including:

- Distribution of marks across cognitive levels/assessment objectives,
- Distribution of marks across content units,
- Types of questions and their weightage.
- A summary section.

A thoughtfully prepared blueprint not only ensures alignment with the prescribed weightages of test design but also serves as a comprehensive guide for creating a balanced and effective question paper.

2.3.1 STEPS TO DEVELOP A BLUEPRINT

The following is a recommended templet for a question-paper blueprint:

FORMAT OF BLUEPRINT (Subject other than language)

Subject	:				Class	:	
Unit/Paper:	:		Max Marks	:	Time	:	

S. No.	Objective →	Knowledge				Understanding				Application				Analysis				Evaluation				Creation				Total Row Wise
	Form of Questions																									
	Content Unit/ Sub Unit	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	
		4 d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	3 c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
	Sub Total																									Question Seq. (marks)
	Total Column →	2 b				b				b				b				b				b				1 a
	Wise	(Marks Total)				(Marks Total)				(Marks Total)				(Marks Total)				(Marks Total)				(Marks Total)				

Notes: Figures within brackets to indicate the marks allotted to a question and figures outside the brackets to sequence no. question in paper.
Denotes that marks have been combined to form one question.

Summary:	Essay (E)	No.		Marks		Pattern of Options
	Short Answer (SA)	No.		Marks		
	Very Short Answer (VSA)	No.		Marks		Scheme of Sections
	Objective (O)	No.		Marks		
Steps:	a, b, c, d, e, f, g					

The development of a blueprint is a sequential process that makes the job a bit organised and easy for a question paper setter. The small letters a, b, c, d, e, f, and g in the given template are not part of the blueprint itself. They are steps to guide question paper setters. The details of these steps are the following:

I. Begin by filling in the top section of the proforma:
   - Mention the subject, class, maximum marks, time, and name of the examination for which the blueprint is being prepared.

II. **Step (a):** Write the maximum marks for the exam in the bottom-right corner of the proforma.

III. **Step (b):** Record the weightages for cognitive levels or objectives, as decided in the test design.

IV. **Step (c):** Enter the weightages assigned to various content units/sub-units in the rightmost column.

V. **Steps (d) to (g):**
- Use these columns to input details for different types of questions (Essay/Long Answer, Short Answer, Very Short Answer, and Objective).
- Fill the cells using the format: Question Sequence (Marks). For example, "Q1 (5)" means question 1 positioned in the paper carries 5 marks.
- Leave blank cells marked with a <-> to indicate no questions.
- Use asterisks (*) and dotted lines to show marks split between two objectives.
- Shade cells where internal choices are provided between the questions.

VI. **Verify the Question Paper Design**
- Count the number of questions for each type (Essay, Short Answer, etc.) and ensure it matches the design.
- Check row and column subtotals to confirm alignment with weightages for cognitive levels/objectives and content units.
- Adjust the blueprint if any discrepancies arise.

VII. **Finalize the Blueprint**
- Fill in the summary section, which includes the question types, section scheme, and pattern of question choice (if applicable).

2.3.2 STEPS FOR QUALITY ANALYSIS OF A BLUEPRINT

After completing the blueprint, review the following:

I. Does it align with the weightages for objectives, content units, and question types as outlined in the design?
II. Are the predisposing conditions, the pattern of question choice, and the number and form of questions followed correctly?
III. Are questions evenly distributed across content areas and cognitive levels?
IV. Are internal choices balanced in terms of content, difficulty, and time requirements?

A blueprint translates the test design into an actionable plan. It supports validity and reliability by guiding question setters to create various types of questions that align with the test's goals. While a good blueprint is crucial, the quality of the final question paper depends on how well the questions adhere to the blueprint.

The blueprint templates for language and non-language subjects are attached below for reference:

2.3.3 FORMAT OF BLUEPRINT (LANGUAGE)

Subject	:	_____				Class	: _____
Unit/Paper:	:	_____	Max Marks	:	_____	Time	: _____

S. No.	Objective → / Form of Questions	Elements of Language				Comprehension				Expression				Appreciation				Total Row Wise
	Content Unit/ Sub Unit	E	SA	VSA	O	E	SA	VSA	O	E	SA	VSA	O	E	SA	VSA	O	
	Sub Total																	Question Seq. (marks))
	Total Column → Wise	**(Marks Total)**				**(Marks Total)**				**(Marks Total)**				**(Marks Total)**				

Note: Figures within brackets to indicate the marks allotted to a question and figures outside the brackets to sequence no. question in paper.

Summary:
- Essay (E) No. _____ Marks _____
- Short Answer (SA) No. _____ Marks _____
- Very Short Answer (VSA) No. _____ Marks _____
- Objective (O) No. _____ Marks _____

Pattern of Options _____

Scheme of Sections _____

2.3.4 FORMAT OF BLUEPRINT (SUBJECT OTHER THAN LANGUAGE)

Subject	:	______________				Class	:	______________
Unit/Paper:	:	______________	Max Marks	:	______________	Time	:	______________

S. No.	Objective → / Form of Questions / Content Unit/ Sub Unit	Knowledge				Understanding				Application				Analysis				Evaluation				Creation				Total Row Wise
		E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	E/LA	SA	VSA	O	
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
		d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	d	e	f	g	c
	Sub Total																									Question Seq. (marks)
	Total Column → Wise	b (Marks Total)				b (Marks Total)				b (Marks Total)				b (Marks Total)				b (Marks Total)				b (Marks Total)				a

Notes: Figures within brackets to indicate the marks allotted to a question and figures outside the brackets to sequence no. question in paper.

Denotes that marks have been combined to form one question.

Summary:	Essay (E)	No.	______________	Marks	______________	Pattern of Options	______________
	Short Answer (SA)	No.	______________	Marks	______________		
	Very Short Answer (VSA)	No.	______________	Marks	______________	Scheme of Sections	______________
	Objective (O)	No.	______________	Marks	______________		
Steps:	a, b, c, d, e, f, g						

2.4 Writing Questions (Step 3)

Writing questions is the most critical step in developing a question paper. The quality of questions determines how well the paper assesses the targeted competencies and ensures balance. Each question must align with the test design and blueprint, adhering to the predetermined cognitive level, content area, and difficulty level established in earlier stages.

As outlined in the blueprint, the required number of questions should be developed and recorded on item sheets. Each question, along with its specifications, is to be written on a separate sheet using a standardized template, as shown below:

Question Writing Sheet

Question Code:		Class:	
Subject:		Unit:	
Cognitive Level:		Marks:	
Mapped Competency:		Estimated Time:	
Type of Question:		Estimated Difficulty Level:	

The process of framing questions is so vital that the overall success of an examination depends on it. Furthermore, the quality of later stages, such as preparing the marking scheme, relies heavily on well-written questions.

While developing questions, item writers must consider several key aspects, including the characteristics of a good question and the strengths and limitations of different question types. Let's explore these aspects in detail.

2.4.1 CHARACTERISTICS OF EFFECTIVE AND FAIR QUESTIONS

Assessment questions are best thought of as communication between the assessor and the learner. Each assessment question should clearly, completely, and directly convey what knowledge the learner needs to deploy and what competencies, skills, or abilities the learner needs to demonstrate. Questions should provide all learners in the given test-taking population with a fair opportunity to communicate their competencies.

In keeping with the goal of clear, complete, and direct communication, all assessment items should display the following characteristics: CLARITY, UNITY, and ECONOMY.

CLARITY:

The prompt (or "stem") of an item should clearly and unambiguously point the learner to the desired competency. In other words, there should be only one way to correctly interpret the prompt. For instance, a question that features a picture of Rani Laxmi Bai and asks, "What is the significance of this picture?" is subject to too many possible interpretations to be a useful question.

Clarity also implies that some formulations of questions and question prompts may be consistently preferable to others. Questions that seek to elicit the same competency can (and in most cases should) have highly similar wording. Question writers should rely on proven formulations (sometimes called "task models") and should not introduce different wordings simply for the sake of interest or variety.

The language of the question should be specific and precise to clearly define the scope of the expected answer.

Additionally, the value points or marks assigned to a question must be clearly indicated. For questions with two or more parts, the marks allotted to each part should also be explicitly mentioned.

UNITY:

All parts of a question as a whole should aim only at the target competency. For instance, a mathematics item in the form of word problem should not include difficult vocabulary or unfamiliar situations, since these are unlikely to be part of a mathematics competency.

"All parts" must include the incorrect answer choices of a selected response question. One frequent mistake made by question writers is to create a prompt and stem that fits the desired competency but then add incorrect answer choices that can be eliminated by other means. For instance, a question intended to test a learner's knowledge of the causes contributing to a

significant historical event which has incorrect answers that can be eliminated because they refer to obviously fictional or concocted occurrences no longer measures the competency of causation.

Unity also implies that there should be a clear fit between the wording of the prompt and characteristics of the desired response. A science question that asks a student to "Evaluate the change in appearance of metallic iron when exposed to water in the presence of oxygen" is a poor fit with the desired response since there is no evaluation needed, only description.

ECONOMY:

Testing time is precious, so each question should only include as much material as is needed to point the learner toward the desired competency. Extraneous material such as decorative images or inessential information should be avoided. That said, it is sometimes necessary to employ more words than one might in a non-testing situation to make the task of the question clear. For instance, "Compare Anand's treatment of the independence movement to Tagore" would probably be understood in everyday speech, but it would be preferable, in a testing context, to explicitly spell out the comparison by asking instead "Compare Anand's treatment of the independence movement to Tagore's treatment of the independence movement."

Economy does not imply that all questions must be short. In order to test certain questions, it will often be necessary to supply a significant amount of material. However, all material supplied should be necessary for a student to display the desired competency.

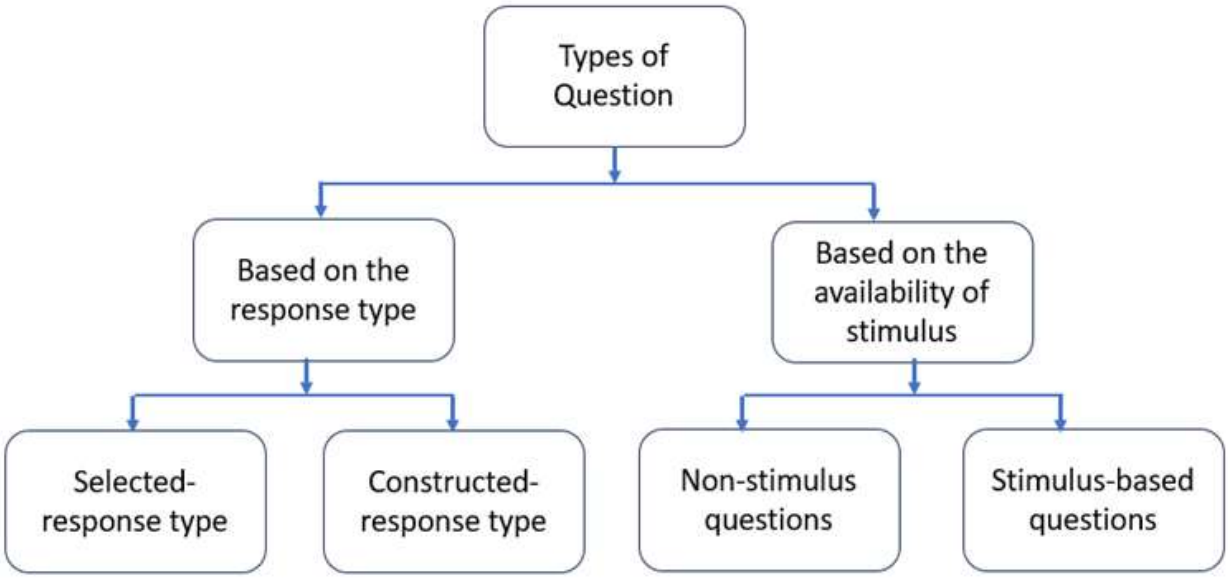
The principle of economy also implies that question writers should avoid trying to "teach through the test." Questions that gratuitously introduce new fundamental concepts or complex information beyond the needs of the target competency can compromise the integrity of the assessment.

2.4.2 STRENGTHS AND WEAKNESSES OF VARIOUS QUESTION TYPES

Question papers for school education board exams contain a variety of question types, ranging from traditional single-answer multiple choice questions to questions requiring the learner to construct an extended response. Different question types can appropriately serve different assessment aims, but the selection of question type or types for a Question paper should be a clear and conscious decision, based on the desired targets of measurement, student capabilities, and available scoring resources. This section of the manual discusses different types of questions, their strengths and weaknesses, and their recommended use cases.

Broadly speaking there are two ways to classify questions used on standardized assessments. The first is how the learner provides evidence of their competency, skill, and/or knowledge: either by **selecting** a response from a provided list of possible answers, or by **constructing** a response in reaction to a prompt.

The second mode of classifying questions pertains to whether the learner is expected to select or construct a response without any material beyond the prompt (a "non-stimulus question") or the learner is provided with additional material (a "stimulus-based question"). Stimulus material can take many forms: authentic texts, passages created explicitly for the test, images, graphics, charts, maps, data sets, etc. The choice of stimulus material (or the choice not to include it) should be guided by the target competency. For more guidance regarding stimulus material see section 3.5 below.



These two modes of classification are independent. In other words, there can be both stimulus-based and non-stimulus selected-response questions and both stimulus-based and non-stimulus constructed response questions. It is important to note that **all** four varieties of questions can be valid, depending on the measurement purposes they are intended to serve.

2.4.2.1 Selected response questions

Strengths or advantages of selected-response questions in general

Selected-response questions have certain advantages when compared to constructed-response questions. The first advantage is that the scoring is objective. In other words, there generally is (or should be) clear agreement as to the correct answer prior to the question being presented to learners. The second advantage of selected-response questions is that they can

almost always be scored and analysed more easily and require less time than constructed-response questions. The third is that they tend to be compact and take up a relatively small amount of testing time. Thus, a group of selected response questions can often be used to "cover" a larger number of competencies in the target domain.

Weaknesses or disadvantages of selected-response questions in general

Selected response questions have certain disadvantages when compared to -constructed-response questions. The first is the possibility of a learner guessing the correct answer given the available choices. The second is the difficulty of creating selected-response questions that address the highest cognitive levels or higher-order thinking skills. Although selected response questions can test more than factual recall, such questions must be carefully vetted and reviewed. The third disadvantage is related to the first two: selected-response questions, even well-crafted examples that truly measure higher-order thinking skills are sometimes dismissed or distrusted by educators, administrators, or the public, even though their use for measuring higher-order thinking skills has been extensively supported by research.

Despite these disadvantages, the use of selected-response questions as a significant part of standardized testing is almost always justified. When establishing the design and blueprint of a question Paper, teachers should strongly consider including a significant proportion of selected-response questions because of their greater coverage possibilities and their greater ease of scoring. Selected response questions come in three main types: single answer-multiple choice, multiple-answer multiple choice or alternate answer questions.

Single-answer multiple choice

This is the most familiar type of question, having been used successfully in a wide variety of subject areas since the middle of the twentieth century. As their name implies, each question of this type should have one and only one defensible correct answer to the prompt.

Recommended use cases: This question type can be used to test recall of discrete facts (though that will not typically be sufficient evidence of the target competency). It can also be used to test comprehension of material, inferences, application of concepts, mathematical operations, the ability to recognize valid explanations, as well as other competencies.

Multiple-answer multiple choice

As its name implies this question type is a multiple-choice question with more than one defensibly correct answer.

Recommended use cases: This question type is best employed to deal with complex concepts, multi-causal explanations, or competencies defined or taught as lists of desirable practices or behaviours.

In most cases, a single-answer multiple choice question can effectively target the same competency as a multiple-answer multiple choice, multiple-answer multiple-choice questions should only be used when necessary. Because this type of question is typically much less familiar to learners and is often mistaken for a traditional single-answer multiple-choice, questions of this type should clearly signal that multiple correct answers are possible. If practicable, these questions should be grouped together in a separate question paper section with a clear header that announces the question type. This type of question tends to function best when the prompt states how many of the answer choices are correct. At a minimum, the prompt should contain language that clearly indicates the need to look for multiple correct answers.

Alternate answer selected response

This question type is essentially a multiple-choice question with only two answer choices, one of which is correct. The most common form is the true-false question. True-false questions are relatively easy to construct and require less time and effort compared to other selected-response questions. True/False questions are effective for testing basic knowledge and understanding of fundamental concepts.

Recommended use cases: Because of the high chance of guessing the correct answer, this question type should be used extremely sparingly, or not at all. In particular, it is not recommended to use true/false questions. It is possible, however, to construct valid questions that consist of two or more alternate choices. For instance, if the target competency was the learner's ability to use the correct sequence of tenses in formal written English, the following example of alternate-choice format could be valid:

"If I (have/had) seen your friend before I left the village, I (would/would have) given her your message."

Note that this has four possible permutations, only one of which is grammatically correct, thus it is functionally equivalent to a single-answer multiple choice question with four answer choices.

2.4.2.2 Constructed-Response Questions

Strengths or advantages of constructed-response questions in general

By requiring the test-taker to produce a correct response, rather than identifying one, constructed-response questions can eliminate the possibility of guessing. Constructed response questions can also more easily assess complex skills such as structured writing, scientific explanations or explications of mathematical or other reasoning. Short or very short constructed response questions can sometimes be written more quickly than selected response questions, since there is no need to craft wrong answer choices.

Weaknesses or disadvantages of constructed-response questions in general

The main disadvantage of constructed -response questions is that the need to score the learners' responses creates a degree of subjectivity. Unless the range and exact wording of possible credited responses is extremely limited, accurate scoring will depend on informed human judgment (either directly by a human scorer or built into a scoring program). Responding to constructed-response questions also typically takes more testing time, thus they usually provide smaller opportunities for content coverage. Apart from this, judging the answer and scoring is a time-consuming and often tiring process. Sometimes factors like handwriting, spelling mistakes and the teacher's impression of the student affects scoring. These types of questions, particularly essay-type ones, are less reusable than selected response questions. They are easier to remember and thus become insecure more quickly.

Importance of a rubric or marking scheme

Any constructed-response question used in a standardized test such as a school board question paper must have a rubric or scoring scheme created as part of the question-writing process. Rubrics may consist of a simple list of possible responses that would receive credit, or they may consist of descriptive statements specifying the characteristics of a successful response (For example: "The paragraph begins with a topic sentence that clearly and accurately states the main points made in the paragraph.").

Rubrics or marking schemes are primarily intended to provide scorers with clear and comprehensive guidance on what should and should not be credited in test-takers' responses. However, rubrics and marking schemes also have the important secondary purposes of signalling the intent of the question and serving as a check on the clarity of the question. Rubrics or marking schemes should be created with the first draft of the question and reviewed (and modified if necessary) alongside the question as part of subsequent reviews.

Types of Constructed Response Questions

The most common type of constructed-response question is according to the approximate length of the response the learner needs to produce to provide evidence of the target

competency. Based on the survey of existing state board QPs, constructed-response items can be usefully divided into the following categories:

Very short answer (VSA):

Responses to this type of question are typically expected to range from a single word or number to a short sentence or two. They sometimes take the form of "fill in the blank" questions. Learners are typically allotted 1-2 minutes of testing time for each question of this type.

Recommended use cases: Very short constructed-response questions are best employed to assess discrete pieces of knowledge, vocabulary, grammatical usage, simple mathematics problems, etc[4].

Short answer (SA):

Responses to this type of question are typically expected to range from 50-60 words. Learners are typically allotted 3 to 5 minutes of testing time for each question of this type.

Recommended use cases: Short constructed response question is best used to elicit multiple pieces of related information (e.g., significant factors contributing to a historical event), complex explanations or, comparisons or evaluations that cannot be adequately conveyed in a sentence or two. Short constructed-response questions can also be employed to elicit simple structured writing such as a sequence of events, a structured paragraph, or steps in a procedure.

Long Answer (E/LA):

Responses to this type of question are typically expected to have more than 100 words. Learners are typically allotted 8 to 10 minutes of testing time for each question of this type.

Recommended use cases: Long answer constructed response questions can be used for many of the same purposes as short answer CRs. The greater length of the expected response provides more opportunity to elicit structured writing.

[4] Please note regarding Very short answer-Constructed response (VSA-CR): Some VSA CR questions are constructed to have one and only one possible correct answer. In this case, it may be more cost-effective to reconfigure them as a selected-response question, since this will obviate the need for human scoring.

Value of coherence in long and short answer CRs:

These two types of questions should be structured to elicit responses that have some degree of coherence. A Long Answer CR that in effect asks learners to simply recount anything they may know or have learned about a topic is unlikely to result in fair or accurate measurement.

2.4.3 VALID USES AND DESIRABLE CHARACTERISTICS OF STIMULUS MATERIAL

The decision to include stimulus material should be guided solely by the demands of the target competency. For instance, if the question is intended to test map-reading skills, then a map is likely to be an appropriate stimulus. However, if the target competency is reading comprehension of text and a travel account is chosen as the text, including a map is likely to be superfluous and possibly distracting.

Stimulus material of any type should avoid extraneous elements. For instance, a picture of items with price tags in a shop used in a mathematics question does not need to include a picture of the shopkeeper. On the other hand, a question intended to target the competency of detecting significant patterns in complex data sets should include some data that does not fit the predominant pattern, since this is integral to the competency.

The length of textual stimulus should take into account available testing time and the number of questions based on the text. Having learners read a 300-word passage and answer only 2 multiple-choice questions based on it would represent a misallocation of testing time and test-taker effort. Text stimulus materials should also be at the appropriate grade level.

Visual stimulus material should be no more complex than necessary. Many learners have some degree of visual impairment and have difficulty processing visually dense images with lots of decorative features, small font sizes, or unconventional scripts. The ability to deal with such features is usually not part of the target competency.

In addition to choice of question type and decisions regarding stimulus material, another important consideration in question writing is that of Fairness and Accessibility

2.4.4 FAIRNESS AND ACCESSIBILITY IN QUESTION PAPER DESIGN

What is meant by “fairness” in the context of assessment? Fairness in testing can by defined by the validity of inferences drawn from test scores across different groups. Fairness is achieved when test scores are valid for all test takers regardless of their gender identity, socio-economic status, or other group affiliation not relevant to the abilities being tested.

A test or question paper can be considered "fair" to the extent that learners answer questions *correctly* because they possess the relevant knowledge and competencies AND the extent that learners answer *incorrectly* because they lack the relevant knowledge and competencies and not because of some other feature of the question. In other words, learners should get questions **right** for the **right reasons** and they should get them **wrong** for the **right reasons** as well! If learners are getting questions wrong for the wrong reasons, a test is said to display "construct-irrelevant variance," i.e. the test is measuring something other than what it was constructed to measure.

Fairness rules aim to prevent test items from causing three groups of barriers

2.4.4.1 Cognitive barriers (measuring irrelevant knowledge)

Cognitive barriers can be introduced by unnecessary complication of language or the use of overly specialized terms. For instance, in a vocabulary test intended for a general population the agricultural term "plow" is probably acceptable, but "thresher" probably is not. Similarly, in the area of information technology "computer" would be an acceptable term for a general population, but "JPEG" is probably too specialized.

Cognitive barriers can also be introduced by assuming familiarity with knowledge or concepts that are not universally known. For instance, a mathematics item that assumed knowledge of how a polo match is scored would risk being biased against learners of socio-economic groups not familiar with polo.

2.4.4.2 Affective barriers (eliciting strong emotions)

Determinations of fairness are directly connected to the purpose and audience of a test. A topic like food contamination might risk triggering an emotional response (an affective barrier) in a reading test for a general population but obviously would be relevant in a restaurant inspector's test. Affective barriers can be caused by the unnecessary introduction of upsetting or controversial topics. Such topics should be avoided unless they are absolutely necessary for valid measurement. These include **abortion**, **genocide**, **ethnic conflicts**, **gun control**, **child abuse**, **euthanasia**, etc.

Even in cases where the mention of such topics is warranted, they should be treated as neutrally as possible. For instance, on a history test, it might be appropriate to deal with past instances of ethnic or religious conflict, but graphic depictions of violence associated with these conflicts should still be avoided.

Some topics require careful presentation to avoid eliciting negative emotions, even if they are construct-relevant:

- **Advocacy**: Avoid taking sides unless needed for valid measurement.
- **Death, disease, and disasters**: Avoid graphic details unless necessary.

Gender Sensitivity

Question writers should ensure **gender neutrality** in test items, avoiding stereotypical language or assumptions (e.g., using "nurse" for both males and females, referring to women and men equally by name or role). Questions should not rely on the assumption that **gender** is universally regarded as binary; test questions should be inclusive of non-binary identities or at least avoid relying on strict binary definitions or linguistic constructions.

Example: "How many students are in the class?" should avoid assuming only male and female students.

Some examples of gender-biased terms to avoid and acceptable replacements:

Acceptable	*Unacceptable, Except in Historical or Literary Material*
Chair, leader	Chairman
Firefighter	Fireman
Human beings, people	Man, mankind
Mail carrier	Mailman
Sales representative	Salesman
Insurance agent	Insurance man
Supervisor	Foreman
Workers, personnel, labor	Manpower

Religion:

Religion has been identified as both a potential cognitive and affective source of construct-irrelevant variance. Test takers may have strong emotional attachments to religious views and could be upset by any perceived endorsement of opposing beliefs. Even subtle positive or negative references to religion—whether general or specific—can provoke strong reactions from certain groups. To ensure fairness, any necessary mention of religion in a test should be presented in an objective, factual, and neutral manner.

Special Considerations for Children

Tests for children (particularly younger children) should, if possible, avoid topics that deal with:

- **Disturbing or frightening content** (e.g., spiders, fire, abductions),
- **Inappropriate behavior** (e.g., cheating, fighting),
- **Family or moral issues** (e.g., divorce, family violence).

Sensitive topics that are construct-relevant for children or young children should be presented in the least upsetting way possible.

Following these principles will ensure that tests are fair, valid, and free from unnecessary emotional or cognitive bias, fostering equitable assessment for all test takers. The aim should be to give all learners the same chance to display the relevant knowledge and competencies.

2.4.5 COMPETENCY BASED ASSESSMENT

Many assessment systems have focused on summative evaluations that test students' ability to recall information through rote memorization. While this approach provides a snapshot of a student's performance, it can fail to foster deep understanding, critical thinking, or the ability to apply knowledge in real-life contexts. The National Education Policy (NEP) 2020 emphasizes the urgency of transitioning from this outdated model to a more formative, competency-based approach.

According to NEP 2020, the new assessment framework must be regular, formative, and designed to promote learning and development. The focus is on higher-order skills such as analysis, critical thinking, and conceptual clarity. By prioritizing assessment for learning, this model ensures that teaching-learning processes are continually adapted to optimize outcomes for all students, fostering meaningful progress across educational stages.

NEP states that

> "The aim of assessment in the culture of our schooling system will shift from one that is summative and primarily tests rote memorization skills to one that is more regular and formative, is more competency based, promotes learning and development for our students, and tests higher order skills, such as analysis, critical thinking, and conceptual clarity. The primary purpose of the assessment will indeed be for learning; it will help the teacher and student, and the entire schooling system, continuously revise teaching learning processes to optimize learning and development for all students. This will be the underlying principle for assessment at all levels of education."
> [NEP 2020, 4.34]

2.4.5.1 Understanding Competency

The National Curriculum Framework (NCF) for School Education defines competencies as **measurable and observable learning achievements** that align with curricular goals. Competency based assessment evaluates not just what a student knows but how effectively they can apply their knowledge and skills in various contexts. This approach emphasizes ongoing development, encouraging students to build and refine their abilities progressively.

NCF School Education suggest:

> "Competencies are **learning achievements that are observable and can be assessed systematically**. These Competencies are derived from the Curricular Goals and are expected to be attained by the end of a Stage."
> NCF-FS 2022 and NCF-SE 2023

2.4.5.2 Why is Competency-Based Assessment Essential?

Competency based assessment aligns with the broader goals of 21st century education, which prioritize skills such as problem solving, adaptability, and creativity. Unlike rote memorization, which often results in shallow and temporary learning, this approach nurtures lifelong learners equipped to face complex challenges.

By focusing on meaningful learning outcomes, competency-based education transforms education into an empowering process, preparing students for real-world success. Through this shift, education systems can ensure that every learner achieves their potential and contributes meaningfully to society.

How learners are tested shapes how they will be taught. When educators know that learners will be faced by a summative exam consisting entirely of factual recall questions, they will emphasize rote memorization. On the other hand, true competency-based **assessment** will encourage competency-based **education**.

2.4.5.3 Assessment framing guidelines as applied to State Board Question Papers

S.no.	Quality check parameters	Mark (✓) if Yes and Mark (✗) if No
1	Assessment should measure achievement of Competencies and Learning Outcomes leading to attainment of Curricular Goals.	
2	Assessments should be Stage appropriate.	
3	Assessments should be constructive, developmental, and learning-focused.	
4	Assessments should focus on testing Higher Order Thinking Skills (HOTS).	
5	Assessments should promote reflective and self-regulated learning.	
6	Assessments should be transparent with clear criteria and expectations.	
7	Assessments should ensure inclusivity and equity in access and participation.	
8	Assessments should be designed to motivate and build confidence in learners.	
9	Assessments should accommodate student diversity	
10	Assessments should employ clarity and simplicity in item language and structure.	
11	Assessments should highlight the integration of technology where appropriate.	
12	Assessments should be aligned with contemporary societal and environmental issues.	
13	Assessments should encourage critical questioning and exploration of multiple perspectives	
14	Assessments should integrate interdisciplinary connections where applicable.	
15	Assessments should incorporate ethical and value-based considerations where applicable.	
16	Assessments should align with the principles of fairness and objectivity.	
17	Assessments should encourage creativity and innovation in responses where applicable.	
18	Assessments should allow for demonstration of skills through performance-based tasks.	
19	Assessments should emphasize problem solving and decision-making abilities.	
20	Assessments should promote ethical reasoning and cultural sensitivity.	
21	Assessments should include diverse question types to address multiple learning styles.	

2.4.5.4 Creating and reviewing competency-based questions

For an assessment to be considered competency-based, the individual questions must obviously be competency-based as well. In some subject areas, particularly mathematics, it can be relatively easy to create competency-based items, in part because there is virtually no distinction between content knowledge and procedural knowledge (or skill) in those fields. For instance, a competency statement that says, “Students can perform basic arithmetic (addition, subtraction, multiplication, division) with numbers up to three digits” immediately

suggests how a question that tests that competency would be structured and what content it would have.

Most subjects, however, require more in-depth consideration and often choice of subject matter before a true competency-based question can be constructed. For instance, a history competency statement that says, "Students can understand the main causes of significant events in Indian history" requires a judgement as to what constitutes a "significant event." This judgment is likely to be informed by a combination of the content of widely used textbooks, the question writer's own experience as an educator, and any specific guidance provided by boards or other administrative bodies. Similarly, a language competency statement that says, "Students can comprehend important points from age-appropriate informative texts" requires several judgments.

Choice of question type

Once the competency has been identified with respect to subject matter and cognitive level, the first choice of a question writer is what type of question can be used to measure this competency. In general, question writers should choose the shortest, simplest question-type commensurate with the target competency. For instance, if the target competency is knowledge of significant terminology in a technical field, measurement of this competency does not require a long-answer constructed-response type question. It could probably be measured adequately with single-answer multiple-choice or very-short-answer constructed-response questions. On the other hand, a test question aimed at the competency of creating and explaining an illustrative scenario for a particular concept or principle would almost certainly require a long-answer constructed-response type question and carefully specified scoring scheme or rubric.

Cognitive level and question type

There is a widespread misconception that selected-response (multiple choice) questions are only suitable for measuring competencies at the most basic cognitive level ("know"). This misconception has been reinforced by the long-standing practice of using multiple-choice question mostly to test factual recall or declarative knowledge. Research has shown, however, that multiple-choice questions can in fact measure higher-order thinking skills. In practice, well-constructed multiple-choice questions have been successfully used to measure competencies at the cognitive levels of "understand," "apply," and "analyse" in addition to the more familiar "know" type.

The key term in the previous assertion is "well-constructed." For a multiple-choice question to provide valid evidence of higher-order thinking skills, the **entire question**, including the

incorrect answer choices, must function together. For instance, a question intended to measure the ability to correctly apply literary terms to a fictional text passage should have incorrect answers that are all valid literary terms that the learner is likely to be familiar with. If, instead, the question writer invents imaginary terms or uses abstruse terms unlikely to have been seen by the test-taking population, then the question will no longer reliably measure the target competency, instead collapsing into a mere recognition task.

Given the greater economy and ease of scoring associated with multiple-choice items, they are often to be preferred to constructed-response items, especially if they can measure competencies at some of the higher cognitive levels. Designers, blueprint makers, and questions writers should make every effort to include multiple-choice question, while also ensuring that they truly measure the competency.

Choice of context or setting

It is common practice to provide a context or setting for test questions. For instance, an arithmetic problem might ask the test-takers to imagine they are making purchases at a store, a reading comprehension question might have the test-takers read and extract information from a safety poster. Question writers typically provide a setting or context for one of the following reasons:

- The explicit requirements of the competency
- The desire to lessen test anxiety by using familiar settings
- The desire to improve test-taker engagement

The first reason is always valid. Whenever a competency calls out a particular setting or context (e.g. "Students can understand basic conversations on everyday topics."), then test questions targeting that competency should, as a matter of course, reflect that setting or context.

Attempting to alleviate test anxiety is a laudable goal and adding some material to place the test-taker in a familiar context may do no harm. However, any contextualizing material should be as brief as possible and should be checked carefully to make sure it is plausible, does not introduce unnecessary distraction, and is truly familiar. For instance, a question that targets test-takers' ability to compute area of rectangles might ask them to calculate the area of a dining table. But if the dimensions of the table were implausibly large (e.g. 8 meters by 3 meters), this could work against the desire to create a familiar setting.

There is some research to support the idea that at least some test-takers are more engaged when test questions supply a concrete context. However, the same potential problems can arise as do when attempting to lessen anxiety by adding familiar contexts.

In general, writers of competency-based questions should be wary of gratuitously adding contexts or settings not required by the competency itself. Reviewers of competency-based questions should always ask themselves if any additional context is truly needed and whether it is likely to prove distracting to some students.

Stimulus-based or non-stimulus

When considering whether to use stimulus material, the question writer should be guided only by the explicit requirements of the competency and the principle of economy. Stimulus material should only be included in support of the target competency and only as much stimulus material as is needed. Writers should avoid "decorative" visual images in questions that can be solely verbal. Text passages or scenario descriptions should avoid providing large amounts of extraneous detail. Setting or scenario descriptions should provide the minimum necessary verbiage to put the test-takers in the scene.

The importance of question review

Even the most basic test questions intended for a large population should receive a thorough content and editorial and fairness review. This is particularly true of competency-based questions. Reviewers of such questions should ask themselves "Does the question as a whole, truly afford test-takers the opportunity to demonstrate the target competency by answering the question correctly?"; "Will test-takers who answer this question incorrectly truly lack the target competency?"; and "Are there features of the question that will interfere with its ability to measure the target competency?" If they can satisfy themselves on those three points, the question is likely to be part of a fair and valid assessment.

2.5 Preparation of a Question Paper (Step 4)

The preparation and assembly of a question paper is a crucial step in the assessment process, requiring careful planning, alignment with instructional goals, and attention to detail. A well-constructed question paper not only evaluates students' knowledge and skills effectively but also promotes a positive testing experience. This process begins with developing a comprehensive design blueprint that outlines the structure, content, and cognitive levels to be assessed. The blueprint serves as a foundation for selecting and sequencing questions, ensuring balanced coverage of topics, alignment with learning objectives, and adherence to prescribed difficulty levels. Additionally, the question paper must incorporate clear instructions and formatting that facilitate smooth administration and enable test-takers to focus on demonstrating their abilities without unnecessary confusion. A thoughtful approach to assembling a question paper ensures fairness, validity, and reliability while providing an opportunity for students to showcase their learning effectively.

There are several criteria for choosing the order in which questions appear in a question paper:

i. Type of questions
ii. Instructional objectives
iii. Content units,
iv. Difficulty level
v. According to the convenience of administration of the examinations

2.5.1 TYPE OF QUESTIONS

It is often desirable to group questions within a paper by type. Selected-response questions are frequently placed first since they are often easier and can serve as a confidence-builder to test takers.

2.5.2 INSTRUCTIONAL OBJECTIVES

It is sometimes desirable to place questions dealing with more basic competencies or skills (such as comprehension) earlier in the paper, so that test-takers encounter them in the same order in which they were instructed.

2.5.3 CONTENT UNITS

The order of questions within a topic or section should align with the logical flow of the content. If a specific sequence is commonly followed during teaching (e.g., chronological events in history), the same order should be maintained in the paper.

2.5.4 DIFFICULTY LEVEL

To the extent possible, it is desirable to present test-takers with easier questions before harder questions, assuming reasonably accurate estimates of difficulty can be made.

2.5.5 CONVENIENCE OF ADMINISTRATION

Administrative considerations can influence the order of questions. Separate sections for different question types (e.g., objective vs. subjective) reduce confusion. Grouping questions by response format minimizes the need for switching between answer sheets.

In addition to the questions themselves, question papers must have clear instructions and information for the test-takers. These directions should include:

- The suggested amount of testing time allotted for the paper as a whole or the section (if applicable)
- Means of indicating answers to questions (e.g. circling the correct answer choice, filling in a bubble on a separate answer sheet, writing an answer on blank lines, etc)

- The approximate length of expected answers to constructed-response questions and any specific required features of the response (e.g. "Write a structured paragraph with a topic sentence.")

If the question paper is divided into major topics or sub-topics (e.g. a science question paper that has different sections dealing with Biology, Chemistry and Physics), it can be helpful to the test-takers if the different sections are labelled as a kind of advance organiser.

By thoughtfully organizing questions and incorporating clear instructions, question papers become more accessible to test-takers, reduce ambiguity, and enhance the accuracy of assessment outcomes.

2.6 Preparing Scoring Key/Marking Scheme (Step 5)

Marking schemes are essential tools in educational assessment, serving as structured guides for teachers to evaluate constructed response questions systematically, fairly, and transparently. They enhance the consistency and objectivity of grading, ensuring that assessments reflect students' true understanding, skills, and effort. To develop robust and comprehensive marking schemes, the following principles should be observed:

2.6.1 ALIGNMENT WITH SPECIFIC COGNITIVE DOMAIN AND CONTENT

Marking schemes should directly correspond to the skills and content targeted by the item. Each item assesses specific learning objectives, whether conceptual understanding, analytical ability, problem-solving, or creativity. Marks should be distributed in a way that emphasizes these intended outcomes, ensuring that the evaluation reflects the focus of the question.

2.6.2 PROPORTIONAL ALLOCATION OF MARKS TO VALID VALUE POINT

The distribution of marks must be proportional to:

- **The scope of Expected Content**: Items requiring detailed explanations, multiple steps, or extensive reasoning should carry more weight.
- **The Time Investment Required**: Marks should also align with the estimated time students need to respond. For example, higher marks should be allocated to items involving multi-step problem-solving or complex analysis.

This approach ensures fairness, as students are appropriately rewarded for effort and depth in their responses.

2.6.3 RECOGNITION OF VALID AND EXTENDED ANSWERS

While the marking scheme should outline correct and valid answers, it should also allow flexibility to reward responses that:

- Go beyond the scope of the syllabus.
- Demonstrate creative thinking, critical analysis, or unique perspectives.

Rubrics should include provisions for alternative valid responses that are not explicitly mentioned but reflect accurate reasoning or advanced understanding. This flexibility encourages innovation and rewards students who think critically or apply knowledge in novel ways.

2.6.4 CLEAR GUIDELINES TO MINIMIZE SUBJECTIVITY

Subjectivity in evaluation can compromise fairness. To eliminate this, marking schemes should provide:

- **Detailed Instructions**: Clear descriptions of what constitutes a correct or complete answer.
- **Explicit Criteria**: A breakdown of marks for different components of an answer (e.g., structure, accuracy, originality).
- **Analytic rather than holistic scoring**: To the extent possible, rubrics should use a series of binary determinations (e.g. Valid thesis statement present or not? Presented as paragraphs organized around topic sentences or not?) rather than relying on sweeping statements about the overall quality of a response (e.g. "Response is a highly persuasive essay.")
- **Exemplar Responses**: Model answers, or sample responses can help evaluators understand the expected standard and reduce ambiguity.

Consistency in grading is achieved when evaluators have a precise reference point, minimizing the impact of personal biases or interpretations.

2.6.5 DETAILED VALUE POINTS

Each marking scheme should specify key value points for every question. These value points:

- Highlight the critical elements of an answer that must be included to earn marks.
- Provide a clear framework for evaluators to assess answers systematically. For instance, a long-answer question in science might include value points for defining terms, explaining principles, and providing relevant examples or diagrams.

By breaking responses into measurable components, value points ensure that the marking process is both systematic and exhaustive.

2.6.6 AVOIDANCE OF NEGATIVE MARKING IN RUBRICS

Marking schemes should focus on rewarding correct or partially correct responses, rather than penalizing errors. Defining negative marks within the rubric can lead to confusion and may discourage students from attempting questions. Instead:

- Errors should be noted in feedback without disproportionately impacting overall marks.
- Rubrics should emphasize use a positive approach that highlights areas for improvement while acknowledging effort and valid attempts.
- Rubrics should distinguish and to the extent possible specify the difference between serious conceptual or procedural errors that rightly prevent a response from receiving full credit and minor errors that do not detract from the overall quality of the response.
- It is good practice to phrase the language of the rubric so that it focuses on the characteristics of the RESPONSE rather than on the abilities or shortcomings of the STUDENT.

2.6.7 SUBJECT-SPECIFIC CONSIDERATIONS

Different subjects demand distinct response formats, reflecting the unique nature of their content and objectives:

- **Mathematics and Sciences**: Responses should be evaluated for accuracy, logical reasoning, and stepwise problem-solving. Diagrams and units, where applicable, must be considered.
- **Languages**: Focus on content, structure, grammar, and creativity. Subjective answers (e.g., essays) require well-defined rubrics with weightage for different aspects like coherence and originality.
- **Humanities**: Assessments should consider argumentation, critical analysis, and use of evidence. Marking schemes should emphasize clarity and depth of reasoning.

Custom templates for each subject should be prepared, incorporating examples and response structures that reflect subject-specific demands.

2.6.8 ENCOURAGING FAIR AND INCLUSIVE ASSESSMENT

Marking schemes should be inclusive and designed to fairly assess the diverse abilities and strengths of students. Some strategies to ensure this include:

- Using language in the rubrics that is clear and accessible.
- Accommodating alternative valid approaches to problem-solving or reasoning.

- Providing additional guidance for evaluators on how to handle unique or unexpected responses.

A well-constructed marking scheme serves as an effective assessment, ensuring that grading is consistent, objective, and reflective of students' efforts and abilities. By adhering to these principles—emphasizing alignment with learning objectives, proportionality, flexibility, and subject-specific nuances—teachers can foster a more transparent and equitable evaluation process. Comprehensive marking schemes not only improve the reliability of assessments but also enhance their role as tools for learning and growth.

2.6.9 CHECKLIST FOR DEVELOPMENT OF MARKING SCHEME

S.No.	Quality check parameters	Mark ✓ if Yes and Mark ✕ if No
1	Are the marks distributed according to the specific skills assessed (e.g., analysis, application, creativity)?	
2	Does the marking scheme explicitly state the learning objectives each question addresses?	
3	Are the criteria aligned with the syllabus and lesson plan?	
4	Is there a distinction between core knowledge and higher order thinking skills in the distribution of marks?	
5	Are instructions for skill-based evaluation (e.g., critical thinking, problem-solving) explicitly defined?	
6	Is the weightage of each question proportionate to its relevance in the curriculum?	
7	Are interdisciplinary skills (e.g., communication in science or logic in humanities) incorporated where relevant?	
8	Are higher marks allocated to questions requiring detailed reasoning, multi-step problem-solving, or analysis?	
9	Is the time required to answer each question considered when assigning marks?	
10	Are questions with similar difficulty levels assigned consistent marks across the paper?	
11	Is the proportionality between the marks and effort required for subjective questions (e.g., essays) reasonable?	
12	Are correct answers, as defined in the marking scheme, included explicitly?	
13	Does the marking scheme account for alternative approaches or methods to solve a problem?	
14	Are evaluators instructed to reward evidence-based reasoning, even if the final answer is incorrect?	
15	Are criteria provided for partial credit in multi-step problems or partially correct responses?	
16	Are extended, well-reasoned answers given appropriate recognition, even if unconventional?	
17	Are clear and detailed marking instructions provided for each question?	
18	Are criteria for awarding marks specific enough to minimize evaluator interpretation?	
19	Are common misconceptions or errors outlined with guidance on how to address them?	
20	Does the scheme explicitly describe the level of detail required in student responses?	

21	Are descriptors provided for awarding full, partial, or no credit in subjective questions?	
22	Are key value points identified for every part of the question?	
23	Is the weightage of each value point clearly defined in the marking scheme?	
24	Are marks allocated for supporting elements (e.g., relevant examples, proper formatting, or diagrams)?	
25	Are correct procedures or methods (e.g., showing work in math) part of the evaluation criteria?	
26	Is the importance of clarity, accuracy, and organization addressed in the value points?	
27	Are templates tailored to the nature of the subject (e.g., numerical accuracy for math, structure for essays)?	
28	Are unique evaluation criteria provided for different types of questions (e.g., objective vs. subjective)?	
29	Are discipline-specific skills (e.g., grammar in languages, experiment design in sciences) explicitly addressed?	
30	Are technical or format requirements (e.g., labelled diagrams in biology) included in the rubric?	
31	Are rubrics differentiated for creative or exploratory disciplines (e.g., art, literature)?	
32	Is the balance between content, structure, and originality defined for subjects like humanities or languages?	
33	Are subject-specific nuances (e.g., stepwise logic in math, evidence-based reasoning in history) emphasized?	
34	Are marking schemes reviewed for inclusivity and sensitivity to diverse evaluator backgrounds?	

2.7 Question paper Analysis (Step 6)

After drafting a question paper, it is important to ensure that it aligns with the blueprint and design. To determine the degree of alignment, the paper reviewer must analyse each question against key criteria, such as the cognitive skills it assesses, the specific skills it measures, the unit it belongs to, the question type, the expected answer length, the average time required by a student to answer, and its estimated difficulty level. The following checklist can be useful in conducting this review:

1. Does the question paper carry weightages for all the cognitive domains same as given in the design?
2. Does the question paper carry weightages for all content areas same as given in the design?
3. Does the question paper carry weightages for all forms of questions same as given in the design?
4. Does the question paper carry weightages for different difficulty levels same as given in the design?
5. Does the question paper follow scheme of sections same as given in the design?
6. Does the question paper carry pattern of options same as given in the design?
7. Does the question paper have allotted time to attempt the paper same as given in the design?
8. Is the content tested the same as indicated in blueprint?
9. Is the unit tested the same as indicated in blueprint?
10. Does the question really test the intended cognitive domains?
11. Does the question measure the intended competency?
12. Is the question set in the same form as shown in blueprint?
13. Does the wording of question evoke the intended response?
14. Does the language of question clarify the scope of answer?
15. Does the question carry the same marks indicated in the blueprint?

Additionally, the proforma given below can also be used for a systematic cross-check. This proforma is divided into two parts:

Part I: Collects general information about the question paper, such as the name of the board, examination details (name and year), class, subject, question paper number, maximum marks, duration, and any sections present.

Part II (Question-wise Analysis): Focuses on reviewing each question to ensure it aligns with the test design and blueprint. This includes an evaluation of the cognitive domain being tested, the content area, question format, marks allocated, estimated time for completion, and

difficulty level. Comments or remarks are also recorded for each question, categorizing them as "OK," "Good," "Needs Improvement," or "Reject," based on the analysis.

2.7.1 PART I: GENERAL INFORMATION

1. Name of the Board:
2. Name of Examination:
3. Year of Examination:
4. Class:
5. Subject:
6. Question Paper Number:
7. Maximum Marks:
8. Time Allotted:
9. Sections, if any:

2.7.2 PART II: GENERAL ITEM ANALYSIS OF A QUESTION PAPER

Ques. No.	Objective	Content	Unit/Sub-Unit No.	Form of Question	Marks Allotted	Estimated time (min)	Estimated Difficulty Level
1							
2							
3							
4							
5							
6							
7							
8							

[1].Performa for question paper Analysis

This analysis helps identify any imbalances in the question paper, enabling necessary modifications to improve its quality.

2.7.3 QUALITATIVE ANALYSIS OF A QUESTION PAPER

A question paper's effectiveness largely depends on the quality of its individual items. The qualitative analysis of these items focuses on various parameters, including language, structure, content, question format, options, allocated time, and marking scheme. Each parameter must be critically evaluated to ensure the following:

- **Language of the Question**: Is it simple, clear, precise, unambiguous, and free from redundancy? Can it be easily translated and comprehended by students?
- **Structure of the Question**: Does it include directional words that clearly define the scope and style of the expected answer? Is the stimulus provided functional and acceptable?
- **Content of the Question**: Is it accurate, relevant to the syllabus, and focused on significant learning points rather than trivial details? Does it include adequate value points or criteria for awarding marks?
- **Form of the Question**: Is it appropriate for the purpose (e.g., selection or construction), marks allocation, and the time required? Does it adequately sample the content?
- **Comparability of Options**: Are the options comparable in terms of objectives, content, difficulty level, time required, and marks allocation?
- **Time Allocation**: Is the time allocated sufficient for answering each question?
- **Marking Scheme:** Is the distribution of marks appropriate and in line with the value points/criteria? Does the marking scheme align with the answers provided by examinees?

Any defects or areas for improvement identified during the analysis should be documented in the table below:

Question no.	Point of Analysis	Detailed Observation/Remark (if any)

These observations guide the refinement process to ensure the question paper is balanced, fair, and of high quality. For reference, a supportive checklist is included to assist in systematically reviewing these parameters.

2.7.4 QUESTION-WISE REVIEW CHECKLIST

S.No.	Quality check parameters	Mark ✓ if yes and mark X if No
1	Item Structure	
1.1	Is the stimulus clear, concise, and directly relevant to the learning objective?	
1.2	Does the stimulus provide sufficient context and information to enable the test-taker to respond effectively?	
1.3	Does the stimulus avoid unnecessary jargon, overly complex language, or irrelevant information?	
1.4	If visuals (e.g., graphs, charts, or images) are used, are they clear, labelled, and relevant to the item?	
1.5	Does the stem clearly communicate what is being asked, without ambiguity?	
1.6	Is the stem phrased as a complete question or task that aligns with the stimulus?	
1.7	Does the stem avoid leading the test-taker to the correct answer through unintended cues?	
1.8	Is the language in the stem accessible and appropriate for the target audience's skill level?	
1.9	Are the distractors plausible and free from obvious errors or impossibilities?	
1.10	Do the distractors reflect common misconceptions or errors that align with the learning objective?	
1.11	Are all options (distractors and the key) written in parallel structure, with similar length and complexity?	
1.12	Do the distractors avoid overlapping content with the key or other options?	
1.13	Is there no use of ambiguous options such as "all of the above" or "none of the above"?	
1.14	Is the key indisputably correct, with no ambiguity?	
1.15	Does the key align precisely with the stimulus and stem?	
1.16	Is the key not overly detailed, technical, or distinct in tone compared to the distractors?	
1.17	Does the key address the specific skill or knowledge being assessed?	
1.18	Is the item clear and free from ambiguity, ensuring a single, consistent interpretation?	
1.19	Does the item encourage detailed, thoughtful, and structured responses?	

1.20	Are expectations for responses clearly communicated (e.g., "Provide at least three reasons," "Analyze the situation using the given framework")?	
1.21	Does the item align with the cognitive level being assessed (e.g., application, analysis, evaluation?	
1.22	Does the case provide relevant, real-world context that aligns with the learning objectives?	
1.23	Is the case information structured logically, with clear and concise details?	
1.24	Are there any extraneous or irrelevant details in the case that might confuse test-takers?	
1.25	Does the case require the application of knowledge to a new or complex situation?	
1.26	Are all key terms and references within the case clearly defined?	
1.27	Are prompts sequenced logically if multiple parts are involved (e.g., "Describe the problem, propose a solution, evaluate its feasibility")?	
1.28	Are response formats specified where applicable (e.g., "Write a short essay," "List three recommendations")?	
1.29	Are scaffolding questions provided if the item is designed to assess complex skills?	
1.30	Do any hints or guiding questions support but do not directly answer the main question?	
2	Content alignment	
2.1	Is the item aligned with the specified cognitive domain?	
2.2	Is the item aligned with the specified instructional objective?	
2.3	Is the content consistent with subject-specific benchmarks?	
2.4	Does each objective have adequate and relevant content supporting it?	
2.5	Does each item have a clear instructional purpose that aligns with assessment objectives?	
3	Content Validity	
3.1	Is the item clear, and factually and conceptually, correct?	
3.2	Does the item reflect current and accurate subject-specific conventions and standards?	
3.3	Does the item avoid cultural, gender, or socioeconomic stereotypes?	
3.4	Is the numerical data in the item realistic and error-free?	

3.5	Are any references or citations used in the item accurate and relevant?	
3.6	Is the content relevant to the learner's context, age, and prior knowledge?	
4	Language	
4.1	Is the vocabulary used in the item contextual, age-appropriate, and sensitive to bias?	
4.2	Is the language of the item concise and free from unnecessary complexity?	
4.3	Is technical terminology, if used, explained or contextualized?	
4.4	Does the item minimize ambiguity to avoid misinterpretation?	
5	Quality Concerns in Item Art (Graphics) and Other Stimuli	
5.1	Is art/stimulus necessary, relevant, and useful to answer the item?	
5.2	Is art/stimulus likely to be interesting and engaging to students?	
5.3	Is art/stimulus pitched at an appropriate grade, age, or reading level?	
5.4	Is the art/stimulus clear, accurate, and sufficient to answer the item?	
5.5	Does the art/stimulus avoid being confusing or overwhelming?	
5.6	Does the art/stimulus contain appropriate and accurate labels?	
5.7	Is the art/stimulus free of errors or inconsistencies that might mislead students?	
5.8	Does the art/stimulus avoid clueing the correct answer to an item?	
5.9	Is the art/stimulus significantly free from copyright or intellectual property issues?	
6	Item Bias and Sensitivity Issues	
6.1	Is the item accessible to the greatest number of test-takers, including those with diverse abilities and backgrounds?	
6.2	Is the item free from bias in the areas of gender, caste, religion, socio-economic class/status, regional diversity, age, culture, or physical appearance?	
6.3	Is the item language unbiased towards a particular linguistic group?	
6.4	Is the item sensitive to the needs of special-needs groups, including physical, cognitive, or sensory disabilities?	

6.5	Does the item avoid offensive, disturbing, or controversial information that may alienate or distress test-takers?	
7	Formatting	
7.1	Is the font size appropriate (e.g., 12-14 pt for text, larger for headings) and consistent throughout?	
7.2	Is a clear and legible font type used (e.g., Arial, Times New Roman, Calibri)?	
7.3	Are items spaced adequately to avoid visual clutter (e.g., 1.15–1.5 line spacing)?	
7.4	Is there sufficient white space around each item to improve readability?	
7.5	Are margins uniform and appropriately sized?	
7.6	Are the stimulus, stem, and options distinctly separated and visually organized?	
7.7	Are all items numbered consistently and sequentially (e.g., Arabic numerals or letters for sub-parts)?	
7.8	For multi-part questions, are sub-questions appropriately labelled (e.g., a, b, c)?	

2.8 Moderation of Question Paper (Step 7)

Moderation is the process of reviewing and refining question papers to ensure they are high-quality, relevant, and practical for exams, whether at the school level or for board examinations. It involves examining the overall structure of the paper to confirm that it aligns with the exam's blueprint and design and meets the required standards.

This step is crucial because errors or biases can sometimes find their way into the paper-setting process, often due to personal preferences or a lack of technical knowledge. Moderation helps identify and correct such issues, making the question paper a reliable and valid tool for assessment.

Effective moderation requires subject experts who have a thorough understanding of paper-setting techniques and the exam's instructional objectives. These experts ensure the question paper serves its purpose, is practical to administer, and can be accurately scored and processed.

Key elements reviewed during moderation include:

1. The questions.
2. The answer key and marking scheme.
3. Instructions provided to students.
4. The paper's format and structure.
5. The clarity and language of questions.
6. Adherence to the specification as laid out in the Design and Blueprint

For school-based exams like term-end or summative assessments, moderation is often done internally by subject teachers and senior educators. However, for board exams, which have higher stakes, external moderators are necessary to maintain confidentiality and ensure unbiased reviews. External moderators must:

- Be different from the paper setter.
- Have deep subject knowledge.
- Be familiar with the syllabus and instructional goals.
- Be trained in paper-setting processes.
- Have proven skills in reviewing exam items.

Moderators should not make arbitrary changes or superficial edits. Moderators should respect the work of the paper setter and only make necessary adjustments to correct typographical errors and repair "fatal flaws" that would prevent the questions from functioning as intended. Untrained moderators who lack expertise in instructional objectives and paper-setting principles can unintentionally compromise the quality of a well-designed question paper.

Moderators should have access to the Exam Design and Blueprint so they can accurately gauge the contents of the questions against the needs of the design. They should also be provided with information about the classification of individual questions with respect to subject areas, specific competencies, and cognitive levels. Classification information can be provided either in the form of a text document or a spreadsheet but should clearly identify the intended classification for each individual question in the question paper.

Moderators should also be given sufficient time to do a thorough review of both the questions and the marking schemes. If the moderator wishes to propose extensive changes, it may be necessary to have a "resolution meeting" during which possible changes are discussed and agreed upon.

Moderation is a vital step in the process of creating high-quality question papers and should never be an afterthought.

3 Conclusion

As this document indicates, the task of creating balanced, high-quality question papers is a complex and multifaceted endeavor. It involves a systematic process that requires the coordinated efforts of numerous individuals, each playing distinct and crucial roles. The creation of assessments should never be an afterthought or conducted in a hurried manner. Instead, it demands careful planning, meticulous attention to detail, and a commitment to educational integrity.

The process begins with the clear definition of learning objectives and the development of a comprehensive blueprint that outlines the scope and structure of the assessment. This blueprint serves as a guide for item writers, who craft questions that are aligned with the specified objectives and designed to measure a range of cognitive skills, from basic recall to higher-order thinking.

Once the questions are developed, they undergo rigorous review to ensure clarity, fairness, and alignment with the intended learning outcomes. This step is critical to identify and eliminate any potential biases or barriers that could disadvantage certain groups of students. Pretesting, when possible, also provides valuable data on the performance of individual items, which can be used to refine and improve the assessment.

The final stages of the process include assembling the question paper, preparing the scoring key or marking scheme, and ensuring that all logistical aspects of test administration are in place. Throughout this process, it is important to maintain a focus on the ultimate goal: creating assessments that provide all learners with a fair opportunity to demonstrate their competencies.

In conclusion, the development of high-quality assessments is a deliberate and collaborative effort that requires expertise, precision, and a deep understanding of educational principles. While decisions about specific aspects of the process and the final product may vary, the overarching objective remains the same: to create assessments that are valid, reliable, and equitable, thereby supporting the goal of achieving comparable learning outcomes for all students.