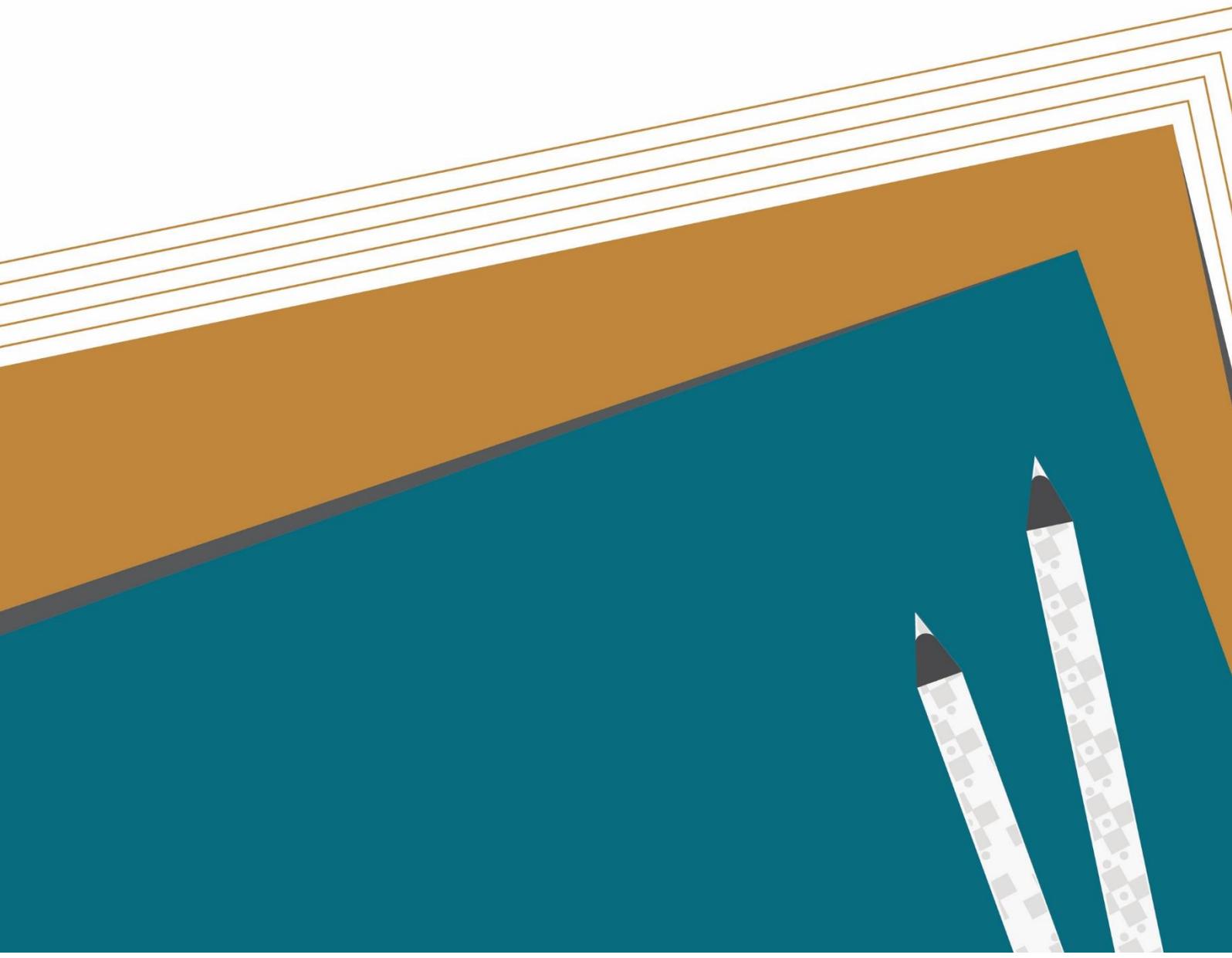


Guidelines for Question Paper Development



Guidelines for Question Paper Developmentⁱ

Introduction

A question paper is the most commonly used assessment tool. Question papers could be used for summative assessments such as conducting an evaluation at the end of a unit, semester or year. It could be used for diagnostic purposes, to understand the prior learning of a student before the teacher decides to teach a new concept. Research organizations and field based organizations may develop a question paper for assessing its program effectiveness. It may be administered as a pre-test or post-test to track the impact of an intervention.

Designing a good question paper is an art. It requires an understanding of the subject that is being tested, the way children learn the subject, the principles of designing good items and the procedure of developing a balanced question paper that matches students' ability levels. To enable this, there are few basic steps to follow in designing a question paper. Assessment is conducted to gather information about students' learning. This information can be used for identifying student's misconceptions, trace learning gaps, report learning levels, certify students and select them for future academic programs or scholarships. When a high-quality assessment is designed, it provides maximum information about students learning. Quality in assessment is characterized by validity and reliability.

Process of Developing a Question Paper

There are 4 clear steps in the development of a question paper. These can be moderated based on the purpose of the assessment and the scale at which the test will be administered.

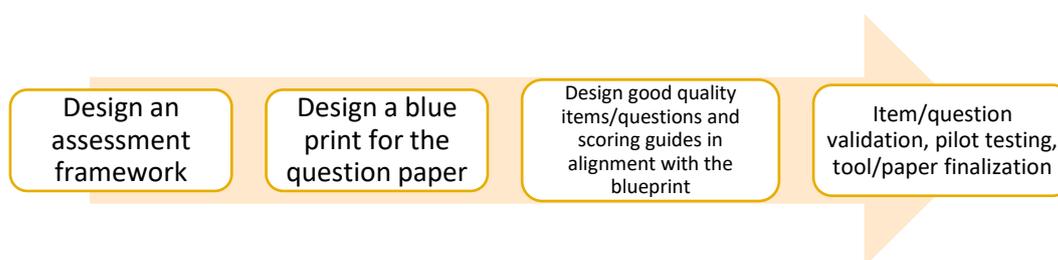


Figure 1. Steps of Question Paper Development

Step 1- Designing an assessment framework

Developing an assessment framework is the first step towards the creation of a question paper. Assessment frameworks ensure a well-articulated basis for making decisions regarding *what to test and what not to*. The frameworks are based on the objective of the test- whether it is a summative assessment, a standardized certification exam, a diagnostic test, etc. In most cases it is developed by articulating the *test construct*, listing out assessable outcomes/competencies to test within the construct from the curriculum and organizing them systematically. An assessment framework details out the learning outcomes, content domains, specific competencies, and skills and /or dispositions to be assessed.

1. Defining a Test Construct

It is the theoretical attribute measured by the test. (AERA standards, NCME,1999). It is based on the purpose of the assessment. The level at which one pitches the test construct should be guided by stage specific curricular considerations, grade specific outcomes and the ability level of the students. Let us evaluate 3 scenarios from mathematics where test construct can be pitched at different levels based on the purpose of the test.

<p>Scenario 1: Purpose: Summative Assessment; end term evaluation. Test construct: Mathematical Ability</p> <p>The Content Domains selected for this test construct are derived from the NCERT syllabus of primary level mathematics.</p>	<ul style="list-style-type: none"> Number sense Measurement Geometry Data handling Patterns
<p>Scenario 2: Purpose: Summative Assessment; at the end of a unit/semester. Test construct: Numerical Ability</p> <p>The Content Domains selected for this test construct are derived from the NCERT syllabus of primary level mathematics.</p>	<ul style="list-style-type: none"> Number concepts Addition and multiplication Subtraction and division Framing problems Solving problems
<p>Scenario 3: Purpose: Diagnostic Assessment; before starting a new topic/concept that is linked to the test construct. Test construct: Data Handling</p> <p>The Content Domains selected for this test construct are derived from the NCERT syllabus of elementary level mathematics.</p>	<ul style="list-style-type: none"> Recording data Tabulating data Representing data Interpreting data Inferring data

Figure 2. Sample Test Construct, Mathematics, Primary Stage

In each of the above 3 scenarios the test developer can tweak the proportion of questions from the content domains selected on the basis of the test objectives and ability of the students. A test construct can also be inter-disciplinary- it can draw its content from multiple subjects.

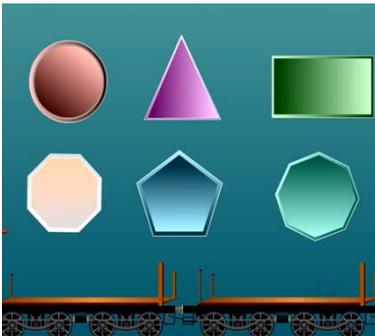
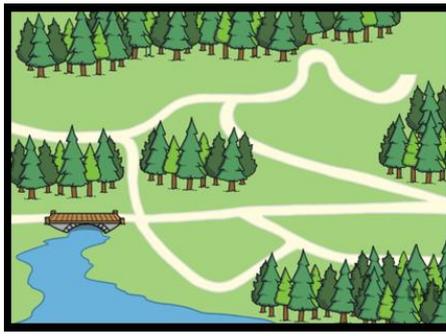
Test Construct: Spatial Understanding		
Language- vocabulary related to understanding space e.g. top, bottom, far, near, etc.	Mathematics- understanding of various 2D & 3D shapes	EVS- interpreting spatial patterns e.g. river/trees/other attributes organized in space
		

Figure 3: Sample Construct, Inter-disciplinary

2. Detailing the framework

Once the test construct is decided and the broad content domains identified, the test developer can detail the other aspects of the framework, namely the sub-domains (if any), learning outcomes, indicators, cognitive levels, etc. The level of detailing in the framework depends on the test developer. The idea is to provide as much information that will help in designing a blueprint to enable item writing (*more details in Annexure*).

Illustratively, template 1

Content Domains & sub-domains	Learning Outcomes	Indicators	Cognitive Levels

A test developer may also detail this further by adding details of the no. of questions and possible context of the questions,

Illustratively, template 2

Content Domains & sub-domains	Learning Outcomes	Indicators	Cognitive Levels	No. of items	Context/Stimulus to be used in the item

3. Review assessment framework

Once the framework is developed, it is useful to review it against these parameters-

1. Are the content domains **adequately representing** the test construct? (Domain representation)
2. Is there any content domain or learning outcome that is **not relevant** to assess the test construct? (Domain relevance)
3. Are the learning outcomes **distributed across different cognitive levels**?

Step 2 - Designing a blueprint

Based on the assessment framework, the test developer creates a blueprint which is a planning document where all the relevant information of a test is listed in the form of a table. The blueprint is usually a working document which undergoes changes during the process of item designing. The information in the blueprint pertains to the content domains that will be tested; particular learning outcomes and indicators that will be tested; item format (Multiple choice, short written answers, others); length of the test; aggregation of the percentages of content areas and cognitive processes. A sample template is attached in the annexure.

Step 3 - Designing good quality items/questions and scoring guides

As a test developer you may decide on the item format/item type that you would like to use in the test. Broadly item formats are of 2 kinds- Selected Response questions like Multiple Choice Questions, True/False where student has to select the correct response from the options provided and Constructed Response questions- where the student has to produce the correct response. These could be as short as one-word responses to long essay type questions. A few quality parameters should be borne in mind while developing test items:

1. **Clarity in language used**- the question should be worded in simple and unambiguous language. The instruction should be clear and there should be no extraneous, irrelevant information in the question. Such additional information may simply confuse the student and not add any real value to the outcomes being tested. It is also important that the language used should not cause any bias towards students or any ethical issues. It should be sensitive and inclusive to all children. The vocabulary used in the question should also be contextual and age appropriate.
2. **Quality of Distractors or response options used in MCQ questions**- it is important that the distractors used in the MCQ's are plausible and represent common misconceptions that students may have about the concept being tested. Very often distractors may be irrational, or

obvious incorrect answers, or subsets of an incorrect/correct option. In such cases, the right answer becomes obvious which in turn affects the quality and difficulty of the item.

3. **Choice of Stimulus Materials Used**- stimulus materials refers to the graphics, illustrations, tables, maps, images, etc. used in a question. These materials aid the student in responding to the question. It is therefore necessary that these are clear, unambiguous, factually and conceptually accurate and pertinent to the question. Including illustrations in the paper unnecessarily does not add any value to the item and should therefore be avoided.
4. **Factual and Conceptual Accuracy**- it is important for the item to be factually and conceptually accurate. Care should be taken that the item tests a competency/skill rather than mere recalling of memorized facts and information.

The scoring guides/marking scheme is as important as the question. For questions, particularly those assessing understanding and other cognitive levels, it is necessary to include variations in student responses. There should be scope for fully correct, partially correct, levels of partially correct responses. Also, the scoring guides should be aligned to the cognitive level of the item. If an item is testing application of a concept, the scoring guide should illustrate the possible responses that could be representing application. At any given point of time the scoring guide would only be indicative – it may not include an exhaustive summary of all possible responses. (*sample attached as Annexure*)

Step 4 - Question paper validation and finalization

Depending on the purpose of the test, various processes for tool validation can be adopted. If it is a standardized test, pilot testing the tool is recommended. Pilot testing means administering a few sample items/full test paper with a representative sample of students. This sample should represent the cohort that would ultimately write the test. During the pilot testing, it must be observed if children are finding it difficult to read the question, or understand the language, or the stimulus materials. The time taken to complete the test paper should also be noted. This in turn helps in fixing the time for the final test. Another process of tool validation is panelling – this involves a review of the question paper with subject experts, teachers and assessment experts. In panelling, the focus is on examining the accuracy of the paper, the alignment of the questions with the cognitive levels, whether the question paper is balanced for diverse learners, clarity of language, images, illustrations, duplication of items, etc. Scoring guides/marking schemes are also reviewed along with the questions to check whether it is exhaustive. This process can be adopted for school based summative assessments- teachers of a grade/subject could collaborate and review these questions. In case the question papers are being developed for standardized assessments, advanced techniques of item validation like Classical Test Theory or Item Response Theory should be used. These theories enable in examining item functioning by providing item level quality indicators and also help in predicting the overall reliability of the test.

Once the test papers are panelled or piloted and all relevant suggestions incorporated, the tool/question paper is ready to be administered.

Summary

The task of question paper development should be approached systematically. There should be alignment across all the steps of question paper development in order to make it a valid and reliable tool of assessment. The process of question paper development is iterative and it requires ongoing scrutiny and review of the test construct, the different elements of the framework, the blueprint and its alignment with the framework/construct; and the quality of items and scoring guides.

Collaboration among teachers in a school or subject matter resource persons in an organization can fetch a high quality paper. However, the most important pre-requisites for developing a good question paper is a sound understanding of the subject being tested, the curricular aims and objectives of teaching the subject, learning outcomes associated with the subject and grade, an understanding of student misconceptions in the concepts being tested and a working knowledge of designing assessments.

Annexure

1. SAMPLE ASSESSMENT FRAMEWORK, TEST CONSTRUCT: LANGUAGE ABILITY (PRIMARY LEVEL)

Content Domain	Learning Outcome	Indicators	Cognitive Level	No. of items	Context/Stimulus
1. Oral expression	1.1 Recites poems, jokes and riddles	1.1.1 Recites familiar poems with gestures 1.1.2 Recites unfamiliar poem that she/he may have learnt from home 1.1.3 Uses appropriate intonations (pauses during commas, voice modulation during !, ?)	Remember/ Understand Remember/ Understand Understand/ Apply	3 – Oral activities	A familiar poem learnt in school
2. Reading Comprehension- Story	2.1 Reads a story independently and summarizes its key ideas	2.1.1 Reads a grade appropriate text with fluency 2.1.2 Responds to retrieval and interpreting questions based on the story 2.1.3 Summarizes the story in her own words(orally/writing /drawing)	Understand Understand Understand	5 items	1 story from the prescribed textbook, 1 unfamiliar story with contextual theme, age/grade appropriate language complexity
3. Written/creative expression	3.1 Writes short paragraphs based on visual clues	3.1.1 Constructs/creates 4-5 sentence paragraph based on visual clues 3.1.2 Can give a	Create Apply/Analyze Apply/Analyze	2 items	2-3 sequence of visual clues. Child to choose any one and perform the task

		suitable title to the paragraph 3.1.3 Can give a suitable theme to the visual clues			
--	--	--	--	--	--

2. SAMPLE BLUEPRINT

Content Domains	Cognitive Levels						No. of questions
	Remember	Understand	Apply	Analyze	Evaluate	Create	
1. Oral expression	1	1	-	-	-	-	2 (15%)
2. Reading Comprehension- Story	-	1	1	1	1	-	4 (35%)
3. Reading Comprehension- Newspaper Report	1	1	1	1			4 (35%)
4. Written/creative expression	-	-	-	-	-	2	2 (15%)
No. of questions	2	3	2	2	1	2	12 Qs.
Marks	2*1=2	3*1=3	2*2=4	2*3=6	1*3=3	2*3.5=7	25 marks

3. CHECKLIST OF ITEM AND TOOL REVIEW/VALIDATION PARAMETERS

1. Simple, clear and unambiguous language

- The language used in items must be accessible to the teachers taking the test
- Technical word/vocabulary as appropriate for the subject must be used and the instructions given in the question must be clear and unambiguous
- There should be no spelling errors or grammatical mistakes
- There should be no language insensitivities/bias favouring/dis-favouring a caste/class/gender/religious/geographic/ethnic group

2. Familiar Context

- The context of the questions should be “real-world/classroom scenarios” that teachers may be encountering frequently

3. Factually accurate and Relevant Graphics

- Diagrams and graphs should be accurate (to scale unless otherwise noted- as *NOT TO SCALE*).
- Diagrams and graphs should be correctly and fully labeled
- Any graphics/illustrations included in an item should be necessary in order to solve the problem or to answer the question and need to be adequately explained and referred to directly within the item
- In questions on policies, programs, or other data related questions, the facts should be accurate

4. Higher Cognitive Level Questions

- The cognitive level of items should be conceptual understanding, application and other higher order thinking skills
- Factual questions testing memorized facts must be minimized
- Care must be taken that the cognitive level of the question corresponds to the learning outcome/indicator

5. Distribution of Item Difficulty

- Items must be of different difficulty levels
- They can range from 30% easy, 40% moderate and 30% difficult

6. Multiple Choice Questions

A. Item Stem

- Sufficient information should be provided in the stem to make the item clear and unambiguous
- The stem should not include extraneous information (Extraneous information is liable to confuse students who might otherwise have determined the correct answer).

- Negative stems to be avoided — those containing words such as NOT, LEAST, WORST, EXCEPT, etc. If it is absolutely necessary to use a negative stem, highlight the negative word. If the stem is negative, use only positive response options—do not use double negatives.

B. Response Options (Distractors)

- There should be only one correct or suitable answer
- The correct option should be distributed across options A, B, C, D almost equally to minimize chances of easy guessing
- The four response options should be independent of each other. For example, response options should not represent subsets of other options. Eg. one option as river and another as water bodies or one option as acids, and another as sulphuric acid
- “None of these” and “All of these” as response options to be avoided
- Common errors/misconceptions should be incorporated as distractors
- Matching words from question stem that make some responses easy to guess/exceptionally long or short responses in comparison to other distractors, must be avoided

7. Constructed Response/Descriptive Questions

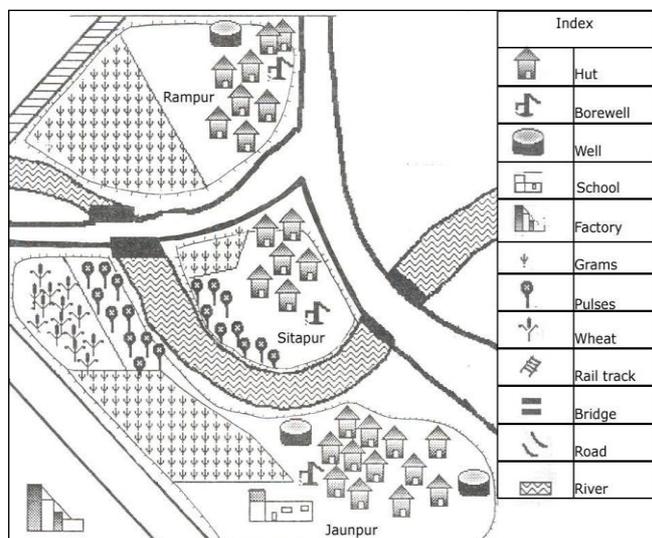
- The scoring rubrics should include all possible correct options
- The performance levels described against each criterion should be properly indicated/graded

8. Overall Tool

- The tool should be sequenced from easy to complex questions
- Questions with diagrams/illustrations should be interspersed with dense text items
- Tool should have a distribution of questions across cognitive levels
- Font size should be appropriate- space to be used judiciously to avoid unnecessary lengthiness
- Questions with illustrations should be on the same page or on adjacent pages
- Scoring guide/marking scheme of the tool should be well distributed between options A, B, C and D

4. SAMPLE SCORING GUIDE

Look at the map and answer the question- Why do you think Jaunpur village has 2 wells?



Source: Learner Assessment Study, Azim Premji Foundation

Correct answer: Code A/Full marks

Child gives a reason related to the village Jaunpur as it appears in the map-

- E.g.:
1. as there are more houses/huts/people/population so there is need for more water
 2. more number/variety of crops, so water is needed for crop cultivation
 3. presence of a factory/school in Jaunpur village, hence need for more water
 4. because the neighbouring village Sitapur has no well, hence people are dependent on Jaunpur village

Partially correct answer: Partial marks

Code B1: child gives a reason that refers to general use of water without relating it to the map

E.g. 1. Because people need water for drinking, washing, bathing, cooking, etc.

Code B2: child refers to scarcity of water:

To avoid drought /famine conditions

To avoid problems arising due to water shortage

Incorrect answer:

Code C1: child gives a reason which shows that the map has been **WRONGLY** interpreted:

E.g. 1. No river/ other source of water in Jaunpur village

2. less number of people live in Jaunpur

3. Because Jaunpur is a large city

Code C2: any wrong/irrelevant answers:

E.g. 1. Children fell down in the village

2. Wells are used as dustbins

3. It makes the village look beautiful

Code D: blank/unattempt.

References

1. Azim Premji Foundation, 2020, Discussions that emerged during the online Certificate Course on Student Assessment in Practice, Oct 5th-30th, 2020 offered by Azim Premji University.
2. Chomal, A. Azim Premji Foundation, 2013, Large Scale Assessments, Guidelines for Design, Implementation and Institutionalization, submitted to MHRD.
3. Azim Premji Foundation, (2007, 2010) Workshop- Processes of Test Development, conducted by Australian Council of Educational Research, Bangalore, India.
4. National Council of Educational Research, & Training (India). NCERT, 2017. Learning Outcomes at the Elementary Stage.

i This article is authored by Aanchal Chomal. She works as Associate Professor in School of Continuing Education at Azim Premji University. She can be reached at aanchal@azimpremjifoundation.org

This article can be cited as –

Guidelines for question paper development, Assessment resources, Azim Premji University

Guidelines for question paper development

