Competency Based Assessment
Test Item Development Guide
for Schools

Department of Curriculum and Professional Development
&
Bhutan Council for School Examinations and Assessment
Ministry of Education
Royal Government of Bhutan

Competency Based Assessment
Test Item Development Guide
For Schools

Department of Curriculum and Professional Development
&
Bhutan Council for School Examinations and Assessment
Ministry of Education
Royal Government of Bhutan

Published by:

Department of Curriculum and Professional Development &
Bhutan Council for School Examinations and Assessment
Ministry of Education
Royal Government of Bhutan
Thimphu, Bhutan.

Advisers
Mr. Tashi Namgyal, Director, DCPD, MoE

Research and writing:
Mr. Wangpo Tenzin, Curriculum Specialist II & Dean, DCPD, MoE, Thimphu
Mr Arjun Kumar Gurung, Educational Assessment Specialist, BCSEA, Thimphu

Reviewers:
Curriculum Developers, DCPD, MoE, Thimphu
Examination and Assessment Officers, BCSEA, Thimphu

Type setting:
Mr. Wangpo Tenzin, Curriculum Specialist II & Dean, DCPD, MoE, Thimphu

Layout and Design:
Mr. Jamyang Drukda, Sr. Programme Officer, Bhutan National Commission for UNESCO, MoE.
Ms. Karma Yangden, Sr. Adm Assistant, Bhutan National Commission for UNESCO, MoE.

ISBN: 978-99936-0-665-9

FOREWORD

The Department of Curriculum and Professional Development as of one the professional bodies in the country has the mandates to develop cutting edge school curricula, informed and inspired by rigorous educational research studies and national values and priorities. Provide transformative reorientation in the effective implementation of the school curricula leveraged on digital technologies and educational programs, as vital support paraphernalia. Through such mandates, the DCPD strives to improve the quality of education and transform the educational materials keeping abreast of the changing time and needs of society and the nation.

Assessment forms a critical component in education as it influences its quality. The cognitive challenging complexity in assessment has direct impact on the teaching and learning in classrooms. For instance, content based or traditional question items in assessment encourage learners to learn through rote memorization, which largely fails in achieving the prescribed learning standards. On the contrary, competency-based assessment is cognitively challenging, and hence encourages metacognitive learning and learners learn through understanding. However, discrepancies in reliability and validity in assessment, including the procedures and assessment utility in enhancing teaching and learning have been generic concern in education system in Bhutan.

Underscoring the existing and emerging issues and concerns in assessment, it is envisaged the critical need for a National Student-learning Assessment Framework (NSAF) and Test Item Development technical guide for teachers and students. It is believed that a systemic and standardised assessment with purposeful and reliable test item guarantees that the marks learners secure commensurate the learners' mastery of various competencies, including the "portable skills." At the same time, understanding that the diversity is the common thread amongst learners, every opportunity must be extended to learners with diverse needs and interest to display their abilities and opportunities through the deployment of authentic and alternative assessment methods and tools.

Thus, to the optimism of the Ministry, the "Test Item Development Guide for

Schools" has been published as resource to augment the effective implementation of authentic and reliable assessment practices in measuring the learners' learning outcomes enshrined in the National School Curriculum Framework. It also informs the education system of the processes and the diverse types of assessment practiced in schools.

Tashi Delek

DIRECTOR

TABLE OF CONTENTS

1. INTRODUCTION

The purpose of school education in Bhutan, according to the draft National Education Policy (2019) is to prepare young men and women with the right values and skills, usefulness, and gracefulness, to meet the ever-growing needs of a developing country against the backdrop of a fast-globalizing world. It is to build a cadre of highly motivated and competent educators who are endowed with an abiding love of children, a deep love of learning, and who passionately value education as a positive instrument of empowerment. Towards this, it is to promote a system of continuous and life-long learning through formal, non-formal as well as informal modes to empower youths with competencies to participate meaningfully and constructively in life and in the society. Thus, the bigger goal of education is to build "…an educated and enlightened society of Gross National Happiness at peace with itself, at peace with the world, built and sustained by the idealism and the creative enterprise of our citizens" (NEP 2019).

In pursuit of realising the intended purposes of education, assessment forms the critical component. The NEP (2019) envisions that "For all their imperfections and inadequacies, examinations and assessments are here to stay at least till the foreseeable future. They influence the scope, delivery, and emphasis in the implementation of the curriculum. Examinations also guide learners' notion of success and failure. What ought to be emphasized though is the integrity of examinations and assessment themselves considering the soundness and rigour which inform selection and preparation of assessment items, and their administration, evaluation, and certification to satisfy professional standards and public scrutiny" (pp 18).

The relevance of competency-based education is quintessential of the 21st century education, and inspires the shift in the purpose of curriculum from knowledge acquisition to knowledge application, and the process shifts

from the teacher to the student. The impetus of the shift has been because the emphasis of traditional tradition is on preparing learners for jobs and labour market, independence, individual fulfilment. Learning emphasises on academic disciplines, and progression in learning is linear and standardized. The competency-based education on the other hand, accentuates how students perform on measures of competencies. It focuses on the real-world aspect of applying the knowledge, rather than just on the ability to recall the knowledge. The role of teacher shifts from knowledge transmitter to that of facilitator, guide, and motivator, while students have the shared responsibility for their learning. Consequently, the focus of the assessment process shifts from emphasis on summative to formative, and evaluation is criterion referenced rather than norm-referenced. In order to effectively measure competencies and determine that a student has mastered them, CBE programs must offer multiple measures of those competencies with multiple assessment methods.

The measure of competencies must not only measure one's knowledge and skills, but also get at one's ability to integrate, synthesize, and use the knowledge and skills necessary to become part of a community of practice. In other words, each competency must be measured more than one time, and in more than one way which may include multiple choice tests, papers, presentations, performance-based real-world assessments, etc. The tests must also focus on clusters of competencies as they might appear in the real world rather than on just one competency.

Test item as defined elsewhere is a specific task that the test takers are asked to perform by which one or more learning objectives are assessed. Test items consist of two components: the content (what is asked) and the format (the way it is asked). The quality of the test item depends as much on the way the test item is constructed as on the content. Important topics that are tested by ambiguous, awkward, or poorly specified test items cannot be considered valid.

Therefore, the selection of valid topics and the careful construction of test items are equally important parts of a valid test development process.

As the post activity of the assessment, test item analysis which is determined by the objective of the assessment is crucial to augment the validity and reliability of test items. It is generally perceived as a statistical analysis of student's responses on exam items and the relationship between them. Collection and summarization of students' responses can provide quantitative objective information that is useful in deciding the quality of the test items and increasing the assessment's efficiency. It provides constructive feedback about items quality, increases the effectiveness of the exam, and supports exam validity and reliability (Rezigalla, A.A, 2021).

The National Curriculum Framework (REC, 2018) makes provisions for national testing to take place during the schooling and at the end of each key stage. This is imperative to ensure that our children develop their full potential according to their individual abilities and interests, considering that they too have an important role to play in their own development.

Purpose of the Guide

Designing tests is an important part of assessing students understanding of course content and their level of competency in applying what they are learning. An effective test development requires a systematic, well-organized approach to ensure sufficient validity evidence to support the proposed inferences from the test scores (Downing, S.M, 2006).

Therefore, the Test Item Development Guide is purported to inform teachers and others on the design of effective and meaningful test items as stated in the following section:

i. Provide comprehensive technical guide to test developers, and teachers and students alike on the test development processes, test item development planning, item specification, identification of test item types and format, item development, and test item validation including the test item analysis.

ii. Based on the outcomes of the item analysis, teachers are informed of competencies developed, efficacy of instruction and interventions in enhancing the quality of students' learning.

iii. Maintain consistency between goals for the course, methods of teaching, and the tests used to measure achievement of learning outcomes. Based on the purpose of learning, focused test items can be designed for students to demonstrate how well they have learned. A multiple-choice test can assess memory and recall, for example, and essay or open-ended problem-solving to demonstrate more independent analysis or synthesis.

iv. Design test items that allow students to show a range of learning. That is, students who have not fully mastered everything in the course should still be able to demonstrate how much they have learned.

v. Ensure that the assessment provides authentic evidence by maintaining the rigour of reliability, validity, authenticity, and flexibility.

vi. Use evidence to generate information about students' learning which inform issues and challenges in learning and suggest appropriate interventions.

2. THE PARADIGM SHIFT IN EDUCATION

"We must revisit our curriculum, pedagogy, learning process, and assessments to either transform or rewrite them in view of the challenges and opportunities of the twenty-first century. Otherwise, continued focus on textbooks and content without integrating technology and social learning risks perpetuating passive modes of learning.

By developing their abilities for critical thinking, creative thinking, and learning to be life-long learners, we have to prepare them to be inquisitive, to be problem-solvers, to be interactive and collaborative, using information and media literacy as well as technological skills. We must prioritize self-discovery and exploration, and involve learners in the creation of knowledge rather than making them mere consumers of it. We must make STEM subjects part of their everyday language.

As we prepare to educate and equip them with competencies for the twenty-first century, we must equally prioritize their holistic development so that they become caring, dependable, and honest human beings as well as patriotic citizens. We need to embed in them the conviction and sense of pride as a Bhutanese by grounding them in our country's history, culture, tradition, and value system. In their thoughts, attitudes, and actions, they must live the very ideals and values which define us as a unique nation and people, who have overcome all odds stacked against our survival".

Druk Gyalpo, 17 December, 2021

Drawing inspiration from the Royal Kasho (Dec 2020), one of the significant priorities in education as aftermaths of the COVID 19 pandemic has been the urgency to rethink and refine the purpose of education, and of its design and modalities of how the teaching and learning in schools must ensue. This culminated the shift in tradition of education from pervasive traditional tradition to competency-based education. Though the traditional or factory model of education served the humanity in the past, it has been increasingly perceived

fraught with limitations in the context of the 21st century and the evolving uncertain world. The content of curricula (knowledge) occupies almost the entire field of the schooling, oblivious of the need to establish positive relationship amongst teachers, students, and parents. Accumulation of a large amount of knowledge to progress in the school system has been the hallmark of the traditional tradition. Hence, at school, teaching aims to bring learners to know, but not or very little to know themselves or to live and create together. The result is a split between thinking, on the one hand, and action and sensitivity on the other. Traditional education tends to prioritize the function of thinking and to condition learners to situate themselves exclusively in the outside world alone. It reminds that they remain alien to the inner world of their emotions, of their sensitivities and of their needs, disregarding the essential pointers based on which they could unleash both the creativity and the relational aptitudes that life requires of them (Morin, 1999) (UNESCO IBE, 2016).

The traditional education model emphasizes the mastery of 3Rs – reading, writing and arithmetic, with the goal of preparing learners to the world of work and employment (DoE, 1974). Accordingly, teacher centred approach dominated the instruction delivery, and it focused on acquisition of facts and conceptual knowledge. Repetitive tests and examinations dominated the assessment of learning outcomes of learners. Tests and examinations assessed what do learners know and remember at the end of the year.

2.1. Competency Based Education

The competency-based education tradition on the other hand is to empower learners with the abilities to relate and apply knowledge, skills, and values to solve problems in diverse situation. It is an attempt to broaden and elevate the purpose of education to the development of cognitive, psychomotor, and behavioural competencies. In essence, the competency-based education is a system of instruction where learning focuses on content mastery rather than

mere content completion. Besides instruction, CBE also focuses on meaningful assessments, fast feedback, and is intensely data driven (FTTS 2020).

The contents of the curricula have undergone considerable changes in recent decades. In this era, human interaction and well-being are priorities. Technology, particularly digital technology that enables communication, collaboration and learning across distance, though not a panacea is a formidable tool and the source of innovation and expanded potentials. As we embrace this exceptional opportunity to transform the education, it is imperative to reimagine the organization of our educational institutions and learning environments. The 1996 Delors report, "Learning the Treasure within", perhaps elucidates the purpose of education for the present world. It's four pillars of learning, "learning to know", "learning to do", "learning to be", and "learning to live together" are the current global ethos of teaching and learning. Therefore, curricula must be increasingly perceived as an integrated, themes based and problems-based orientation that allows learners develop a strong base of knowledge about one's self and about the world, and find purpose of life and be better able to participate in social and political milieu.

In tandem with the shift of ideology, a student-centred method of teaching and learning is to focus on mastering specific skills or standards rather than completing course work over a specific period. Competency-based learning is more open to interdisciplinary learning and focuses more on problem solving and critical thinking than traditional "sit and get" learning. Students continue to work on specific skills or knowledge until they can demonstrate their understanding and the ability to apply them. This however calls for shift in culture and thinking about how to educate kids for today, anywhere, anytime. Teacher is no longer as the sole source of the information for the students impact on the necessity to adjust their role from a 'dictator' to a facilitator. This also brings changes to the method on how the teaching–learning process should be conducted. It entails teachers to act and position themselves as facilitator and

motivator to inspire students to be active, critical, independent, and creative to face the challenges in their future (Anggraeni, K.A., 2017, pp. 1).

Through the mastery focused and student-centred approach, students are facilitated to understanding about the journey of their education and develop a wide range of competencies inclusive of cognitive, psychomotor, and affective, as articulated by the CBE Model in Figure 1.

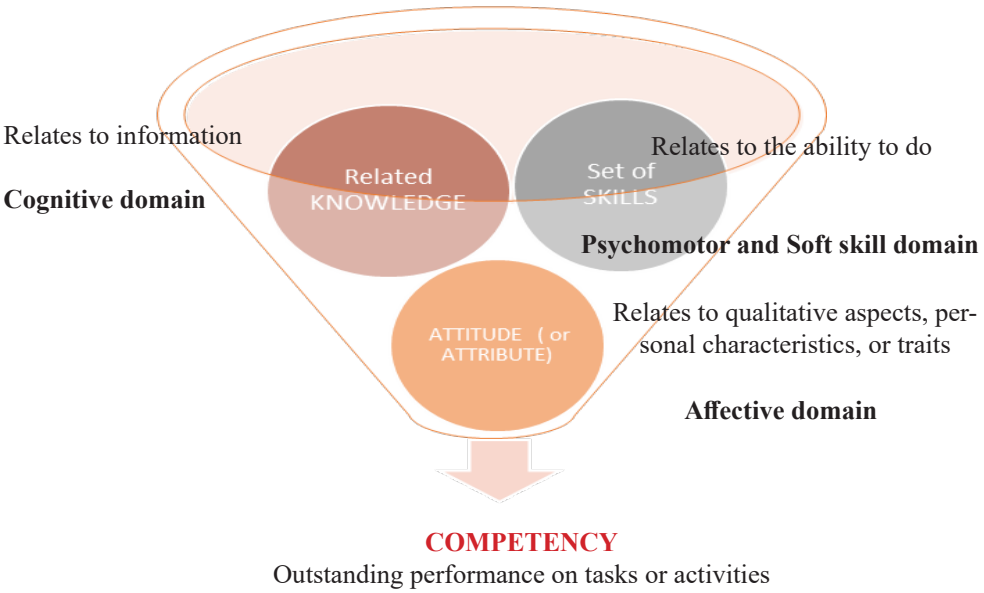


Figure 1: CBE Model

The competency-based learning model recognizes that learning is progressive, "cross-curricular" and integrative. Espouses that learning experiences are built on foundational strengths to develop both tacit knowledge ("know how") and explicit knowledge ("know that"). Knowledge comprises of a set of facts, concepts, ideas, theories, events, and procedures about aspects of the world acquired or learned through experience or instructions. Skill is an ability to choose and perform the right technique in diverse situation – solving problems and generate new knowledge and ideas. Learning experiences provide avenues and opportunities for their development through active engagement in learning and practice. Attitude or attribute is an inherent characteristic or quality which is often expressed through what individual value, think, feel, and respond to motivation appropriately. The CBE integrates academic content and skills with "soft skills" such as critical analysis, creativity, communication, problem solving, citizenship responsibilities and cultural sensitivity so that learners can apply or demonstrate those skills. Therefore, CBE is designed to create self-directed, future-focused, lifelong learners. It premises that the student is the engine of learning.

The ERI_NEQMAP, UNESCO (2016) describes the range of competencies expected to be developed in learners as seen in Table 1.

Table 1: Framework on competencies adapted from ERI_NEQMAP, UNESCO (2016)

Competencies	Description
Critical and Innovative Thinking	Creativity, entrepreneurship, resourcefulness, application skills, reflective thinking and reasoned decision-making
Interpersonal Skills	Communication skills, organizational skills, teamwork, collaboration, sociability, collegiality, empathy, and compassion

Intrapersonal Skills	Self-discipline, ability to learn independently, flexibility and adaptability, self-awareness, perseverance, self-motivation, compassion, integrity, spirituality, and self-respect
Global Citizenship	Awareness, tolerance, openness, responsibility, respect for diversity, ethical understanding, intercultural understanding, ability to resolve conflicts, democratic participation, conflict resolution, respect for the environment, national identity, and sense of belonging
Media and Information Literacy	Ability to obtain and analyse information through ICT, ability to critically evaluate information and media content, sensitivity, and responsiveness to social ills of media, and ethical use of ICT
Other (Physical health, religious values)	Appreciation of healthy lifestyle, respect for religious values

From the above, it may be perceived that the CBE is a shift from the "sage-on-the-stage" model to a "guide-on-the-side" approach. CBE, as a model, focuses on mastery of content rather than completion (FITT, 2020) and prioritises the learning of "how" through the engagement of learners in procedural skills as opposed to teaching of the prescribed content knowledge. This is well articulated in the National School Curriculum (REC, 2021) designed based on the concepts and principles of competency-based education as seen in Figure 2.

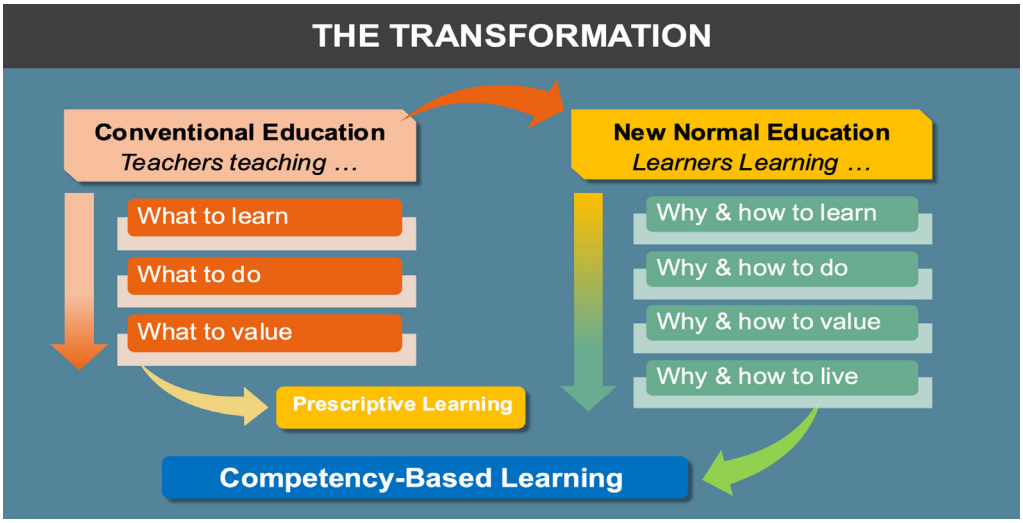


Figure 2: NSC CBE Model

The design and development of National School Curriculum (NSC), is informed by the following tenets of competency-based education (CBE):

i. It is student-centred, and students profit from a competency-based course design by focusing on mastery of the chosen subject rather than mere completion.

ii. It offers opportunities for all students, not just the advanced and those who are struggling. By this arrangement, bright students can move ahead and not be bored.

iii. Students can work on specific skills or knowledge until they can demonstrate their understanding and ability to apply them. Hence, students develop the abilities to apply a broad set of skills, and develops positive dispositions.

iv. Students learn how to learn, a skill they need throughout life which nurtures a growth mindset and the culture of learning.

v. Explicit and transparent learning objectives empower students and improve instruction. Instruction focuses on standards and skills sets priorities that students must develop.

vi. Teachers become skilled at facilitating differentiated learning environments which benefits all students through the deployment of personalised and individualised learning.

vii. Assessment is also aligned and rooted in the cycle of teaching-learning, which enables student to progress and accelerate learning at their own pace and succeed; hence lowers student retention.

viii. Enhances diversity in learning by offering programs in varying modalities for mastery of content and skills and accelerate the achievement of learning outcomes. Upon completion, successful students become productive members of the society ensuring better pathways to employment.

ix. Facilitates schools to adopt multiple modalities of teaching and learning – hybrid, in-class, and online, which allow meaningful interaction and engagement amongst teachers and students, and accelerate learning in their own pace.

x. Students develop confidence that the skills they learn in school will translate to qualifications sought by employers. Therefore, students develop intrinsic motivation which inspire students to put their best efforts forward.

xi. Students receive timely and differentiated support to succeed in achieving the learning outcomes, which ultimately helps to minimise retention and save both time and money.

xii. CBE emphasises on personalized learning. Wherein, students establish learning goals based on personal, academic and career interests. The five essential elements of personalized learning include:

- flexible, anytime/everywhere learning
- redefined teacher role and expand "teacher"
- project-based/authentic learning opportunities
- student-driven learning path
- mastery/competency-based progression/pace

There can be no denying the influence of competencies on the development of school and its curricula. It is increasingly the case in the current time as learners need to imbibe to take part in the fabric of socio-economic development, whether locally or globally, learners– male or female, by relating and using their knowledge and know-how at the service of action. This empowers them with the ability to deal with complex situations of daily and professional life through the art and science of transferring their knowledge and know-how. The school and its stakeholders must therefore be tooled to be able to handle this novelty: conducting learning processes in terms of competencies, but also assessing learners in terms of competencies (IBE-UNESCO, 2016).

2.2 Competency Based Assessment

With the paradigm shift in education, in the context of competency-based education, rethinking of assessment in the two traditions of education is of paramount importance. Traditional assessment systems generally focus on learning outcomes, and predominantly assesses lower order skills, such as Remembering, Understanding and Applying. On the other hand, competency-based assessment is defined as a process which emphasises on the following procedures and characteristics:

- It is based on standards that describe the competency levels. And, standards include competence criteria that provide details regarding the world of work.
- Learning outcomes are mainly higher order skills, such as Analysing, Evaluating and Creating.
- The assessment is individualised and ongoing process rather than a snapshot; there is no comparison among students.
- It is done, preferably, in real working situations - based on direct observation or simulations to evaluate and discriminate competent from not competent.

- It includes the recognition of acquired competencies and identification of needs for improvement and development of skills and abilities.

Competency-Based Assessment is an assessment of student's competence usually inferred from their performance in the assigned tasks, which is generally measured against the prescribed standard and competency indicators. It measures the demonstration of one's knowledge, skills, values, and attitude through proper actions and evaluated against the set of prescribed competencies or learning outcomes defined for respective curriculum or programme.

The geneses of competencies of Bhutan are inspired by the Nine Student Attributes, and the five extraordinary qualities of Bhutanese: Sincere, Mindful, Astute, Resilient and Timeless, which are also embedded in the National School Curriculum Framework (draft 2021). These wide spectra of competencies are articulated in the CBE model, deliberated as the following broad competencies:

2.2.1 Intellectual competencies

Comprises of two components namely, academic mastery and intellectual resourcefulness. Academic mastery in general includes the subject knowledge and literacy, while intellectual resourcefulness fundamentally involves solving problems in diverse context. The assessment of intellectual competence is the assessment of the mastery of learner on the subject knowledge and the abilities of active cognitive processing of the understanding, applying, analysing, generalizing, evaluating, synthesizing information, and creating new knowledge. The assessment determines the ability of student to eventually solve authentic problems of different levels of complexity effectively, using personal qualities and experience in a specific subject area. Assessment of intellectual competency requires a student to demonstrate the following abilities.

i. Acquire conceptual and practical knowledge
ii. Apply knowledge, reasoning, and innovative thinking
iii. Approach challenges with creativity, flexibility, and determination

iv. Solve meaningful, real-life, and complex problems
v. Construct new knowledge
vi. Provide creative and innovative solutions to problems
vii. Work cooperatively and value honesty, fairness, and open-mindedness
viii. Design and manage projects and address issues
ix. Connect and transfer learning from one situation to another
x. Recognize opportunities and imagine possibilities to apply ideas in new ways
xi. Demonstrate initiative, entrepreneurship, resourcefulness, perseverance, and courage to pursue new ideas.

(BCSEA, 2022, pp. 5-6).

2.2.2 Social, cultural and Global Citizenship competencies

Recognizing the increasingly globalized society, it is imperative to emphasize more on social, cultural, and global citizenship competencies in education and workplace. The assessment of these competencies involves the measurement of students' ability to effectively interact, work, and develop meaningful relationships with people in various social, cultural, and cross-cultural settings.

In the competency-based assessment, performance of students is assessed based on the following criteria:

i. Ability to acquire and apply sound knowledge on language, values, traditions, citizenship, and civic rights.
ii. Foresee possible consequences of the actions and accept consequences.
iii. Recognize the influence of beliefs on emotions manifested in behaviours in expression, communication, and sharing.
iv. Handle conflict and recognize common interest.
v. Identify, understand, and evaluate local and global problems, or challenges and act willingly for a common good.
vi. Demonstrate care, kindness, love, gratitude, respect, trust, equality,

fairness, harmony, loyalty, patriotism, national identity, and respect for environment.

vii. Bear sense of belongingness, respect for diversity and the potential for a collective identity that transcends individual economic, cultural, religious, ethnic, or other differences.

(*BCSEA, 2022, pp. 6)*

2.2.3 Information and Technological competencies

In the modern society, everyone needs an increasingly sophisticated set of competencies for finding, handling, and using information. The assessment of these competencies involves the application of a set of knowledge, skills, and attitudes to determine when and what information is needed, how to extract information, organize, interpret, and understand, and evaluate its credibility and authenticity. Assessment of information and technological competencies requires individual to demonstrate the following abilities.

i. Critically evaluates information in terms of nature, benefits, and risk it may pose to oneself and the society.

ii. Recognizes and assess the need for information and determines the nature and extent of the information needed.

iii. Manages information collected using ICT to generate new information, concepts, or understandings.

iv. Uses technology to retrieve, assess, store, produce, present and exchange information in learning and to communicate and participate in collaborative networks.

v. Employs technology to access and use information ethically, legally, and respectfully.

(BCSEA, 2022, pp. 6-7).

2.2.4 Well-being and lifelong learning competencies

The psycho-social and emotional health influence the behaviours and learning of students. Assessment of well-being and lifelong learning involves measurement

of ability of student to manage emotional, intellectual, physical, social, and spiritual aspects of life and living. Assessment also determines capability to organize or structure one's learning, effective management of time, learning information management, and problem solving. The assessment of these competencies is inferred to the attributes of being reflective, resourceful, and optimistic to develop interest, skills and talents and strive for personal excellence.

Although the well-being and lifelong learning of students cannot be assessed objectively, they can be assessed based on demonstration of the following attributes.

i. Demonstrate optimism, flexibility, resilience when adapting to new situations and transitions.
ii. Explore, select, or adapt strategies and resources that support personal growth, safety and well-being of self and others in life.
iii. Identify and understand one's interests, values, skills, emotions, strengths, and weaknesses in setting learning, life, and career goals.
iv. Be self-motivated, exercise discipline, work hard, manage time, and display strong goal setting, leadership, and organizational skills.
v. Identify and analyse a situation and reflect upon the implications of decisions made, based on personal, moral, and ethical considerations.
vi. Recognise duty to him/herself, family, community, nation, and the world, and fulfil responsibilities with love and commitment, and stand up for what is right when deemed.
vii. Act with kindness and compassion for inner peace and happiness, and promote social cohesion by appreciating the unity and diversity of a multicultural society with integrity and perseverance in the face of challenges with learning-to learn attitude.

(BCSEA, 2022, pp. 6-7).

2.2.5 Collaborative and communicative competencies

Communication is crucial means of expression of opinion, views, and ideas in the journey of learning, and to establish social bond and harmony in the society. It is comprised of attributes like information acquisition, vetting its benefit, clarity and sharing of information. The collaborative competencies are the ability to conflict resolution, decision making, problem solving, and negotiation during communication. Therefore, assessment of these competencies to assess the personal disposition of collective intelligence, co-construction of meaning, sharing ideas through various media with respect, empathy and responsibility while working with others to achieve a common goal.

The assessment of these competencies is to assess and evaluate the following abilities of students.

i. Analyse, clarify, and interpret complex information and issues effectively and respond appropriately.
ii. Provide others with clear, concise, accurate, and timely knowledge, information, and ideas by using appropriate media, technologies and communication style that effectively engages and involves audiences.
iii. Ensure full participation and interact with others in a courteous, pleasant, and helpful manner with empathy, understanding and patience in all communications.
iv. Accept and discuss ideas, opinions, concepts, and feedback positively and constructively in a receptive and congenial manner while treating all with dignity, respect, and fairness.
v. Establish and maintain healthy and rewarding relationships through communication, resolving issues, and where necessary, provide assistance effectively.

(BCSEA, 2022, pp. 6-7).

Hence, competency assessment deals with the defined performance criteria and the processes associated with specific tasks. That means that, during an assessment, several performance criteria may be assessed simultaneously for some tasks. To accentuate the effectiveness of the assessment, the CBE programs clearly define competencies and link those competencies to wide range of tasks covered in the assessment of learning outcomes. It is envisaged that the valid test-score interpretations are empirically linked to external measures such as future outcomes, which can be also used in the standard-setting process so that job providers develop benchmark that truly differentiate masters from non-masters.

3. MODES OF ASSESSMENT

Assessment is defined as the systematic collection, interpretation, and use of information about learning. It gives teachers better understanding and awareness of what students know and understand, what learning experiences enable them to do, and what their skills and personal capabilities are. Decades ago, Benjamin Bloom and others constructed what was called the Taxonomy of Educational Objectives. This list divides educational experiences into two main domains, cognitive and affective. Within the cognitive area, it includes matters such as acquiring factual knowledge, being able to comprehend, developing the ability to analyse, to synthesise and to prepare evaluations as being a useful way to structure what we teach and hence what we test (UNESCO, 2005, pp. 13). The assessment hence is a systematic basis for making inferences about the learning and development of students. It encompasses the processes of defining, selecting, designing, collecting, analysing, interpreting, and using information to increase students' learning and development. *(Erwin, 1991)*.

3.1 Assessment Principles

Assessing students' competence in a specialized area requires a given set of skills. Multiple assessment tools in addition to traditional examinations are to ensure the right balance between the uses of formative assessment approaches combined with the summative approach of high-stakes examinations. The marks or grades a certain test item fetches ought to give a true picture of the attainment level of the learner in enhancing the reliability and validity of assessment to ascribe the health of the education system.

Drawing insights from diverse sources (e.g., WMO 2018, Clay, B. 2001), fundamentals of assessment ascertain that the assessments are valid, reliable, flexible, and fair in order to provide authentic and reliable information of

learner's learning outcomes, and of the efficacy of educational practices which are inclusive of curricular materials and instructions.

i. **Validity** refers to the extent to which the interpretation and use of an assessment outcome can be supported by evidence. An assessment is "valid" if it integrates the required knowledge and skills with the practical application in performing the task, which is supported by the evidence gathered.

ii. It is also crucial that the assessment is consistent and "**reliable**". Another way to think of reliability is in terms of "replicability." Is there a general consistency in students' overall performance on an exam? In other words, reliability refers to the level of consistency and accuracy of the assessment outcomes, which provides similar outcomes for learners with equal competence at different times or places, despite being assessed by different assessors. It ascertains that learner can demonstrate competence on more than one occasion, and in more than one context.

iii. In the real-world classroom setting, learners are with diverse needs and disabilities and operate in different cognitive levels. It is justifiable that assessments are "**flexible**" so that learners are provided with the opportunity to negotiate certain aspects of their assessment, resources needed and timing for example, with their assessors. In this context, it is important that through the assessment plan, all learners are fully informed of the purpose of assessment, the assessment criteria, and the details of methods and tools used, including the context and timing of the assessment.

iv. The "**cognitive complexity**" refers to the various levels of learning that can be tested. Test questions need to focus on appropriate intellectual activity ranging from simple recall of facts to problem solving, critical thinking, and reasoning. If the assessor is mainly concerned with students memorizing facts, the test should ask for simple recall of material. On the other hand, if the assessor is trying to test analytic skills, a test that asks

for recall is inappropriate. Tests are to stimulate students to utilize higher order level of questions. These questions require much more brain power.

v. The test questions must permit students to demonstrate their in-depth knowledge of important subject matter in the context of "**content quality**". While planning a test, one must outline the actual course content that the test is to cover. In addition, it is imperative that other considerations are taken care to ensure the content quality. They include matters related to what are the test specifications? What skills do they indicate will be tested? How many questions and how many areas will be covered? How many sections will there be? What formats will be used to test?

vi. In order to motivate students, it is vital that they recognize and understand their value as "**meaningful**" and worth of their time to take part in the process of assessment. Students should not be forced to guess what will be on a test, instead students are provided with specific study questions from which the tests are designed. A few of the research studies show that the less able students are heavily penalized by a failure to realize what is required for a test. The more able students seem to sense what the teacher wants, hence perform better.

vii. The "**language is appropriate**" to the assessment tasks and to students so that students understand the expectations of the assessment task. It is better assured when the test questions reflect the language used in the classroom. Therefore, test items are stated in simple, clear language, free of non-functional material and extraneous clues, and are free of race, ethnic, sex bias, and regional bias.

viii. Successful performance on the test must test the generalization ability of students as evidence of "**transfer and generalizability**". Valid generalizations about achievement can be made by using authentic and performance assessments, such as presentations, scenarios, projects, and portfolios add dimensions. Generally, well-constructed tests, whether they are objective or performance oriented, allow assessors to understand

the needs of students and influence their decision what needs to be taught next.

ix. Finally, it must be noted that the assessment is "**fair**" and does not advantage or disadvantage particular students or groups of students. This mandates that assessment methods are adjusted to ensure that the methods used lend levelled platform to every student to demonstrate their competencies. It is also to ensure that unnecessary demands are not placed on them which may negatively influence abilities to demonstrate their competencies.

It is also acknowledged that the assessments are not random events. They should be embedded in instruction in order to support and enhance learning. If an assessor wants students to solve problems, answer open-ended questions, and perform, then tests must be developed that measure performance.

3.2 Assessment Processes

Assessment is an effective way for the learner to communicate their learning to their teacher, and for the teacher to determine the progress of learning and communicate back to the learner to let them realize where they are and where they are heading.

Evidence is the information gathered which, when matched with the requirements of the competency, provides proof of achievement of competence. Evidence can take many forms and be gathered from several sources. And, assessment methods are the means of collecting the evidence required to demonstrate satisfactory performance. It can be direct, indirect, or third-party supplementary. In fact, applying direct, indirect, and third-party supplementary evidence in combination can be the most effective (and fair) means of assessing an individual's competence. Once the method has been selected, the means for collecting and analysing the evidence are then chosen or designed. These means are called assessment tools. In general, the term assessment tool is used

to describe a document that contains both the instrument and the instructions for gathering and interpreting evidence (WMO, 2018).

For achieving the purposes of assessment in education effectively, it is vital to incorporate the following procedural elements in the design of assessment in respective learning areas.

i. The assessments are norm-referenced assessment, criterion-referenced assessment or standard-referenced assessment determined by the purpose of assessment and evaluation.
ii. Assessment, both informal and formal, explicitly state learners the knowledge, skills and competences that are being assessed in each learning areas, and feedback are provided to learners after assessment to enhance the learning experience.
iii. Diverse assessment techniques and tools are included that provide accurate information of students learning and learning styles; thereby indicating the areas of improvement for learners' progress.
iv. Assessment in each learning area provide teachers with a systematic way of evaluating how well learners are progressing in a particular subject in a grade. It is crucial that assessment criteria are clear to the learners before the assessment process.
v. The forms of assessment used are appropriate to the age and the developmental level of the learners.
vi. Assessment plan in each grade is developed by school and are provided to learners, parents, and relevant officials.
vii. Assessment is demonstrably aligned with learning outcomes in each course and is transparent and what is being assessed is clear to students in a timely manner. Students receive their work back with constructive and timely feedback.
viii. Carry out testing in a manner that is fair, valid, robust, manageable and contribute to the process of student learning.
ix. Ensure that students with disabilities are provided with appropriate

opportunities to demonstrate their achievement of learning outcomes.

x. Establish procedures to ensure the levels of consistency of assessment, levels of equity, and the standard and quality of feedback.

3.3 Assessment Modes

Based on the intrinsic nature of the subject, some form of assessment is more effective than others in specific discipline of study or assessment task. Therefore, the following assessment types are practised in most of the educational institutions.

3.3.1 Formative Assessment

The Formative Assessment refers to all those activities undertaken by teachers, and students in assessing students and themselves of the learning. This information is used to provide constructive feedback in improve the students' learning, and to modify the teaching and learning activities in which they are engaged. According to Black and William (1998), such assessments become formative when the evidence is actually used to adapt the teaching to meet the needs.

Formative assessment is assessment ***for*** learning and is an integral part of teaching and learning. It does not contribute to the final mark given for the term or end of year examinations, instead, it contributes to learning through providing feedback. It should indicate what is good about a piece of work and why this is good; it should also indicate what is not so good and how the work could be improved. Effective formative feedback affects what the student and the teacher do next.

Formative assessment ***as*** learning is fundamental in helping students become more effective and engaged learners. They understand the learning goals, and aim for them through the meaningful use of assessment evidence. According to Brookhart, S. & Lazarus, S. (2017), three key formative assessment questions

students generally ask in their journey of learning include; where am I going? where am I now? what do I need to do next? They evaluate evidence of their own learning and generate ideas about their learning and strategize ways of improving the achievement of learning outcomes.

Some of the assessment methods that can assess the level of competence of learners include the following:

i. **Direct observation**: Direct observation evaluates the individual performing a task in real time. It enables the assessment of actual processes employed by individuals undertaking the assessment task activities. This method focuses on the process and the outcome. Therefore, the performance assessments examine the learners' actual application of knowledge to solve problems and using it in learning.

   Video recordings can be used to record the direct observation where appropriate. This has the minimal disturbance at the learning place by the assessor and remote accessibility. However, the assessor cannot assess the thinking process of the learner through video recording. This short coming can be remedied through following up interviews and questioning.

ii. **Simulations**: A simulation is where the learner is given a real or hypothetical situation and asked to respond as if he or she were on the real time situation and problems. Simulations range from simple questions, such as experiential questions, case replay and demonstration, to mock briefings or full operational simulators. It can cover any of the aspects that were not covered through direct observation. However, it may be difficult and time consuming to create the scenario in a few cases.

iii. **Tests**: Tests are a more traditional method of assessing knowledge, skills, and values at the cognitive level. Test could be in the form of interview, written quizzes, or self-evaluation. Depending on the focus of the assessment, varied styles of questioning can be used including multiple choice, short

answers, and open-ended questions. This can help to deal with the aspects of competence that cannot be assessed through direct observation. However, being focused on cognitive knowledge, assessment of skills and values as dictated by the learning outcomes based on the Blooms taxonomy can be elusive.

iv. **Experiential Questions**: Experiential questions are like tests but the question asked is of the form "what would you do if…?". Experiential questions can be posed in written quizzes or oral interview. In the case of verbal questions, it is very important that answers and results are documented. It has the advantage in completing the competence that cannot be assessed through direct observation.

v. **Portfolio**: A portfolio contains evidence of knowledge, ability or competence based on past experiences. Portfolio evidence can be powerful in demonstrating competency as it provides clear evidence of what an individual has done. Testimonials from learners, self-assessment can be included in the portfolio. A portfolio could even describe unsuccessful examples along with remedial work that the learner has done to remedy the deficiency. This method promotes self-evaluation, reflection, and critical thinking, and measures performance based on genuine samples work of learners. However, this method is subjective to priorities of the assessors, and may prove time consuming for assessors to mark if large number of learners require assessment.

The 'continuous assessment' and 'coursework' signify the use of regular tests and assignments throughout a course of study, where results for each piece of work contribute to the final mark or grade for each student.

3.3.2 Summative Assessment

The main purpose of summative assessment is to make a judgment regarding each student's level ***of*** achievement at the end of term, year, or completion of the course. The summative evaluation is directed toward a much more general assessment of the larger outcomes achieved by students over the entire course or some substantial part of it. The results of this type of assessment are generally expressed as marks, percentages, grades, or qualifications.

The summative assessment is essentially passive and does not normally have immediate impact on learning, although it often influences decisions which may have profound educational and personal consequences for the student. The primary distinction between formative and summative assessment relates to purpose and effect, not to timing (Sadler,1989, (p. 120).

A. Examinations

Examinations are assessment methods used to determine whether a student meets the established criteria usually set by the course or by an external agent. They are generally perceived as ways to test students' comprehension of the taught content and abilities to relate and apply them in solving problems. Examinations can also facilitate the development of metacognitive abilities in students and inspire them to take the ownership of their learning. Being aware of why testing is conducted, both teachers and students are reminded of what to teach and the learning outcomes students expected to achieve.

Examinations may lead to the award of certificates for promotion or selection for an employment. Further, results of the exams inform policy makers and parents of the status of quality of education of the school, and issues related to classroom instruction and of the curriculum materials.

i. Term/Year-End Written Examination

Written examination is a common assessment method at all levels of education. The written examination addresses a question/problem or set of questions/ problems defined by the examiner. The examination is conducted as a formal, sit-down exam lasting from one to three hours. The invigilated sit-down written exams are typically designed to test students' basic knowledge and analytical skills in a subject, as well as their ability to work under pressure. Written questions should be unambiguous and allow students to demonstrate their command of the material in respect to the learning objectives and outcomes of the course.

The examination shall assess the competencies identified in respective learning areas under the following broad domains (BCSEA,2021).

Knowledge or how it is used: Design test questions to assess students' knowledge and ability to relate and apply material learnt class to learning and solve problems.

Process or product: Test students' reasoning skills and evaluate the process by focusing the marks and other feedback on the process they follow to arrive at a solution. Alternatively, the product can be evaluated.

The communication of ideas: Evaluate students' communication skills their ability to express themselves - whether this is by writing a cogent argument, or creating an elegant mathematical proof.

Convergent thinking or divergent thinking: Test students' ability to draw a single conclusion from different inputs (convergent thinking), or generate diverse possible answers (divergent thinking).

Absolute or relative standards: The success of students is defined by the achievement of the set learning outcomes through the demonstration of

competencies in diverse areas of learning and in their life. The exams provide evidence of their progress and achievements over the duration of the course?

The conduct of exams in schools is informed by the National Student-learning Assessment Framework (NSAF) (DCPD, 2022).

ii. Oral examination

Oral examination is a variable genre that can consist of presenting a paper or an analysis, answering questions, analysing texts or other material, translating, debating, demonstrating, or performing. It is therefore important to make the assessment criteria clear, both to the student and to the co-examiner. Several parameters can be used to adapt this exam type to different learning objectives. They include:

- the use of preparation time just before the examination, with or without specific material to analyse or questions to answer;
- the production of a written report or synopsis to be defended orally. In this case, the assessment criteria usually indicate how much the oral defence counts in the grading of the whole performance with reference to the course objectives, which serve as the basis for the grading;
- the time allotment for each element of the oral exam. It is common to start with a student presentation of the subject area, question, or material, then move on to a discussion. But an oral exam can also consist of precise questions for the student to answer at the beginning, or there might be a text, picture, or case to analyse, with or without preparation and access to the student's own notes, a textbook, or the internet.

The actual knowledge, skills, or competences to be assessed at oral exams are multiple. Presentation of work and the student's dialogical skills and competencies can be weighted at the level of disciplinary knowledge and

understanding. These are usually indicated as the course or learning objectives in the academic regulations.

iii. Practical Work Examinations

Practical Work are learning experiences in which students interact with materials or with secondary sources of data to observe and understand the content knowledge. For example: photographs of plants to examine nutrient deficiency in plants and analyse the roles of nutrients in plants.

All science related examinations should include practical work assessed by a combination of:

- Questions in the written paper which are designed to assess candidates' experimental and investigative skills and competencies developed through their familiarity with performing, and understanding of, certain experiments identified in the learning outcomes (for example, investigating the effect of temperature on enzyme action).

- Teachers' assessments of certain technical and scientific skills identified in the learning outcomes for example, using an ammeter to measure the current in a circuit. Evidence of participation in this work are collected for every student. Teachers would give each student an overall mark across the full range of skills, thus allowing a deficit in one skill to be compensated for by proficiency in others.

4. COMPETENCY BASED ASSESSMENT TEST ITEMS

The competency-based assessment is a rigorous, ongoing process, which aims in testing and building the knowledge, skills, and abilities of the learner (Mary Clotilda, 2021). The need for CBE assessment is gearing up in today's dynamic world, where the stress is on sharpening current capabilities and developing new ones in order to stay ahead. Rather than developing new skills, CBA is defining the competency levels at different stages of expertise, which helps the learners gradually progress from novice to expert, which is the cardinal of a CBA.

The conversion of a learning objective into a test item is, in part, a creative process. Test items are the building blocks in the assessment, and they are the tasks or stimulus presented to students to demonstrate their abilities to apply knowledge and skills in solving problems and task in hand. There are certain procedures and guidelines that can help in writing a test item and in ensuring that the test item measures the **knowledge, skill, or ability that it is intended to measure. Traditionally, test items that require the learner to supply an answer (e.g., short-answer, essay) have been considered "subjective"; test items** requiring the learner to select an answer (e.g., multiple-choice, matching) have been considered "objective" (Department of Energy, n.d).

Competency-Based Assessment is an assessment of student's competence usually inferred from their performance in the assigned tasks, generally measured against the prescribed standard and competency indicators. It measures the demonstration of one's knowledge, skills, values and attitude through proper actions and evaluated against the set of prescribed competencies or learning outcomes defined for respective curriculum or programme. (BCSEA, 2022)

The following guidelines provide insights that the test developers need to be mindful in selecting, constructing, and scoring different formats of written test items (e.g., short-answer, multiple-choice, matching, and essay).

4.1. Guiding Principles on Competency Based Assessment

The competency-based education focuses on the real-world aspect of applying the knowledge, rather than just on the ability to recall the knowledge. The competency requirement outlines the minimum skills, knowledge and behaviours students need to demonstrate to perform the task. This mandates that competencies are stated in specific and measurable terms, focusing on the skills and abilities that are required in the real world (PEARSON 2015). Therefore, the best measures of competence involve complex, real-world, hands-on performance-based assessments (context based) (Baughman et al., 2012; Drisko, 2014).

There are many forms of categorizing competencies, resting more or less on sound bases: cross-cutting competencies, organizational competencies, functional competencies, psycho-emotional competencies, basic competencies, disciplinary competencies, essential competencies and others (UNESCO IBE, 2016). Collection of evidence and documentation of results in competency-based assessment is equivocally critical to do justice of the CBA process. For instance, mastery of subject knowledge can be tested using multiple choice or forced response items, but students must demonstrate what they can do by completing real-world, performance-based measures.

Consistent with the principles of the generic assessment and the evidence-based assessment, CBA is generally context based and is informed by the following guiding principles, which are grounded on the fundamental attributes of competency-based education and competency-based assessment (BCSEA, 2022):

Principle 1: The CBA tests the abilities to transfer knowledge, skills, and values/attitudes to a new situation

In the competency-based education tradition, assessments used to measure

competencies must not only measure one's knowledge and skills, but also get at one's ability to integrate, synthesize, and use the knowledge and skills necessary to become part of a community. This inspires the shift of the assessment process from an emphasis on summative to formative, and evaluation is criterion referenced rather than norm-referenced (Carraccio, Wolfsthal, Englander, Ferentz, & Martin, 2002 in PEARSON, 2015).

In order to assess the high order thinking and the ability to transfer to the new situation or life realities, students are engaged in a series of realistic stimuli that they must interpret or analyse accurately to answer the items correctly. The possible stimuli may include, textually in the form of claim, statement, passage, mini-case, quote, report, text-based data set, description of an experiment or a project; graphic based in the form of chart, graph, table, map, picture, model, diagram, drawing, schematic map, or spreadsheet (Nilson, L. B., n.d)

Therefore,

- ensure that the concept being measured has a direct relationship to the ability to perform the task. The construction of the test item should clearly reflect the enabling objective, word the test item so that it would be considered valid and reasonable to learning outcomes in the learning areas.

- tests used to measure competencies should focus on clusters of competencies as they might appear in the real world rather than on just one competency. The more real-world the assessment is, the more value the results hold for determining students' competency (Drisko, 2014) (PEARSON 2015)

- Each competency must be measured more than one time, and in more than one way (that is, multiple choice tests, papers, presentations, performance-based real-world assessments, etc.).

Example 1: Look at the Drukair time table given below.

	**DEPARTURES**		**10: 55**
Scheduled time	Flight No	Destination	Gate
07:40	KB 300	Dhaka	1
11:40	KB 400	Kathmandu	1
17: 00	KB 121	Bangkok	1
13:25	KB 401	Kolkata	1

You are traveling to Kolkata on flight KB 401. The time now is 10: 55 AM.

i. How much time is there before your flight's departure?

i. If duration of the flight 30 minutes, by what time will you arrive at the destination?

Example 2:

i. What message is the speaker trying to convey in the poem? Explain in your own words with reference to the poem.

ii. What kind of person do you think would need the message conveyed in the poem? Give any THREE reasons to support your answer.

Principle 2: The CBA is context based, task based and issue-based including the real-life situation

The very idea that CBE prepares its students for the workforce, the assessments are planned in line with real-life experiences. Choosing the type of assessment is a crucial element in achieving the learning outcome. Therefore, assessments used to measure competencies are sensitive enough to recognize personal, national, global, social, political, economic, and geographical context, and the

measurement of the competencies occur more than once. The best measures of competence involve complex, real-world, hands-on performance-based assessments (Baughman et al., 2012; Drisko, 2014). While assessments in the CBE program are crucial for determining if a student has the knowledge required for the competencies, the performance-based assessment is often the deciding factor of whether the student demonstrates competencies.

Therefore,

- emphasise the deployment of authentic assessment or evidence-based assessment in measuring the display of multitude of competencies by students in diverse situation. Unlike the conventional testing, the authentic assessment renders greater scope to assess the impact of both explicit and hidden curriculum which are inherent in all forms of curricula.

- assessment must provide opportunities for students to demonstrate real-world skills and abilities. For example, students could be asked to write a marketing plan for a fictitious product in order to demonstrate mastery of the competencies for a marketing course or program.

- use formative assessment with lots of activities that help students to demonstrate their mastery level, irrespective of the time/place/pace of learning. On achieving this with confidence, they then continue by the summative assessment.

Example 1: Read the following text and the answer the questions that follow:

> Nationwide Vitamin B complex supplement for students: Measure being taken to prevent further outbreaks of peripheral neuropathy in schools

Nutrition: …

A press release from the health ministry stated that, in the last two decades, Bhutan has seen a series of peripheral neuropathy outbreak among students. "The affected students of Gomphu LSS showed breathlessness, chest pain, palpitations, tingling sensation and numbness in lower limbs, which were symptomatic of nutritional deficiency illness," the release stated. "A detailed investigation by the health team inferred that the outbreak was due to deficiency of vitamin B1 or Thiamine in the students' diet." On October 29, representatives from the health and education ministries met with the Prime Minister to discuss issues regarding the outbreak and its immediate interventions.

"Of the several interventions discussed, Lyonchhoen instructed immediate supplementation of Vitamin B-complex to all students for November and December, in order to prevent any possible outbreaks," the ministry's press release states……."

(Kuenselonline.com NOV 19, 2014)

Answer the following questions with reference above situation.

i. What could be the reason for the peripheral neuropathy in the school? Mention the symptoms of this disease.

ii. If you were one of the investigating teams, what diet would you suggest to the school to prevent the disease?

iii. Do you think Vitamin B-complex supplement for students would prevent the peripheral neuropathy disease in all the schools in Bhutan? Is the government doing enough to prevent this disease?

iv. Design a poster on healthy eating habit and its effects.

Example 2:

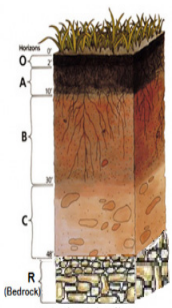
i. We meet a lot of people in our lives but do not get close to all of them. However, there are some who become close friends as we get to know each other. Write about how you met one of your close friends and how he or she has changed your life.

OR

ii. We celebrate different events in life in different ways. It could be by throwing parties, going on picnics, visiting places, or having a family get-together at home. Write about an event you celebrated. Mention the reasons for the celebration.

Example 3:

A vertical cross section of the soil as shown in the diagram is:



Graphic: Soil Profile

Graphic courtesy USDA [1]

D Horizon

A Horizon

B Horizon

C Horizon

A Soil Texture

B Soil Profile

C Soil Structure

D Soil Property

Principle 3: Test the use of knowledge rather than knowledge itself

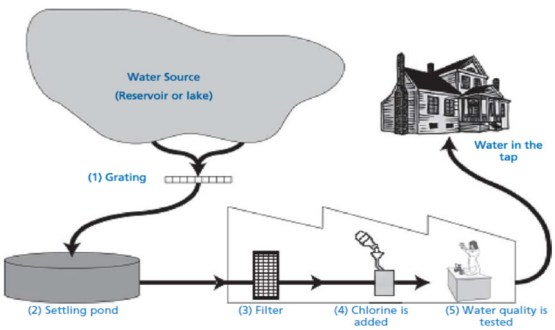
Assessments developed to measure competencies are written specifically and vetted to ensure that they are measuring what is claimed to be measuring. Knowledge can be tested using multiple choice or forced response items, but students must demonstrate what they can do by completing the real-world, performance-based measures (PEARSON, 2015) based on the assigned task and problems.

Therefore,

- mastery of subject knowledge can be tested using multiple choice or forced response items, but students must demonstrate what they can do by completing real-world, performance-based measures.
- use rubric to evaluate student's performance based on the total range of criteria rather than a numerical score. The rubric is almost like a working guide both for teachers and students, which is usually handed out before the commencement of assessments in order to let the students know the criteria on what their work will be judged.

Example 1:

Study the diagram below and answer the questions that follow:



The figure above shows how water supplied to houses in cities is made fit for drinking.

i. It is important to have a source of good drinking water. Water found underground is referred to as ground water.
Give one reason why there is less bacteria and particle pollution in ground water than in water from surface sources such as lakes and rivers.

ii. The cleaning of water often happens in several steps, involving different techniques. The cleaning process shown in the figure involves four steps (numbered 1–4). In the second step, the water is collected in a settling pond. In what way does this step make the water cleaner?

A The bacteria in the water die.
B Oxygen is added to the water.
C Gravel and sand sink to the bottom.
D Toxic substances are broken down.

Example 2:

Karma was surprised to notice his legs appear shorter while crossing a river. Upon investigation, he found out that it was due to the phenomenon called

A reflection.
B diffraction.
C dispersion.
D refraction.

Principle 4: Assess the ability to learn and generate new knowledge

Competency-based assessments provide insight in: what the students know, still need to learn, have learned, and where they need for improvement. In the competency-based education systems, students continue to work on specific skills or knowledge until they can demonstrate their understanding and the ability to apply to construct their understanding and generate new knowledge.

In maintaining the academic integrity of the course, one should start with the

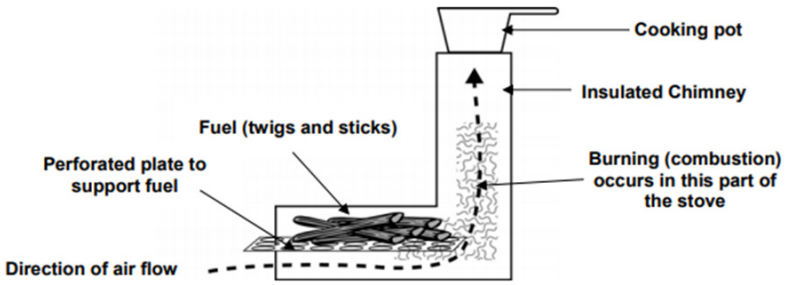
outcomes to derive competencies to be measured. Ensure that competencies are specific and measurable, while focusing on the skills and abilities that are required in the real world.

Therefore,

- competency assessment is a continuous process, by which student's knowledge and skills are constantly developed. It progresses an individual from being a novice to an expert by rigorously constructing their understanding and generating new knowledge vital for today's competitive world.
- collection of evidence and documentation of results is critical in competency-based assessment. Use process that integrates knowledge and skills with practical applications in a workplace task (holistic approach).
- assessments used to measure competencies must not only measure knowledge and skills, but also must stimulate students to integrate, synthesize, and use the knowledge and skills necessary to generate new knowledge.

Example 1:

A type of stove known as a "rocket stove" has been designed. A cross section of the stove is shown below.



i. Rocket stoves can be used to replace open fires for cooking. The cooking pot being heated is placed on top of the insulated chimney rather than in the fire itself. The fuel burnt in the rocket stove is sticks and twigs.

Some studies have found that rocket stoves are more efficient than open fires. What does "more efficient" mean in this case?

A Rocket stoves are safer.

B Rocket stoves are easier to light.

C Rocket stoves produce more usable heat.

D Rocket stoves use smaller pieces of wood.

ii. The amount of carbon dioxide in the air is increasing in Thimphu due to the growing number of vehicles. The Thimphu *Thrompen* wants to plant more tree. Do you agree? Justify your answer with two points.

iii. Design a campaign plan as how would you control the air pollution in Thimphu?

Example 2:

a. Landforms are formed by the interaction of various forces. What is the most dominant force of landforms in a mountainous terrain? Support your answer with at least TWO reasons.

b. Landform

Observe the picture given and explain the significance of landforms on the lives of the people in the community. Cite at least TWO examples of landforms showing their significance.

Principle 5: Assess the ability to think beyond literal questions and promote critical thinking

Competency-based education requires a high degree of critical thinking, problem solving and personal responsibility on the part of the student, which are the key attributes of CBE and much sought after by employers.

Consider the cognitive level at which the objective being tested is written. For example, if you are writing multiple-choice test items for three enabling objectives, two of which are written at the knowledge level of Bloom's Taxonomy to support a third which is written at the comprehension level, then the point value assigned to the higher cognitive level would be higher than the point value assigned to the lower cognitive level. For example,

ACTION VERB	COGNITIVE LEVEL	POINTS E.O
List	Knowledge	(2.0 pt)
State	Knowledge	(2.0 pt)
Determine	Comprehension	(4.0 pt)

(Department of Energy, n.d.)

Therefore,

- determine the ease or difficulty of a test item based on the criticality of the objective that is being tested to the overall mastery of the terminal objective which will allow successful performance of the job task

- test items assess diverse cognitive levels within the spectrum of LOT to HOT in order to discriminate between those who have mastered the objective and those who have not.

- assessment focuses on student learning, not student grades. Student's progress is often categorized in three or five levels that capture the competency levels – Exceeding, advancing, meeting, attaining, and beginning or novice or still working toward proficiency.

Example 1:

Students interview students and/or teachers to identify a problem (e.g., bullying on the playground) and as a group brainstorm creative ways to address the problem (producing a play that examines the issue, hosting a contest for best ideas). Develop creative solutions to a class or school problem.

Example 2:

Trongsa Poenlop Ugyen Wangchuck forgave and accommodated many rivals. He used force to suppress the rivals and rebellions when required. If you were Ugyen Wangchuck, how would you have handled the situation differently?

Principle 6: Assessment is meaningful and a positive learning experience for students

One of the fundamentals of CBE is to provide life enriching learning experiences in the real time, so that students realise the purpose of learning and engagement in diverse learning experiences. This enforces that the assessment is authentic through the rigorous formative assessment and feedback is provided without the time lapse. The authentic assessment with timely feedback empowers students in understanding the status of their learning and hone their abilities to apply the knowledge, skills, and values in diverse situation.

Formative assessments are aligned with learning objectives. Students receive immediate feedback when assessment occurs. This is used to encourage students to return to difficult concepts and skills until they achieve mastery. Thus, assessments are student-centred (meaningful and positive learning experience) in which students are assessed on material with which they are familiar.

Therefore,

- teachers collaborate with students to develop understanding of what is an adequate demonstration of proficiency. Teachers must share a clear understanding of what students need to demonstrate before they advance to the higher levels.
- assess skills or concepts in multiple contexts and multiple ways to assess proficiency demonstrations of learning by students. It is understood that students must demonstrate proficiency multiple times to ensure that they are completely comfortable with the material.
- avoid copying text directly from training or other reference material. Test items written in this way generally encourage rote memorization. Further, copying from reference material can cause confusion in test items because the material lifted often draws its meaning (and importance) from its surrounding context. Therefore, important assumptions or conditions stated elsewhere in the material are often omitted from the test item.
- state the test item as concisely as possible, but provide all necessary information. The test item should be clear, grammatically correct, and free of clues to the correct answer. It should be written at a reading level appropriate for the students. Often the individuals who develop a test item assume that certain conditions are inherent in the question when, in fact, they are not.

Example 1:

When flash flood occurs, water is made undrinkable. People filter the muddy water and boil it for drinking.

i. What is the purpose of filtering the muddy water?
ii. Why do you need to boil the water before drinking?
iii. Suggest an alternative for treating the muddy water for drinking.

Example 2:

Make a 5 minutes presentation on the important changes that have occurred in your city/community over the last 10 years. Your argument must be strong as to why these changes are important. Your presentation must include appropriate photographs, maps, graphs, and other visual aids for the audience

Teacher use rubric for assessment and reporting based on the following suggestive criteria:

- Knowledge
- Argument regarding importance
- Communication
- Use of Aids
- Oral Presentation skills

When it comes to implementing assessments, computer-based testing is most effective in diagnostics and personalized learning. The concept of personalized learning is possible because of the technology used to assess students' knowledge in a more formative way, allowing everyone involved to better address education gaps and needs.

Therefore,

- Integrate technology in assessment as it can shorten the feedback loop. (Google classroom, Kahoot, Emodo, iMocha, etc.)

- use digital technologies to provide students, parents, and educators with timely information vital to create an increased understanding of individual student needs related to differentiation and individuation.

4.2. CBA Test Item Development Guidelines

Test is an instrument for collecting data and is used in instructional programs to assess behaviours, monitoring learning process, diagnose learning difficulties, and measure performance at the end of the instruction. Its procedure is designed to reveal specific behaviours of an individual based on which inferences can be drawn about the traits of an individual. A student is required to respond to a sequence of tasks, as a result of which measures of a specific trait of a student is made. Hence, a test is defined as task or a series tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes of a person (Lestari, H. 2022).

In the context of competency-based assessment, consider the following guidelines in framing the test items.

i. Understand the content and cognitive domain descriptions in the subject frameworks and the assessment protocols. Align items with the content topics and cognitive domains designed to collect evidence about what learners know, can do, and apply them.

ii. Order the selected questions logically from a content standpoint (e.g., chronologically, in conceptual groups, etc.). Consider to place some simpler items at the beginning to ease students into the exam. Group them together under common instructions to save reading time (Kalim, T., 2009).

iii. Consider the time required for students to complete the required task. For example, a multiple-choice item on TIMSS 2015, it is expected to require about 1 minute or less to complete, and constructed-response items are allocated 1-3 minutes (IEA, 2013).

iv. Assign scoring guides with well-defined categories and a detailed description of the kind of responses that belong in each category.

v. The reading level of items should be at an elementary level for the target grade. In general, the amount of reading should be kept to a minimum,

given the context of the problem (IEA, 2013).

vi. Write questions in the active voice (i.e., doer of action (subject) before action (verb) and avoid conditional words, clauses, and tenses (e.g., if, suppose, when) (IEA, 2013).

vii. Make sure that item validity is not affected by factors that unnecessarily increase the difficulty of the item, such as unfamiliar or overly difficult vocabulary, grammar, directions, contexts, or stimulus materials (IEA, 2013).

viii. It is desirable that there be some relatively easy items and some challenging items. However, items that almost all students or almost no students can answer correctly reduce the effectiveness of the test to discriminate between groups with high achievement and groups with low achievement. Typically, most items used in the test should be ones that are answered correctly by 30 to 70 percent of the students on average in the class (IEA, 2013).

ix. When preparing assessment items, be sensitive to the possibility of unintentionally placing particular groups of students at an unfair disadvantage. It warrants that special attention is considered to the diversity of environments, backgrounds, beliefs, gender, and cultures among students in the participating countries (IEA, 2013).

x. Ensure that diagrams and graphs are drawn accurately (to scale unless otherwise noted), and are correctly and fully labelled. Any graphics included in an item should be in order to solve the problem or to answer the question, and should be adequately explained and referred to directly within the item, as indicated by the principles of universal design for assessment items.

xi. Use plausible distracters (incorrect response options) that are based on likely student errors or misconceptions. This reduces the likelihood of students arriving at the correct response by eliminating other choices and, equally important, may allow identification of widespread student

misunderstandings or tendencies that could lead to curricular or instructional improvements.

xii. Consider the following ideas in the selection of test item format (types) used at various cognitive levels (Kalim, T., 2009).

Factual Knowledge	Application	Analysis and Evaluation
• Multiple choice • True/False • Matching • Completion • Short answer	• Multiple choice • Short answer • Problems • Essay	• Multiple choice • Essay

xiii. Measure a single construct

- Content should be independent from item to item. If an examinee incorrectly answers an item that has multiple constructs, it is impossible to know which construct is not mastered.
- Compound items heighten test anxiety and can lower the perceived validity of the exam.

xiv. Avoid opinion-based items

- Never ask "What would you ... do", " ... use", " ... try", etc. The examinee's answer can never be wrong.
- Use caution when asking for the "best" thing, or the "best" way of doing something, unless it is clearly the best amongst the options.
- If differences exist between any experts' opinion about what the "best" is, then avoid using it.
- Qualify the standard for "best" (i.e., according to ...).

xv. Avoiding excessive verbiage

- "Verbosity is an enemy to clarity." (Haladyna, 2004)
- Wordy items take longer to read and answer, meaning fewer items

can be presented in a fixed amount of time, reducing reliability.

- Write items as briefly as possible without compromising the construct and cognitive demand required.
- Get to the point in the stem and present clean, clear options for the examinee to choose.
- Avoid unnecessary background information.

xvi. Use novel content

- Do not repeat exact wording from textbooks or other reference materials.
- Repeated wording tends to test recall and recognition, rather than learning.

xvii. Mix It Up!

Being mindful of students' diverse aptitude and needs, it is often advantageous to mix types of items - multiple choice, true-false, essay, on a written exam or to mix types of exams - a performance component with a written component. Weaknesses connected with one kind of item or component or in students' test taking skills will be minimized.

xviii. Test Frequently

Frequent testing helps students to avoid getting behind, provides assessors with multiple sources of information to use in computing the final course grade (thus minimizing the effect of "bad days"), and gives students regular feedback. It is important to test various topics in proportion to the emphasis given in class.

xix. Proofread Exams

It is important to proofread written exams test carefully and, if possible, have another person proofread them. Tiny mistakes, such as misnumbering the

responses, can create confusion to examinees. Further, do the collation check since missing pages can cause a great deal of trouble.

(Adapted from "Is This a Trick Question? A Short Guide to Writing Effective Test Questions", Clay, B. 2001, pp 3-4)

4.3 Test Item Review

The value of test items to what extent they achieve their objective purposes is vetted against the targeted curricular objectives and process used in the test items. In the broader sense, best items are highly discriminating and moderately difficult. Programs can calculate Discrimination Index and Difficult Index for each item. Conversely, an item is poor if, it fails to differentiate among more able or better prepared from the weak or less prepared students; almost all students get it either right or wrong, especially if the more able students get it wrong.

Once the test items are developed, it is fundamental that they are vetted for their values based on the principles of assessment. By this, reliability, validity, currency, flexibility, and fairness of the test items are checked and validated. Although there may be plethora of ideas on review of test items, one practical method is proposed by Hopkins an Antes (1990: 156-7). They propose three main areas of review, which they call 'Balance', 'Specificity' and 'Objectivity'. A pair of questions are attached with each area to inform the process of review of test items by reviewers and facilitate the formation of mental schema of how good the test is (IIEP UNESCO, 2005).

- **Balance**
  i. Are the items selected for the test representative of the achievement (content and behaviours) which is to be assessed?
  ii. Are there enough items on the test to adequately sample the content which has been covered and the behaviours as spelled out by the objectives?

- **Specificity**
  i. Do the test items require knowledge in the content or subject area covered by the test?
  ii. Can general knowledge be used to respond to the test items?

- **Objectivity**
  i. Is it clear to the test taker what is expected?
  ii. Is the correct response definite?

The intention of the test item review is to ensure that more of the top-scoring students can provide correct answer than the low-scoring students. This is a method by which discrimination of top performing students are differentiated from low performers. In case, if top-scoring students do not know the answer while low performing students can answer them correctly, it warrants the further review of the test items to improve the quality so that the test items provide evidences for the intended purposes or objectives (IIEP UNESCO, 2005).

Guidelines for Review of Test Items

The following guidelines are recommended for reviewing individual test items. During the review make notes about the test items and suggestions with changes, if any.

i. Consider the item as a whole and whether …

   a. it measures knowledge or a skill component which is worthwhile and appropriate for the examinees who will be tested;
   b. there is a markedly better way to test what this item tests;
   c. it is of the appropriate level of difficulty for the examinees who will be tested.

ii. Consider the stem and whether it …

   a. presents a clearly defined problem or task to the examinee;

b. contains unnecessary information;
c. could be worded more simply, clearly, or concisely.

iii. Consider the alternatives and whether …

a. they are parallel in structure;
b. they fit logically and grammatically with the stem;
c. they could be worded more simply, clearly, or concisely;
d. are they inclusive that they logically eliminate another more restricted option from being a possible answer.

iv. Consider the key and whether it …

a. is the best answer among the set of options for the item;
b. actually, answers the question posed in the stem;
c. is too obvious relative to the other alternatives (i.e., should be shortened, lengthened, given greater numbers of details, made less concrete).

v. Consider the distractors and whether

a. there is any way you could justify one or more as an acceptable correct answer;
b. they are plausible enough to be attractive to examinees who are misinformed or ill-prepared;
c. distractor should not merely state the reverse of the key or resemble the key very closely unless another pair of choices is similarly parallel or involves opposites).

(Source: Cohen, A. S. and Wollack, J. A. (n.d). Handbook on Test Development: Helpful Tips for Creating Reliable and Valid Classroom Tests)

5. TEST ITEMS WRITING

Within the field of test development, the tasks and/or questions that are used to construct tests and examinations are referred to as 'items', and the range of techniques involved in preparing those items are collectively referred to as 'item writing' (Withers, G., 2005). Traditionally, test items that require the student to supply an answer (e.g., short-answer, essay) have been considered "subjective"; test items requiring students to select an answer (e.g., multiple-choice, matching) have been considered "objective" (Department of Energy, n.d).

Test Item Types

Decades ago, Benjamin Bloom and others constructed what was called The Taxonomy of Educational Objectives. This list divided educational experiences into two main domains: cognitive and affective. Within the cognitive area, focus is on matters such as acquiring factual knowledge, being able to comprehend, developing the ability to analyse, to synthesise and to prepare evaluations as being a useful (and important) way to structure what we teach, and hence what we test. Use these elements to structure the classification of test items, to ensure that we obtain a good coverage of what went on (or should have gone on) during the learning process. Hence, IIEP UNESCOS (2005) propounds that test items can be usefully classified into three main categories:

III. Objective / Selected response items

The choice is 'closed': that is to say, only one option will receive credit.

- a. True-false
- b. Matching item
- c. Classification items
- d. Multiple-choice items

II. Constructed response items /expanded response types

In items of this kind, the candidates construct or prepare their own responses on the spot, based on knowledge developed during a course, or by rewriting bits of given stimulus material in a new way.

- d. Short-answer items
- c. Fill-in-the-blank and sentence completion items
- b. Cloze items
- a. Extended responses (paragraphs and essays)

I. Problems for solution

The problems which test-writers set for solution by candidates represent a sub-category of constructed responses.

For instance, setting an essay task for extended response has a problem element built into it, especially if the candidate is required to read a large amount of informative or manipulative stimulus and choose personal response to be developed during the test.

- c. Requiring words as a response as in Language test.
- b. Quantitative reasoning and expressions for their solution as in Mathematics.
- a. Manipulation of materials that is required, as in a test conducted in a laboratory or workshop or art studio.

Figure 3: Classification of Test Items

(Adapted from "Item writing for tests and examinations Module 5" IIEP UNESCO, 2005, pp 22-27)

The test item writing is a task that requires imagination and creativity, but at the same time demands considerable discipline in working within the curriculum and assessment frameworks and base the item construction on the set of principles and norms. For instance, *Clay, B. (2001) in "Is This a Trick Question? A Short Guide to Writing Effective Test Questions",* proposes the following:

When to use any type test items:

i. Essay tests are appropriate when:

- the number of students to be tested is small and the test is not to be reused.
- there is need to encourage and reward the development of student comprehension and analytical skills in writing.
- the assessor is interested to explore students' attitudes and measure their achievements.

ii. Objective tests are appropriate when:

- the number of students to be tested is large and the test may be reused.
- highly reliable scores must be obtained as efficiently as possible.
- impartiality of evaluation, fairness, and freedom from possible test scoring influences are essential.

iii. Either essay or objective tests can be used to:

- measure almost any important educational achievement a written test can measure.
- test understanding and ability to apply principles.
- test ability to think critically.
- test ability to solve problems.

The following section espouses concepts, ideas, and the construction of diverse test item writing commonly used in schools. To facilitate better understanding, the description for each test format is presented in the sequence of what – concepts, salient features, types; why – utility or purpose; and finally of how – the process of writing test items.

5.1 Objective/Selected Response Types

Test items necessitate students to select an answer (e.g., multiple-choice, matching) are considered "objective", they are also called selected response (SR) types. An objective test is one in which a students' performance is measured against a standard and specific set of answers (i.e., for each question there is a right or wrong answer). When composing test questions, it is important to be direct and use language that is straightforward and familiar to the students. In addition, the answer choices provided on the test should be challenging enough that students cannot guess the correct answer simply by comparing how all the options are written (Callahan, M. & Logan, M.M., 2022).

5.1.1 Multiple Choice Test Item

Multiple-choice test item is one of the versatile objective or selected response (SR) test item formats common in the global educational arena. They can be used to test factual recall as well as levels of understanding and ability to apply and analyse the learning in diverse context. This test format consists of the following parts:

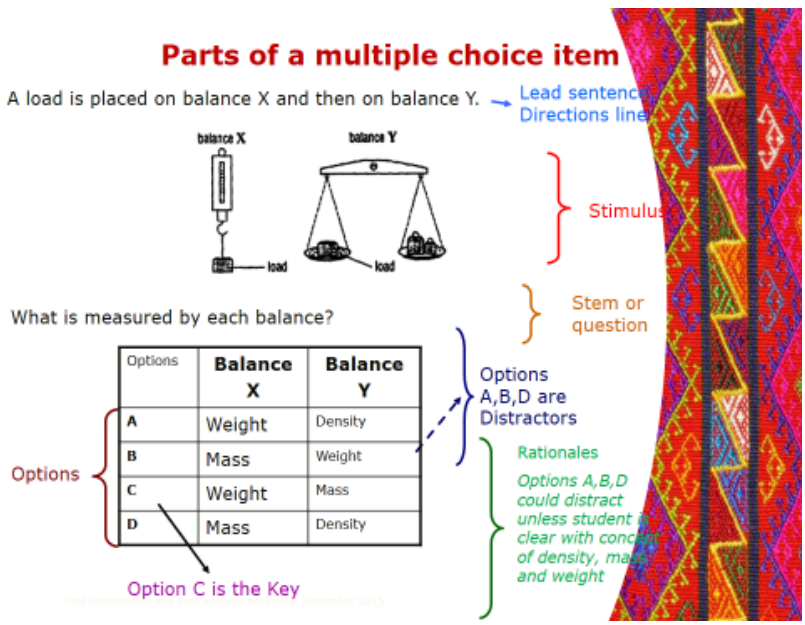
- *Stem:* This states the problem for the examinee, which can be either in the form of question, a set of direction or statement or stimulus with lead sentence.
- *Response options/alternatives:* They are the choices given for the item. The item choices are generally numbered from A to D.
- *Key:* This is the correct choice for the item.
- *Distractors:* These are the incorrect choices for the item.

Guidelines for writing distractors

i. Use logical misunderstandings or misconceptions.
ii. Use common errors.
iii. Use familiar terms, key-words, structures, or ordering.
iv. Use statements that are correct or true but do not answer or address the stem (question).
v. Avoid opposing statements.
vi. Use correct concepts, but "mixed up."

(Source: Certification Management Services, 2006)

Example



It is generally recommended that 4 or 5 choices are used in a question. It is on the premise that fewer alternatives can lead to guess work, and conversely, if more responses are provided it confuses examinees and entails more reading time (Cohen, A. S. and Wollack, J. A.)

In comparison to subjective test items, although multiple-choice item can be difficult to write, especially if one wants students to go beyond recall of information, scoring objective test items is considerably reliable and the time needed to answer is less. Because less time is needed to answer, more test items can be constructed and cover more content, which is perceived to enhance the test validity.

The multiple-choice test item is considered versatile because, it assesses students'

- critical thinking to discriminate and make choices.
- comprehension of concepts, principles, and generalization.
- abilities to analyse and evaluate the phenomena or situation to make judgement about and make choices of course of action.
- abilities to infer, interpret new data or information and make generalization.
- abilities to apply information and knowledge to solve the problem of the task.

It is of paramount importance that the test items have direct link to the learning outcomes and learning objectives in the curriculum. The development of multiple-choice test items from learning objectives involves three steps:

- Write the stem of the test item in the form of question of an incomplete sentence that implies a question
- Write the correct answer to the stem in a few words

- Write distractors that are plausible to those students lacking the degree of knowledge and understanding the test item is purported to assess.

In order to assist the test developers in designing multiple choice test items, the following four basic models are presented as suggested in Test Development (US Department of Energy, n.d). Other models may be developed and used in combination with one another.

a. Models of multiple-choice test items

Model A:
(a) Correct answer
(b) Incorrect answer
(c) Incorrect answer
(d) Incorrect answer

Model A is the traditional multiple-choice design format which shows one correct single word/phrase with three incorrect single word/ phrase options. The length of all options should be similar.

Example: Bhutan joined the United Nations Organization in 1972 with the objectives to

A ascertain its sovereignty.

B conserve its unique culture.

C maintain as a neutral state.

D represent as an independent country.

Model B:
(a) Correct answer
(b) Plausible misconception
(c) Incorrect answer
(d) Incorrect answer

Model B is a variation of Model A, where a plausible misconception is used as an incorrect answer. Again, the length of all options should be similar.

Example: In the plant body, the trade-off of loss of water for the ascent of ascent of sap is a life sustaining phenomenon. The physiological process involved is

A transpiration.

B osmosis.

C diffusion.

D assimilation.

Model C:

(a) Correct answer with correct condition (e.g., because, since, when, if, etc.)

(b) Correct answer with incorrect condition

(c) Incorrect answer with incorrect condition

(d) Incorrect answer with incorrect condition

Model C is with an acceptable multiple-choice design that uses conditions with answers (a condition in a setting, event, cause/effect that may make the answer correct or incorrect). The Model C shows only one correct answer with its correct condition.

Example: The vegetation in southern region is generally dense and green. The vegetation is

A deciduous forest due to higher precipitation.

B deciduous forest due to more sunlight.

C temperate forest because of rich soil.

D alpine forest due to low temperature.

Model D:

a) Correct answer

b) Incorrect answer

c) Correct answer with incorrect condition

d) Incorrect answer with incorrect condition.

Model D illustrates an acceptable model when it may not be possible to create all options in uniform length. This model shows paired lengths with two long and two short options avoiding setting any single option apart (either too long or too short) from the remaining options.

Example: Based on the scientific ideas, climate change can stimulate one of the following:

A Extreme weather condition
B Desertification of a place
C Extreme weather condition due to receding glaciers on the mountains
D Biological disaster because of the decline in biodiversity

General Guidelines to multiple-choice test item writing

The general rules used for writing multiple-choice items are described below. It must be noted that these are general rules and may not apply to all types of testing.

i. When possible, state the stem as a direct question rather than as an incomplete statement

- The stem should contain the problem and any qualifications followed by alternatives.
- Each item should be as short and verbally uncomplicated as possible. Give as much context as is necessary to answer the question, but do not include superfluous information.
- Keep each item independent from other items. Don't give the answer away to another item. If items require computation, avoid items that are dependent on one another.

Example:

Undesirable: Alloys are ordinarily produced by…

Desirable: How are alloys ordinarily produced?

2. Present a definite, explicit, and singular question or problem in the stem.

- To test understanding of a term or concept, present the term in the stem

followed by definitions or descriptions in the alternatives.

- Use only correct grammar in the stem and alternatives.

Example:

Undesirable: Physiology …

Desirable: The study of metabolic process in the body is called ….

3. Eliminate excessive verbiage or irrelevant information from the stem.

- If an omission occurs in the stem, it should appear near the end of the stem and not at the beginning.

Example:

Undesirable: Nitrates from agricultural land into ponds polluting the water, many plants grow resulting to suffocation and death of plants in the pond. This phenomenon is

Desirable: Which of the following phenomena explains the death of plants due to excessive inflow of nitrates into the pond?

4. Include in the stem any word(s) that might otherwise be repeated in each alternative.

Example:

Undesirable: Botany is

A the study of microbes.
B the study of animals.
C the study of fossils.
D the study of plants.

Desirable: Botany is the study of

A microbes.
B animals.
C fossils.
D plants.

5. *Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.*

- Avoid terms such as "always" or "never," as they generally signal incorrect choices.

Example:

Undesirable: Which of the following is not the statement of monopoly?

Desirable: Which of the following is NOT the statement of monopoly?

5. *Make all alternatives plausible and attractive to the less knowledgeable or skilful student.*

- Make all incorrect alternatives (i.e., distractors) plausible and attractive. It is often useful to use popular misconceptions and frequent mistakes as distractors.

- Use a logical sequence for alternatives (e.g., temporal sequence, length of the choice). If two alternatives are very similar (cognitively or visually), they should be placed next to one another to allow students to compare them more easily.

- Avoid wording directly from a reading passage or use of stereotyped phrasing in the key.

Example:

Undesirable: What process is most nearly the opposite of photosynthesis?

A Digestion
B Assimilation
C Respiration
D Catabolism

Desirable: What process is most nearly the opposite of photosynthesis?

A Digestion
B Relaxation
C Respiration
D Exertion.

7. Make the alternatives mutually exclusive.

- Try to avoid "all of the above" or "none of the above" as a last option. If an examinee can eliminate any of the other choices, this choice can be automatically eliminated as well
- If one or more alternatives are partially correct, ask for the "best" answer.
- All alternatives should be homogeneous in content, form, and grammatical structure.
- Make all alternatives grammatically consistent with the stem.
- Alternatives should not overlap in meaning or be synonymous with one another.

Example:

Undesirable: The daily minimum required amount of milk that a 10-year-old child should drink is

A 1-2 glasses.
B 2-3 glasses.
C 3-4 glasses.
D at least 4 glasses.

Desirable: What is the daily minimum required amount of milk a 10-year-old child should drink?

A 1 glass.
B 2 glasses.
C 3 glasses.
D 4 glasses.

8. Make alternatives approximately equal in length.

- The length, explicitness and technical information in each alternative should be parallel so as not to give away the correct answer.
- Use 4 or 5 alternatives in each item.
- Avoid repeating words between the stem and key.

Example:

Undesirable: The most general cause of unemployment of youths is:

A Lack of entrepreneurial knowledge and skills
B Unwillingness to work.
C Recruitment.
D Inflation.

Desirable: What is the most general cause of unemployment of youths?

A Lack of entrepreneurial knowledge and skills.
B The population's overall unwillingness to work.
C The nation's weak recruitment policy.
D An increasing national level of inflation.

9. Avoid items based on personal opinions unless the opinion is qualified by evidence or a reference to the source of the opinion (e.g., According to the author of this passage, . . .).

(Adapted from "Is This a Trick Question? A Short Guide to Writing Effective Test Questions", Clay, B. 2001, pp 15-16).

5.1.2 Match the following

In the matching test items, students match each word, sentence, or phrase in one column with a word, sentence, or phrase in another column. The test items in the first column are called "premises" or stems or problems to be answered, and the answers in the second column are called "responses."

Matching questions is a most efficient way to test knowledge when idea, events, dates, names, and places are important. Matching questions are also appropriate for the sciences in which numerous experiments, experimenters, results, and special terms and definitions are to be remembered. Hence, this test format allows maximum coverage at knowledge level in a minimum amount of space and valuable in content areas that have a lot of facts. However, homogeneity in the items can target testing higher level of learning.

Guidelines on matching items

Consider the following guidelines when constructing matching test items.

i. Provide clear directions of how the matching the stimuli with the responses are to be attempted

Clarify whether a response can be used more than once and indicate where to write the answer.

Example:

Poor: Directions: Match the following.

Better: Directions: On the line to the left of each identifying location and characteristics in Column I, write the letter of the country in Column II that is best defined. Each country in Column II may be used more than once.

ii. Use only items that share the same foundation of information (homogenous matching type)

Unrelated topics included in the same matching item may provide cues for obvious matches and mismatches.

Example:

Poor: Directions: Match the following.

1. Water	A. NaCl
2. Discovered Radium	B. Fermi
3. Salt	C. NH3
4. Ammonia	D. 1942
5. Year of the first	E. H20
Nuclear Fission	F. Curie
	G. 1957

Better: Directions: On the line to the left of each compound in Column I, write the letter of the compound's formula presented in Column II. Use each formula only once.

Column I	Column II
____1. Water	A. H_2SO_4
____2. Salt	B. HCl
____3. Ammonia	C. NaCl
____4. Sulfuric Acid	D. H_2O
	E. H_2HCl

iii. Avoid grammatical or other clues to the correct response

Example:

Poor: Directions: Match the following in order to complete the sentences on the left.

_____1. Plato insisted that government was	A. The Prince.
_____2. Machiavelli wrote about achieving political unity in	B. desirable and inevitable
_____3. Hobbes argued that human nature made absolute monarchy	C. a science requiring experts.
_____4. Marx was a German philosopher and economist who founded absolute monarchy	D. organized along industrial lines.
	E. communism

Better: Directions: On the line to the left of each statement write the letter of the philosopher from the right-hand column that the statement describes. Use each philosopher once.

____1. Thought government was a science requiring experts.	A. Hobbes
____2. Described methods of achieving political unity.	B. Marx
____3. Founded Communism.	C. Machiavelli
____4. Believed that human nature made absolute monarchy desirable and inevitable	D. Durkheim
	F. Plat

iv. The column of stimuli on the left should set the question clearly.

Example:

Poor: Directions: Match the following.

____1. City dwellers	A. Wild animals
____2. Hunter-gatherers	B. Farm
____3. Pastoral nomads	C. Apartment buildings
	D. Graze animal

Better: Directions: On the line to the left of each definition, write the letter of the term in the right-hand column that is defined. Use each term only once.

____1. Live in areas of high population density.	A. Pastoral nomads
____2. Move from one place to another in search of wild animals.	B. Ranchers
____3. Move from one place to another with grazing animals.	C. Hunter-gatherers
____4. Till land for cash crops.	D. City dwellers
	E. Farmers

v. Consider the following possibilities for pairing premises and responses for homogenous matching type:

PREMISES	RESPONSES
Terms or words, or	Definitions
Short test items, or	Answers
Symbols, or	Proper names
Causes, or	Effects
Principles	Situations in which they apply

vi. Each response should be a plausible answer for each premise

Example: Match the alarm (letter) with its appropriate tone (number). Responses (numbers) may be used more than once:

PREMISES	RESPONSES
A. Fire	1. Pulse Tone _____
B. Reactor Building Evacuation	2. Wailing Siren _____
C. Site Evacuation	3. Steady Tone _____

vii. Present an unequal number of premises and responses, or allow responses to be used more than once.

This will inhibit the examinee from obtaining clues to a correct answer through a process of elimination.

(Adapted from "Is This a Trick Question? A Short Guide to Writing Effective Test Questions", Clay, B. 2001, pp 30-31 & "Development of Test Items", US Department of Energy, n.d. pp 21-22)

5.1.3 Text Completion

A text completion item demands the student to provide response to the incomplete statement by filling in a blank with the correct word or phrase. Text completion items provide a wide sampling of content. They minimize guessing compared with multiple-choice and true false.

Completion items are especially useful in assessing mastery of information when a specific word or phrase is important to know. Since they require a definite response rather than simple recognition of the correct answer, they prevent guessing that is possible on limited-choice items. Because only a short answer is required, their use on a test can enable a wide sampling of content. However, they cannot be written to measure more than simple recall of information. It is more time consuming to score than other objective types and difficult to write as there is only one correct answer and no irrelevant clues.

Guidelines to Writing Text Completion test item

i. *Start with a direct question, switch to an incomplete statement, and place the blank at the end of the statement.*

Example:

Poor: What is another name for cone-bearing trees? (Coniferous trees)

Better: Cone-bearing trees are also called_____. (Coniferous trees)

ii. *Leave only one blank. This should relate to the main point of the statement.* Provide two blanks if there are two consecutive words.

Example:

Poor: The ___ is the ratio of the ___ to the___.

Better: The sine is the ratio of the ___ to the ___ (opposite side, hypotenuse).

iii. Make the blanks in uniform length and avoid giving irrelevant clue to the correct answer.

Example:

Poor: The laws of motion in Physics were formulated by _____. (Two words - Isaac Newton)

Better: The laws of motion in Physics were formulated by _____ _____.

iv. Omit only significant words from the statement.

Example:

Poor: Every atom has a central _____ called a nucleus.

Better: Every atom has a central core called a(n) _____

v. Do not omit so many words from the statement that the intended meaning is lost.

Example:

Poor: The _____ were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.

Better: The Pharaohs were to Egypt as the _____ were to Persia and as _____ were to the early tribes of Israel.

vi. Avoid obvious clues to the correct response.

Example:

Poor: Plants and animals that live in water constitute a(n) _____ ecosystem.

Better: The ecosystem in which plants and animals live is _____

vii. Be sure there is only one correct response.

Example:

Poor: Trees which shed their leaves annually are _____.

Better: Trees which shed their leaves annually are called _____

viii. If possible, put the blank at the end of a statement rather than at the beginning.

Asking for a response before the student understands the intent of the statement can be confusing and may require more reading time.

Example:

Poor: _____ is the measure of central tendency that is most affected by extremely high or low scores.

Better: The measure of central tendency that is most affected by extremely high or low scores is the _____.

(Adapted from "Guideline for Constructing Effective Test Items", Kalim, T. 2009, pp 8 -9, & "Is This a Trick Question? A Short Guide to Writing Effective Test Questions", Clay, B. 2001, pp. 36-37)

Remember to avoid using a long quote with multiple blanks to complete. When working with definitions, supply the term, not the definition, for a better judge of student knowledge.

5.1.4 True-False

The "true-false" test items require the students to select a response - true or false, based on their recognition of correct or incorrect information presented to them. There are many situations which call for whether a specific solution is right or wrong, whether to continue or to stop, whether to use a singular or plural construction, and so on. For such situations, the true-false item is an ideal measuring device.

True-false questions can test student's recall or comprehension of the conceptual knowledge. Students can generally respond to many questions, covering a lot of content, in a short amount of time. They are generally easy to construct and score. Because they can be objectively scored, the scores are more reliable than for items that are at least partially dependent on the teacher's judgment.

The true-false test can also be used effectively as an instructional test to promote interest and introduce points for discussion. Despite its values in objectivity and easy scoring, they are difficult to discriminate between students that know the material and students who do not because students have a 50-50 chance of getting the right answer by guessing. Hence, a large number of items is needed for high reliability of the test.

Guidelines on Writing True-False Test Items

One method for developing true-false items is to write a set of true statements that cover the content, then convert approximately half of them to false statements. Remember: When changing items to false (as well as in writing the true statements initially), state the items positively, avoiding negatives or double negatives.

1. *Base true-false items upon statements that are absolutely true or false, without qualifications or exceptions.*

Include directions that tell students how and where to mark their responses.

Example:

Poor: The solution of salt contains sodium and chlorine.

Better: Table salt is a compound of sodium and chlorine elements.

2. Express the item statement as simply and as clearly as possible.

Keep language as simple and clear as possible as extremely long or complicated statements will test reading skill rather than content knowledge.

Example:

Poor: The Social Studies is a subject taught in schools in the primary level and students learn about how people in the society interact and develop the society.

Better: The Social Studies is a subject taught in primary school to teach about the dynamics of society in the developmental process.

3. Express a single idea in each test item

Randomize the sequence of true and false statements to break the pattern of responses.

Example:

Poor: Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.

Better: Water will boil at a higher temperature if the atmospheric pressure on its surface is increased.

1. *Include enough background information and qualifications so that the ability to respond correctly to the item does not depend on some special, uncommon knowledge.*

If a proposition expresses a relationship, such as cause and effect or premise and conclusion, present the correct part of the statement first and vary the truth or falsity of the second part.

Example:

Poor: The second principle of education is that the individual gathers knowledge.

Better: According to John Dewey, the second principle of education is that the individual gathers knowledge.

5. *Avoid lifting statements from the text, lecture, or other materials so that memory alone will not permit a correct answer.*

Avoid the tendency to add details in true statements to make them more precise. The answers should not be obvious to students who do not know the material.

Example:

Poor: Nationally rooted, globally competent is the goal of education.

Better: Education is to empower children as individuals who manifest Bhutanese values and competent to apply knowledge in their daily life.

6. *Avoid using negatively stated item statements.*

Make the incorrect response more plausible or attractive to those without the specialized knowledge being tested.

Example:

Poor: The repeated addition is not multiplication.

Better: The repeated addition is multiplication.

7. Avoid the use of unfamiliar vocabulary.

Use popular misconceptions/beliefs as false statements.

Example:

Poor: The market values of shares, inter-alia, is influenced by inflation rate at the time of selling

Better: The market values of shares are influenced by inflation rate at the time of selling.

(Adapted from "Is This a Trick Question? A Short Guide to Writing Effective Test Questions", Clay, B. 2001, pp 20-23)

8. Forced choice questions related to the content

While true-false and other forced choice questions are generally used to measure knowledge and understanding, they could also be used at higher levels. Students could be provided with a set of information new to them, perhaps a portfolio, set of data, or a written work of some type, then asked various forced choice questions related to the content or the presence/absence of certain characteristics in the work.

Example

The table below shows the revised rate of parking fees collected in Thimphu Thromde.

Time	Parking fees (Nu)
First one hour	Nu 40
Every subsequent ½ an hour	Nu 15

State true or false for each statement given below based on the table given above.

	Statement	True/False
	Dophu parked his car from 2.30 p.m to 7 p.m and paid Nu 200.	
	Pema parked his car for 5½ hours and paid Nu 175.	

5.2 Expanded Response / Constructed Response Items

This test items require examinees to construct or prepare responses based on the knowledge developed during a course, or by rewriting bits of given stimulus material in a new way. Many such items are 'open', in that equal credit during scoring might be given to several totally different responses. These subjective tests unlike objective tests for which there is a definitive standardized or formulated answer, these tests are evaluated based on the judgment or opinion of the examiner. Tests of this nature are often designed in a way the student is presented with a few questions or writing prompts for which students demonstrate mastery of the learning objective in response to the question. When composing prompts as test questions, it is crucial that the prompt is clearly and precisely stated. Ensure that prompt elicits the type of thinking skill that one wants to measure and that the students' task is clear. For example, if students are to compare two items, provide or list the criteria to be used as the basis for comparison.

When grading subjective tests or test items, the use of an established set of scoring criteria or a well-developed rubric helps to level the playing field and increase the test's reliability

Examples of subjective test items include the following: short answer and essay.

5.2.1. Short Answer

Short answer questions stimulate students to write brief answers which may vary in length from one or two words to a few sentences. Short answer questions are often used to test basic knowledge of key facts and terms.

Short answer questions have the following advantages:

- relatively easy to construct and can be constructed faster than multiple choice questions.
- make it difficult for students to guess the answer.
- provide students with more flexibility to explain their understanding and demonstrate creativity than they would have with multiple choice questions.
- provide more structure than essay questions and thus are often easy and faster to mark and often test a broader range of the course content than full essay questions.

Despite the above advantages, scoring of essay test items is relatively laborious and can be quite subjective.

Different Varieties of Short answer questions

a. The question Variety:

Example:

(1) In which year was Druk Gyalpo Jigme Dorji Wangchuk born? ……
(2) What is the sum of the three angles of a Triangle? ……..

b. The Completion Variety:

Example:

(1) Druk Gyalpo Jigme Dorji Wangchuk was born in the year ……
(2) The sum of the three angles of a triangle is ……

c. The Association variety:

Example:

Direction: In the bank provided, write the name of the mathematical law of equality, which is demonstrated.

(1) 3 + 2 = 2 + 3 (2) (2 + 3) + 5 = 2 + (3 + 5)

……. ……..

Guidelines for writing good short answer items:

i. A direct question is generally preferable to an incomplete sentence.

Example:

Poor: The angle incident is equal to the angle of reflection is ……....

Better: State the first law of reflection of light.

ii. Structure an item so that the required response is concise. Be sure that only a unique word, phrase or number can answer the item.

Example:

The Earth revolves around the …….

iii. For incomplete-statement types, restrict the number of blanks to one, or at most, two. Be sure that only a unique word, phrase or number can answer the item.

Example:

The earth revolves around the …….

iv. Place the blank near the end of an incomplete sentence or in the margin for a direct question.

v. Blanks for answers should be equal in length.

vi. Provide sufficient answer space. Provide as many as spaces or blanks as there are words in the answer.

Example: The First King of our Country was _____ _____

vii. Do not delete too many words from a completion question.

Example: The sun _____ temperature and _____ the _____ shines we have more _____

viii. In computational problems, specify the degree of precision and the units desired.

Example: Direction: In the bank provided, write the name of the mathematical law of equality, which is demonstrated.

(1) 3 + 2 = 2 + 3 (2) (2 + 3) + 5 = 2 + (3 + 5)

……. ……..

ix. *Consider having a mix of questions, some that supply the term and students are to provide the definition, or supply the definition and prompt students provide the term.*

When students are expected to provide the terms, the item can be structured as fill-in-the-blank questions. This mix of formats will better test student knowledge because it doesn't rely solely on recognition or recall of the term.

Example:

(1) What are anthropogenic activities? _____

(2) Elements combined in definite proportion to form _____

(3) The part of speech that talks more about the subject is _____

(4) Define sales tax. _____

5.2.1 Essay Type

Essay tests lend opportunities for students to demonstrate their understanding and ability to think critically, organize their thoughts, and be creative and original and communicate. They are the best measure of students' skills in higher-order thinking and written expression." (Barbara Gross Davis, Tools for Teaching, 1993, 272). A typical essay test usually consists of a small number of questions to which the student is expected to relate and organize knowledge and skills in logical, integrated response.

An essay test item can be an extended response item or a short answer item. The main advantages of essay and short answer items are that they permit students to demonstrate achievement of such higher-level objectives as analysing and critical thinking. Written items offer students the opportunity to use their own judgment, writing styles, and vocabularies.

The Center for Teaching and Learning (2022) provides insights on fundamentals of constructing essay test items as described in the following section.

When are essay exams appropriate?

- When there is need to measure students' ability to analyse, synthesize, or evaluate;
- When these skills are being taught as the focus of the teaching or the content lends itself to more critical analysis as opposed to recalling information.

How do we design essay test items is deliberated under the "Guidelines to Writing Essay Test Items" in the following section.

Guidelines on Essay Type Test Items

An essay test item can be an extended response item or a short answer item. Generally, the words "how" and "why" are the best, followed by "what" and "when," although the latter two can more easily lead to a single word or very short answers. The level of the answer expected is important in determining whether a more specific statement is useful.

While respecting the ideas of using words and phrases that alert students to the kind of thinking, for example, identify, compare, or critique, and conscious of time for writing responses, effective essay test items can be developed by considering the following guidelines:

1. Formulate the question so that the task is clearly defined for the student

Use words that aim the student to the approach you want them to take. Words like discuss and explain can be ambiguous without being specific of what points to be discussed.

Example:

Poor: Discuss the Middle path development philosophy.

Better: Discuss the Middle path development philosophy in the context of conserving the forest of 62% land coverage.

In order to make the purpose of the test clear and specific, the assessor must remember to -

- Define the behaviour the examinee is expected to exhibit or describe the process to be exhibited before beginning to write the essay-format test item.
- Determine the level of learning expected in response to a test item on this topic from the learning objective. Use the words "discuss," "explain," or "compare" only when deliberately aiming at higher cognitive levels.
- Make essay questions comprehensive rather than focused on small units of content.
- Define the problem explicitly and provide clear directions as to the expectations of an essay-format test item.
- Have another test item developer review the test item for accuracy, format, and readability. The subject matter experts should also agree on the relevance or importance of a particular test item to course completion, as well as the correct response.

2. Pay attention to the number of items

In order to obtain a broader sampling of course content, use a relatively large number of questions requiring shorter answers (one-half page) rather than just a few questions involving long answers (2-3 pages).

Example:

Question 7 *(Pictures and tables omitted)*

i. Culture and tradition have influence on the design and structure of the buildings in our society. Support the statement with TWO valid reasons by referring to the pictures given, how culture and tradition have influenced design and structure in our country? [2]

ii. Figure A shows Thimphu in late 1980s and figure B, Thimphu in 2000s. Compare these two photographs and, explain how humans have impacted the environment? [2]

iii. Draw a bar graph with the help of population data given in the table (Population development in Bhutan since 1960). [3]

iv. The change in population has been less than 2.0 % since 2003 as per the information given in Question 7 c. What do you think are some of the reasons for this? Mention at least TWO reasons. [2]

(Source: Sample Class VIII Common Examination – Geography)

Assessor must remember to -

- Allow students an appropriate amount of time. It is helpful to give students some guidelines on how much time to use on each question, as well as the desired length and format of the response, such as full sentences, phrases only, outline, etc.
- Ensure that all the data (stimulus material) needed by the student is provided and clearly displayed together.
- Referring to the learning objectives, select a topic from which the test item is structured. Ask only for important relevant information.
- Ensure the test item is within the educational maturity level of examinees.

- Limit the problem - the scope of a content area should not be too large. Ask questions that are relatively specific or focused, and which require relatively brief responses.

3. Write essay items at different levels of learning

The purpose of the essay items is to measure the higher cognitive processes and abilities. The item should be the task that tests the student's ability to relate and use knowledge in order to analyse, justify, explain, contrast, evaluate, and apply them in solving problems. Therefore, use verbs that elicit thinking abilities that one wants students to demonstrate. Assessors often must use their best judgment about what cognitive skill each question is measuring; the easier way is to refer the Bloom's taxonomy.

Assessor must remember to -

- Require students to demonstrate command of background information by asking them to provide supporting evidence for claims and assertions.
- Ask test items that require the student to demonstrate the ability to use essential knowledge and do so in situations that are new or novel for them, rather than simply recalling information.
- Define the problem explicitly, an essay-format test item is of no use if the student cannot comprehend the test item.

4. Choose a scoring model

The major task in scoring essay tests is to maintain consistency, to make sure that answers of equal quality are given the same number of points. There are two approaches to scoring essay items: (1) analytic or point method and (2) holistic or rating method.

- **Analytic:** Before scoring, prepare an ideal answer in which the major components are defined and assigned point values. Read and compare the

student's answer with the model answer. If all the necessary elements are present, the student receives the maximum number of points. Partial credit is given based on the elements included in the answer. In order to arrive at the overall exam score, the instructor adds the points earned on the separate questions.

- **Holistic:** This method involves considering the student's answer as a whole and judging the total quality of the answer relative to other student responses or the total quality of the answer based on certain criteria that you develop.

Assessor must remember to -

- Decide how to treat irrelevant or inaccurate information contained in students' answers.

- Develop the test item grading criteria. Identify the key responses expected from the student. Identify each necessary response (e.g., by underlining key words, phrases, or steps, or by circling important parts of graphs or diagrams). Determine the percentage value to be allocated for each test item subpart (based on their relative importance).

- Ensure that the point value for each test item or part of a test item is stated.

- Check the test item against the grading criteria:

  ✓ Does the test item clearly request what is considered important in the grading criteria?

  ✓ Is the time necessary to answer the test item proportional to the importance of the test time? (Critical objectives justify more time if required.)

  ✓ Do trivial details (e.g., an extended calculation) take too much time to address?

  ✓ How could the test item and/or the grading criteria be revised to be more consistent?

5. How to write HOT items with example

- Write items where the "load" is on the problem to solve rather than on the content. Generally, if content is needed to respond to the item that is more difficult than the problem embedded in the item, it should be provided as part of the item.
- Write items that require students to predict the outcome of a situation rather than simply labelling or listing the specific elements associated with a response.
- Give examples, and ask for the principle or theory they illustrate.
- Design items that permit multiple interpretations or solutions; this could require for constructed response items students being asked to justify or explain their answers.
- Design constructed response or performance items in which students have to explain their reasoning and the items have rubrics that reward higher order thinking.

Examples:

Analyzing

i. Social Science

Analyse the effects of the introduction of railways in different countries in the world.

ii. English

Read aloud the description of 'A Beautiful' man (Chapter:7 A visit to Cambridge Honeydew). What is the most beautiful sentence in the description? Justify your choice.

Evaluating

i. Science

Justify your choice of equipment or resources to perform a first-hand investigation to draw a longitudinal section of xylem tissue.

ii. Social Science

Assess the influence of the ideologies of communism and capitalism on the origins and development of the Cold War to 1968.

iii. English

What did the author mean by saying "Before you, like a lantern whose walls are worn so thin you glimpse only the light inside, is the incandescence of a man. The body, almost irrelevant, exists only like a case made of shadows". Do you agree or disagree with author? Give reasons for your decision. (Chapter:7 A visit to Cambridge Honeydew)

Creating

i. Science

Develop a safe work practice procedure for using a piece of chemical application equipment. In your answer, include at least five steps in a logical sequence and explain why they are a significant component of your solution.

ii. Social Science

Write the script of a street play giving the importance of tree plantation and try to enact it in your locality.

5.2.3 Problem Solving

An essay is not the only form of a subjective test item. Another form is the problem solving or computational exam question. Such items present the student with a problem situation or task and require a demonstration of work procedures. Problem solving is classified as subjective due to the procedures used to score item responses. Assessors can assign full or partial credit to either correct or incorrect solutions depending on the quality and kind of work procedures presented.

This test items are relatively easy to construct and serve as the means of assessing higher-order skills and can measure complex learning outcomes and

an extensive amount of content or objectives. They are most appropriate in measuring objectives which focus on the ability to apply skills or knowledge in the solution of problems. Further, as these test items require students to provide an original response rather than to select from several alternatives, guessing is minimised.

Unfortunately, these test items require an extensive amount of instructor time to read/grade. There is a perceived likelihood of subjectivity and bias in scoring the responses of examinees.

Guidelines on Problem solving Items

1*. Provide directions which clearly inform the student of the type of response called for*

State in the directions whether the student must show the work procedures for full or partial credit. Ask questions on which experts could agree that one solution and one or more work procedures are better than others. Work through each problem before classroom administration to double check accuracy.

Example:

3.a. A farmer borrowed money from two banks to start a farming business as shown in the table given below. [3]

Bank	Annual Interest Rate (%)	Principal (Nu)	Loan Time (Years)
BDBL	10.25%	45,000	3
BOB	10.85%	60,000	2

i. What is the total amount he must pay to each bank at the end of their term period?

ii. Which bank would you suggest him to borrow money in the future? Explain your answer.

b. A school is organizing a cultural programme at the end of the year. For a Dzongkha dance, 48 students volunteered. The ratio of boys to girls is 10 to 14. How many girls will not participate in the Dzongkhag dance, if the required ratio of boys to girls is 1 to 1? [2]

(Source: Sample Class VIII Common Examination – Mathematics)

2. Separate item parts and indicate their point values

For each test item, a correct answer should be supplied in order to develop an answer key. The answers should be unambiguous so that the grading of the test item will be consistent. The minimum response for full credit should be given, and indications of the relative value of partial responses should be made. This kind of information is obvious in multiple-choice and matching test items, but short-answer and essay test items do leave room for interpretation.

To prevent interpretive inconsistencies in grading short-answer or essay test items, the designated instructor should include the minimum answer required for full credit. If the exact wording or specific concept should be included to fulfil the requirement of a correct answer, one method that may be used is to underline the necessary parts and assign point value to the individual parts. If the exact underlined wording is not necessary, include a note to the grader that the content is necessary, but the wording is left to the grader discretion.

Example:

Test item: State the purpose of the emergency diesel generator. (4 points)

Answer: Provides emergency power (2.0 points) to essential loads (1.0 point) in the event of a total loss of power (1.0 point). NOTE: *Alternate wording acceptable—grader discretion requested.*

After the test item and answer key are written, the point value should be assigned in a clear and consistent manner. The assignment should be based on a point scale with the "easiest" test item assigned lower point value and the "more difficult" test item assigned higher point value. The point value should be placed in parentheses at the end of the test item stem.

Problem Solving

Problem solving is conceptualisation of a problem, considering options, implementing a plan and monitoring to reach a final outcome.

Skill Strands	Skill Sub Strands
Identify problem	Collect information
	Understand problem
	Analyse problem
Explore options	Identify alternatives
	Consider from other perspectives
Strategize	Select strategy
	Make plan
	Implement the plan
	Persevere
Monitoring	Evaluate the implementation
	Provide feedback
	Try alternatives

5.2.4 Scoring Essay-type Test Items

A major difficulty in the use of essay-format test items is the inability to maintain consistent scoring. The answer should clearly identify the key points of each test item, and the grading criteria should be followed closely to ensure consistency of scoring. Effective grading criteria eliminate bias toward students (whose answers appear plausible, have neat handwriting, or are known to the instructor) by removing subjectivity and promoting objectivity.

How do we grade essay exams?

- Develop criteria for appropriate responses to each essay question
- Develop a scoring guide that tell what you are looking for in each response and how much credit you intend to give for each part of the response

- Read all the responses to question 1, then all the responses to question 2, and on through the exam. This will provide a more holistic view of how the class answered the individual questions

During scoring, it is easy to drift from the established grading criteria. It is, therefore, necessary to check each answer against the answer and grading criteria for every test. It is also advisable for the same person to score one test item across all tests before scoring the next test item. This increases the consistency of the scoring process. Also, during scoring, the examiner may find that a large percentage of students incorrectly answer a test item. The examiner should then re-evaluate the test item to determine if the test item could have been interpreted differently than what was intended or if there may have been other potentially correct answers (US department of Energy, n.d, pp 29).

6. TEST ITEM ANALYSIS

One way to assess how well your test is functioning for the purpose of providing evidence of students' performance in learning, is to look at how well the individual items do so. The basic idea is that a good item is one that good students get correct more often than do poor students. One might end up with a big spread in scores, but what if the good students are no more likely than poor students to get a high score? Despite the proper instruction, the test has not really assessed what they have learned, it indicates that the test is "not working."

In the current scenario, test items are grossly subjective as they are designed and developed based on the priorities of the assessors concerned and schools. Some test items are easy while others are difficult, which may be the precursor to assessment fallacy. If the assessment is to do justice to what it is purported, technical elements of test discrimination power, test difficulty and Cronbach alpha correlation among test items, due consideration and statistical treatment of evidence gathered are vital for reliable and valid judgement of students' learning performance. The item or psychometric analysis parameters include difficulty index, reliability, discrimination index, distractor efficiency. The descriptive statistics of the exam are important and can provide helpful generalized information. The descriptive statistics include scores frequency, the mean, the mode, the median, and the standard deviation.

6.1 What is Item Analysis?

Item analysis is one process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test in totality. Item analysis is especially valuable in improving items which will be used repeatedly in later tests, and can also be used to eliminate ambiguous or misleading items in a single test administration.

6.2 Parameters of Test Item Analysis

Two principal measures are generally used in item analysis; item difficulty and item discrimination.

i. Item Difficulty

The difficulty of an item in a test is the percentage of the sample taking the test that answers that question correctly. This metric takes a value between 0 and 1. High values indicate that the item is easy, while low values indicate that the item is difficult. For example, if out of 100 students only 30 students respond to an item correctly then its difficulty will be 0.3 (30/100).

Item difficulty is denoted by letter 'p' and its interpretation for dichotomous item is summarised in Table 2.

Table 2: Summary of interpretation of p-Value for dichotomous item

p-Value	Interpretational for items
1.00	Items are extremely easy (everyone gets it right)
0.80	Items are easy (80% get it right)
0.50	Items are of medium difficulty (half get it right; half get it wrong)
0.30	Items are difficult (70% get it wrong)
0.00	Items are difficult (everyone gets it wrong)

Generally, items with 'p' values between 0.21 to 0.79 are considered for inclusion in tests to have good spread of items over the test.

ii. Item Discrimination

It is a measure of how well an item distinguishes between the high achiever and low achievers. A good test item should have a characteristic that high-ability students may more frequently answer it correctly than lower-ability students. Item discrimination expresses how well an item can differentiate among students with different ability levels. This is judged by studying the correlation

between the right or wrong scores that students receive when their scores are summed up across the remaining items.

For a test, item having discrimination values ranging between 0.4 to 0.7 are considered.

Another measure of item discrimination is the point-biserial correlation between the scores on the entire test and the scores on the single item. The table below shows the interpretation and discrimination values with respect to quality of an item.

Table 3: Interpretation and discrimination values with respect to quality of an item

Discrimination Value	**Interpretation**
> 0.40 (> 40%)	Strong, positive discrimination
0.25 – 0.40 (25% – 40%)	Moderate, positive discrimination
0.10 – 0.25 (0% – 25%)	Weak discrimination
= 0.00 (0%)	No discrimination
<0.00 (<0%)	Negative discrimination

This is measured by selecting two groups: high skill and low skill based on the total test score. For example, one can assign the high skilled group to be those subjects whose score on the entire test is in the top half, and the low skilled group to those in the bottom half. Alternatively, assign the high skilled group to be those subjects whose total score is in the top 33%, and the low skilled group those in the bottom 33%. The discrimination index is then the percentage of subjects in the high skilled group who answered the item correctly minus the percentage in the low skilled group who answered the item correctly.

Example 1:

A 20-question test is given to 18 students. The table …. shows the results for question 1 and for the whole test. Calculate the difficulty of question 1, its discrimination index (using the top third vs. the bottom third) and its point-biserial correlation.

Table 4: item analysis for MCQ items

	A	B	C	D	E	F	G	H
1		MCQ ITEMS KEY						
2	stds	Q1	Q2	Q3	Q4	Q4	Q5,,,,,,	TOT
3	1	1						20
4	2	0						19
5	3	1						18
6	4	1						17
7	5	1						17
8	6	1						17
9	7	1						16
10	8	1						16
11	9	0						15
12	10	1						15
13	11	0						14
14	12	1						14
15	13	0						14
16	14	1						13
17	15	0						12
18	16	1						12
19	17	0						8
20	18	0						5
21	Item difficulty of Q1			0.61				
22	Item disc of Q1:			0.5				
23	Point Biserial of Q1			0.4701				

Based on the analysis, it may be inferred that

- the difficulty for Q1 is given by = SUM (B4:B20)/ COUNT (B4:B20) = 11/18 = .611.
- since 5 of the top 6 students got question 1 right and 2 of the bottoms 6 got the question right (top 33% and bottom 33%, (=6/18*100)), the discrimination index = 5/6 – 2/6 = 3/6 = 0.5.
- the point-biserial correlation coefficient = CORREL (B1:B20, C4:C20) = 0.405.

The conclusion of the Q1:

- item (Q1) is medium difficulty
- has strong, positive discrimination
- the correlation between scores on the entire test and the score on the single item (Q1) is moderate.

Example 2:

Find the difficulty, discrimination index and correlation coefficient for each of the five open ended questions in a test based on the data in Table 5, where row max shows the maximum credit assigned to each question from Q1 to Q5 (maximum score for the whole test is the sum of these values, i.e. 10).

Table 5: Item analysis of partial credit items/ERQs

	A	B	C	D	E	F	G
1		Open ended response question -item					
2	stds	Q1	Q2	Q3	Q4	Q5	TOT
3	1	1	2	2	3	2	10
4	2	0	2	2	3	2	9
5	3	1	1	2	2	2	8
6	4	1	1	2	2	2	8
7	5	1	1	2	2	2	8
8	6	1	1	2	2	1	7
9	7	1	1	2	1	1	6
10	8	1	1	2	1	1	6
11	9	0	2	2	1	1	6
12	10	1	2	1	0	1	5
13	11	0	2	1	1	1	5
14	12	1	2	1	0	1	5
15	13	0	1	1	1	1	4
16	14	1	1	1	1	1	5
17	15	0	1	1	1	1	4
18	16	1	0	1	1	1	4
19	17	0	1	2	0	0	3
20	18	0	0	0	0	2	2
21	max	1	2	2	3	2	10
22	difficulty of Q5					0.638889	
23	disc of Q5					0.416667	
24	Corr of Q5 with total score					0.611477	

The above example is focused on Q5. Ut may be inferred that,

- difficulty for Q5:

= SUM (F3:F20)/ (COUNT (F3:F20) *F21)

- the discrimination index for Q5:

= (SUM (F3:F8)-SUM (F15:F20))/ (COUNT (F3:F8) *F21)

- the correlation coefficient between the scores for Q5 and the total scores:

= CORREL (F3:F20, $G3: $G20)

Conclusion of the Q5:

- Item is medium difficulty
- Moderate, positive discrimination
- The correlation between scores on the entire test and the score on the single item (Q5) is moderate

Apart from the above two principal measures, it is also important to check the other test statistics such as Reliability coefficient of the test.

The reliability of a test refers to the extent to which the test is likely to produce consistent scores. The reliability coefficient reflects three characteristics of the test:

- **Inter correlations among the items** — the greater the relative number of positive relationships, and the stronger those relationships are, the greater the reliability. Item discrimination indices and the test's reliability coefficient are related in this regard.

- **Test length** — a test with more items will have a higher reliability, all other things being equal.

- **Test content** — generally, the more diverse the subject matter tested and the testing techniques used, the lower the reliability.

Reliability coefficients theoretically range in value from zero (no reliability) to 1.00 (perfect reliability). In practice, their approximate range is from .50 to .90 for about 95% of the classroom tests scored is considered.

Example 3:

How to find the Reliability coefficient of the test?

Table 6: Reliability coefficient of the test using Cronbach's Alpha method.

	A	B	C	D	E	F	G
1		Open ended response question -item					
2	stds	Q1	Q2	Q3	Q4	Q5	TOT
3	1	1	2	2	3	2	10
4	2	0	2	2	3	2	9
5	3	1	1	2	2	2	8
6	4	1	1	2	2	2	8
7	5	1	1	2	2	2	8
8	6	1	1	2	2	1	7
9	7	1	1	2	1	1	6
10	8	1	1	2	1	1	6
11	9	0	2	2	1	1	6
12	10	1	2	1	0	1	5
13	11	0	2	1	1	1	5
14	12	1	2	1	0	1	5
15	13	0	1	1	1	1	4
16	14	1	1	1	1	1	5
17	15	0	1	1	1	1	4
18	16	1	0	1	1	1	4
19	17	0	1	2	0	0	3
20	18	0	0	0	0	2	2
21				variance of total score			4.617647
22	varience	0.251634	0.418301	0.382353	0.888889	0.330065	
23				sum of variance of each item:			2.271242
24				Number of item			5
25				Cornbach's Alpha			0.206619

Cronbach's Alpha = $(\frac{k-1}{k}\frac{k-1}{k})$ $(s^2_y - \sum s^2_i)/ s^2_y$; k is number of item, s^2_y variance of total score and $\sum s^2_i$ sum of variance of each item.

Cronbach's Alpha = ((5-1) / 4) * (4.617647- 2.271242)/ 4.6176647 = 0.206619.

iii. Interpretation guidelines

The following general guidelines can be used to interpret reliability coefficients for classroom exams:

Reliability	Interpretation
0.90 and above	Excellent reliability; at the level of the best standardized tests
0.80 – 0.90	Very good for a classroom test
0.70 – 0.80	Good for a classroom test; in the range of most. There are probably a few items which could be improved.
0.60 – 0.70	Somewhat low. This test needs to be supplemented by other measures (e.g., more tests) to determine grades. There are probably some items which could be improved.
0.50 – 0.60	Suggests need for revision of test, unless it is quite short (ten or fewer items). The test needs to be supplemented by other measures (e.g., more tests) for grading.
0.50 or below	Questionable reliability. This test should not contribute heavily to the course grade, and it needs revision.

The measure of reliability is Cronbach's Alpha. In general, for any assessment, the following factors play a critical role in the test during the test item development and test evaluation:

- Validity: Is it assessing an appropriate knowledge, skill, values and attitude?

- Reliability: Does it work consistently over the period or with the different cohort?
- Discriminating: Does the chance of answering it correctly correlate with ability of students?
- Authentic / worthwhile: Is it worth asking?
- Difficulty: Does the assessment comprise of range of item difficulty?

In short, Item analysis is a statistical analysis of the student's responses on a test. Collection and summarization of students' responses can provide quantitative objective information that is useful in deciding the quality of the test items and increasing the assessment's efficiency. Also, Item analysis "investigates the performance of items considered individually either in relation to some external criterion or the remaining items on the test" (Rezigalla, A.A., 2022).

In the area of assessment, Item analysis is a statistical analysis of student's responses on examination items and the relationship between them which provides constructive feedback about items quality. This ensures in enhancing the effectiveness of the examinations, and supports examination validity and reliability. The feedback of item analysis can support modification of instruction methods. Therefore, it is envisaged that teachers endure to initiate the process of doing Test Item Analysis for any test and examinations conducted in one's own teaching subject and generate an annual report detailing what practices worked well and identify imperatives in improving students' learning performance on a regular basis.

In General, item writers should have;

- an understanding of the content standards of the curriculum and the performance of individuals;
- an understanding of the learning objectives, experience and competency;

- a thorough knowledge of the subject matter - including knowledge of popular fallacies and misconceptions;
- good written communication and technical item writing skills;
- imagination and ingenuity- effective item writers are trained, not born;
- clearly connect knowledge to different context- real life situation;
- avoid linguistic complexity of content-based test items that may threaten the construct validity of assessment;
- a detail test blueprint as it provides a major source of validity evidence for the test;
- always avoid regional, geographical, racial, ethnical and gender test bias; and
- involve professional editing for test items.

In essence, item analysis gives a way to exercise additional quality control over the tests. Well-specified learning objectives and well-constructed items give you a head-start in the process, but item analyses provide feedback on the test items. It answers the question of whether the test is working by asking the same question of all individual items—how well does it discriminate? If there are lots of items that do not discriminate learners, test items are not functioning to measure the target learning outcomes. Item analyses can also help to diagnose why some items do not work especially well, and thus suggest ways to improve them.

7. TERMINOLOGY USED IN TESTING

The following are some common terms used in testing:

Answer sheet or booklet: A piece of test stationery separates from the question booklet, on which candidates record personal details and the answers to the test items.

Assessment: Methods to determine achievement of educational outcomes or goals. Formative assessments are interim measures that guide teacher and student; they often measure the acquisition of content and skills. Summative assessments measure the student's ability to transfer content and skills. End-of-course exams are an example of a summative assessment.

Competency: Competency is more than the acquisition of knowledge and skills. Competency implies the ability to analyze and synthesize information and transfer learning from one subject to another, and to real life context.

Competency-based Learning: Flexible methods of teaching that enable students to progress as they demonstrate mastery of academic content, regardless of time, pace, or place of learning. Performance-based learning is an alternate term for the same concept.

Constructed response item: A test item which allows or requires candidates to produce individual responses rather than merely select from a list of given options.

Criterion score: The mean facility of an item considering only the performance of those candidates who attempted to answer: can also be calculated for whole tests or groups of items, using the same rule.

Discrimination: The ability of an option to distinguish between those groups

of candidates who had greater and lesser ability as indicated by their performance on the whole test. The indices used are usually *expressed as* positive or negative fractions of 1.0 and can be derived using a number of different formulae (e.g., point bi-serial; phi coefficients, etc.)

Distractor: In a multiple-choice item, an option for choice which is not the keyed answer, but which has been written in such a way as to distract weaker candidates from selecting that key.

Distractors: These are the incorrect choices for the item.

Editing: Preparation of refined versions of tests or items after other key stages in the development process, such as panelling or trial-testing. It is usually performed by the original item-writer(s).

Evidence of Learning: Materials, assignments, projects, and other artifacts that students may use to demonstrate that they have made progress or mastered certain knowledge or skills. Evidence of learning may be compiled in a portfolio that can be in hard copy or a digital file.

Extended response item: Any test item which requires the production of a personal response by the candidate which is longer than a sentence or two.

Facility: The index obtained by a multiple-choice item during testing which indicates the number of candidates who got it right: expressed as a percentage of the total number of candidates who sat the test. The index for an item to be used in a final test should always lie within the range 20-80 percent

Formative assessment: Assessment ***for*** learning and is an integral part of teaching and learning. Formative assessment provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring.

Item development: The term item is used as a shorthand for questions on the test. Item development can proceed only when a clearly agreed upon set of objectives is available. To as large an extent as possible, an item should measure only a single objective. Each objective, however, should be measured by one or several items, depending on the test specifications.

Item format: The item types derived from the blueprint. The selection of and test format should be based on the kinds of skills to be measured and not on some personal like or dislike for a particular item format.

Item: An individual task which forms one component of a test instrument: usually applied in the context of a multiple-choice test to indicate a single question, but can be used more broadly.

Key order: In a multiple-choice test, the sequence of letters attached to keyed answers, as in 1 D, 2 B, 3 C, etc. The same key letter should be used for no more than two consecutive items: viz. 3 C, 4 C, 5 A.

Key, keyed answer: In a multiple-choice item, the option which is designated to be correct, and for which a score is awarded.

Key: This is the correct choice for the item.

Learning Objective: A statement of what students should know or be able to do at the end of a course that supports the goal of the course. Some experts talk about objectives in terms of SKA – Skills, Knowledge, and Attitudes. Skills: what students should be able to do at the end of the course. Knowledge: what students should know and understand. Attitudes: what students' opinions will be about the subject matter of the course.

Moderation: Sometimes used to describe the process whereby expert panels meet to discuss and offer critical comment on test materials.

Multiple-choice: An item format whereby a restricted number of optional responses is offered to candidates, from which they must select one as their answer.

Omit rate: A tally of the number of candidates who did not answer a test item: especially important in estimating the performance of trial test candidates in the later items of the test, with a view to establishing an acceptable test length.

Option: In a multiple-choice item, a set of responses (usually four or five) from which the candidates select their answer.

Options/Alternatives: These are the choices given for the item.

Panel, panelling: A group of experts called together to discuss and evaluate draft items proposed for use in a test instrument.

Question book(let): A printed test instrument which contains instructions, stimulus material and test items for students to work through during the test session. Answers may be recorded in this book, or on a separate answer sheet.

Selected response item: Any item which prints a limited range of options from which candidates must select their answers.

Specification: A document which specifies in some detail the nature and composition of a test program or instrument: sometimes called a 'blueprint'.

Standard: Standards identify specific content and skills students should learn to be successful. There is not a one-to-one correspondence between standards and competencies. Multiple related standards may comprise a competency.

Stem: In a multiple-choice item, the sentence(s) or part-sentence which indicate the testing point or question, which candidates use to select their answer from the options which follow.

Stem: This is the part of the item in which the problem is stated for the examinee. It can be a question, a set of directions or a statement with an embedded blank.

Stimulus - directive: Any information in a test which candidates need to understand the specific task which they are being asked to perform (e.g., the stem of a multiple-choice item, or a detailed essay topic).

Stimulus - instructive: Any information printed in a question booklet which candidates are expected to refer to when answering the specific questions which relate to it.

Summative assessment: Summative assessment takes place after the learning has been completed and provides information and feedback that sums up the teaching and learning process. Typically, no more formal learning is taking place at this stage, other than incidental learning which might take place through the completion of projects and assignments.

Test Blueprint: The specification that identifies the objectives and skills which are to be tested and the relative weight on the test given to each. This statement necessarily precedes any development of the test. In absence of such a blueprint, test development can potentially proceed with little clear direction. The development of such a set of specifications is the crucial first step in the test development process.

Trial form: The test instrument after it has been developed, panelled and edited ready for administration to a trial population in the field.

Vetting: The process of editing and arriving at a draft test form, using the discussions and evaluations of draft items by a panel as a guide.

BIBLIOGRAPHY

ADEA (2006). *Item Writing Guidelines/Rules/Suggestions/Advice as derived from 46 Authoritative in Textbooks*, Applied Research in Education

Albano, T. (2020). Introduction to Educational and Psychological Measurement Using R, College of Education and Human Sciences, University of Nebraska – Lincoln https://www.thetaminusb.com/intro-measurement-r/introduction.html

Anggraeni, K. A. & Yusnita, M. R. (2017). *Teachers' Role in 21st Century: Teacher is a Facilitator, not a Dictator,* LUNAR, Vol. 1 No. 1; May 2017 ISSN; 2541-6804

BCSEA (2022). Competency Based Assessment Guidelines, Royal Government of Bhutan, Thimphu

Brookhart, S. & Lazarus, S. (2017) Formative Assessment for Students with Disabilities. Commissioned by the Council of Chief State School Officers State Collaboratives on Assessing Special Education Students and Formative Assessment, Washington, DC

Callahan, M. & Logan, M.M. (2022). How Do I Create Tests for my Students? Teaching, Learning, & Professional Development Center, Texas Tech University, USA https://www.depts.ttu.edu/tlpdc/Resources/Teaching_resources/TLPDC_teaching_resources/createtests.php 4/11

Cathy (2013). Competency-Based Education & Assessments. https://commission.fiu.edu/helpful-documents/competency-based-courses-degrees/competency-based-education-assessments-skilledup.pdf

Center for Teaching and Learning (2022). Constructing tests, University of Washington, Seattle, WA

https://teaching.washington.edu/topics/preparing-to-teach/constructing-tests/

Center for Teaching Excellence (n.d). Preparing Tests and Exams, *University of Waterloo* https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/exams/exam-preparation

Certification Management Services (2006). Item Writing Guidelines, 435-654-5975.

Clay, B. (2001). Is this a Trick Question? A short guide Writing effective test questions. Kansas Curriculum Centre, Department of Education, USA.

Clotilda, M. 2021). 10 things you need to know about Competency-based Assessments. https://www.creatrixcampus.com/blog/10-things-you-need-know-about-competency-based-assessments

Cohen, A. S. and Wollack, J. A. (n.d). Handbook on Test Development: Helpful Tips for Creating Reliable and Valid Classroom Tests, Testing & Evaluation Services University of Wisconsin-Madison

DCPD (2021). National Student learning Assessment Framework, Ministry of Education Royal Government of Bhutan, Thimphu

DCPD (2022). National School Curriculum: Curriculum Frameworks (All subjects), Ministry of Education Royal Government of Bhutan, Thimphu

DCPD (2022). National School Curriculum: Instructional Guides (All subjects), Ministry of Education, Royal Government of Bhutan, Thimphu

Department of Energy (n.d). Development of Test Items, U.S. Department of Energy Washington, D.C. 20585 https://sites.ntc.doe.gov/partners/

tr/Training%20Best%20Practices/2-Best%20Practices%20for%20Examinations%20and%20Testing/1-Development%20of%20Test%20Items%

FTTS (2020). What is CBE? Competency Based Education Consortium, https://sites.austincc.edu/cbec/what-is-cbe/

Hanna & Dettmer (2004). Formative and Summative Assessment, Northern Illinois University, Faculty Development and Instructional Design Center. http://facdev.niu.edu

IBE UNESCO (2016). *A Conceptual Framework for Competencies Assessment*, Current and Critical Issues in the Curriculum and Learning, IBE/2016/WP/CD/04.

IEA (2013). TIMSS 2015: Item Writing Guidelines, TIMSS & PIRLS International Study Center, Boston College

International Institute for Educational Planning/UNESCO (2005). *Item writing for tests and examinations: Module 5*, Quantitative research methods in educational planning, International Institute for Educational Planning/UNESCO, Paris.

Kalim, T. (2009). Guidelines for Constructing Effective Test Items, Bangladesh Education Article. https://bdeduarticle.com/guideline-for-constructing-effective-test-items/

Kentucky Department of Education (2013). Competency-based Education: Helping All Kentucky Students Succeed, Final Report / January 2013

Lawrence Erlbaum Associates (2006). Handbook of Test Development, 10 Industrial Avenue Mahwah, New Jersey

Lestari, H. (2011). An item analysis of discriminating power of English summative test. Department of English Education, Faculty of Tarbiyah and Teachers' Training, Jakarta.

McClarty, K.L. and Gaertner, M.N. (2015). Measuring Mastery: Best Practices for Assessment in Competency-Based Education, AEI Series on Competency-Based Higher Education, American Enterprise Institute.

Nilson, L. B. (n.d). Writing Objective Test Items that Assess Thinking Skills, Office of the Teaching Effectiveness and Innovation, Clemson University. (Lecture note) www.linkedin.com/in/lindabnilson

PEARSON (2015). Defining Competencies and Outlining Assessment Strategies for Competency Based Education Program, Pearson Education, Inc. or its affiliate(s).

PEARSON (2016). Competency-Based Learning, https://creativecommons.org/licenses/by-sa/4.0/legalcode

Peters, B. (2017). Competency Based Learning, State of the US K-12 Market

REC (2020). National School Framework, Royal Government of Bhutan, Paro

Rezigalla, A.A. (2022). Item Analysis: Concept and Application, Medical Education for the 21st Century, Edited by Michael S. Firstenberg and Stanislaw P. Stawicki. https://www.intechopen.com/chapters/81018

Sturgis, C. (2016). What Is Competency Education? iNACOL, 1934 Old Gallows Road, Suite 350 Vienna, VA 22182

Sturgis, C. and Patrick, S. (2010). When Success Is the Only Option: Designing Competency-Based Pathways for Next Generation Learning, International Association for K-12 online learning.

WMO (2018). Competency Assessment Toolkit, WMO-No.1205 Guide to Competency. https://worldweather.wmo.int/tt_cat/home.php

WMO (2018). Competency Assessment Toolkit: Competency Assessment Philosophy, WMO-No.1205 Guide to Competency. https://worldweather.wmo.int/tt_cat/home.php

WMO (2018). Competency Assessment Toolkit: Developing a Competency Assessment Plan, WMO-No.1205 Guide to Competency. https://worldweather.wmo.int/tt_cat/home.php

WMO (2018). Competency Assessment Toolkit: Introduction to Competency Assessment System, WMO-No.1205 Guide to Competency. https://worldweather.wmo.int/tt_cat/home.php

ANNEXURES

ANNEXURE 1: Bloom's Taxonomy

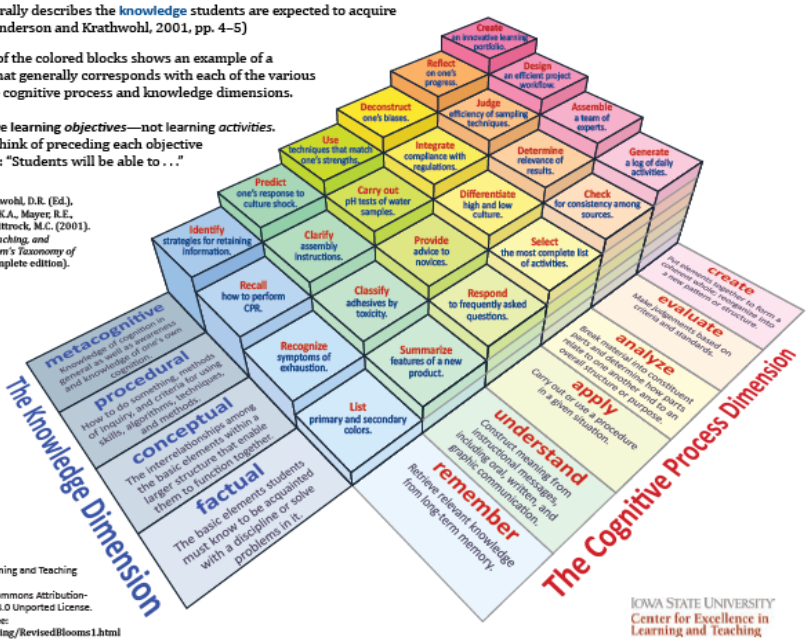
A statement of a **learning objective** contains a **verb** (an action) and an **object** (usually a noun).

- The **verb** generally refers to [actions associated with] the intended **cognitive process**.
- The **object** generally describes the **knowledge** students are expected to acquire or construct. (Anderson and Krathwohl, 2001, pp. 4–5)

In this model, each of the colored blocks shows an example of a learning objective that generally corresponds with each of the various combinations of the cognitive process and knowledge dimensions.

Remember: these are learning *objectives*—not learning *activities*. It may be useful to think of preceding each objective with something like: "Students will be able to . . ."

*Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.



Model created by: Rex Heer
Iowa State University
Center for Excellence in Learning and Teaching
Updated January, 2012
Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
For additional resources, see:
www.celt.iastate.edu/teaching/RevisedBlooms1.html

IOWA STATE UNIVERSITY
Center for Excellence in Learning and Teaching

ANNEXURE 2: Cognitive Complexity

COGNITIVE COMPLEXITY

(Clay, B. (2001). Is this a Trick Question? A short guide Writing effective test questions. Kansas Curriculum Centre, Department of Education, USA.)

Cognitive Complexity	Sample Question Frames
1. Knowledge Recognizing and recalling information, including dates, events, persons, places; terms, definitions; facts, principles, theories; methods and procedures	Who invented the…? What is meant by…? Where is the…?
2. Comprehension Understanding the meaning of information, including restating (in own words); translating from one form to another; or interpreting, explaining, and summarizing.	Restate in your own words…? Convert fractions into…? List three reasons for…?
3. Application Applying general rules, methods, or principles to a new situation, including classifying something as a specific example of a general principle or using a formula to solve a problem.	How is...an example of... ? How is...related to... ? Why is...significant?
4. Analysis Identifying the organization and patterns within a system by identifying its component parts and the relationships among the components.	What are the parts of... ? Classify ...according to... Outline/diagram...
5. Synthesis Discovering/creating new connections, generalizations, patterns, or perspectives; combining ideas to form a new whole.	What would you infer from... ? What ideas can you add to... ? How would you create a... ?
6. Evaluation Using evidence and reasoned argument to judge how well a proposal would accomplish a particular purpose; resolving controversies or differences of opinion.	Do you agree…? How would you decide about... ? What priority would you give... ?

ANNEXURE 3: Selection of Test Item Types

SELECTION OF TEST ITEM TYPES

(Adapted from "Item writing for tests and examinations Module 5" IIEP UNESCO, 2005, pp 22-27)

CRITERIA FOR CHOICE OF SELECTED RESPONSE ITEM FORMATS		
ITEM TYPE	**ADVANTAGES**	**DISADVANTAGES**
A. True-false	• easy to write • easy to mark • easy to sample variety within a course	• guessing factor very high (50%) • limited to unequivocal choices • cannot test higher order skills
B. Matching items	• useful for testing relationships • useful for testing factual information • easy to construct a large number	• the cluster approach destroys item independence • difficult to word instructions
C. Classification items	• relatively easy to construct • easy to mark • useful for testing factual information • useful for testing simple relationships	• the cluster approach destroys item independence to some degree • limited to factual sorting • limited to unequivocal facts
D. Multiple-choice items	• reduces the guessing factors • versatile – can be used to measure a wide range of cognitive processes • reduces problem of subjective scoring • analysis of results can provide much diagnostic information • easy to mark	• little, if any, stimulus given to creative thought • expensive and time-consuming to construct • difficult to measure organization and presentation of ideas • plausible distractors hard to write • presents wrong information as if it was right

CRITERIA FOR CHOICE OF CONSTRUCTED RESPONSE ITEM FORMAT		
ITEM TYPE	**ADVANTAGES**	**DISADVANTAGES**
1. Short-answer items	• excellent for testing factual knowledge • successful guessing is reduced • easy to write • easy to mark	• unsuitable for measuring complex learning • easy to respond in inappropriate ways
A. Fill-in-the-blank sentence completion	• easy to test a range of factual knowledge • guessing factor is reduced • easy to write • easy to mark	• hard to measure higher-order skills • easy to respond in inappropriate ways
B. Cloze, modified cloze	• easy to construct • a good measure of word knowledge • tests passage understanding	• some ambiguities hard to decide on • many opportunities for choice have little value • often little more than guesswork
C. Extended responses	• a means of assessing higher-order skills • relatively easy to construct • stimulate creative and critical thought as well as learned responses • can measure learning in affective domain	• sometimes lead to inadequate sampling of learning done • time-consuming and expensive to mark • difficult to achieve inter-marker reliability
D. Problem solutions	• a means of assessing higher-order skills • can measure complex learning outcomes • relatively easy to construct	• can be time-consuming to mark • sometimes difficult to establish stable assessment criteria