Course: Assessment in Science Education (697)
Semester: Spring, 2022

ASSIGNMENT No. 1

Q.1 What is a Norm-Referenced test? Discuss its purposes, content, item Characteristics and scoring interpretation.

A **norm-referenced test** (**NRT**) is a type of test, assessment, or evaluation which yields an estimate of the position of the tested individual in a predefined population, with respect to the trait being measured. Assigning scores on such tests may be described as **relative grading**, **marking on a curve** (BrE) or **grading on a curve** (also referred to as **curved grading**, **bell curving**, or using **grading curves**). It is a method of assigning grades to the students in a class in such a way as to obtain or approach a pre-specified distribution of these grades having a specific mean and derivation properties, such as a normal distribution (also called Gaussian distribution).[1] The term "curve" refers to the bell curve, the graphical representation of the probability density of the normal distribution, but this method can be used to achieve any desired distribution of the grades – for example, a uniform distribution. The estimate is derived from the analysis of test scores and possibly other relevant data from a sample drawn from the population. That is, this type of test identifies whether the test taker performed better or worse than other test takers, not whether the test taker knows either more or less material than is necessary for a given purpose. The term normative assessment is used when the reference population are the peers of the test taker.

Norm-referenced assessment can be contrasted with criterion-referenced assessment and ipsative assessment. In a criterion-referenced assessment, the score shows whether or not test takers performed well or poorly on a given task, not how that compares to other test takers; in an ipsative system, test takers are compared to previous performance. Each method can be used to grade the same test paper.

Many college entrance exams and nationally used school tests use norm-referenced tests. The SAT, Graduate Record Examination (GRE), and Wechsler Intelligence Scale for Children (WISC) compare individual student performance to the performance of a normative sample. Test takers cannot "fail" a norm-referenced test, as each test taker receives a score that compares the individual to others that have taken the test, usually given by a percentile. This is useful when there is a wide range of acceptable scores, and the goal is to find out who performs better.

IQ tests are norm-referenced tests, because their goal is to rank test takers' intelligence. The median IQ is set to 100, and all test takers are ranked up or down in comparison to that level.

As alternatives to normative testing, tests can be ipsative assessments or criterion-referenced assessments.

Ipsative

In an ipsative assessment, the individuals' performance is compared only to their previous performances. For example, a person on a weight-loss diet is judged by how his current weight compares to his own previous weight, rather than how his weight compares to an ideal or how it compares to another person.

Criterion-referenced

A test is criterion-referenced when the performance is judged according to the expected or desired behavior. Tests that judge the test taker based on a set standard (e.g., everyone should be able to run one kilometre in less than five minutes) are criterion-referenced tests. The goal of a criterion-referenced test is to find out whether the individual can run as fast as the test giver wants, not to find out whether the individual is faster or slower than the other runners. Standards-based education reform focuses on criterion-referenced testing.[6][7] Most everyday tests and quizzes taken in school, as well as most state achievement tests and high school graduation examinations, are criterion-referenced. In this model, it is possible for all test takers to pass or for all test takers to fail.

Methods

One method of grading on a curve uses three steps:

1. Numeric scores (or possibly scores on a sufficiently fine-grained ordinal scale) are assigned to the students. The absolute values are less relevant, provided that the order of the scores corresponds to the relative performance of each student within the course.
2. These scores are converted to percentiles (or some other system of quantiles).
3. The percentile values are transformed to grades according to a division of the percentile scale into intervals, where the interval width of each grade indicates the desired relative frequency for that grade.

For example, if there are five grades in a particular university course, A, B, C, D, and F, where A is reserved for the top 20 % of students, B for the next 30 %, C for the next 30–40 %, and D or F for the remaining 10–20 %, then scores in the percentile interval from 0 % to 10–20 % will receive a grade of D or F, scores from 11–21 % to 50 % will receive a grade of C, scores from 51 % to 80 % receive a grade of B, and scores from 81 % to 100 % will achieve a grade of A.

Consistent with the example illustrated above, a grading curve allows academic institutions to ensure the distribution of students across certain grade point average (GPA) thresholds. As many professors establish the curve to target a course average of a C,[clarification needed] the corresponding grade point average equivalent would be a 2.0 on a standard 4.0 scale employed at most North American universities.[1] Similarly, a grade point average of 3.0 on a 4.0 scale would indicate that the student is within the top 20 % of the class. Grading curves serve to attach additional significance to these figures, and the specific distribution employed may vary between academic institutions.

Advantages and limitations

The primary advantage of norm-reference tests is that they can provide information on how an individual's performance on the test compares to others in the reference group.

A serious limitation of norm-reference tests is that the reference group may not represent the current population of interest. As noted by the Oregon Research Institute's International Personality Item Pool website, "One should be very wary of using canned 'norms' because it isn't obvious that one could ever find a population of

which one's present sample is a representative subset. Most 'norms' are misleading, and therefore they should not be used. Far more defensible are local norms, which one develops oneself. For example, if one wants to give feedback to members of a class of students, one should relate the score of each individual to the means and standard deviations derived from the class itself. To maximize informativeness, one can provide the students with the frequency distribution for each scale, based on these local norms, and the individuals can then find (and circle) their own scores on these relevant distributions."

Norm-referencing does not ensure that a test is valid (i.e. that it measures the construct it is intended to measure).

Another disadvantage of norm-referenced tests is that they cannot measure progress of the population as a whole, only where individuals fall within the whole. Rather, one must measure against a fixed goal, for instance, to measure the success of an educational reform program that seeks to raise the achievement of all students.

With a norm-referenced test, grade level was traditionally set at the level set by the middle 50 percent of scores. By contrast, the National Children's Reading Foundation believes that it is essential to assure that virtually all children read at or above grade level by third grade, a goal which cannot be achieved with a norm-referenced definition of grade level.

Norms do not automatically imply a standard. A norm-referenced test does not seek to enforce any expectation of what test takers should know or be able to do. It measures the test takers' current level by comparing the test takers to their peers. A rank-based system produces only data that tell which students perform at an average level, which students do better, and which students do worse. It does not identify which test takers are able to correctly perform the tasks at a level that would be acceptable for employment or further education.

The ultimate objective of grading curves is to minimize or eliminate the influence of variation between different instructors of the same course, ensuring that the students in any given class are assessed relative to their peers. This also circumvents problems associated with utilizing multiple versions of a particular examination, a method often employed where test administration dates vary between class sections. Regardless of any difference in the level of difficulty, real or perceived, the grading curve ensures a balanced distribution of academic results.

However, curved grading can increase competitiveness between students and affect their sense of faculty fairness in a class. Students are generally most upset in the case that the curve lowered their grade compared to what they would have received if a curve was not used. To ensure that this does not happen, teachers usually put forth effort to ensure that the test itself is hard enough when they intend to use a grading curve, such that they would expect the average student to get a lower raw score than the score intended to be used at the average in the curve, thus ensuring that all students benefit from the curve. Thus, curved grades cannot be blindly used and must be carefully considered and pondered compared to alternatives such as criterion-referenced grading.

Furthermore, constant misuse of curved grading can adjust grades on poorly designed tests, whereas assessments should be designed to accurately reflect the learning objectives set by the instructor.

Q.2 Discuss different domains of learning objectives as conceived by Bloom, with reference to science education.

Bloom's taxonomy is a set of three hierarchical models used for classification of educational learning objectives into levels of complexity and specificity. The three lists cover the learning objectives in cognitive, affective and psychomotor domains. The cognitive domain list has been the primary focus of most traditional education and is frequently used to structure curriculum learning objectives, assessments and activities.

The models were named after Benjamin Bloom, who chaired the committee of educators that devised the taxonomy. He also edited the first volume of the standard text, Taxonomy of Educational Objectives: The Classification of Educational Goals.

The publication of Taxonomy of Educational Objectives followed a series of conferences from 1949 to 1953, which were designed to improve communication between educators on the design of curricula and examinations.

The first volume of the taxonomy, Handbook I: Cognitive was published in 1956, and in 1964 the second volume Handbook II: Affective was published. A revised version of the taxonomy for the cognitive domain was created in 2001

In the 1956 original version of the taxonomy, the cognitive domain is broken into the six levels of objectives listed below.[10] In the 2001 revised edition of Bloom's taxonomy, the levels have slightly different names and their order was revised: Remember, Understand, Apply, Analyze, Evaluate, and Create (rather than Synthesize).[9][11]

Knowledge

Bloom's Taxonomy

Knowledge involves recognizing or remembering facts, terms, basic concepts, or answers without necessarily understanding what they mean. Some characteristics may include:

- Knowledge of specifics—terminology, specific facts
- Knowledge of ways and means of dealing with specifics—conventions, trends and sequences, classifications and categories
- Knowledge of the universals and abstractions in a field—principles and generalizations, theories and structures

Example: Name three common varieties of apple.

Comprehension

Comprehension involves demonstrating an understanding of facts and ideas by organizing, summarizing, translating, generalizing, giving descriptions, and stating the main ideas.

Example: Summarize the identifying characteristics of a Golden Delicious apple and a Granny Smith apple.

Application

Application involves using acquired knowledge to solve problems in new situations. This involves applying acquired knowledge, facts, techniques and rules. Learners should be able to use prior knowledge to solve problems, identify connections and relationships and how they apply in new situations.

Example: Would apples prevent scurvy, a disease caused by a deficiency in vitamin C?

Analysis

Analysis involves examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations. Its characteristics include:

- Analysis of elements
- Analysis of relationships
- Analysis of organization

Example: Compare and contrast four ways of serving foods made with apples and examine which ones have the highest health benefits.

Synthesis

Synthesis involves building a structure or pattern from diverse elements; it also refers to the act of putting parts together to form a whole or bringing pieces of information together to form a new meaning. Its characteristics include:

- Production of a unique communication
- Production of a plan, or proposed set of operations
- Derivation of a set of abstract relations

Example: Convert an "unhealthy" recipe for apple pie to a "healthy" recipe by replacing your choice of ingredients. Argue for the health benefits of using the ingredients you chose versus the original ones.

Evaluation

Evaluation involves presenting and defending opinions by making judgments about information, the validity of ideas, or quality of work based on a set of criteria. Its characteristics include:

- Judgments in terms of internal evidence
- Judgments in terms of external criteria

Example: Which kinds of apples are suitable for baking a pie, and why?

Affective domain (emotion-based)[edit]

A scaffolding hierarchy of the affective domain related to learning

Skills in the affective domain describe the way people react emotionally and their ability to feel other living things' pain or joy. Affective objectives typically target the awareness and growth in attitudes, emotion, and feelings.

There are five levels in the affective domain moving through the lowest-order processes to the highest.

Receiving

The lowest level; the student passively pays attention. Without this level, no learning can occur. Receiving is about the student's memory and recognition as well.

Responding

The student actively participates in the learning process, not only attends to a stimulus; the student also reacts in some way.

Valuing

The student attaches a value to an object, phenomenon, or piece of information. The student associates a value or some values to the knowledge they acquired.

Organizing

The student can put together different values, information, and ideas, and can accommodate them within their own schema; the student is comparing, relating and elaborating on what has been learned.

Characterizing

The student at this level tries to build abstract knowledge.

Psychomotor domain (action-based)

A scaffolding hierarchy of the psychomotor domain related to learning

Skills in the psychomotor domain describe the ability to physically manipulate a tool or instrument like a hand or a hammer. Psychomotor objectives usually focus on change and/or development in behavior and/or skills.

Bloom and his colleagues never created subcategories for skills in the psychomotor domain, but since then other educators have created their own psychomotor taxonomies.[7] Simpson (1972)[12] proposed the following levels:

Perception

The ability to use sensory cues to guide motor activity: This ranges from sensory stimulation, through cue selection, to translation.

Examples: Detects non-verbal communication cues. Estimate where a ball will land after it is thrown and then moving to the correct location to catch the ball. Adjusts heat of the stove to correct temperature by smell and taste of food. Adjusts the height of the forks on a forklift by comparing where the forks are in relation to the pallet.

Key words: chooses, describes, detects, differentiates, distinguishes, identifies, isolates, relates, selects.

Set

Readiness to act: It includes mental, physical, and emotional sets. These three sets are dispositions that predetermine a person's response to different situations (sometimes called mindsets). This subdivision of psychomotor is closely related with the "responding to phenomena" subdivision of the affective domain.

Examples: Knows and acts upon a sequence of steps in a manufacturing process. Recognizes his or her abilities and limitations. Shows desire to learn a new process (motivation).

Keywords: begins, displays, explains, moves, proceeds, reacts, shows, states, volunteers.

Guided response

The early stages of learning a complex skill that includes imitation and trial and error: Adequacy of performance is achieved by practicing.

Examples: Performs a mathematical equation as demonstrated. Follows instructions to build a model. Responds to hand-signals of the instructor while learning to operate a forklift.

Keywords: copies, traces, follows, reacts, reproduces, responds.

Mechanism

The intermediate stage in learning a complex skill: Learned responses have become habitual and the movements can be performed with some confidence and proficiency.

Examples: Use a personal computer. Repair a leaking tap. Drive a car.

Key words: assembles, calibrates, constructs, dismantles, displays, fastens, fixes, grinds, heats, manipulates, measures, mends, mixes, organizes, sketches.

Complex overt response

The skillful performance of motor acts that involve complex movement patterns: Proficiency is indicated by a quick, accurate, and highly coordinated performance, requiring a minimum amount of energy. This category includes performing without hesitation and automatic performance. For example, players will often utter sounds of satisfaction or expletives as soon as they hit a tennis ball or throw a football because they can tell by the feel of the act what the result will produce.

Examples: Maneuvers a car into a tight parallel parking spot. Operates a computer quickly and accurately. Displays competence while playing the piano.

Key words: assembles, builds, calibrates, constructs, dismantles, displays, fastens, fixes, grinds, heats, manipulates, measures, mends, mixes, organizes, sketches. (Note: The key words are the same as in mechanism, but will have adverbs or adjectives that indicate that the performance is quicker, better, more accurate, etc.)

Adaptation

Skills are well developed and the individual can modify movement patterns to fit special requirements.

Examples: Responds effectively to unexpected experiences. Modifies instruction to meet the needs of the learners. Performs a task with a machine that was not originally intended for that purpose (the machine is not damaged and there is no danger in performing the new task).

Key words: adapts, alters, changes, rearranges, reorganizes, revises, varies.

Origination

Creating new movement patterns to fit a particular situation or specific problem: Learning outcomes emphasize creativity based upon highly developed skills.

Examples: Constructs a new set or pattern of movements organized around a novel concept or theory. Develops a new and comprehensive training program. Creates a new gymnastic routine.

Q.3 In what way the assessment of knowledge objectives in science education is important? Discuss its need and significance.

The assessment standards provide criteria to judge progress toward the science education vision of scientific literacy for all. The standards describe the quality of assessment practices used by teachers and state and federal agencies to measure student achievement and the opportunity provided students to learn science. By identifying essential characteristics of exemplary assessment practices, the standards serve as guides for developing assessment tasks, practices, and policies. These standards can be applied equally to the assessment of students, teachers, and programs; to summative and formative assessment practices; and to classroom assessments as well as large-scale, external assessments. The assessment process is an effective tool for communicating the expectations of the science education system to all concerned with science education. Assessment practices and policies provide operational definitions of what is important. For example, the use of an extended inquiry for an assessment task signals what students are to learn, how teachers are to teach, and where resources are to be allocated.

Assessment is a systematic, multistep process involving the collection and interpretation of educational data.

As science educators are changing the way they think about good science education, educational measurement specialists are acknowledging change as well. Recognition of the importance of assessment to contemporary educational reform has catalyzed research, development, and implementation of new methods of data collection along with new ways of judging data quality. These changes in measurement theory and practice are reflected in the assessment standards.

In this new view, assessment and learning are two sides of the same coin. The methods used to collect educational data define in measurable terms what teachers should teach and what students should learn. And when students engage in an assessment exercise, they should learn from it.

This view of assessment places greater confidence in the results of assessment procedures that sample an assortment of variables using diverse data-collection methods, rather than the more traditional sampling of one variable by a single method. Thus, all aspects of science achievement—ability to inquire, scientific understanding of the natural world, understanding of the nature and utility of science—are measured using multiple methods such as performances and portfolios, as well as conventional paper-and-pencil tests. Another important shift is toward "authentic assessment." This movement calls for exercises that closely approximate the intended outcomes of science education. Authentic assessment exercises require students to apply scientific knowledge and reasoning to situations similar to those they will encounter in the world outside the classroom, as well as to situations that approximate how scientists do their work.

Another conceptual shift within the educational measurement area that has significant implications for science assessment involves validity. Validity must be concerned not only with the technical quality of educational data, but also with the social and educational consequences of data interpretation.

An important assumption underlying the assessment standards is that states and local districts can develop mechanisms to measure students' achievement as specified in the content standards and to measure the opportunities for learning science as specified in the program and system standards. If the principles in the assessment standards are followed, the information resulting from new modes of assessment applied locally can have common meaning and value in terms of the national standards, despite the use of different assessment procedures and instruments in different locales. This contrasts with the traditional view of educational measurement that allows for comparisons only when they are based on parallel forms of the same test.

Q.4 Discuss comprehension in the context of its three levels. Develop three objectives of each level on any topic of science.

Reading comprehension is the ability to process information that we have read and to understand its meaning. This is a complex process where skills are built upon one another like the blocks used to make Billy's tower. There are three levels of understanding in reading comprehension: **literal meaning**, **inferential meaning**, and **evaluative meaning**.

Let's take a closer look at each of these different meanings.

Literal Meaning

Literal meaning is simply what the text says. It is what actually happens in the story. This is a very important level of understanding because it provides the foundation for more advanced comprehension. Without understanding the material on this level, you could not go any farther.

Let's use our story about Billy to provide an example. The literal meaning of the story was that Billy built a tower out of blocks. The answers to questions based on literal meaning will always be found in the text. For example: Who was building the tower? The answer is Billy.

Here are examples of the type of information that could be identified as literal meaning:

- The main idea
- Stated facts
- The sequence of events
- Characters in the story

Inferential Meaning

Inferential meaning involves determining what the text means. You start with the stated information. This information is then used to determine deeper meaning that is not explicitly stated. Determining inferential meaning requires you to think about the text and draw a conclusion.

Getting back to Billy again, what inferential meaning could we get from our story? We could infer that Billy is good at building towers! A question about inferential meaning will typically make you provide examples from the text that back up your thinking. For example: Why could you assume that Billy is good at building towers? You assume this is true because the story says that Billy's tower got higher and higher, and the design became more intricate with each block.

Examples of the type of information that could be identified as inferential meaning include:

- Generalizations
- Cause and effect relationships
- Future predictions
- An unstated main idea

Evaluative Meaning

Evaluative meaning is what the text is telling us about the world outside the story. Readers must analyze what they have read. Then, they must form an opinion based on the information.

Q.5 What are higher ability skills? In what manner these are dissimilar from other skills?

Creativity, innovation and problem-solving are thought processes that are sometimes referred to as "higher-order thinking skills" In this article, we discuss higher-order thinking skills, why they're important and how to use higher-order thinking skills in class plans. Higher-order thinking skills are thought processes that help you connect information in meaningful ways and use those connections to solve problems. These thought processes are sometimes innovative in that they apply knowledge in new ways. Critical thinking is one example of higher-order thinking skills, as well as synthesis and metacognition. Higher-order thinking skills can help you solve problems efficiently by anticipating connections between different ideas. Some cognitive researchers organize the ways they understand thought processes using taxonomies, another word for categories of ideas. One of these ways of organizing thinking, Bloom's Taxonomy, identifies skills such as making connections as more challenging but potentially more rewarding than skills like memorization by repetition. This is why they are called "higher-order" thinking skills.

Using higher-order thinking skills may help you address complex problems with creative solutions. Consider making connections between pieces of information when you are trying to solve a complicated or nuanced problem.

Teachers, school leaders and education researchers often discuss the role of higher-order thinking skills in learning and development. Some educators believe that students must master lower-level thinking skills, such as memorization, before they can connect those ideas using higher-order thinking skills.

Others think that higher-order thinking can happen at any stage of learning and growth. However, many professional educators agree that higher-order thinking skills are important to consider when developing lesson plans, providing instruction and assessing student growth.

If you want to learn more about higher-order thinking skills, here is a list of seven types to help you get you started:

1. Critical thinking

Critical thinking means using your own best judgment to understand and evaluate other people's ideas. For example, if you are reading an article in a business journal, you might ask yourself who wrote the article, what their credentials are, what other writing they have done in the past and other questions that can help you assess

their ideas. You may want to encourage your students to apply critical-thinking skills when they're reading industry periodicals or online resources to help evaluate what they've just read.

2. Metacognition

Metacognition involves an awareness of how you think. When students engage in metacognition, they closely examine the processes they are using in order to learn and retain new information. This involves understanding their own strengths (such as note-taking) and weaknesses (such as procrastination) as students.

For example, you may have a student who excels at memorizing grammar rules but doesn't always understand how to correctly apply the rules. In this instance, the student may wish to supplement their learning process to include a wide variety of examples so that they understand what they are memorizing.

3. Comprehension

Comprehension refers to the process of internalizing material and understanding the importance of content. Comprehension is a necessary first step for many other higher-order thinking skills because it ensures that you are making connections between ideas you have mastered. For instance, a student in law school needs to understand not only which laws exist for certain situations but how those laws can be applied to new situations.

4. Application

Application as a higher-order thinking skill happens when you apply a piece of information you have attained to a similar issue or project. For example, if a student learns the woodworking techniques necessary to craft a bench, they might also be able to apply those same woodworking techniques to craft a similarly-designed coffee table.

5. Evaluation

Evaluation and critical thinking often overlap because they both have to do with assessing new information based on ideas or concepts you already know. Evaluation allows you to place a relative value on a piece of information, which can help you make decisions based on reasoning and evidence. Students in law school and in the medical field often need to use evaluation to apply the knowledge they are learning in new ways.

6. Synthesis

Synthesis involves combining two or more ideas to generate a new idea that is more meaningful and productive than any of the original ideas were on their own. For example, if your student gathers relevant information about every way to study for a big exam and then develops a new study plan, they are engaging in a kind of synthesis. This skill can help you identify overall themes in a wide array of circumstances. Synthesis can take the form of metaphors, analogies, visualizations and other creative forms.

7. Inference

Inference is a higher-order thinking skill in which you use available information to make a reasonable estimate of information that is unknown. You might use inference to determine the context of an email message from a colleague or anticipate an expected response from a student during finals week. You can use inference skills to understand and anticipate classroom dynamics and reevaluate as more information becomes available.