CHAPTER II

THEORETICAL REVIEW

A. Evaluation

1. The Nature of Evaluation

The teaching and learning process covers the activities carried out by the teacher starting from the planning, implementation of activities to evaluation and follow-up programs that take place in an educational way to achieve certain objectives, namely teaching, while what is meant by the ability to manage teaching and learning processes is the ability of the teachers to create an educational communication environment. Between teachers and students which include cognitive, affective, and psychomotor aspects as an effort to learn something based on planning up to the evaluation and follow-up stages in order to achieve teaching objectives (Suryosubroto, 2002: 19). In broad terms, according to Mehrens and Lehmann (in Purwanto, 2008: 3) evaluation is a process of planning, obtaining, and providing information that is needed to make alternative decisions. In accordance with this understanding, each evaluation activity is a process that is deliberately planned to obtain information or data based on the data and then try to make a decision. Of course the information or data collected must be data that is suitable and supports the planned evaluation objectives.

2. The Function of Evaluation

The evaluation function in education cannot be separated from the purpose of the evaluation itself. The purpose of educational evaluation is to obtain evidence data that will show the extent to which the ability and success of students in achieving curricular goals. In addition, teachers and education supervisors can also be used to measure or assess the effectiveness of teaching experiences, learning activities, and teaching methods used. In more detail, in Purwanto (2008: 5-7) the evaluation function in education and teaching can be grouped into four, namely:

a. To know the progress and development and success of students after experiencing or doing learning activities for a certain time. The results of the evaluation can then be used to improve students' learning methods (formative functions) and or to fill in report cards or graduation certificates which also means to determine grade or graduation or not a student from a particular educational institution (summative function).

b. To find out the level of success of teaching programs. Teaching as a system consists of several components which are interrelated to one another. The intended components include objectives, teaching materials or methods, teaching methods and activities, learning tools and resources, and evaluation procedures and tools.

c. For Guidance and Counseling (BK) purposes. Evaluation results that have been carried out by teachers to their students can be used as a

source of information or data for counseling services by school counselors or other supervisory teachers such as: to make a diagnosis of weaknesses and strengths or abilities of students, to find out in what a person or a group of students requires remedial services, as a basis for handling certain cases among students, and as a reference in serving the needs of students in the context of career guidance.

d. For the purposes of developing and improving the curriculum of the school concerned. A dynamic teacher will not simply follow what is stated in the curriculum; they will always try to determine which material is in accordance with the conditions of students and the environmental situation and the development of society at that time. Curriculum material that is no longer relevant to the development and needs of the community will be abandoned and replaced with material that is deemed appropriate.

3. The Techniques of Evaluation

Broadly speaking, the evaluation techniques used can be classified into two kinds, namely: test techniques and non-test techniques.

a. Non technical testing

There are several non-test techniques, namely:

1) Rating scale
2) Questionnaire
3) Check-list
4) Interview

5) Observation

6) Biography

b. Test technique

Understanding the test according to Bukhori (in Daryanto, 2012: 35) in his book: Evaluation Techniques, say: "The test is an experiment conducted to find out whether or not there are certain learning outcomes on a student or group of students". Meanwhile according to Collegiate (in Daryanto, 2012: 35) said: "The test is a series of questions or exercises or other tools used to measure the skills of knowledge, intelligence, ability or talent possessed by a group or individual". From the description above it can be concluded that the test is a tool information gatherers but when compared to other tools, test it is more official because it is full of restrictions.

4. Learning Outcomes Test

Designing evaluation tools is one step that should not be left behind in planning and learning design. Appropriate evaluation not only can we determine the success of students achieving learning goals, but also at the same time can see the effectiveness of the design program that we have planned (Sanjaya, 2008: 231).

During this time the test is a measuring tool that is often used to measure the success of students achieving competence. In certain cases test results are often used as the only success criteria. Tests measuring success or what we often know Criterion Referenced Test (CRT) are tests

that consist of items that directly measure the behavior that must be achieved by a learning process. This test of success is also known as "Penilaian Acuan Patokan" (PAP). The behavior that must be achieved is described in basic competencies as a translation of competency standards. Because the formulation of behavior in basic competencies is still general, the behavior is described in a number of indicators of learning outcomes. Thus, the test items are arranged after the indicator has been established. It could be that one indicator is measured more than one or two test items. Thus the test items must be parallel with the formulation of learning outcomes indicators.

5. Types of Learning Outcomes Tests

Learning outcomes tests can be divided into several types, including the following:

a. In Arikunto (2008: 33) in terms of usefulness to measure students abilities, the test is divided into three tests, namely:

1) Diagnostic Test

Diagnostic test is a test used to find out the weaknesses of students so that based on these weaknesses can be given appropriate treatment.

2) Formative Test

Formative tests are assessment activities aimed at seeking feedback, which can then be used to improve teaching and learning processes that are currently or have been carried out. Actually

formative assessment is not only done at the end of each lesson, but also when the lesson is ongoing. So the formative test is not only in the form of a written test and is only done at the end of each lesson but can also take the form of oral questions or assignments given during the lesson or after the lesson. In this connection, the usual pretest and post test conducted in the PPSI teaching system is included in formative assessment.

3) Summative test

The function and purpose is to determine whether the value obtained by the student can be declared passed or not passed. So summative assessment is not only an assessment carried out at the end of each quarter or at the end of each semester, but also carried out for example at the end of each module, at the end of the teaching year, evaluation of the final learning phase, and the entrance examination of higher education known as the selection of new student admissions.

b. Based on the method of preparation, the test can also be divided into two, among others:

1) Standardized Test

Standardized test is a test that has been experienced standardization process, namely the process of validation and reliability so that the test is truly valid and reliable for a purpose and for a particular group. A test can be called valid if the test is really

capable of assessing what must be assessed. The test if used can achieve the target in accordance with the objectives that have been planned beforehand. In other words as an evaluation tool, the test is a surefire and careful tool because it has undergone try-outs and improvements so that ultimately it is a standard test. Standardization tests are generally made by psychologists and are widely used in government institutions. At school, at Standardization tests generally have not been widely used (Purwanto, 2008: 33-34).

2) Teacher-made test

Teacher-made tests are prepared to gather information about the level of mastery of the subject matter students they teach, or to see the effectiveness of the learning process that has been implemented. Teacher-made tests usually do not pay much attention to the level of validity and level of reliability because teacher-made tests only cover limited material. (Sanjaya, 2008: 239).

c. Based on how it is implemented, the test can be divided into three, including:

1) Oral Test

Oral test is a form of test that uses language verbally. Through verbal language, assessors can find out in depth student understanding of something being evaluated, which is not just an understanding of concepts. Oral tests are only possible when there

are a small number of students, and assessing something is not too broad but profound.

2) Behavioral Test

Behavioral tests are in the form of demonstrations. This test is suitable when you want to learn someone's abilities and skills about something. For example demonstrating the movement, running a device, and so on (Sanjaya, 2008: 240).

3) Written Test

Written tests can be divided into essay tests or essay examinations and objective tests or short-answer tests. What is meant by an essay test is a test in the form of written questions, the answers of which are essays or long sentences. The length of sentences or tests relative, according to the skills and knowledge of the answerer. Essay tests require a long answer and a long time, usually the number of essay test questions is very limited, generally numbering about five to ten questions (items) only. Whereas an objective test is a test made in such a way that the results of the test can be objectively assessed, judged by anyone to produce the same score. This test is called a short-answer test because the answer is short and concise. The only answer is to choose, fill, match, and so on by using the signs as stated in the problem.

B. Kinds of Assessment in learning

Summative assessment is an assessment conducted to obtain data or information to which the mastery or learning achievement of students of the lessons they have learned over a period of time. So summative assessment is not only an assessment carried out at the end of each quarter or at the end of each semester, but it is also carried out for example at the end of each module, at the end of the teaching year, evaluation of the final learning phase, and the entrance examination of higher education known as the selection of new student admissions (Purwanto, 2008: 26-27). In Peraturan Menteri Pendidikan Nasional Nomor 20 tahun 2007 regarding the standard of education assessment it is stated that the Examination is a process carried out to measure the achievement of students' competencies in a sustainable manner in the learning process, to monitor progress, make improvements to learning, and determine the learning success of students.

1. Final Test

   The final test are activities carried out by educator to measure the achievement of student competencies at the end semester. Measuring here means to judge, which is done by way of testing the achievement of student competencies in the form of learning outcomes.

2. Mid Test

   Mid Test is a test conducted in every middle of one unit of time which includes more than one subject, and is intended to determine the extent to which students have been able to move from one unit to the next

unit (Mulyasa, 2007: 259). Coverage of tests includes all indicators that represent all basic competencies in the semester.

C. HOTS

1. Definition of HOTS

Applying HOTS to students means getting students used to practice Thinking to solve a problem they are facing and being able to apply the knowledge they have. From the development of their lives, they must be able to process information, make generalizations, solve problems even though they are simple, draw data conclusions, explain causality, and link the basic concepts of science with everyday life and community problems. Schools can equip them with thinking skills, namely Higher Order Thinking Skills (HOTS) (Nugroho, 2018: 4)

According to Bookhart (in Nugroho 2018: 16) states that HOTS is the ability of students to be able to connect learning with other elements beyond what the teacher teaches to be associated with real life. This is in line with the opinion of Hamidah (2018: 62) which states that HOTS is a thinking skill that not only requires the ability to remember but requires other higher abilities.

2. Criteria of Hots

Hots Assessment is implemented in knowing student learning outcomes. To inspire teachers to compile Hots questions, the following according to Widana (2017: 3) describes the characteristics of Hots

questions, including (a) Measuring critical thinking skills, (b) contextual problem-based and (c) using diverse types of questions.

a. Critical Thinking

The Australian Council for Eduvationa Research (ACER) states that high-level thinking is a process: analyzing, reflecting, giving opinions, applying learning to different situations, composing, and creating. With the demilian answer about the hots is not explicitly expressed in the stimulus.

Higher order thinking ability is the ability to solve problems, critical thinking skills the ability to give opinions.

Creativity solve problems in HOTS, consisting of:

1) Ability to solve problems that are not familiar
2) The ability to evaluate the strategies used to solve problems from different points of view.
3) Find a different new ways model.

Higher-order thinking skills can be trained in the learning process of the class. Therefore, for students to have high-level thinking skills, the learning process also provides space for students to solve the concept of activity-based knowledge. Learning activities must be able to encourage students to build creativity and critical thinking

b. Contextual Based Problems

HOTS questions are assessment based on real situations in daily life, where students are expected to apply learning concepts in class to

solve problems. Contextual problems faced by the world community today are related to the environment, health, earth and space, social life and the use of science and technology in various aspects of life.

The following Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah (2017: 5) outlines 5 assessment characteristics that are abbreviated to REACT

1) Relating, judgments are directly related to the context of real life experiences
2) Experiencing, assessment that is emphasized on exploration, discovery, and creation.
3) Appliying, assessment that demands students' ability to apply the knowledge gained in the classroom to solve a real problem.
4) Communicating, assessments that demand the ability of students to be able to communicate the conclusions of the model at the conclusion of the context of the problem.
5) Transfering, assessment that demands students' ability to transform the concept of knowledge in the classroom into new contexts.

Other than that, Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah (2017:5) express the characteristics of contextual assessment based on authentic assessment namely,

1) Students construct their own responses, not just choose the answers available.
2) The task is a challenge that is faced in the real world.

3) The assignment given not only has one particular answer that is correct, but allows many correct answers or all correct answers.

The following is a Comparative Traditional Assessment and Contextual Assessment

Tabel 2.1 Comparison of Traditional and Contextual Assessments

Tradisional Assessment	**Contextual Assessment**
Students tend to choose the response given.	Students express responses
World-class context	Real world context
Measuring aspects of memory (recalling)	Measuring task performance (Hots)
Indirect proof, tends to be theoretical.	Proof directly through the application of knowledge and skills with real context.
Separate from learning	Integrated with learning

c. Using Various Types of Questions

The various forms of questions in a test set (HOTS questions) as used in PISA, aim to provide more detailed and comprehensive information about the ability of test takers. (Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah 2017:3)

There are several alternative forms of questions that can be used to write items, as follows.

1) Multiple Choice

In general, HOTS questions use stimuli that originate in real situations. Multiple choice questions consist of the main problem (stem) and choice of answers (options). The choice of answers consists of an answer key and a distractor. The answer key is the

correct answer. Deceiver is an incorrect answer, but it allows someone to be fooled to choose it if they do not master the subject matter properly.

2) Multiple Choice Complex (true/false)

The question of the form of complex multiple choice aims to examine students' understanding of a problem comprehensively related to one statement to another. As a matter of ordinary multiple choice, HOTS questions in the form of complex multiple choice also contain stimulus originating in contextual situations.

3) Essay

The question of essay form is a question which the answer requires students to organize the ideas or things they have learned by expressing or expressing the ideas using their own sentences in written form.

d. Bloom's Taxonomy

Bloom's taxonomy has influenced education both directly and indirectly in curriculum development, learning design and teacher education. Taxonomy in education is used for the classification of instructional purposes; learning goals, performance goals, which are classified in three general classifications or domains, namely: (1) cognitive domains, related to learning goals oriented to thinking abilities; (2) the affective domain relates to feelings, emotions, value

systems, and heart attitudes); and (3) psychomotor domain (oriented to motor skills or use of skeletal muscle).

So that Bloom's taxonomy in the realm has been revised by Anderson and Krathwohl (2001: 66-88) namely: remember, understand, apply, analyze , evaluate, create.

Tabel 2.2 Bloom's Taxonomy of Anderson and Krathwohl

Remembering	Recognizing or recalling knowled ge from memory. Remembering is when memory is used to produce or retrieve definitions, facts, or lists, or to recite previously learned information.
Understanding	Constructing meaning from different types of functions be they written or graphic messages or activities like interpreting, exemplifying, classifying, summarizing, inferring, comparing, or explaining.
Applying	Carrying out or using a procedure through executing, or implementing. Applying relates to or refers to situations where learned material is used through products like models, presentations, interviews or simulations.
Analyzing	Breaking materials or concepts into parts, determining how the parts relate to one another or how they interrelate, or how the parts relate to an overall structure or purpose. Mental actions included in this function are differentiating, organizing, and attributing, as well as being able to distinguish between the components or parts. When one is analyzing, he/she can illustrate this mental function by creating spreadsheets, surveys, charts, or diagrams, or graphic representations.
Evaluating	Making judgments based on criteria and standards through checking and critiquing. Critiques, recommendations, and reports are some of the products that can be created to demonstrate the processes of evaluation. In the newer taxonomy, evaluating comes before creating as it is often a necessary part of the precursory behavior before one creates something.
Creating	Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing. Creating requires users to put parts together in a new way, or synthesize parts into something new and different creating a new form or product. This process is the most difficult mental function in the new taxonomy.

D. Steps for Preparing HOTS Question

In the preparation of problems, HOTS generally uses stimulus. the stimulus used must be interesting and contextual and can raise problems that occur. in line with what was revealed by Hamidah (2018: 82), that in the context of HOTS, the stimulus presented should be contextual and attractive. Stimulus can be sourced by glogbal problems such as information technology, economic science, health, education and infrastructure.

To write an assessment in the form of questions, the question maker is asked to determine the behavior that is intended to be measured by formulating the material that will be used as a basis for questions (stimulus) in context (which requires high reasoning) not always available in textbooks. Therefore, in the preparation of HOTS questions required mastery of teaching materials, skills in writing questions (problem construction), and teacher creativity in choosing stimulus questions in accordance with the situation and conditions of the area around the education unit.

The following are the steps for preparing hots according to Kemendikbud(2019:47-51)

1. Analyzing Basic Competency

   Not all basic competency models can be made about Hots model. Therefore, educators must be able to analyze basic competency questions that can be made about hots through the MGMP forum.

2. Arranging Blueprint Question

   The writing of the grid helps the educator make the test hots being tested. in general, the grid is needed to guide the teacher in a). choose BC that can be made about HOTS, b). determine the scope of material and materials related to BC to be tested, c). formulate the problem indicators, d). determine the number of questions, and e). determine cognitive level.

3. Choose stimulus that are interesting and contextual

   The stimulus used should be interesting and contextual, meaning that it can encourage students to read the stimulus. Interesting and contextual stimuli are usually something completely new and that is present in the school environment by students. So it makes me curious. According to Widana, (2017: 17) an interesting stimulus is a stimulus that has never been read by students. While contextual stimulus means that stimulus is in accordance with reality in daily life, interesting for students reading.

4. Write Questions According to The Blueprint Questions

   The items are written according to the HOTS question writing rules. The rules for writing HOTS items are quite different from the rules for writing items in general. The difference lies in the material aspects, while in the construction and language aspects are relatively the same. Each item is written on a question card.

5. Make Answer Key / Score Rubric

   Every HOTS item written should be accompanied by a scoring guideline or answer key. Scoring guidelines are made for the form of description

items. Whereas the answer key is made for multiple choice questions, complex multiple choices (true / false, yes / no), and short entries.

E. Relevant Research

Relevant research conducted by the author is research conducted by Anggi Lestari with the title "Development of HOTS-Based Test Questions on Learning Training Models in Elementary Schools". Her research explains that most teachers are familiar with HOTS-based questions, but most teachers are more likely to take LOTS questions because they are adjusted according to indicators. The research conducted by Anggi is similar to the research that the author will do. The Anggi Study analyzes multiple choice test items and essays. the difference is only slight, namely the analysis conducted by Anggi including validity, reliability, differences in strength, level of difficulty, and fraud. whereas this research only focuses on the level of questions created and analyzed in terms of HOTS criteria.