INSPECTION AND PROCESSING TECHNIQUES OF ASSESSMENT RESULTS IN LEARNING

Hambali Alman Nasution

UIN Sunan Kalijaga Yogyakarta, Indonesia
hambalialmannasution@gmail.com

ABSTRACT

The purpose of this paper is to determine the technique of examination and processing of assessment results. This research uses a type of *library research,* namely research that uses various literature sources such as books and journal articles to serve as writing guidelines. This article's results indicate that there are steps that must be carried out in the assessment, namely when examining and processing the assessment results. In the stage of checking the assessment results, correcting the students' answers carefully and thoroughly according to the existing scoring guidelines. The raw scores obtained so that they become more meaningful and must be interpreted based on specific and relevant assessment references and needs, namely by benchmarking reference assessments (PAP) and norm-referenced assessment (PAN). For example, to determine whether students' scores have met the passing *grade predetermined*, then use a benchmark reference assessment. Meanwhile, to determine the comparison of students' scores, it uses the norm reference assessment.

Keywords: Inspection Techniques, Processing, Assessment

INTRODUCTION

According to Omar Hamalik quoted by Juhairiyah, assessment is the entire process of measurement (data and information collection), processing, interpretation, and consideration to decide the level of learning outcomes to achieve predetermined learning objectives (Juhairiyah 2017). In general, teachers conduct classroom assessments to collect data, facts, and student learning documents to improve the learning program. Appropriate assessment will reflect or reflect students' learning process (Ridwan Abdullah Sani, Riza Dwi Aningtyas 2016).

Assessment is an essential part of the learning process because it can determine a learning activity's quality. In addition to measuring and assessing the level of competency attainment, assessment is also used to determine strengths and weaknesses in the learning process and diagnose and improve the learning process (Ridwan Abdullah Sani, Riza Dwi Aningtyas 2016). The main thing that must be considered in education is planning, implementation, and evaluation (Hidayat and Asyafah 2019). With the assessment, we can also determine the extent to which the methods' effectiveness is used as the material's success; then the assessment can be done to improve the learning process (Setemen, Erawati, and Purnamawan 2019).

Because a fair and careful assessment will provide a description of the process and output of objective learning outcomes, therefore the assessment system used in educational institutions must be able; 1) provide accurate information, 2) encourage students to learn, 3) motivate teaching staff in teaching, 4) improve institutional performance, and 5) improve educational quality (Kete 2017).

However, to avoid misunderstandings, it is necessary to distinguish between the meaning of score and value. The score results from the work by providing the numbers achieved by adding them to each item that has answered the *testee* correctly, then calculating the correct answer's weight. Whereas the value is a number or letter resulting from a change from the scores collected into one, then the settings are adjusted to a certain standard. Score or raw value will be meaningful if interpreted based on a standard or norm (called an assessment). Processing values into final grades can be done by referring to specific criteria or benchmarks (Alfath and Raharjo 2019). So that in this article, various inspection techniques (scoring) of the assessment results will be presented along with the processing steps based on two approaches, namely benchmark reference assessment (PAP) and norm reference assessment (PAN).

RESEARCH METHODS

The method of analysis used in this research is descriptive qualitative analysis. Qualitative descriptive analysis is a method of analysis that uses words to describe and explain the phenomena or data obtained.

The type of research used is library research *(library),* the research used to solve a theoretical conceptual problem, whether about educational figures or certain educational concepts such as objectives, methods, techniques, and educational environment (Suwardi et al., 2012). In simple terms, library research attempts to collect data from various literature and make it the main object of analysis (2012). In this study, the authors wanted to research and analyze the development of techniques and instruments to assess technology-based aspects of knowledge.

The author collects data utilizing documentation. Documentation or documentary studies *(documentary study)* is a data collection technique that collects and analyzes documents, written documents, images, and electronics (Nana Syaodih Sukmadhinata 2012). The books, journals, and internet sources used in this research aim to obtain data regarding the development of techniques and assessment instruments for aspects of technology-based knowledge.

FINDING AND DISCUSSION

A. Technique for Examination of Assessment Results

The teacher's task after the test is carried out is to correct student answers. Of course, when correcting student answers, the teacher must examine carefully and thoroughly according to the answer keys or scoring guidelines that already exist. At this early stage, the teacher gets a raw score (raw score) (Endrayanto and Harumurti 2014). The following is the scoring of the student evaluation results:

1. Cognitive Domain Assessment

    a. Determining the score on the essay questions

    The scoring on the essay test questions is generally based on the *weight* given to each item. The weights are also given based on difficulty or the number of elements that must be contained in the best or most correct answer. For example, a test of an essay test consists of five items. Each question is given a score of 10. It is based on the fact that the compiler has determined that the five questions have difficulty, and the answer elements for each item are made the same. So that if the *testee* (student) answers completely and correctly to all questions correctly,

they will be given a score of 10, if it is true, they will be given a score of 8, and so on. Then the total score obtained by students is the sum of the scores on each question. For example, Rianti scored seven on question number 1, four on question number 2, three on question number 3, eight on question number 4, and 9 on question number 5. Then the score that Rianti managed to achieve was 7 + 4 + 3 + 8 + 9 = 31.

If the test items in the form of description have different levels of difficulty and the number of elements, the scores' weighting must also be differentiated. For example, of the five questions above, question number 1 has a low or easy level of difficulty, so it is given a score of 3, while questions on number 2 have a higher difficulty level and more elements then the weight of the score is 10, and so on (Sukiman ed. Arifin 2012).

b. Determine the score on objective questions

1) Multiple choice test

MCQs construction is a matter that has subject matter*(stem)*and the alternative answers*(option).* One alternative is the correct answer, while the other alternative is a *distractor*(Asep Ediana Latip, Engkus Kuswandi 2018). Meanwhile, there are two ways to score the multiple-choice form test, namely, without applying an adequate system and by applying a sufficient system to guess answers.

a) Scoring without a penalty system

There are two possibilities for scoring: considering and without considering the weight of the score on each question. The first way is to calculate the correct answer and then multiply the weight of each question's score. This method can be formulated as follows:

$S = \sum R \times Wt$

Information:

S   : Score (the score being sought)

$\sum R$: Right (number of correct answers)

Wt   : Weight (weight of the score for each question)

Example: The

 number of questions on a test is 50 items. There are 30 correct answers. Each item has a weight of 2, then the score obtained is 30 x 2 = 60.

The second way is to calculate the correct answer, and each item that is answered correctly is given a score of 1, then the total score obtained by students is the number of items answered correctly. This method can be formulated as follows:

$S = \sum R$

Example: The

number of questions on a test is 50 items. There are 30 correct answers. Then the score obtained is 30 (Sukiman 2012).

b) Scoring with an adequate system

$S = \sum R - \dfrac{\sum W}{O-1}$

Description:

S : Score (the score being sought)

$\sum R$: Right (number of correct answers)

$\sum W$: Wrong (number of wrong answers)

O : Option (number of options for answers to questions)

1 : Constant (fixed)

Example: The

numbers number of questions in multiple-choice consists of 50 items. The *options* number of answer(choices) is 4, the number of correct answers is 35, the number of wrong answers is 12, and the answer is not 3, then the score obtained is:

$S = 30 - \dfrac{12}{3-1}$

$= 30 - 6$

$= 24$

2) Short answer and match-making test

In general, the two tests the scoring does not consider the penalty in the form of a fine. Namely, the correct answer is given a score of 1, and the wrong answer is given a score of 0. The formula used for scoring is:

$S = \sum R$

Example:

Test in the form of short answers with 30 items. There are 28 correct answers. Then the score obtained is 28 (Sukiman 2012).

3) The test is false(true-false)

The problem is on a true-false test of statements*(statement)*. Some of these statements are true, and some are wrong. How to calculate this test using the following formula:

a) With a fine

$S = R - W$

Description:

S   : score obtained

R   : the correct answer

W   : wrong answer

Example: The number of questions on a test is 30 items. 26 correct answers, four wrong answers, then the score obtained is 26 - 4 = 22

b) Without penalty

With the formula:

$S = R$

S   : Score obtained

R   : Correct score

So only correct answers are counted.

4) Oral test

with the formula:

$S = R$

S: Score obtained

R: Correct score (Alfath and Raharjo, 2019).

2. Affective Domain Assessment Affective

learning outcomes are evaluated using instruments in the form of an assessment scale and observation guidelines, which generally use a Likert scale model with a scale or range of 3, 4, or 5. Interpretation uses verbal categories such as very high, high, moderate, low, and so forth. Example:

EVALUATION SCALE

CLASS DISCUSSION APPEARANCE

Class / Semester: ....

Number of Students   : ....

Subject: ....

No.	Indicator / Student's Name	Asking Actively	Answering Actively	Giving Opinion	Responding Idea	Communication	Score
1.	Ana Rosmalina	5	4	5	5	4	23
2.	Eka Vidiana	5	3	5	3	4	20
3.	Indah Indriana	4	5	4	3	5	21
And so on							

Table 1.1

Guidelines:

ScoringScore: 5 = Very Good   2 = Poor

                4 = Good   1 = Very Poor

                3 = moderate

   Based on the example above, the way to score each student is to write down the score on each indicator or aspect that has been determined based on the teacher's observations by referring to the existing scoring guidelines. Furthermore, students' total score is the sum of all scores from each indicator or aspect being assessed (Sukiman 2012).

Then to give meaning to the score achieved by each student, it is necessary to compile an interpretation guide with the following steps:

a. Calculating the lowest score or *lowest score* that students might achieve by multiplying the lowest score of each indicator being assessed, then multiplying it by the number of indicators assessed. In the example above, the lowest score on each indicator is 1 (very poor), and the number of indicators assessed is five indicators, namely actively asking questions, answering, putting forward ideas, responding to ideas, and establishing communication so that the lowest score is 1 x 5 = 5.

b. Calculating the highest score or *highest score* that students might achieve, namely by multiplying the highest score of each indicator being assessed, then multiplying it by assessing the number of indicators. In the example above, the highest score for each indicator is 5 (very good), and the number of indicators assessed is also 5. So the highest score is 5 x 5 = 25.

c. Calculating the difference between the highest score and the lowest score, namely the highest score minus the lowest score, 25 - 5 = 20.

d. Determines the number of categories that will be used to interpret each student's score. Preferably, the number of categories is comparable to the initial scoring guidelines. As in the example above, there are 5 categories, namely: very good (5), good (4), moderate (3), poor (2), and very poor (1). Then determine the number of categories and 5, like the scoring guidelines above: very good, good, moderate, lacking, and very poor.

e. Determine the range for each category employing the difference between the highest score and the lowest score divided by the number of categories, or it can also be formulated as follows:

$$\text{Range} = \frac{Higher\ Score - \text{Lower Score}}{Number\ of\ Categories}$$

$$= \frac{25-5}{5}$$

$$= 4$$

Then the range in each category is 4. It means that each category contains four scores.

f.   Determining the score for each category, the number of scores for each category is 4 (point e), then to determine the score for each category can be started from the lowest score or the highest score. Here is an example:

Very good   : 21 - 25
Good   : 17 - 20
Enough   : 13 - 16
Poor   : 9 - 12
Very Poor   : 5 - 8

g. The final step is to provide meaning or interpretation of the score. For example, the score obtained by Eka Vidiana is 20; this score can be consulted with the criteria above (point f), is in the range 17-20, it can be concluded that Eka Vidiana's appearance in participating in the class activities is categorized as useful.

3. Psychomotor Domain Assessment

A *rating scale* is an instrument to measure learning outcomes in the psychomotor domain, using a range from very imperfect to very perfect. For example, it is made on a scale of 5; then scale one is at least perfect, and scale 5 is the most perfect.

The scoring technique is by writing the score on each aspect/indicator of the ability as displayed by the evaluated student and then adding up the overall score to obtain each student's total score (Sukiman 2012). The following is an example of a rating scale in the practice of *fardhu* prayer:

EVALUATION SCALE
FARDHU PRAYER PRACTICE

Education Unit: Madrasah Ibtidaiyah
Class / Semester   : III / 2
Subject   : Fiqh

No.	Name	Evaluated Aspect						Total Score
		A	B	C	D	E	F	
1.	Amelia Farha	5	4	4	5	4	3	25
2.	Farhan Yodia	5	3	4	5	3	3	23

| 3. | Ina Pratiwi | 5 | 4 | 4 | 4 | 4 | 3 | 24 |
| And so on. | | | | | | | | |

Table 1.2

Description:

· Aspects assessed:

A = Ability to move and recite *takbiratul ihram*

B = Ability to move and recite *ruku'*

C = Ability to move and recite *i'tidal* recitation

D = Ability to move and recite prostration recitation

E = Ability to move and recite recitation sitting between two prostrations

F = Ability to move and recite *tahiyyat*

· Scoring Guidelines:

Very Good   : 5

Good   : 4

Enough   : 3

Poor   : 2

Very Poor   : 1

Based on the example above, each student's scoring method is to write a score on each aspect that has been set based on the results of teacher observation refers to the existing scoring guidelines. Furthermore, the total student score will be obtained by summing the entire score from each aspect assessed.

B. Assessing Results Processing Techniques

In the first chapter, the scoring technique (scoring) has been explained; then, the next step is to process the student evaluation results. Because raw scores do not yet have meaning, they cannot make decisions about student achievement. For example, Andi's students answered correctly 20 items in the math review from the 30 existing questions. If each question weighs 1 point if correct, then the score obtained is 20. However, is a score of 20 a good score? How does the score of 20 compare to the scores of other students in his class ? Therefore, teachers should interpret the raw score based on individual relevant assessment models to become more meaningful (Endrayanto, Herman Yosep Sunu, and Yustiana Wahyu

Harumurti. 2014). Here are the score processing techniques based on two approaches:

1. Processing Techniques Using Benchmark Reference Assessment (PAP)

    a. The Concept of Benchmark Reference Assessment (PAP)

    Standard reference assessment (PAP) or known as the *Criterion-Referenced Test*, is an assessment that refers to the criteria for achieving the learning objectives formulated previously. The focus, the value obtained, is associated with the level of mastery of teaching materials following the learning objectives set.

    Educational institutions that make assessment criteria or standards based on percentages with a value scale of 0-100, students who get a score of 75 are deemed to have 75% mastery of the knowledge and skills of the subject or material concerned, and so on. From these values, it is then transformed into letter values with specific criteria as well. The values 80 - 100 are transformed into A values, 70 - 79 values are transformed into B values, and so on—furthermore, the stipulated *passing grade*, for example, 65.

    If a student gets a minimum score of 65 or above, the student has passed the *passing grade* predetermined limit. Then these students can be given enrichment to broaden their knowledge further. If it is below the limit *passing grade*, the student has not mastered the material's knowledge and skills, so remedial activities need to be given (Asrul, Rusydi Ananda, and Rosnita, 2014).

    b. Purpose and Benefits of PAP

    The purpose of LAP is to measure with certainty objectives or competencies that have been defined as criteria for success. At the same time, the benefits of improving the quality of student learning outcomes to achieve predetermined standards and the degree of achievement are known. The ideal score is used as a reference or benchmark to determine the *passing grade* that students may achieve (Sukiman ed. Arifin 2012). In the application of LAP, it can be used as follows:

    1) Placing a person in learning activities

    2) To examine (diagnose) a person's ability in learning

    3) It Can be used to monitor each student's progress in the learning process if it is done periodically, whereas if it is done continuously, it can be

known someone's status in a series of learning activities so that it can motivate the enthusiasm of students in learning.

4) The ability of each student to complete the curriculum cumulatively will determine the implementation of the curriculum (Alfath 2019).

c. Score Processing Techniques With PAP

The processing of student learning evaluation results uses a benchmark reference assessment approach. The value to be given is based on absolute standards or predetermined criteria by comparing the raw score of students' evaluation results with the ideal maximum score students may achieve. The formula used is:

$$\text{Range} = \frac{\text{Real Score}}{\text{Ideal Maximum Score}} \text{ X } 100$$

Information:

Real score : The score achieved by each student.

Maximum Ideal Score: The score that each student could achieve if he were able to answer all the questions correctly

100 : The scale used is a scale with a range ranging from 0 - 100.

Example 1: Multiple choice model test The number of questions is 50 items on the subject of Islamic Civilization History. The number of correct answers is 30, so the score achieved is 30. How many points did you get?

Answer: Score $= \frac{30}{50}$ x 100

= 60

Example 2: Psychomotor test (Table 1.2)

In the *fardhu* prayer practice test, a student named Ina Pratiwi got 24; how much did the student score?

Answer:

Value $= \frac{24}{30}$ x 100

= 80.

2. Processing Techniques with Norms Reference Assessment Approach (PAN)
    a. The Concept of Reference Assessment (PAN)

    Normative reference assessment or norms is an assessment with a classical approach because it compares student learning outcomes in groups (classes) by describing the relative position to other students (Endrayanto, Herman Yosep Sunu, and Yustiana Wahyu Harumurti. 2014). The measurement method using competitive learning principles is used in this measurement. The purpose of using this test approach is usually more general and covers a large area of content and study assignments. The norm reference test is intended to determine the relationship between one test taker's status and other test takers (Rapono, Safrial, and Wijaya, tt), which is to distinguish the test taker's ability level from the lowest to the highest. Ideally, the distribution of ability levels in a group describes a standard curve (Alfath 2019).

    Norm reference assessment or also known as group reference assessment, is based on relative standards; this is motivated in determining the value of the evaluation results of the raw score of the evaluation results achieved by students compared to the raw scores of evaluation results achieved by other students so that the quality of the group will significantly determine the quality of a student. Student graduation also depends on the achievements of the group.

    The use of relative standards will show that students in group 1 are classified as "high achievers or smart" if included in group 2, it turns out that they are only in the "moderate," "sufficient," or "ordinary" group, so that the position of the student in question is relative (Sukiman 2012).
    b. Criteria for Formulation of Norms Reference Assessment (PAN)
        1) PAN questions are not based on students' teaching or skills or behavior identified as relevant to student learning (*behavioral objectives*).
        2) In the learning process, the assessment is generally carried out by the teacher.
        3) Emphasis in assessment refers to the provisions of norms that apply in schools.
        4) Teachers can use the national normative value reference.
    c. Characteristics of Norms Reference Assessment (PAN)

1) Used to determine students' ability in the community, such as in class, school, etc.

2) Using relative criteria.

3) The PAN value is used to indicate the position (rank) of students in the community or group.

4) Tends to use the range of one's level of mastery over his group, from the very special to those experiencing severe difficulties (Alfath 2019).

d. Score Processing Techniques with PAN

The raw scores from the learning evaluation result in values based on the group's average value and the standard deviation or standard deviation. After knowing the average value or mean and standard deviation, the raw scores are converted into traditional values. A wide range of standard values can be used to do the conversion, but which is used in the base rate is eleven common value*s (standard level).* Following are brief steps for processing scores using the norm reference assessment approach (Sukiman ed. Arifin 2012). The results of the evaluation on Indonesian subjects were followed by ten students with the following scores: 8 4 5 9 9 7 5 6 8 7

1) Calculating the average (*mean*) value of the raw score

$$\bar{x} = \frac{\sum xi}{N}$$
$$= \frac{68}{10}$$
$$= 6,8$$

Information::

$\bar{x}$: The average value sought (*mean*)

$\sum xi$: The sum of all scores achieved by the group

N : Number of students evaluated

2) Calculating the standard deviation (standard deviation)

Before looking for standard deviation, first look for the score deviation. Namely, by subtracting each student's score (x_i) from the *mean* ($\bar{x}$) obtained by 6.8, then each deviation's results are squared. Like the following example:

No.	Score (x_i)	Deviation $(x_i - \bar{x})$	Squared Deviation $(x_i - \bar{x})^2$
1.	8	1,2	1,44
2.	4	- 2,8	7,84
3.	5	-1,8	3,24
4.	9	2,2	4,84
5.	9	2,2	4,84
6.	7	0,2	0,04
7.	5	-1,8	3,24
8.	6	-0,8	0,64
9.	8	1,2	1,44
10.	7	0,2	0,04
	68		27,6

Table 1.3

The formula for finding the standard deviation (Sugiyono, Endang Mulyatiningsih 2007):

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

$$= \sqrt{\frac{27,6}{10}}$$

$$= \sqrt{2,76} = 1,67$$

3) Processing the score into numbers using eleven standard values

Description:

M   : average (*mean*) = 6.8

SD   : Standard deviation = 1.67

Score		Score Range	Score Number
M + 2,25 SD	6,8 + 2,25 x 1,67 = 10,55	10,55 ke atas	10
M + 2,25 SD	6,8 + 1,75 x 1,67 = 9,72	9,72 − 10,54	9
M + 2,25 SD	6,8 + 1,25 x 1,67 = 8,88	8,88 − 9,71	8
M + 2,25 SD	6,8 + 0,75 x 1,67 = 8,05	8,05 − 8,87	7
M + 2,25 SD	6,8 + 0.25 x 1,67 = 7,21	7,21 − 8,04	6
M + 2,25 SD	6,8 − 0,25 x 1,67 = 6,38	6,38 − 7,20	5

M + 2,25 SD	6,8 − 0,75 x 1,67 = 5,54	5,54 − 6,37	4
M + 2,25 SD	6,8 − 1,25 x 1,67 = 4,71	4,71 − 5,53	3
M + 2,25 SD	6,8 − 1,75 x 1,67 = 3,87	3,87 − 4,70	2
M + 2,25 SD	6,8 − 2,25 x 1,67 = 3,04	3,04 − 3,86	1

Table 1.4

4) Converting the raw score into numerical values / standards using the standards above. For example, a student named Hilma gets a score of 9, then the number value obtained is 8.1

3. Differences and advantages and disadvantages of PAP and PAN

a. Difference between PAP and PAN

PAP	PAN
Assessment by comparing student learning outcomes with predetermined benchmarks/criteria.	Assessment by comparing student learning outcomes with the learning outcomes of other students in the group/class.
Student success is determined based on benchmarks/criteria.	Their position in the group determines student success.
Criteria are fixed.	Passing limits are variable (not fixed)
Passing *grades are* linked to mastery/competency levels.	Student success is not linked to the quality of mastery/competence.
Assessment does not use normal curves or statistical calculations.	Assessment is based on a normal curve using mean and standard deviation (Ali and Khaeruddin 2012).

Table 1.5

b. Strengths and Weaknesses of PAP and PAN

Benchmark reference assessment and norm reference assessment have advantages and disadvantages. The following is a brief description of the advantages and disadvantages of the two assessment references:

Assessment Matrix	Strengths	Weaknesses
Benchmark Assessment (PAP)	• Assist the teacher in deciding whether students are ready to continue with the next level/stage of learning. • The learning achievement of each student does not depend on group work (class). • It is implemented on the principle of complete learning (*mastery learning*). • Each student can be evaluated based on different learning materials/materials depending on their achievement or learning outcomes. • Allows teachers to check learning progress and learning difficulties. • It can be used to compare achievement/learning outcomes of all different subjects in the same student.	• Difficult to develop meaningful criteria because of the application of the method *cut-off* • Reliability tests made based on benchmarks/criteria are difficult to do. • Teachers find it difficult to make comparisons between students.

Norms Reference Assessment (PAN)	• It can be used to make comparisons between students. • The results of one class can be compared with other classes. • It can be used to make decisions or determine whether students can continue the learning program or the next level of education.	• If students get good results in general, some students who get bad grades still do not experience progress in learning achievement. • If all students in a class get low learning achievement, giving quality marks is difficult to account for. • Often it does not motivate students to improve learning achievement. • Each student in the class should be evaluated using the same measurement and assessment instruments under the same conditions.

Table 1.6

School administrators and teachers can indeed apply the two assessment references to make decisions about student achievement, but the two usually have differences in the context of decision making. Norm reference assessment can be applied when a certain degree of selectivity or relative comparison between students is required to determine student rankings in class.

(Kete 2017a; Endrayanto, Herman Yosep Sunu, and Yustiana Wahyu Harumurti. 2014) Meanwhile, the benchmark reference assessment can be applied if the teacher is more interested in knowing whether students master the learning material/material. Based on specific standards and criteria, the teacher assesses student learning success and evaluates the teaching methods given. Teachers can make decisions on student learning outcomes to make *judgments* against students such as complete and incomplete (pass and not pass) (Endrayanto, Herman Yosep Sunu, and Yustiana Wahyu Harumurti. 2014).

CONCLUSION

Assessment in learning aims to determine the quality of learning activities. Its function is to measure and assess competency attainment, identify strengths and weaknesses, and diagnose and improve the learning process. In the assessment, some steps must be done. Namely, the stages that exist when examining and processing the assessment results. At the stage of checking the assessment results, it is correcting the results of the students' answers carefully and thoroughly according to the existing scoring guidelines (giving scores). The intended score is the raw value that must be processed to have meaning and make decisions about student achievement. The teacher must interpret the raw score based on specific relevant assessment references to become more meaningful. Two approaches can be used to process raw scores, namely the benchmark reference assessment (PAP) and the norm reference assessment (PAN). The assessment approach is used as needed. For example, to determine whether students' scores are good or have met the *passing grade* predetermined, then use a benchmark reference assessment. Meanwhile, to determine the comparison of students' scores with each other, it uses the norm reference assessment.

REFERENCES

Alfath, Khairuddin. 2019. "TEKNIK PENGOLAHAN HASIL ASESMEN: TEKNIK PENGOLAHAN DENGAN MENGGUNAKAN PENDEKATAN ACUAN NORMA (PAN) DAN PENDEKATAN ACUAN PATOKAN (PAP)." *Journal Al-Manar* 8 (1): 1–28.

Asep Ediana Latip, Engkus Kuswandi. 2018. *Evaluasi Pembelajaran Di SD Dan MI,*. Bandung: PT. Remaja Rosdakarya.

Asrul, Rusydi Ananda, dan Rosnita,. 2014. *Evaluasi Pembelajaran*. Bandung: Citapusttaka Media.

Endrayanto, Herman Yosep Sunu, and Yustiana Wahyu Harumurti. 2014. *Penilaian Belajar Siswa Di Sekolah,*. Yogyakarta: Kanisius.

Endrayanto, Herman Yosep Sunu, dan Yustiana Wahyu Harumurti. 2014. *Penilaian Belajar Siswa di Sekolah*. Disunting oleh C. Erni Setiyowati dan Yosef Bayu Anangga. Yogyakarta: Kanisus.

Hidayat, Tatang, dan Abas Asyafah. 2019. "KONSEP DASAR EVALUASI DAN IMPLIKASINYA DALAM EVALUASI PEMBELAJARAN PENDIDIKAN AGAMA ISLAM DI SEKOLAH." *Al-Tadzkiyyah: Jurnal Pendidikan Islam* 10 (1): 159–81. https://doi.org/10.24042/atjpi.v10i1.3729.

Juhairiyah. 2017. "Assesmen Konten Isi Bidang Studi." *Jurnal Pedagogik* Vol 4, No: 63.

Kete, Sukadir. 2017. "Implementasi Evaluasi Program Model Formatif Mata Pelajaran Bahasa Indonesia Pada SMP Negeri 4 Kendari Kota Kendari." *Al-Ta'dib* 10 (1): 109–126.

Nana Syaodih Sukmadhinata. 2012. *Metode Penelitian Pendidikan*. Bandung: PT. Remaja Rosdakarya.

Rapono, Muhammad, Safrial Safrial, dan Candra Wijaya. t.t. "Urgensi Penyusunan Tes Hasil Belajar: Upaya Menemukan Formulasi Tes Yang Baik dan Benar." *JUPIIS: JURNAL PENDIDIKAN ILMU-ILMU SOSIAL* 11 (1): 95–104.

Ridwan Abdullah Sani, Riza Dwi Aningtyas. 2016. *Penilaian Autentik*. Jakarta: PT. Bumi Aksara.

Setemen, Komang, Luh Joni Erawati, dan I. Ketut Purnamawan. 2019. "MODEL PEER ASSESSMENT BERBASIS TEKNOLOGI INFORMASI DAN KOMUNIKASI." *Jurnal Pendidikan Teknologi Dan Kejuruan* 16 (1): 55–64. https://doi.org/10.23887/jptk-undiksha.v16i1.16619.

Sugiyono, Endang Mulyatiningsih. 2007. *Statistika Untuk Penelitian,*. Bandung: CV. Alfabeta.

Sukiman. 2012. *Pengembangan Sistem Evaluasi*. Disunting oleh Arifin. Yogyakarta: Insan Madani.

Sukiman ed. Arifin. 2012. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.

Suwardi dkk. 2012. *Panduan Penulisan Skripsi*. Yogyakarta: FITK UIN Sunan Kalijaga.