



## Evaluation of the implementation of educational assessment standards at *Madrasah Tsanawiyah* Modern Islamic Boarding School

Nurul Ngarifillaili<sup>1\*</sup>; Badrun Kartowagiran<sup>1</sup>; Umwari Yvette<sup>2</sup>

<sup>1</sup>Universitas Negeri Yogyakarta, Indonesia

<sup>2</sup>University of Kibungo, Rwanda

\*Corresponding Author. E-mail: [aqbilinaa@gmail.com](mailto:aqbilinaa@gmail.com)

### ARTICLE INFO

#### Article History

##### Submitted:

4 September 2021

##### Revised:

17 November 2021

##### Accepted:

19 November 2021

#### Keywords

educational assessment standards; MTs within the modern Islamic boarding school; evaluation program

#### Scan Me:



### ABSTRACT

This study aims to collect, analyze, and present evaluation results and assess them by comparing the evaluation indicators. The evaluation focuses on achieving and implementing seven components of the Educational Assessment Standard in Islamic middle school (*Madrasah Tsanawiyah* or MTs) of modern Islamic boarding schools (*pesantren*) in the Kebumen Regency. The study employed a descriptive quantitative approach. The evaluation model was the discrepancy evaluation model. The study subjects were 227 students of grade VIII class 2020/2021. Data collection was performed using a questionnaire and brief interviews. The study result shows that 70.55% of students stated the assessment comprises three aspects, i.e., knowledge, attitude, and skills. The assessment principles of valid, objective, fair, integrated, open, thorough and sustainable, systematic, based on criteria, and accountable have been implemented well. About 67.82% of students assert that all principles have been reflected during assessment by educators and education units. As much as 72.35% of students explained that the assessment had utilized an appropriate form to measure students' competency achievement. The assessment of MTs in modern *pesantren* in Kebumen Regency used an instrument following the regulation, where 74.82% of students revealed that the study instrument had followed the empirical validity requirement.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



#### How to cite:

Ngarifillaili, N., Kartowagiran, B., & Yvette, U. (2021). Evaluation of the implementation of educational assessment standards at Madrasah Tsanawiyah Modern Islamic Boarding School. *REID (Research and Evaluation in Education)*, 7(2), 106-117. doi:<https://doi.org/10.21831/reid.v7i2.43672>

## INTRODUCTION

Education in Indonesia is dominated by formal education from the government and non-formal education with attached existence in the population, i.e., *Pesantrens*. The *Pesantren* term came from the word santri (students) with a *pe-* prefix and *-an* suffix, acting as the student residence in studying religion (Takdir, 2018, p. 156). *Pesantren* culture has been applied using its particular method. Based on this explanation, it is clear that a *pesantren* is a place to study, particularly in Islam, offering many advantages both in the method and studying process.

Rapid *pesantren* development should be responded to wisely. Currently, many *pesantren* develop following globalization. At first, *pesantren* are merely defined as a place to study religion; however, today's definition has been expanded.

Currently, many *pesantren* teach not only religion but also general science. In the past, most *pesantren* were of the *salaf* or religious type, now they have changed to the *Khalaf* (modern) type. The *pesantren* began to develop slowly by establishing public schools so that the learning process was a combination of religious learning within the *pesantren* and learning in schools.

In Indonesia, *pesantren* with primary schools amounted to 2072, with middle schools for 2721, open middle schools for 224, high schools for 1580, vocational schools for 35, and religious high schools for 176. *Pesantrens* continue to accept all advances, including building formal education institutions (Istikomah, 2017, p. 57). *Pesantren* that combines two curricula have more subjects than traditional ones. Students should simultaneously learn two sciences, i.e., general science and religious teachings. General science such as Indonesian, Mathematic, and English lessons are taught in more lesson hours than others. Therefore, the studying process in several subjects is expected to be better than in others. In this study, modern *pesantren* are defined following the integrated curriculum utilization and applying the 2013 Curriculum in its schools.

In Kebumen Regency, ten of *Madrasah Tsanawiyah* (MTs) apply the 2013 Curriculum. However, field observation revealed that mathematics teachers complained about the new curriculum utilization in assessment. Teachers consider the assessment complicated; hence, troublesome for processing scores. Also, teachers acknowledge that the process of attitude assessment is affected by subjectivity (Retnawati, 2015). Based on this finding, although mathematics has more lesson hours than others, its assessment process demonstrates various problems.

The 2013 Curriculum implements a Higher Order Thinking Skills (HOTS) based assessment. However, the fact demonstrates that teacher assessment has not shown the HOTS level and still tends to be Lower Order Thinking Skills (LOTS). This is seen from their knowledge concerning HOTS implementation, HOTS characteristics, and steps to arrange HOTS questions. Teacher perceptions are classified to agree and disagree to HOTS assessment. This study result revealed that the implementation of assessment had not followed the government's expectations. Teachers have limitations in implementing a HOTS-based 2013 Curriculum assessment.

Evaluation of the government's policy relating to education assessment is crucial, given that many problems in the assessment process still exist. The possible measures are identifying the implementation of the Educational Assessment Standards on the field particular to educators and education units, identifying discrepancies on the field against the policy, and finding the discrepancies. In this case, the Discrepancy Evaluation Model has matching characteristics to the problem and is applicable. It is expected to be the standardized assessment process, improving the school quality. Based on this explanation, it is necessary to conduct an evaluation study of the implementation of educational assessment standards at *madrasah tsanawiyah* modern Islamic boarding schools (*Pesantren*), particularly to Indonesian, Mathematics, and English subjects.

## METHOD

The study approach was descriptive-quantitative. The evaluation model employed was the Discrepancy Evaluation model, since the evaluation study was defined as a compatibility process of programs against program standards and whether any discrepancy occurs between program aspects on the field against the predetermined standard. Thus, the evaluation was specific by comparing actual things on the field and expected things from the standard. The steps in this Discrepancy Evaluation are (Wirawan, 2016, p. 140): (a) developing a design and standards specifying characteristics of the ideal implementation of evaluation objects, (b) determining information required to compare the actual implementation and standard defining the evaluation object performance, (c) capturing evaluation object performance, including program implementation and quantitative and qualitative outcomes, (d) identifying discrepancies between standards and the actual implementation of evaluation objects, (e) determining the discrepancy cause, and (f) eliminating discrepancies by making changes on the evaluation object implementation.

Table 1. MTs Data in Kebumen Regency

Madrasah Tsanawiyah	Amount
Public School	8
Private School (Outside Islamic Boarding School)	76
Private School (Including the Modern Islamic Boarding School)	10

This study was conducted on three MTs in modern *pesantren* in Kebumen Regency selected based on the large student amount with the purposive sampling technique. The large student amount category on the study object was over 100 students. Another study object selection category was middle school location distributed in three different districts. Table 1 presents the data of MTs in Kebumen Regency.

The criterion for determining the object of the study was MTs in modern *pesantren* applying an integrated curriculum, i.e., *pesantren* and governmental curricula. The schools included in this category combine school and *pesantren* in the same complex where all students live within the area. These schools are under *pesantren* institutions with the *pesantren* curriculum and school curriculum under the government, i.e., Ministry of Religion. MTs fulfilling the criteria amounted to ten; those with over 100 students are MTs Plus Nururrohmah, MTs YAPIKA, and MTs Salafiyah Wonoyoso.

The population of grade VIII students on the three MTs was 587. The samples as study respondents were selected using the random sampling technique, obtaining 227 students. The sample selection used the Cohen and Morrison table with a 95% confidence level and a 0.05 confidence interval (Cohen et al., 2018, p. 206). Samples were collected using a questionnaire and brief interviews with the schools. Instrument validation should involve content analysis and empirical analysis of the test scores and response data to items by test takers. The content analysis of the test is related to the validity of the content, which furthermore requires empirical analysis to determine the validity of the construct. Both of these analyzes are indispensable in the world of education so that the instrument meets the standard requirements (Retnawati, 2016, p. 18). The study employed a content validity test with expert judgment and a construct validity test with EFA (Exploratory Factor Analysis) for the student's questionnaire. The content validity index determination utilized the Aiken formula, as shown below in Formula (1), where  $V$  = Aiken validity index,  $S$  =  $r$ -Io (score given by rater - lowest validation score),  $n$  = the number of panelists, and  $c$  = the number of categories.

$$V = \Sigma s/n(c-1) \dots\dots\dots (1)$$

The construct was measured by a trial on 44 students. The analysis utilized Exploratory Factor Analysis (EFA). Validity of a question item is assigned using several criteria (Retnawati, 2016, p. 43), including (a) A Kayser Mayer Oikin (KMO) score over 0.5, (b) the significance value of the Barlett's Test of Sphericity analysis under 0.05, (c) the anti-image correlation over 0.5, (d) the Eigenvalue price in Total Variances Explained over 1.0, (d) the Rotated Component Matrix Coefficient is over 0.4, and the loading value of such a factor is bigger than other factors with a minimum difference of 0.1 to discover the item group. In addition, the grid of student questionnaire instruments is presented in Table 2, and the result of the analysis carried out with EFA is presented in Table 3.

Table 2. Grid of Student Questionnaire Instruments

No.	Aspect	Indicator	Question Item
1.	Scope of Assessment	Teachers evaluate students' knowledge Teachers evaluate students' attitude Teachers evaluate students' skills	1-4 5-7 8-10
2.	Principle of Assessment	The assessment meets most principles (Valid, Objective, Fair, Integrated, Open, Thorough, and Sustainable) The assessment does not exclude three other principles (Systematic, Based on criteria, and Accountable)	11-15
3.	Form of Assessment	Teachers conduct daily tests, observation, assignments	16-19
4.	Instrument of Assessment	Teachers use an instrument of assessment including the knowledge, attitude, and skill aspects	20-22

Table 3. KMO Value and Bartlett's Test

<b>Kaiser-Meyer-Olkin Measure of Sampling Adequacy Indicator</b>	<b>.567</b>
Bartlett's Test of Sphericity	613.395
	231
	.000

Table 4. MSA Value

Items	Value	Items	Value
1	0.731	12	0.574
2	0.489	13	0.569
3	0.816	14	0.513
4	0.384	15	0.347
5	0.323	16	0.501
6	0.593	17	0.589
7	0.539	18	0.586
8	0.415	19	0.533
9	0.651	20	0.514
10	0.242	21	0.565
11	0.684	22	0.607

After calculating KMO and Bartlett's Test, the Keiser Meyer Measure of Sampling value was 0.567. Therefore, KMO met the requirement by having a  $> 0.5$  value. It indicates that the samples were sufficient. The subsequent analysis was searching for the MSA value, as presented in Table 4. The analysis results show six reduced (excluded from the study) items for not following the MSA  $> 0.5$  requirements. Therefore, an analysis was performed by excluding the reduced items, leaving the questionnaire with only 16 questions.

The instrument reliability calculation used the SPSS program based on the Cronbach Alpha coefficient. The reliability value is declared good if close to 1 or with a coefficient  $> 0.7$  (Hair et al., 2010, p. 21). Based on the measurement results, the Aiken validity index obtained was 0.97, categorized as highly valid. The construct validity analysis result, referring to previous requirements, demonstrates that the student's questionnaire with 16 question items was declared valid. The reliability reached 0.889, indicating that the student assessment questionnaire is reliable as a study instrument.

## FINDINGS AND DISCUSSION

Following the Regulation of the Minister of Education and Culture No. 23 of 2016 on the Educational Assessment Standards, educational assessment is a process to collect and process information to measure students' competency achievements, including authentic assessment, self-assessment, portfolio-based assessment, tests, daily tests, mid-semester tests, end-semester tests, competency level tests, national tests, and Islamic middle school tests. The assessment of competency achievement involves attitude, knowledge, and skill performed equally to determine each student's relative position against the predetermined standard (Hidayah, 2020, p. 101).

Educational Assessment Standards is a criterion on mechanisms, procedures, and instruments of students assessment. The assessment standard by educators, according to *Badan Standar Nasional Pendidikan* (BSNP) or the Agency of National Education Standard, includes the general standard, planning standard, implementation standard, and report of assessment outcomes and assessment finding utilization standard. Meanwhile, an education rapport by education units has two fundamental standards, i.e., standard for determining grades and standard for graduation determination (Salamah, 2018, p. 287). Based on several explanations, the Educational Assessment Standards applies a criterion including several education components' particular and general standards. Good learning should have a good assessment process by comparing to the predetermined standards or criteria.

Evaluation Research of Assessment Standards Implementation on Indonesian, English, and Mathematics was performed in Kebumen Regency. Islamic middle schools in the modern *pesantren* area were the primary target of this study. In addition, three MTs were the subjects: MTs Plus Nururrohmah, MTs YAPIKA, and MTs Salafiyah Wonoyoso.

The study scope components include: (a) student learning outcome assessment on primary and middle schools including attitude, knowledge, and skills, (b) attitude assessment, an activity carried out by educators to acquire descriptive information regarding students' behaviors, (c) knowledge assessment, an activity to measure students' knowledge mastery, (d) skill assessment, an activity to measure students' ability in implementing knowledge to carry particular assignments. Furthermore, quantitative data processing for the Likert scale can be interpreted into score ranges using the normal distribution criteria (Mardapi, 2017, p. 10). The categories of the score ranges are presented in Table 5. Meanwhile, the assessment scope component results on three MTs in *pesantren* in Kebumen Regency are presented in percentage (%). The achievement is illustrated in Table 6.

Table 6 shows that 70.55% of students admitted that assessment implementation in the three MTs had covered three aspects, i.e., knowledge, attitudes, and skills. It proves that the components of the scope of the assessment have been implemented well. The difference in the achievement of the assessment scope component at the three MTs can be seen in Figure 1.

Table 5. Assessment Score Range Category

No.	Value Interval	Category
1.	$X \geq \bar{X} + 1.5 SBx$	Very good
2.	$\bar{X} \leq X < \bar{X} + 1.5 SBx$	Good
3.	$\bar{X} - 1.5 SBx \leq X < \bar{X}$	Poor
4.	$X < \bar{X} - 1.5 SBx$	Very poor

Table 6. Component Achievement Result of Assessment Scope (%)

Component Achievement Result of Assessment Scope (%)				
No.	Subjects	Madrasah		
		MTs Nururrohmah	MTs YAPIKA	MTs Salafiyah
1.	Indonesian	64.7	71.42	72.15
2.	English	63.55	82.54	69.62
3.	Mathematics	63.53	73.02	74.42
<b>Overall Average</b>		<b>70.55 (Good)</b>		

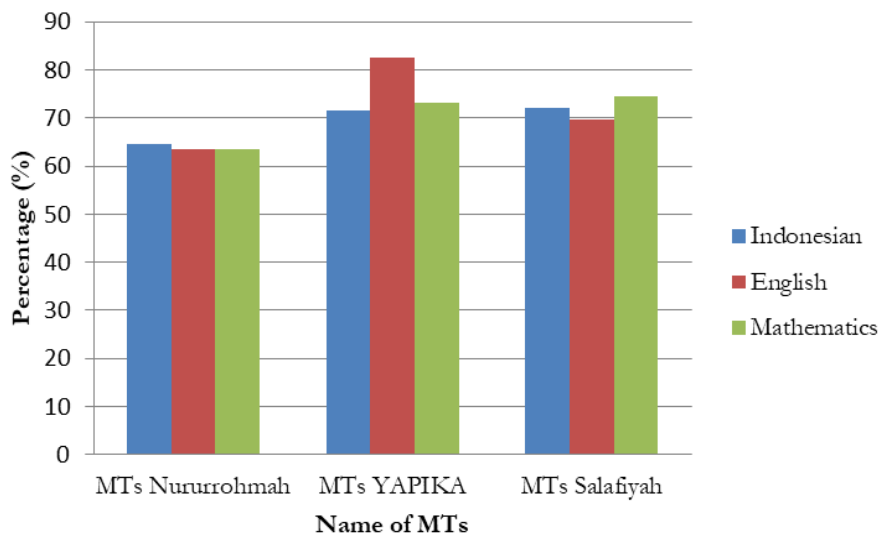


Figure 1. Achievement Diagram of the Assessment Scope



Based on Figure 1, it is discovered that the percentage of students acknowledging that the assessment had been good and covering three aspects at MTs YAPIKA and MTs Salafiyah Wonoyoso is more significant than MTs Nururrohmah. The assessment scope is one of the critical standards in determining the quality of a school. The assessment focuses on three aspects: knowledge, attitudes, and skills. Student learning outcomes, i.e., the achievement during the learning activities, become one of the critical benchmarks in the assessment scope (Jihad & Haris, 2013, p. 14). The assessment of student learning outcomes in three MTs in modern *pesantren* in Kebumen Regency conducted by educators and education units was excellent. It is indicated by 70.55% of student respondents stated that knowledge, attitude, and skill aspects had been reflected in the assessment activities in MTs.

According to Khuriyah et al. (2016), from a managerial perspective, the basis of tradition in managing an institution, including *pesantren*, causes management products not having a focused strategic focus. Personal dominance is too large and tends to be exclusive in its development. It shows that *pesantren* require improvement in their management. This condition is a different obstacle in implementing student assessment in MTs in *pesantren*.

Furthermore, the skill aspect has its obstacles where students could not be assessed objectively. Specific skills that students should achieve have not been implemented due to several obstacles such as limited space for movement and facilities. For example, there was no laboratory for the English subject; thus, it is challenging for students to show their skills to the maximum. Improvement efforts need to be made in order to maximize the assessment, especially concerning improving the implementation of educational assessments. This percentage can be increased to reach the optimum result.

The non-optimal part of the scope of this assessment is caused by several obstacles related to the existence of madrasas that are one with the dormitory. From the observations, the knowledge and attitude aspects have been implemented well. However, students' skills are limited by the rules of the *pesantren* regarding the procurement of tools and materials. Generally, MTs in *pesantren* are different from public schools, which are free to enter and leave. MTs are bound by the rules of the *pesantren* and school rules.

The following evaluation component is the assessment principle. Based on the Regulation of the Minister of Education and Culture No. 23 of 2016, the nine components of the assessment principles evaluated are: (a) valid, meaning that the assessment is based on data that reflects the measured ability; (b) objective, meaning that the assessment is based on clear procedures and criteria, not influenced by the subjectivity of the rater; (c) fair, meaning that the assessment is not beneficial or detrimental to students because of special needs and differences in religious, ethnic, cultural, customs, socioeconomic status, and gender backgrounds; (d) integrated, meaning that assessment is an inseparable component of learning activities; (e) open, meaning that interested parties can know the assessment procedure, assessment criteria, and basis for decision making; (f) comprehensive and continuous, meaning that the assessment covers all aspects of competence by using various appropriate assessment techniques to monitor and assess the development of students' abilities; (g) systematic, meaning that the assessment is carried out in a planned and gradual manner by following standard steps; (h) based on criteria, meaning that the assessment is following the achievement of the specified competence; and (i) accountable, meaning that the assessment can be accounted for both mechanisms, procedures, techniques, and results.

Table 7. Component Achievement Result of Assessment Principle

Component Achievement Result of Assessment Principle (%)				
No.	Subjects	Madrasah		
		MTs Nururrohmah	MTs YAPIKA	MTs Salafiyah
1.	Indonesian	58.82	63.49	69.62
2.	English	65.88	69.84	73.42
3.	Mathematics	67.05	65.08	77.21
<b>Overall Average</b>		<b>67.82 (Good)</b>		

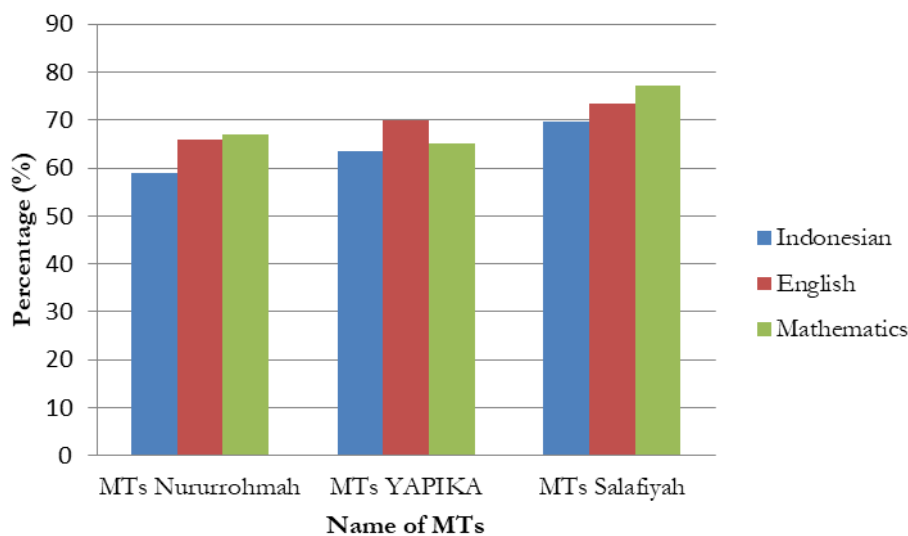


Figure 2. Achievement Diagram of the Assessment Principle

In general, the results of the achievement of the principal assessment components shown in Table 7. It shows that the principal components of the assessment carried out at MTs in modern *pesantren* in Kebumen Regency were in the good category. It is indicated by 67.82% of students asserting that the nine principles had been implemented in the assessment process in MTs.

A bar chart showing the implementation of assessment principles in Indonesian, English, and Mathematics subjects in three MTs in the Kebumen area is shown in Figure 2. The bar chart in Figure 2 shows that the achievement of the assessment principle was in a good category. At MTs Salafiyah, the percentage of respondents who stated that the assessment principle had been implemented was higher than the other two MTs.

The achievement of the principal assessment components out in three MTs in Kebumen Regency was in a good category, with 67.82% of respondents admitting that the nine principles have been reflected in the assessment implementation. However, the results of these achievements remained poor compared to the components of the assessment scope. It is evident that in implementing the assessment principles, students still had a significant level of deficiency.

The lack of high achievement results for the components of the assessment principles is due to several principles that were not appropriately implemented. Although the nine assessment principles are easy to theorize, they are complicated to implement.

The principle of objective and valid still needs to be improved in its implementation. The objectivity of an educator is required to carry out evenly distributed assessment to all students. However, in practice, the assessment was still subjective to specific students. The assessment of the knowledge aspect was related to the attitude and other aspects. According to [Mardapi \(2017, p. 5\)](#), learning outcomes in the three aspects are not summed or affected by each other because they measure different dimensions. [Ayu and Marzuki \(2017, p. 78\)](#) also wrote about the importance of internal competencies possessed by educators, especially the character of educators. However, in the assessment implementation in schools, the subjectivity of educators, e.g., combining both assessment aspects, still occurs, and efforts need to be made to eliminate this.

Furthermore, the validity principle should be improved where the assessment must reflect the ability being measured. In practice, students' ability has not been appropriately measured since the subjectivity of educators influences it. These two principles must be improved to perform educational assessment correctly to meet applicable standards in improving the quality of a school. The assessment form components included assessing learning outcomes by educators in the form of tests, observations, assignments, and other necessary forms. In addition to educators, assessment was carried out on education units in school/MTs exams. The results of the assessment components carried out in three MTs in *pesantren* in Kebumen Regency is shown in Table 8.

Table 8. Component Achievement Result of Assessment Form

Component Achievement Result of Assessment Form (%)				
No.	Subjects	Madrasah		
		MTs Nururrohmah	MTs YAPIKA	MTs Salafiyah
1.	Indonesian	65.87	77.77	67.08
2.	English	69.41	84.12	68.35
3.	Mathematics	78.83	71.42	68.36
<b>Overall Average</b>		<b>72.35 (Good)</b>		

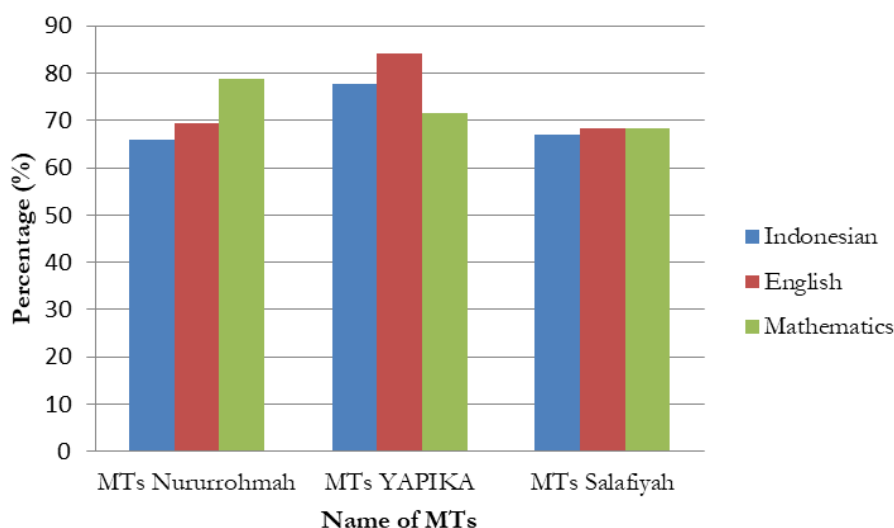


Figure 3. Achievement Diagram of the Assessment Form

Based on Table 8, 72.35% of students stated that the assessment in MTs had used various forms according to the characteristics of the material being taught. It shows that the assessment implementation had been going well, although improvements are appreciated. The difference in specific achievement for student respondents is presented in Figure 3.

From Figure 3, it is observed that the form of the assessment done at MTs Nururrohmah and MTs YAPIKA was better in approaching the standard than MTs Salafiyah Wonoyoso. Assessment of learning outcomes by educators was conducted in tests, observations, assignments, and other necessary forms. In contrast, the assessment by the education unit was in the form of MTs or school exams. Educators must provide an assessment form following the competencies to be measured.

The achievement of the assessment form components in MTs in modern *pesantren* in Kebumen Regency showed promising results, where 72.35% of respondents stated that the assessment form applied was under the [Regulation of the Minister of Education and Culture No. 23 of 2016](#). These results still need to be improved to be more optimal in its implementation. Various obstacles remained a problem, particularly for educators. Indonesian, English, and Mathematics subjects have different characteristics so that the assessment form carried out must also be different. The implementation in schools has the same tendency where educators are still centered on tests and assignments.

The assessment does not merely consist of tests and assignments in the linguistic field, but there must be other forms of measuring the skills possessed. [Utami \(2018\)](#) stated that assessment in Indonesian covers various abilities such as listening, speaking, reading, and writing. Based on the aforementioned various linguistic abilities, educators must look for appropriate assessment forms, for example, the use of an assessment form on portfolio is essential. However, in practice, the portfolio assessment has not been done for linguistic subjects. Various obstacles still occur related to project and portfolio assessment, primarily time constraints where this assessment is centered on the development of student's abilities so that it requires a more extended period.



In its development, efforts to assess various complex abilities still have to be improved so that the implementation of the assessment form components becomes optimal. It aims so that the competence of students can be adequately measured. The last component that is evaluated in this study was the assessment instrument. The components of the assessment instrument evaluated include: (a) the assessment instruments used by educators are tests, observations, individual or group assignments, and other forms following the characteristics of students' competence and level of development, and (b) the assessment instrument used by the education unit in the form of a final evaluation and/or school/MT's exam meets the requirements for substance, construction, and language and has evidence of empirical validity.

The results of the achievement of the assessment instrument components in three MTs in *pesantren* in Kebumen Regency are displayed in Table 9. Based on Table 9, 74.82% of students stated that the assessment instrument was under the [Regulation of the Minister of Education and Culture No. 23 of 2016](#). With a large percentage of 74.82%, it revealed that the assessment instrument had been appropriately implemented. The difference in the level of achievement of the three MTs can be seen in Figure 4. It shows that at MTs Salafiyah, the assessment instrument did not follow the existing rules of the other two MTs. Based on the results of a brief interview with the *madrasah*, there are still many teachers who do not have an educator certificate and are relatively young in age so that in compiling the assessment instrument they are still inexperienced. This is different from the other two MTs where the number of certified teachers is higher and most senior teachers have more extensive experience.

The assessment instruments can be tests, observations, assignments, and other necessary instruments. An education unit instrument must possess requirements that include substance, construction, language, and empirical validity. In this case, the education unit had carried out excellently due to preparing assessment instruments through school studies and various outside parties to guarantee the instrument quality.

Table 9. Component Achievement Result of Assessment Instrument

Component Achievement Result of Assessment Instrument (%)				
No.	Subjects	Madrasah		
		MTs Nururrohmah	MTs YAPIKA	MTs Salafiyah
1.	Indonesian	72.94	85.71	67.09
2.	English	71.76	77.78	74.68
3.	Mathematic	81.17	71.42	70.89
<b>Overall Average</b>		<b>74.82 (Good)</b>		

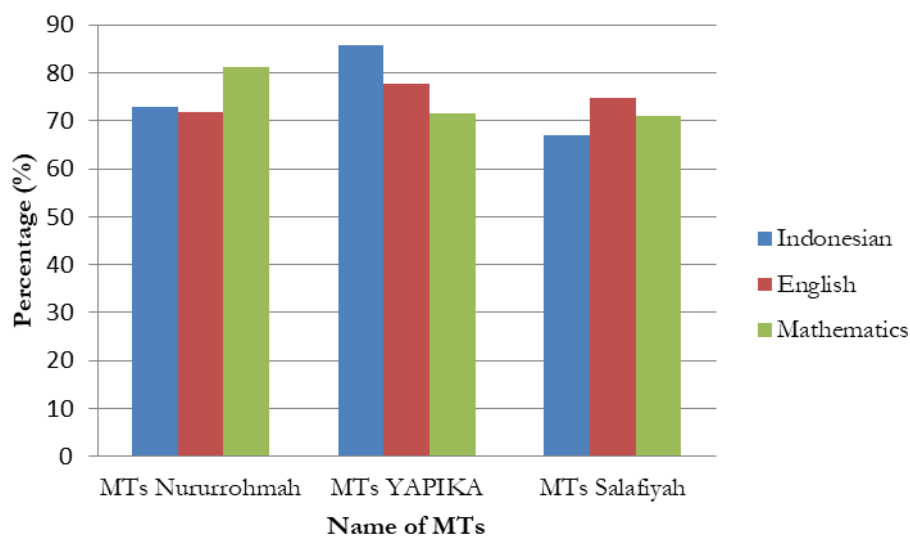


Figure 4. Achievement Diagram of the Assessment Instrument

Table 10. Overall Average of The Research Object

No.	MTs	Average (%)
1.	MTs Nururrohmah	69.01
2.	MTs YAPIKA	83.21
3.	MTs Salafiyah	76.18

Furthermore, the achievement of the results of the assessment instrument components in the three study objects was in the good category where 74.82% of respondents stated that the assessment had implemented various instruments following the characteristics of the subjects being assessed. However, this number still needs improvement to make it better. Various obstacles faced by educators in making assessment instruments still occur. Among the obstacles in the field is an assessment instrument that does not follow the aspect characteristics to be assessed.

Language subjects have a variety of components that are more complex than other subjects. The instrument prepared should measure these aspects. For example, an assessment using projects and portfolios should be carried out to assess linguistic materials. Educators only provide assessment instruments in tests and assignments, leaving students' abilities not appropriately measured. Therefore, it needs to be conveyed back to educators to carry out the assessment process as well as possible so that the progress of students' abilities can be adequately measured.

Educators and education units' implementation of the assessment on the three study objects had been carried out well despite many perceived obstacles. It is based on the assessment implementation where either the mechanism or the previous assessment instrument has been processed together with other MTs. Also, it is equipped with supervision from a higher level.

The assessment process is processed internally within the school and together with other MTs, including state MTs. It is a support for the achievement of the assessment implementation at the education unit level. In the next calculation, the three MTs gave different results. These results are presented in Table 10. Based on Table 10, the results on YAPIKA MTs show the best of the three MTs. This shows that the assessment implementation at the MTs was classified as good and following the standards. Other MTs were lower than MTs YAPIKA.

At MTs YAPIKA, the learning applied follows the 2013 Curriculum, where the assessment is carried out thoroughly. For example, in skills assessment, students are given the freedom to prepare the equipment needed. The interests of *pesantren* do not limit the need for learning. Various activities to hone students' skills are also available in schools, such as scout, drum bands, and other activities such as performing arts. These activities have significant benefits that impact learning in schools to train the courage and mentality of students. This is separate support for the assessment implementation in schools.

In the other two MTs, the regulations applied are more stringent so that students do not have wider scope in developing their skills. Furthermore, poor communication between MTs and *pesantren* managers remained a big problem in implementing the assessment. The competence of students had not been appropriately measured. Also, at MTs YAPIKA, most of the educators came from the surrounding area near the location of the MTs. The teachers seemed to be more active in teaching than other MTs. This is evidence that learning at MTs YAPIKA is more active so that the assessment process becomes better.

This study also found that the principal component of the assessment had a lower level of achievement than the other components. This is because assessment principles such as valid and objective tend to be challenging to implement. The assessment carried out by the teacher is not under the competencies to be achieved. Teachers have difficulty in assess student attitudes. In assessing student attitudes, teachers must observe student behavior during learning. It is challenging since students do not entirely obey the rules conveyed by the teacher (Suciati et al., 2017, p. 70).

Based on this study, the objective principle is challenging to implement in assessment. Therefore, teachers sometimes still assess with high subjectivity since they are affected by their closeness to students.

## CONCLUSION

Based on the study results, the components of the scope, principles, forms, and assessment instruments had been reflected in the assessment activities in MTs. There were 70.55% of students asserting that the assessment had covered three aspects of the assessment scope, i.e., knowledge, attitudes, and skills. In addition, the assessment component implementation was in a good category, with 67.82% of students stating that the nine principles had been reflected in the assessment implementation by educators and education units. A total of 72.35% of students stated that the assessment had used the appropriate form to measure the achievement of student competence. Assessment at MTs in *pesantren* in Kebumen Regency utilized instruments that comply with regulations where 74.82% of students stated that the assessment instrument had met the linguistic requirements and empirical validity.

Based on the study results, several suggestions can be made, such as increasing training for educators to increase competence, especially in carrying out standardized assessments. Another recommendation is the importance of further research to analyze the obstacles in the assessment implementation, especially in schools under *pesantren*, to find solutions in improving the achievement of the assessment component.

## REFERENCES

- Ayu, S. M., & Marzuki, M. (2017). An assessment model of Islamic religion education teacher personality competence. *REID (Research and Evaluation in Education)*, 3(1), 77–91. <https://doi.org/10.21831/reid.v3i1.14029>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: Global perspective* (7th ed.). Pearson Education.
- Hidayah, I. (2020). Analisis standar penilaian pendidikan di Indonesia. *AL-IMAN: Jurnal Keislaman Dan Kemasyarakatan*, 4(1), 85–105. <http://ejournal.kopertais4.or.id/madura/index.php/aliman/article/view/3851>
- Istikomah, I. (2017). Modernisasi pesantren menuju sekolah unggul. *Halaqa: Islamic Education Journal*, 1(2), 53–62. <https://doi.org/10.21070/halaqa.v1i2.1246>
- Jihad, A., & Haris, A. (2013). *Evaluasi pembelajaran*. Multi Pressindo.
- Khuriyah, K., Zamroni, Z., & Sumarno, S. (2016). Pengembangan model evaluasi pengelolaan pondok pesantren. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1), 56–69. <https://doi.org/10.21831/pep.v20i1.7529>
- Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Parama Publishing.
- Regulation of the Minister of Education and Culture No. 23 of 2016 Concerning the Educational Assessment Standards. (2016). [https://bsnp-indonesia.org/wp-content/uploads/2020/12/Permendikbud\\_Tahun2016\\_Nomor023.pdf](https://bsnp-indonesia.org/wp-content/uploads/2020/12/Permendikbud_Tahun2016_Nomor023.pdf)
- Retnawati, H. (2015). Hambatan guru matematika sekolah menengah pertama dalam menerapkan kurikulum baru. *Jurnal Cakrawala Pendidikan*, XXXIV(3), 390-403. <https://doi.org/10.21831/cp.v3i3.7694>
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian*. Parama Publishing.
- Salamah, U. (2018). Penjaminan mutu penilaian pendidikan. *Evaluasi: Jurnal Manajemen Pendidikan Islam*, 2(1), 274–293. <https://doi.org/10.32478/evaluasi.v2i1.79>
- Suciati, R. M., Nurhaidah, N., & Vitoria, L. (2017). Pelaksanaan penilaian hasil belajar siswa pada subtema hidup rukun dengan teman bermain di kelas II SDN 14 Banda Aceh. *Jurnal Ilmiah*

*Pendidikan Guru Sekolah Dasar FKIP Unsyiah*, 2(1), 59–72.  
<http://www.jim.unsyiah.ac.id/pgsd/article/view/2532>

Takdir, M. (2018). *Modernisasi kurikulum pesantren*. IRCiSoD.

Utami, S. (2018). Pengaruh kemampuan berbicara siswa melalui pendekatan komunikatif dengan metode simulasi pada pembelajaran bahasa Indonesia. *Jurnal LikhitaPradnya*, 18(2), 58–66.  
<https://likhitapradnya.wisnuwardhana.ac.id/index.php/likhitapradnya/article/view/59%0A>

Wirawan, W. (2016). *Evaluasi: Teori, model, metodologi, standar, aplikasi dan profesi* (3rd ed.). Rajawali Pers.