

EVALUATION OF TEACHER ASSESSMENT CAPABILITIES THROUGH MULTI-STAGE ADAPTIVE TESTING

Syarifah Eva Surya Habib^{1*}, Zahrila Ismail²

¹ Madrasah Ibtidaiyah Negeri 9 Nagan Raya, Aceh, Indonesia

² SMK Negeri Penerbangan Aceh, Banda Aceh, Indonesia

Corresponding Author: evasurya751@gmail.com

Abstract

The objective of this study is to assess MSAT (MULTI-STAGE ADAPTIVE TEST) performance of 21 Teachers and was carried out at Madrasah Ibtidaiyah Negeri 9 Nagan Raya, with the research subjects being Madrasah teachers and performed with three cycle tests. In accordance with the assessment results, the average score increased from pre-cycle to cycle I to cycle II, where the pre-cycle observation average score only reached 6.4 while in cycle I it reached 11.4, indicating that the teacher's ability to the assessment process using the multi-stage adaptive testing (MSAT) application is good. Cycle II then received a score of 15.8, indicating that the teacher's proficiency in the assessment process utilizing the multi-stage adaptive testing (MSAT) application was excellent.

Keywords: Multi-Stage Adaptive Test, Teacher, assessment, In House Training,

Abstrak

Tujuan dari penelitian ini adalah untuk menilai kinerja MSAT (MULTI-STAGE ADAPTIVE TEST) dari 21 Guru dan dilakukan di Madrasah Ibtidaiyah Negeri 9 Nagan Raya, dengan subjek penelitian adalah guru Madrasah dan dilakukan dengan tes tiga siklus. Sesuai dengan hasil penilaian, skor rata-rata meningkat dari pra siklus ke siklus I ke siklus II, dimana skor rata-rata observasi pra siklus hanya mencapai 6,4 sedangkan pada siklus I mencapai 11,4, hal ini menunjukkan bahwa kemampuan guru dalam proses penilaian. menggunakan aplikasi multi-stage adaptive testing (MSAT) bagus. Siklus II kemudian memperoleh skor 15,8 yang menunjukkan bahwa kemampuan guru dalam proses penilaian dengan menggunakan aplikasi multi stage adaptive testing (MSAT) sangat baik.

Kata Kunci: Multi-Stage Adaptive Test, Teacher, assessment, In House Training,

INTRODUCTION

Education is one of the principles taught in Madrasas and plays a significant part in the intellectual development of students (Parker & Raihani, 2011). In general, person who are knowledgeable with learning patterns deal in more information about concepts that rely solely on memorization, therefore students with this learning system approach tend to get bored (Ahmad, 2013). As a result, assessment is a crucial aspect of the learning process when it comes to achieving student learning outcomes. A good assessment pattern can improve the teaching and learning process and influence student learning results (Broadfoot & Black, 2004).

The learning process at Madrasahs demands active participation of students and aims to form cognitive, emotional, and psychomotor mastery in students, hence measuring learning results with

objective or subjective assessments is insufficient (Nurulloh et al., 2020). The appraisal of students' abilities to carry out activities, both as conducting experiments and generating works, cannot be revealed in this manner. Similarly, student actions while carrying out teacher tasks, both in the form of assignments to carry out experiments, demonstrations, and observations (Sholihuddin, 2020).

The phenomenon described above demonstrates that the assessment system used to measure student learning outcomes has a significant impact on the learning techniques devised and implemented by teachers. Obviously, the appropriate evaluation method must be corresponded with the goals and learning process (Faroh Kamaliya et al., 2023). In order for learning outcomes to be disclosed as a whole, measuring instruments that can determine students' abilities from the aspect of scientific work (scientific skills and attitudes) and how fit students can apply the knowledge information they receive must be used in addition to objective and subjective tests. Evidently, an alternative assessment tool is required, including the assessment process using the multi-stage adaptive testing (MSAT) program, which is presumed to be capable to implemented (Ali, et.al, 2002).

The evaluation of student performance has a good impact in the classroom because it provides teachers with pedagogical references that assist in the development of effective instructional strategies. Furthermore, the exam provides detailed information about student learning progress, including strengths and limitations (Ghorbani & Ghousi, 2020). Accreditation policies and the implementation of accompanying professional standards are being utilized by a variety of governments and regulatory authorities across the globe in an effort to achieve the goal of ensuring that teachers have a consistent level of preparation, readiness, and performance (Aminah, N., K., & Rusilowati, A., 2022). Although numerous articles have been written about the commodification of education, its influence on teachers, educational institutions, and classrooms remains largely undocumented (Nasir, 2021).

Teachers are required to evaluate their students in a variety of settings throughout the course of what they do in order to make assessments on their students' education. This includes circumstances such as lesson planning, modifying a course of instruction, selecting assignments, providing feedback, and determining grades, placement, and monitoring (Grossman et al., 2019). To put it into perspective with various examples: At this point, teachers may have to determine whether or not they can move on to a new component of a subject or whether or not the pupils require additional practice. They require information about their students' present levels of comprehension so that they may make this option (Amhag et al., 2019).

The process of handling tests that is based on algorithms is called multistage testing. This is somewhat similar to computer adaptive testing, in which things are chosen interactively for each examinee by an algorithm. However, rather than choosing individual items, groups of items are chosen, and the test is built up incrementally (Akram, 2019). This collection is referred to as a testlet or panel. Theoretically, people are capable of performing multistage tests; however, due to the significant

computation necessary (which frequently involves item response theory), multistage tests are typically carried out by computers. There may be a variable number of phases or testlets. If the testlets are only a few things each, for example five, it should not be difficult to employ 10 or more of them in the test. Some multistage examinations are constructed using a Madrasah Ibtidaiyah consisting of at least two stages (one of these stages will be a standard fixed form examination). The adaptive testing paradigm can solve the shortcomings of the current educational evaluation model. This concept enables for the use of on-target assessments, i.e. tests whose level of difficulty corresponds to students' abilities. The utilization of these tests will yield ideal results, and the test will end as soon as information on students' abilities can be estimated. The Item Response Theory (IRT) approach is used in the estimate technique for measuring the adaptive testing model(Landrum et al., 2020).

Based on the initial assessment that researchers conducted at Madrasah Ibtidaiyah Negeri 9 Nagan Raya, the average teacher at Madrasah Ibtidaiyah Negeri 9 Nagan Raya did not understand the assessment process using the multi-stage adaptive testing (MSAT) application. Multistage testing is an algorithm-based approach to managing tests. This is very similar to computer adaptive testing where items are selected interactively for each examinee by an algorithm, but instead of selecting individual items, groups of items are selected, building the test incrementally. This group is called a tested or panel. While multistage tests can theoretically be performed by humans, the extensive computation required (often using item response theory) means that multistage tests are performed by computers.

LITERATURE REVIEW

Teacher capability

The capability of a teacher can be defined as their ability to successfully carry out good and right instruction in order to generate quality graduates in the areas of faith, knowledge, and charity. Competence is one of the needed competencies in order to achieve this level of performance. These competencies include both one's professional competence and their personality competence. Professional competence is an ability that is related to the mastery of learning material in the field of study in a broad and deep manner. This includes the mastery of the mastery of the mastery of the substance of the content of the curriculum material in school subjects and the scientific substance that overshadows the curriculum material, as well as increasing scientific insight as a professional teacher(Gay & Howard, 2010).

Furthermore, Stephen P. Robbins and Timonthy A. Judge (2009: 57-61) suggest that an individual's overall ability consists of two groups of characteristics, namely:

a. Intellectual capability

The skill required to conduct various mental activities such as thinking, reasoning, and

problem solving which is can be practice in appropriated method of classroom process.

b. Physical ability

Physical ability refers to the ability to do duties that necessitate the use of Madrasah Ibtidaiyaha staff, skills, strength, and other comparable attributes.

Student Assessment Process

The ability of the educational unit to control the learning process has a substantial impact on the quality of education. Assessment is an essential component of learning. Educators, as supervisors of learning activities, can perform assessments to determine students' abilities, the accuracy of the teaching methods utilized, and students' progress in achieving set competencies. Educators can make the best judgment about what steps to take next based on the evaluation results. The assessment scores might also serve as encouragement for students to improve their performance. Depending on the competency being tested, many assessment procedures can be used in a complimentary (complementary) manner. Depending on the basic competencies that must be acquired, numerous evaluation methodologies can be used to measure learning results. The technique divides the assessment into two parts: tests and non-tests(Prus & Johnson, 1994).

A. Test method

The test methodology is a method of conducting tests in the form of questions that must be answered, queries that must be responded to, or tasks that must be completed by the person being tested. The ability of pupils to master the lessons taught covering areas of knowledge and abilities is to be tested in terms of the learning outcomes test.

In general, assessment tools with test approaches can be classified based on their implementation tools as follows:

a. Written Exam

The written test is a type of evaluation that needs written responses, either in the form of alternatives or entries.

b. Oral Examination

The oral test is a tool for evaluating learning outcomes in which questions and answers, assertions, and reactions are provided orally and spontaneously. This type of test necessitates a set of questions as well as grading standards.

c. Practice Exams / Actions

Practice/deed tests are approaches for assessing learning outcomes that require students to

demonstrate their skills or display their learning outcomes through performance.

B. The non-test method

The non-test method is a strategy for gaining an overview, particularly of attributes, attitudes, or personality. Non-test assessment procedures are classified as follows:

a. Observation

Observation/observation is a strategy used by educators to assess students directly through their senses. Observations were taken out with already constructed instruments.

b. Project

Assessment by assignment is a way of assessing pupils that requires them to do particular activities outside of class. Individual or group assignments can be used to assess students. Assignment-based assessment might take the shape of assignments or projects.

c. Inventive

Product assessment is an evaluation of the ability to produce a product within a specific time frame while adhering to predefined criteria in terms of both process and ultimate outcome.

d. Portfolio

A portfolio is a collection of student work that has been taken during the learning process and is organized in a systematic and organized manner. Portfolios are used by educators and students to track students' knowledge, skills, and attitudes in certain areas. The portfolio describes the student's attainment development, as well as the student's strengths and limitations in areas such as work creation and other student work. Title pages, table of contents, papers, portfolio documents, document groups, educator and parent notes are all included in the portfolio.

C. Multi-Stage Adaptive Testing (MSAT) Applications

The adaptive testing paradigm can solve the shortcomings of the current educational evaluation model. This concept enables for the use of on-target assessments, i.e. tests whose level of difficulty corresponds to students' abilities. The utilization of these tests will yield ideal results, and the test will end as soon as information on students' abilities can be estimated. The Item Response Theory (IRT) approach is used in the estimate technique for measuring the adaptive testing model.

The IRT method is test item oriented rather than test orientated (Hambleton, 1991). As a result,

using the IRT approach, the performance of an individual or group of individuals in an item can be predicted(Huff & Sireci, 2001).

METHOD

This study is conducted in the form of action research, with the goal of enhancing the capacity of teachers to participate in the assessment procedure by means of the multi-stage adaptive testing application and through in-house training program (IHT), in order to implement process standards. These standards include the following: three cycles, with each cycle consisting of four stages: the planning stage program; program implementation; observation and evaluation. This study was carried out in Madrasah Ibtidaiyah Negeri 9 Nagan Raya. conducted for 6 months in semester 1 of 2021, 21 teachers represented the total number of participants.

Table 1 timetable on research activities of multi-stage adaptive testing (MSAT).

Activities	Time
Preliminary Observations	July
Briefing the teacher on the assessment process's capacity to use the multi-stage adaptive testing (MSAT).	September
Meetings 1 and 2 of the in-house training programs (IHT)	October
Meetings 3 and 4 of the in-house training programs (IHT)	October
Evaluation	November

RESULT

According to the initial assessment conducted at the Nagan Raya 9th Madrasah Ibtidaiyah, the average teacher did not comprehend the assessment procedure using the multi-stage adaptive testing (MSAT) application. Multistage testing is an algorithm-based testing management strategy. This is very similar to computer adaptive testing, in which items are selected interactively for each examinee by an algorithm, but instead of selecting individual items, groups of items are chosen, allowing the test to be constructed in stages. This collection is known as a testlet or panel. Theoretically, humans are capable of administering multistage tests, but the extensive computation required (often using item response theory) means that multistage tests are administered by computers.

In the pre-cycle, researchers simply observed the teacher's ability in the assessment process using the multi-stage adaptive testing (MSAT) application, and the results were very low, with an

average score of only 6.4%, indicating that the teacher's ability in the assessment process using the multi-stage adaptive testing (MSAT) application is not good.

In accordance with the results of the observations, the average score increased from pre-cycle to cycle I, from a score of 5 in pre-cycle to a score of 11.4 in cycle I, indicating that the teacher's proficiency in the assessment process using multi-stage application adaptive testing (MSAT) is good. Moreover, in cycle II, the observation results attained a score of 15.8, indicating that the teacher's proficiency with the multi-stage adaptive testing (MSAT) application was excellent.

The in house training (IHT) program is carried out through observation and monitoring in the in house training (IHT) program. From the results of the analysis, it was found that the in-house training program (IHT) was very effective in increasing the ability of teachers in the assessment process using the multi-stage adaptive testing (MSAT) application, because teachers had the opportunity to discuss together to study and solve problems based on the conditions in field, then can fix it or follow up on the next cycle continuously if the problem has not been resolved.

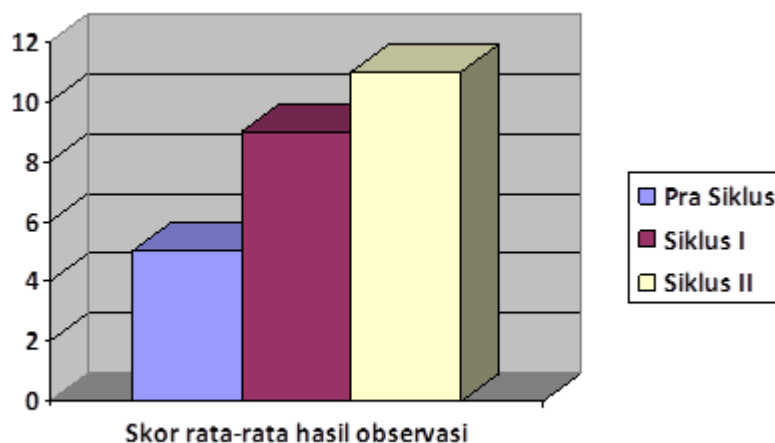


Figure 1. Multi-Stage Adaptive Testing (MSAT) application from Pre-Cycle, Cycle I to Cycle II

As shown in figure 1. The in-house training program (IHT) can enhance the assessment skills of instructors using the multi-stage adaptive testing (MSAT) software. Therefore, teachers must continually hone their skills in mastering and implementing the steps for using learning media in the classroom. Increasing teachers' ability to assess students using the multi-stage adaptive testing (MSAT) application must be constantly upgraded in order for learning to function smoothly in order to increase educational quality. The in-house training (IHT) program that applied at Madrasah Ibtidaiyah Negeri 9 Nagan Raya can increase teachers' abilities in the assessment process by using the multi-stage adaptive testing (MSAT) application. As shown the improvement in teacher performance from pre-cycle, cycle I to cycle II.

CONCLUSSION

This research was carried out at Madrasah Ibtidaiyah Negeri 9 Nagan Raya, with the research subjects was Madrasah teachers, and it was completed using three cycle tests. The purpose of this study is to evaluate the performance of 21 teachers on the MSAT (MULTI-STAGE ADAPTIVE TEST). The pre-cycle observation average score only reached 6.4 while in cycle I it reached 11.4, indicating that the teacher's ability to the assessment process using the multi-stage adaptive testing (MSAT) application is good. In accordance with the results of the assessment, the average score increased from pre-cycle to cycle I to cycle II. After that, Cycle II was given a score of 15.8, which indicated that the teacher's skill in the assessment process using the program for multi-stage adaptive testing (MSAT) was remarkable.

REFERENCE

- Ahmad, S. (2013). FAKTOR PENENTU KEBERHASILAN KEPALA SEKOLAH. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 17(1), 127–147. <https://doi.org/10.21831/PEP.V17I1.1365>
- Aminah, N., K., & Rusilowati, A. . (2022). Pengaruh Supervisi Akademik Dalam Meningkatkan Kompetensi Pedagogik Guru. *JURNAL HURRIAH: Jurnal Evaluasi Pendidikan Dan Penelitian*, 3(3), 28-42. <https://doi.org/10.5806/jh.v3i3.94>
- Akram, M. (2019). Relationship between Students' Perceptions of Teacher Effectiveness and Student Achievement at Secondary School Level. *Bulletin of Education and Research*, 41(2), 93–108.
- Amhag, L., Hellström, L., & Stigmar, M. (2019). Teacher Educators' Use of Digital Tools and Needs for Digital Competence in Higher Education. <https://doi.org/10.1080/21532974.2019.1646169>, 35(4), 203–220. <https://doi.org/10.1080/21532974.2019.1646169>
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy and Practice*, 11(1), 7–26. <https://doi.org/10.1080/0969594042000208976>
- Effect Of Metacognitive Skills On Prospective Teachers' Oral Task Performance | Pakistan Journal of Society, Education and Language (PJSEL)*. (n.d.). Retrieved June 14, 2023, from <https://www.pjsel.jehanf.com/index.php/journal/article/view/946>
- Faroh Kamaliya, A., Sumarni, S., Putranta, H., Lestari, A. I., Nurhayati, F., Ayuningtiyas, N. A., Sunan, U., & Yogyakarta, K. (2023). ANALYSIS OF ISLAMIC SUBJECT TEST QUALITY USING THE ITEM RESPONSE THEORY OF RASCH MODEL. *JURNAL HURRIAH: Jurnal Evaluasi Pendidikan Dan Penelitian*, 4(1), 117–124. <https://doi.org/10.5806/JH.V4I1.118>
- Gay, G., & Howard, T. C. (2010). Multicultural teacher education for the 21st century. <http://dx.doi.org/10.1080/08878730009555246>, 36(1), 1–16. <https://doi.org/10.1080/08878730009555246>

- Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, 8, 67899–67911. <https://doi.org/10.1109/ACCESS.2020.2986809>
- Grossman, P., Dean, C. G. P., Kavanagh, S. S., & Herrmann, Z. (2019). Preparing teachers for project-based teaching. *Https://Doi.Org/10.1177/0031721719841338*, 100(7), 43–48. <https://doi.org/10.1177/0031721719841338>
- Huff, K. L., & Sireci, S. G. (2001). Validity Issues in Computer-Based Testing. *Educational Measurement: Issues and Practice*, 20(3), 16–25. <https://doi.org/10.1111/J.1745-3992.2001.TB00066.X>
- Landrum, B., Bannister, J., Garza, G., & Rhame, S. (2020). A class of one: Students' satisfaction with online learning. *Https://Doi.Org/10.1080/08832323.2020.1757592*, 96(2), 82–88. <https://doi.org/10.1080/08832323.2020.1757592>
- Nasir, M. (2021). Curriculum Development and Accreditation Standards in the Traditional Islamic Schools in Indonesia. *Journal of Curriculum Studies Research*, 3(2), 37–56. <https://doi.org/10.46303/JCSR.2020.3>
- Nurulloh, A., Aprilianto, A., Sirojuddin, A., Anas, M., & Arif, M. (2020). THE ROLE OF THE HEAD OF MADRASAH'S POLICY IN IMPROVING TEACHER PROFESSIONALISM. *Nidhomul Haq : Jurnal Manajemen Pendidikan Islam*, 5(3), 334–346. <https://doi.org/10.31538/NDH.V5I3.963>
- Parker, L., & Raihani, R. (2011). Democratizing Indonesia through Education? Community Participation in Islamic Schooling. *Http://Dx.Do.Org/10.1177/1741143211416389*, 39(6), 712–732. <https://doi.org/10.1177/1741143211416389>
- Prus, J., & Johnson, R. (1994). A critical review of student assessment options. *New Directions for Community Colleges*, 1994(88), 69–83. <https://doi.org/10.1002/CC.36819948811>
- Sholihuddin, M. (2020). INTERNALIZATION OF PRINCIPAL CURRICULUM MANAGEMENT IN PRIMARY SCHOOL AND MADRASAH IBTIDAIYAH. *International Journal of Islamic Education, Research and Multiculturalism (IJIERM)*, 2(3), 222–233. <https://doi.org/10.47006/IJIERM.V2I3.118>