International Journal of Community Service Learning
Volume 7, Issue 3, 2023, pp. 272-281
P-ISSN: 2579-7166 E-ISSN: 2549-6417
Open Access: https://doi.org/10.23887/ijcsl.v7i3.63063

Training and Mentoring in the Development of Test Instruments for Measuring Learning Outcomes of Muhammadiyah School

Hari Sugiharto Setyaedhi[1*], Rusijono[2], Khusnul Khotimah[3]

[1,2,3] Educational Technology, Surabaya State University, Surabaya, Indonesia

ARTICLE INFO

Article history:
Received June 27, 2023
Revised July 05, 2023
Accepted August 10, 2023
Available online August 25, 2023

Kata Kunci :
Pelatihan dan pendampingan, instumen tes, hasil belajar.

Keywords:
Training and mentoring, test instruments, learning outcomes.

ABSTRAK

Program pembelajaran pengabdian masyarakat ini merupakan kerjasama antara Departemen Kurikulum dan Teknologi Pendidikan Fakultas Ilmu Pendidikan. Permasalahan utama yang dihadapi sebagian besar guru sekolah SMA dan SMK Muhammadiyah adalah kurangnya pengetahuan dan pemahaman mereka tentang cara mengembangkan instrumen tes yang valid dan reliabel. Penelitian ini bertujuan untuk meningkatkan pengetahuan guru SMA dan SMK Muhammadiyah dalam mengembangkan instrumen penilaian pembelajaran. Metode Blended Learning diterapkan selama 32 jam pembelajaran yang meliputi ceramah, diskusi tanya jawab, penugasan, kerja praktek, dan tes. Sebanyak 12 guru dari sekolah SMA dan SMK Muhammadiyah mengikuti program tersebut. Metode penelitian yang digunakan adalah One-Group Pretest-Posttest Design yang merupakan jenis desain penelitian pra-eksperimental. Data yang diambil dari hasil pretest dan posttest dianalisis menggunakan uji-t. Setelah pelatihan dan pendampingan berjalan lancar, kesimpulan utama yang dapat diambil adalah bahwa program pembelajaran pengabdian kepada masyarakat telah berhasil mencapai outcome yang diharapkan yaitu peningkatan kemampuan guru dalam mengembangkan instrumen tes. Hal ini dibuktikan dengan peningkatan skor posttest yang signifikan dibandingkan hasil pretest pada Sig. (2-ekor) = 0,001 < 0,05.

ABSTRACT

This community service learning program is collaboration between the Curriculum and Educational Technology Department of the Faculty of Education Sciences. A key issue facing most SMA and SMK Muhammadiyah school teachers is their lack of knowledge and understanding on how to develop valid and reliable test instruments. This research aims to enhance the knowledge of SMA and SMK Muhammadiyah teachers in developing learning assessment instruments. The Blended Learning method was applied for 32 hours of lessons that involved lectures, Q&A discussions, assignments, practical work, and tests. A total of 12 teachers from SMA and SMK Muhammadiyah schools took part in the program. The research method used was the One-Group Pretest-Posttest Design, which is a type of pre-experimental research design. Data drawn from the pretest and posttest results were analyzed using the t-test. Following the smooth implementation of training and mentoring, a key conclusion is that the community service learning program has successfully met its expected outcome of enhancing the teachers' ability to develop test instruments. This is evidenced by a significant increase in posttest scores compared to pretest results at Sig. (2-tailed) = 0.001 < 0.05.

1. INTRODUCTION

Every program involves an assessment and evaluation. The assessment of learning outcomes focuses on individuals, while the assessment of learning outcomes centers on groups or programs. This has implications for the improvement of programs, individuals and groups/institutions. Learning assessments are fundamental and part of an educator's duty and competence (Istiyono et al., 2020; Setyawan et al., 2021). Article 45 of Indonesian Law No. 14/2005 on Teachers and Lecturers stipulates that educators should have at least 4 (four) relevant qualifications. The four competencies are 1) subject mastery, 2) mastery of methodologies, 3) mastery of good evaluation techniques, and 4) good understanding, internalization, and application of moral values and professional ethics (Anetha & Hasriyanti, 2019; Perangin-angin et al., 2020). Evaluation is key in learning development that involves

*Corresponding author
E-mail addresses: harisetyaedhi@unesa.ac.id (Hari Sugiharto Setyaedhi)

teacher-student interaction within a learning environment. Pursuant to Chapter XVI Article 58 Paragraph 1 of Law No. 20/2003 on the National Education System, learning evaluation is necessary to monitor learning outcomes (Riswanda & Burhan, 2022; Warju et al., 2020). An assessment is conducted to gauge the extent to which students have been able to develop their abilities based on what they have learned. An assessment is a planned activity that uses a tool or instrument to determine the condition of the object. The results of which can be compared with predetermined standards to draw conclusions. Learning assessments are inextricably linked to teaching activities. Classroom assessments are crucial as they determine students' level of success in terms of understanding the materials taught (Idrus, 2019; I. Magdalena et al., 2020). The purpose of an assessment is to kindle students' learning interest, motivate teachers to improve the quality of learning, and inspire campuses to continuously upgrade their facilities and infrastructure. Assessments can help enhance student academic achievements and improve learning quality in schools to further broaden learning opportunities. The main benefit of learning assessment is the improvement of learning quality, allowing for high-quality learning and education (Mahirah, 2017; Umami et al., 2021). The quality of learning can be measured with a test instrument, which refers to a set of tests to obtain data as the basis for measuring learning outcomes.

In learning evaluation, teachers rely on tests as a measurement instrument. Evaluation tools may be in the form of tests or non-tests (Jumrah et al., 2023; Widiyawati et al., 2019). Tests form an important part of the learning assessment process. They can be used to measure the success of a particular program (Fahrurrozi & Laili Rahmawati, 2021; Sholihah et al., 2017). In addition, test instruments play a key role in determining the effectiveness of the learning process (Manfaat & Nurhairiyah, 2021; Setiyawan & Wijayanti, 2020). Test are used to determine student learning outcomes achieved throughout the learning process (Aisyah et al., 2021; Birch et al., 2017; Nurfillaili et al., 2016; Sa'diyyah et al., 2021). They are prepared in accordance with how students answer test questions. In such circumstances, students are encouraged to do their best. Tests are administered for various purposes, including student admission, course completion, selection, and others (Litna et al., 2021; Ma'rifah et al., 2021; Sa'diyyah et al., 2021).

The requirements for a good test are a) reliability, b) validity, c) objectivity, d) discrimination power, e) comprehensiveness, and f) usability. A test instrument that qualifies as a learning evaluation tool is having high reliability, among others (Hanifah, 2014; Karlina & Jamilah, 2020). Several aspects that need to be taken into account before using a test instrument are as follows: 1) validity: the degree of accuracy of the measurement instrument towards what is intended to be measured, 2) reliability: giving relatively the same results when a test is administered to students over different points in time, providing the object being measured remains unchanged, 3) objectivity: the test results are not affected by the researcher's subjective opinions, 4) balanced: the difficulty index of test items should be line with test objectives, 5) discrimination power: the test should be able to differentiate between high- and low-achieving students within a group, and 6) norms: the test results should be straightforward in accordance with specific standards (Anastasi. Anne and Urbina, 2017; Farida & Musyarofah, 2021; Sulistianingsih, 2020). It is important to note that a test instrument should meet the requirements of validity and reliability for the accuracy and consistency of measurement results (Nengsi & Efrina, 2019; Sugiono et al., 2020). An instrument is considered of high quality when it has been tested for validity and reliability (Dewi & Sudaryanto, 2020; Puspasari & Puspita, 2022). Test reliability can be affected by factors such as test-taker characteristics, test-taking conditions, variations in test administration, scoring errors and differences, test length, homogeneity of student ability, and test item difficulty level (Alfiatunnisa et al., 2022; Bashooir & Supahar, 2018; Putri & Nahadi, 2019). Furthermore, it was found that test reliability is closely associated with how a test is presented, the test-taker's mood, the test-taker's attitude towards the test, motivation, testing site conditions, and others.

The procedure for developing a reliable test instrument is as follows: 1) develop test specifications, 2) develop test items, 3) review test items, 4) pilot-test the test items, 5) item quantitative analysis, 6) revise the test items, 7) prepare the test, 8) administer the test, and 9) interpret the test results (Muliyani & Huriaty, 2016; Purnomo & Maria Sekar Palupi, 2016). In the educational context, an instrument helps measure student achievement, the teaching and learning process, and program achievement (I. Magdalena et al., 2020; Muttaqin & Kusaeri, 2017). Tests serve as a tool for evaluating student learning achievement, the effectiveness of the teaching and learning process, and program achievement. The types of test instruments most commonly used include learning assessments, formative assessments, summative assessments, and others. Tests are often used as a learning assessment tool. They should be specifically adapted to the learning objectives in accordance with predetermined rules. A quality test is required in the evaluation process to determine the quality of the information generated. Educational tests are tools to measure specific skills in order to differentiate one skill from another. This calls for the need to develop the best possible test instrument. Previous study state a test qualifies as a measurement tool when it meets the requirements of validity and reliability (Arikunto, 2012). It is not

easy to create a good test instrument as it involves several procedures. Other study showed that school teachers were more likely to draw from problems or exercises provided in textbooks rather than developing them on their own (Hodyanto & Saputro, 2018). Textbook problems are not necessarily in line with the learning objectives. This often results in inappropriate test questions that may even diverge from what students are learning, rendering the assessment ineffective. The evaluation of learning outcomes is associated with test instruments. The quality of a test instrument is therefore key to a proper assessment of learning outcomes that can truly measure what has been set out in curriculum objectives. Assessment is a crucial part of the learning process. Trainers should not only be qualified to develop assessment tools to determine the achievement of learning outcomes, but also to consider whether the tools can effectively measure learning outcomes of high quality (Alenezi, 2020; Khaerudin, 2015). An evaluation is considered successful when student performance is consistent with educational goals. Conversely, if students' learning activities are misaligned with educational goals, it is considered unsuccessful. Learning assessments provide the basis for making decisions about student progress and the future directions, especially in the educational process. Educators aspire to consistently improve outcomes, which makes evaluation important by comparing results before and after the learning process (Idrus, 2019; Wulan & Rusdiana, 2015). A good evaluation will foster student motivation and provide teachers with valuable feedback to improve the quality of learning. It is therefore important to enhance knowledge and understanding of the basic concept of learning evaluation.

Evaluation initially involves measurement and evaluation. Training evaluation seeks to identify the learning outcomes achieved following the implementation of a learning process. Learning outcomes are the competencies that students have mastered upon completing the learning process. To analyze learning outcomes, a systematic learning assessment is required using tests as an assessment instrument (Syachlani & Setyorini, 2021; Tarigan, 2008). Tests are measurement tools that involves data collection. These instruments, both tests and non-tests, help gather information. Based on observations and interviews conducted by a team of lecturers from the Curriculum and Educational Technology Department of the Faculty of Education Sciences, Surabaya State University (UNESA) with Mr. Nawachid from the Central Board of Genteng Subdistrict, Banyuwangi as well as Mr. Taslim, Principal of SMK Muhammadiyah 1, and Mr. Tamyis Rosidi, Principal of SMK Muhammadiyah 2, it was revealed that a proportion of SMA (High School) and SMK (Vocational High School) teachers in Genteng Subdistrict did not know how to develop test instruments for measuring learning outcomes. In view of this, the training course provided is based on the issues facing SMA and SMK Muhammadiyah schools in Genteng for the purpose of knowledge transfer that will enable local teachers to more effectively perform their day-to-day duties at school, while minimizing errors in learning assessments. With this in mind, the Community Service Learning program introduced to SMA and SMK Muhammadiyah schools in Genteng seeks to enhance teachers' knowledge of the concept of learning evaluation. It is hoped that through training, the SMA and SMK Muhammadiyah school teachers in Genteng can gain a better understanding of the basic concept of learning assessment. The purpose of test development is to create valid and reliable tests that genuinely reflect the learning outcomes of students after their participation in learning activities. Objective test questions are often used to assess learning outcomes. This is partly due to the scope of topics covered in the exam and the ease with which the answers are assessed (Sudaryono, 2013; Sukardi, 2008). This was similarly highlighted in another study measuring student learning outcomes,whereby the more comprehensive the test content and test items covered in a test, the more reliable the scores tend to be.

2. METHOD

A key issue facing partner organization, Banyuwangi Regional Board's Primary and Secondary Education Council, is the need to enhance knowledge for developing valid and reliable learning assessment instruments to support learning evaluation in SMA and SMK Muhammadiyah schools in Genteng Subdistrict, Banyuwangi. The procedure for implementing the community service learning (CSL) program is show in Figure 1.

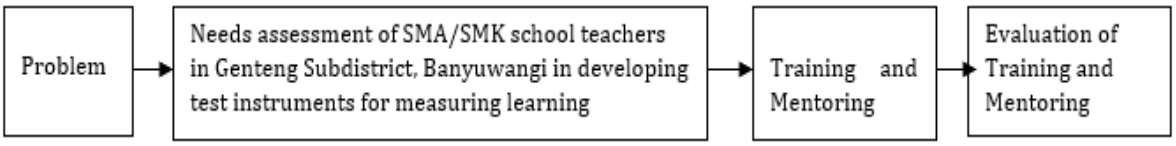

Figure 1. CSL Implementation Procedure

Prior to training and mentoring under the CSL program, the CSL team from the Curriculum and Educational Technology Department of Surabaya State University, led by Hari Sugiharto, has coordinated with partner, Banyuwangi Regional Board's Primary and Secondary Education Council of Genteng Subdistrict, Banyuwangi. Coordination involved interviews and discussions, both by phone and in-person, with relevant individuals including Banyuwangi's local government leaders, as well as SMA and SMK Muhammadiyah school principals and teachers in Genteng. The purpose of coordination is to identify the problems facing SMA and SMK Muhammadiyah school teachers. The CSL team immediately identified and developed a CSL plan to address the situation on the ground. SMA and SMK school teachers have lamented about their lack of knowledge in developing learning assessment instruments. Muhammadiyah high school teachers have been used to developing test instruments by simply writing test questions and then assessing them. The research subjects were 12 SMA and SMK school teachers. Data were collected from lectures, Q&A sessions, assignments, practical work, and tests. The CSL team conducted a pretest before treatment (training and mentoring) to determine the preexisting skills and abilities of SMA and SMK Muhammadiyah school teachers in developing test instruments. A posttest on the other hand was administered upon completion of the treatment (Susilo & Ernawati, 2018). This is an experimental research that aims to identify the causal effect of treatment on a variable (Montolalu & Langi, 2018). The type of experimental design used in this study is the One-Group Pretest Posttest design. This is because the researchers were focused on looking at the differences in skills after (posttest) and before (pretest) the experiment (Maharani et al., 2019; Rahmawati & Hardini, 2020).

Data from pretest and posttest were examined by running a t-test in SPSS. The t-test helps determine how effective (impactful) a teaching process. The type of t-test used is the paired sample t-test where the same individuals undergo 2 different treatments, i.e., pretest and posttest. The paired sample t-test compares the means of two variables in a group where the two variables are correlated (Palimbong et al., 2022). The paired sample t-test is part of comparative hypothesis testing. The data used in the paired sample t-test is a ratio scale. A paired sample t-test determines whether there is a difference in the means of the two paired or correlated samples. Activities were conducted offline and online that followed a schedule agreed by both parties. Training and mentoring in the development of test instruments under the community service learning program took place offline on a weekly basis. Activities were evaluated by SMA and SMK Muhammadiyah schools in Genteng Subdistrict with the CSL team from the Faculty of Education Sciences. An evaluation of mentoring support was conducted by means of a posttest at the end of the activity. A posttest was necessary to determine whether participants have understood and mastered the training materials on test instrument development provided by the UNESA CSL team.

The blended training method is necessary for this research. It refers to the use of both offline instruction and e-learning (Lin et al., 2020; Marchalot et al., 2018). It is a learning model that combines face-to-face and online teaching methods. As part of the community service learning program, in-person learning classes were held on Saturday, July 16, 2022, while online sessions took place every Monday. Both the offline and online approaches involve 1) interactive discussions of problems facing schools on a daily basis and how to deal with them, 2) assignments, by taking the pretest and posttest, formulating objective test questions, etc., and 3) practical work for developing a test instrument. The agreed schedule for online learning activities was every Monday over 4 (four) virtual sessions through Zoom meetings, making it a total of 32 hours of lessons. The topics taught in the first online session on July 25, 2022 were instrument development procedure, guidelines for test item formulation, and the characteristics of online learning assessment instruments. For the second online session on August 1, 2022, the topic was on the basic concept and functions of learning evaluation. This was followed by instructional materials on fun learning assessment techniques. The third online meeting was held on August 8, 2022 for administering a posttest. The training topic, methods, and resource persons are provided in Table 1.

Table 1. Training Topic, Method, and Resource Person

Day/Date	Topic	Method
Saturday, 16 July 2022	Opening of Training on Learning Assessment Pretest Basic concept of learning evaluation	Online Offline test Lecture, discussion, and presentation
Monday, 25 July 2022	Test development procedure, guidelines for test item formulation, and characteristics of online learning assessment instruments	Lecture, discussion, and presentation

Day/Date	Topic	Method
Monday, 1 August 2022	Basic concept and functions of learning evaluation	Lecture, discussion, and presentation
Monday, 8 August 2022	Fun learning assessment techniques	Lecture, discussion, and presentation
Monday, 15 August 2022	Posttest	Online test

3. RESULTS AND DISCUSSION

Results

The partner organization has been strongly supportive of the Community Service Learning Program, including in ensuring the availability of the necessary facilities and infrastructure. The partner team that included the principals and teachers of SMA and SMK Muhammadiyah schools in Genteng Subdistrict, Banyuwangi, took part in the program. The teachers recorded perfect attendance throughout the training course. The documentation of lecturers from UNESA's curriculum and educational technology department provide training and mentoring on learning evaluation in high schools and vocational high schools is sho win Figure 1.

Figure 1. Training and Mentoring on Learning Evaluation in High Schools and Vocational High Schools

Base on Figure 1, the training course began with an opening session that took place offline on July 16, 2022, at high schools and vocational high schools in Genteng Subdistrict, Banyuwangi. This was followed by a pretest to determine pre-existing skills and abilities regarding learning evaluation, after which Prof. Rusijono was on hand to explain the concept of learning assessment. Training participants also learned about the differences between assessment and evaluation, types of assessment based on their functions, norm-referencing and criteria-referencing, assessment methods, the definition of validity and reliability, and others. The day continued with a Q&A session and discussion to elicit responses from the teachers who gave positive reactions as demonstrated in the many questions posed. Given the time constraint where not all topics could be fully covered with many questions left unanswered, the session continued online. Nearly 90% of participants were actively engaged in discussions, asking a wide range of questions.

The CSL program managed to deliver multiple outputs and outcomes, as follows: a) teachers have gain a better understanding of the concept of learning assessment, such as task analysis and learning assessment objects, which consist of difficulty level, discrimination power, distractor effectiveness, as well as validity and reliability tests, b) teachers have an enhanced understanding of the test development procedure, which includes guidelines for developing objective and descriptive tests, characteristics of online learning evaluation tools, c) teachers have the ability to apply fun learning evaluation techniques, d) SMA and SMK school teachers in Genteng Subdistrict, Banyuwangi, have earned certificates signed by the Dean of the Faculty of Education Sciences, Surabaya State University. A posttest was administered to determine the teachers' level of understanding and to evaluate the implementation of the CSL program. A t-test was then conducted to compare the pretest and posttest scores. The mean values of the pretest and posttest are provided in Table 2.

Table 2. Pretest dan Posttest Mean Values

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-test	65.00	12	13.314	3.844
	Post-test	76.67	12	9.838	2.840

Base on Table 2 show descriptive statistics provide a summary of the results of the two samples examined, i.e., the pretest and posttest. The pretest mean value is 65, while the posttest mean value is 76.67. A total of 12 teachers took the tests. As the mean value of the pretest (65) is lower than the posttest (76.67), a mean difference is descriptively observed in the pretest and posttest learning outcomes. To determine whether the difference is significant, it was necessary to interpret the results of the paired sample t-test as show in Table 3.

Table 3. T-Test Results

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Condifence interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttes	-11.667	8.659	2.499	-17.168	-6.165	-4.668	11	0.001

Based on Table 3 show paired t-test output value, a Sig. (2-tailed) value of 0.001<0.05 signifies a difference between pretest and posttest mean scores. This confirms the positive effect of training and mentoring in enhancing the competence of SMA and SMK school teachers in Genteng Subdistrict, Banyuwangi. Training and mentoring in the development of test instruments for measuring learning outcomes by SMA and SMK school teachers in Genteng can therefore be considered a success.

Discussion

Given the research findings, the community service learning program in overall can be said to have went smoothly, meeting all expectations. SMA and SMK Muhammadiyah school teachers were keen and motivated to conduct proper assessments in their respective schools. In the evaluation of learning, a teacher uses tests as a measurement tool. A good test instrument should be valid and reliable (Arif et al., 2022; Mappalesye et al., 2021; Setiyawan & Wijayanti, 2020). To design a good test, an analysis is required. Test instruments should be developed through the proper procedure to produce effective measurement tools (Aisyah et al., 2021; Iskandar & Rizal, 2018; Setiyawan & Wijayanti, 2020). Test instruments will only be of high quality and suitable if they are based on the applicable principles of test development. A test instrument can be considered good if an item analysis is conducted. In evaluating student learning outcomes, teachers seldom perform a quantitative (empirical) analysis (Hasyim et al., 2021; R. Magdalena & Angela Krisanti, 2019). Test instruments rarely undergo a quantitative review. A good test instrument should meet the following requirements: 1) Moderate difficulty of test items, 2) item discrimination, 3) distractors, 4) high validity, and 5) high reliability (Arifin, 2017; Srika Ningsih Pasi; Yusrizal, 2018; Widayanti et al., 2021).

Teachers measure learning outcomes to determine students' knowledge and understanding after one semester of studying. Tests are normally used as an instrument to measure learning outcomes. As an assessment tool, a learning test helps determine whether a student has achieved the learning objectives. This is because tests are instruments that measure the extent to which educational goals have been achieved (Hikmah & Muslimah, 2021; Ina Magdalena et al., 2021). In addition, tests can determine the success of a learning program (Apsari & Acep Haryudin, 2017). Previous study state function as a tool to inform students about their mastery of a particular subject or topic (Wenno et al., 2021). Other study state tests can shed light on student achievement in order to make the right decisions (Ulfah et al., 2020). There is study state part of an educator's deliberate effort to show students their learning outcomes (Kurniawati, 2019). Test instruments will only be of high quality and suitable if they are based on the applicable principles of test development. A test instrument can be considered good if an item analysis. As state by previous study in evaluating student learning outcomes, teachers seldom perform a quantitative (empirical) analysis (Elviana, 2020). Other study state a test instrument is usually tested qualitatively, but

never quantitatively (Erawati, 2018). An evaluation has shown the successful implementation of training and mentoring by the UNESA CSL team. This is evidenced by the difference in the pretest and posttest mean scores (65 < 76.67) based on a paired t-test, whereby the Sig. (2-tailed) value of 0.001 < 0.05, which indicates a significant (real) difference between the pretest and posttest mean scores (Montolalu & Langi, 2018; Palimbong et al., 2022; Prameswari & Rahayu, 2020). There was however a host of challenges that affected training and mentoring activities, such as weak signals for online activities and participants' tardiness due to various reasons. Nevertheless, despite the snags, the participating teachers remained focused and eager to learn. The training and mentoring activities in which teachers from different subjects representing SMA and SMK Muhammadiyah schools in Genteng Subdistrict, Banyuwangi, have had a profound impact on schools. Many of the teachers consulted on the problems that they face in school, specifically regarding how test instruments can help improve the quality of education in schools and why this is important. The knowledge and skills that they have acquired from training will be put into practice in evaluation processes at their respective schools. The participating teachers are expected to share their training experiences with others, thereby improving the quality of education in Genteng Subdistrict, Banyuwangi. The training and mentoring activities met the expectations of SMA and SMK Muhammadiyah school teachers in Genteng Subdistrict, Banyuwangi, who have been facing difficulty in conducting proper learning evaluation, and as such found the training materials highly relevant and useful for their day-to-day duties and obligations. There were requests for the CSL Partner to facilitate schools in establishing good internet connection for continuous training and mentoring.

4. CONCLUSIONS

The community service learning program that centers on the provision of training and mentoring support in the development of test instruments for measuring learning outcomes has been beneficial for educators, especially SMA and SMK Muhammadiyah school teachers. It has succeeded in enhancing teachers' skills and understanding of learning evaluation, which included the following: a) teachers have gain a better understanding of how to develop test instruments, b) teachers have a good grasp of the concept of learning evaluation and item analysis, c) teachers are more knowledgeable of the test development procedure, which includes guidelines for developing objective and descriptive tests, and the characteristics of online learning evaluation tools, and d) teachers have the ability to apply fun learning evaluation techniques.

5. REFERENCES

Aisyah, S. N., Taena, L., & Ili, L. (2021). Pengembangan Tes Hasil Belajar Ekonomi Kelas X SMA melalui Google Formulir. *Jurnal Online Program Studi Pendidikan Ekonomi*, *6*(3), 106--114. http://repository.uinsu.ac.id/id/eprint/13576.

Alenezi, A. (2020). The role of e-learning materials in enhancing teaching and learning behaviors. *International Journal of Information and Education Technology*, *10*(1), 48–56. https://doi.org/10.18178/ijiet.2020.10.1.1338.

Alfiatunnisa, E., Khairunnisa, H. Z., Hayati, S., & Maulida, V. L. (2022). Uji Validitas dan Reliabilitas Terhadap Kemandirian Siswa Sekolah Dasar Kelas 1. *Jurnal Hurriah: Jurnal Evaluasi Pendidikan Dan Penelitian*, *3*(2), 29–36. https://doi.org/10.56806/jh.v3i2.81.

Anastasi. Anne and Urbina, S. (2017). *Psichological Testing* (Seventh Ed). New Jersey: Prentice-Hall, Inc.

Anetha, & Hasriyanti. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *8*(1), 57–68. https://doi.org/10.15408/jp3i.v8i1.13068.

Apsari, Y., & Acep Haryudin. (2017). The Analysis Of English Lecturers ' Classroom-Based Reading Assessments To Improve Students ' Reading Comprehension. *Journal Of English Language Teaching In Indonersia*, *5/1*, 35–44. https://doi.org/10.22460/eltin.v5i1.p35-44.

Arif, N., Yuanita, P., & Maimunah. (2022). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Matematis Berbasis Taksonomi SOLO pada Materi Barisan dan Deret. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *06*(02), 2318–2335. https://j-cup.org/index.php/cendekia/article/view/1498.

Arifin, Z. (2017). Kriteria Instrumen dalam Suatu Penelitian. *Jurnal THEOREMS (The Original Research of Mathematics)*, 28–36. https://doi.org/10.31949/th.v2i1.571.

Arikunto, S. (2012). *Prosedur penelitian suatu pendekatan praktek*. Rineka Cipta.

Bashooir, K., & Supahar. (2018). Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 168–181.

https://doi.org/10.21831/pep.v22i2.20270.
Birch, C., Lichy, J., Mulholland, G., & Kachour, M. (2017). An enquiry into potential graduate entrepreneurship: Is higher education turning off the pipeline of graduate entrepreneurs? *Journal of Management Development*, *36*(6), 743–760. https://doi.org/10.1108/JMD-03-2016-0036.
Dewi, S. K., & Sudaryanto, A. (2020). Validitas dan Reliabilitas Kuesioner Pengetahuan, Sikap dan Perilaku Pencegahan Demam Berdarah. *Junal Pendidikan Jasmani Indonesia*, *4*(1), 73–79. https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/11916.
Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal Mudarrisuna*, *10*(2), 58–74. https://doi.org/10.22373/jm.v10i2.7839.
Erawati, N. K. (2018). Analisis Tes Penilaian Pencapaian Kompetensi Pada Mahasiswa Kebidanan. *Jurnal Penjakora*, *5*(2), 111–120. https://doi.org/10.23887/penjakora.v5i2.17287.
Fahrurrozi, M., & Laili Rahmawati, S. N. (2021). Pengembangan Model Instrumen Evaluasi Menggunakan Aplikasi Kahoot Pada Pembelajaran Ekonomi. *Jurnal PROFIT Kajian Pendidikan Ekonomi Dan Ilmu Ekonomi*, *8*(1), 1–10. https://doi.org/10.36706/jp.v8i1.13090.
Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, *I*(1), 34–44. https://doi.org/10.32923/al-muarrib.v1i1.2100.
Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Jurnal SOSIO E-KONS*, *6*(1), 41–55. https://doi.org/10.30998/sosioekons.v6i1.1715.
Hasyim, A. F., Munawar, B., & Ma'arif, M. (2021). Penggunaan Media Video Untuk Meningkatkan Pemahaman Karakteristik Arus Searah Dan Bolak-Balik Pada Peserta didik MAN 1 Pandeglang. *Jurnal Pendidikan*, *9*(1), 5–24. https://doi.org/10.36232/pendidikan.v9i1.545.
Hikmah, & Muslimah. (2021). Validitas dan Reliabilitas Tes dalam Menunjang Hasil Belajar PAI. *Proceedings*, *1*, 345–356. https://jurnal.ar-raniry.ac.id/index.php/mudarrisuna/article/view/7839.
Hodyanto, H., & Saputro, M. (2018). Workshop Pembuatan Dan Analisis Butir Soal menggunakan ITEMAN Pada Madrasah Aliyah Miftahul Huda Kecamatan Sungai Ambawang. *Jurnal Transformasi*, *14*(2), 85–90. https://doi.org/10.20414/transformasi.v14i2.578.
Idrus. (2019). Evaluasi Dalam Proses Pembelajaran. *Evaluasi Dalam Proses Pembelajaran*, *2*, 920–935. https://garuda.kemdikbud.go.id/documents/detail/1655265.
Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(1), 12–23. https://doi.org/10.21831/pep.v22i1.15609.
Istiyono, E., Setiawan, R., & Harun. (2020). Pelatihan Penyusunan Instrumen Tes dan Analisnya Secara Modern Bagi Guru-Guru SMP IPA. *Jurnal Pengabdian Masyarakat MIPA*, *4*(1), 102–108. https://doi.org/10.21831/jpmmp.v4i2.37499.
Jumrah, Rukli, & Sulfasyah. (2023). Pengembangan Instrumen Tes Berbasis HOTS dengan Pendekatan Pengukuran Rasch pada Pelajaran Matematika Topik Bangun Ruang untuk Siswa Sekolah Dasar. *Jurnal Basicedu*, *7*(1), 11–27. https://doi.org/10.31004/basicedu.v7i1.4207.
Karlina, L., & Jamilah, J. (2020). Pengembangan Buku Ajar Berbasis Katalog Materi Plantae. *Jurnal Al-Ahya*, *2*(3), 103–114. https://repository.uinjkt.ac.id/dspace/handle/123456789/60934.
Khaerudin. (2015). Kualitas instrumen hasil belajar. *Jurnal Madaniyah*, *2*, 212–235. https://journal.stitpemalang.ac.id/index.php/madaniyah/article/view/26.
Kurniawati, A. (2019). Analisis Hasil Tes Evaluasi Pendidikan Pada Mahasiswa Ditinjau Dari Perbedaan Gender. *Jurnal Ilmiah DIDAKTIKA*, *19*(1), 89–106. https://doi.org/10.22373/jid.v19i1.4196.
Lin, C.-Y., Huang, C.-K., & Ko, C.-J. (2020). The impact of perceived enjoyment on team effectiveness and individual learning in a blended learning business course: The mediating effect of knowledge sharing. *Australasian Journal of Educational Technology*, *36*(1). https://doi.org/10.14742/ajet.4446.
Litna, K. ., Mertasari, N. M. ., & Sudhirta, G. (2021). Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) Matematika SMA Kelas X. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *11*(1), 10–20. https://repo.undiksha.ac.id/6637/.
Ma'rifah, U., Algiovan, N., & Sutarsyah, C. (2021). An Item Analysis of English Test During Online Learning. *International Journal of Multicultural and Multireligious Understanding*, *8*(12), 647–654. https://doi.org/10.18415/ijmmu.v8i12.3396.
Magdalena, I., Hifziyah, M., Aeni, V. N., & Rahayu, R. P. (2020). Pengembangan Instrumen Tes Siswa Tingkat Sekolah Dasar di Kabupaten Tangerang. *Nusantara: Jurnal Pendidikan Dan Ilmu Sosial*, *2*(2), 227–237. https://doi.org/10.33506/jq.v8i1.389.
Magdalena, Ina, Syariah, E. N., Mahromiyati, M., & Nurkamilah, S. (2021). Analisis Instrumen Tes Sebagai Alat Evaluasi Pada Mata Pelajaran SBDP Siswa Kelas II SDN Duri Kosambi 06 Pagi. *Jurnal*

Pendidikan Dan Ilmu Sosial, *3*, 276–287. https://ejournal.stitpn.ac.id/index.php/nusantara/article/view/1264.

Magdalena, R., & Angela Krisanti, M. (2019). Analisis Penyebab dan Solusi Rekonsiliasi Finished Goods Menggunakan Hipotesis Statistik dengan Metode Pengujian Independent Sample T-Test di PT.Merck, Tbk. *Jurnal Tekno*, *16*(2), 35–48. https://doi.org/10.33557/jtekno.v16i1.623.

Maharani, N. M. A. P., Ardana, I. K., & Putra, D. K. N. S. (2019). Pengaruh Metode Bercerita Berbantuan Media Gambar Berseri Terhadap Keterampilan Berbicara Anak Kelompok A. *Jurnal Pendidikan Anak Usia Dini Undiksha*, *7*(1), 25. https://doi.org/10.23887/paud.v7i1.18742.

Mahirah. (2017). Evaluasi belajar peserta didik (siswa). *Jurnal Idaarah*, *1*(2), 257–267. https://journal3.uin-alauddin.ac.id/index.php/idaarah/article/view/4269.

Manfaat, B., & Nurhairiyah, S. (2021). Pengembangan Instrumen Tes untuk Mengukur Kemampuan Penalaran Statistik Mahasiswa Tadris Matematika. *Jurnal Pembelajaran Matematika*, *1*(1), 1–19. https://repository.syekhnurjati.ac.id/1668/.

Mappalesye, N., Sari, S. S., & Arafah, K. (2021). *Pengembangan Intrumen Tes Kemampuan Berpikir Kritis Dalam Pembelajaran Fisika*. *1*, 69–83. http://localhost:8080/xmlui/handle/123456789/8261.

Marchalot, A., Dureuil, B., Veber, B., Fellahi, J.-L., Hanouz, J.-L., Dupont, H., Lorne, E., Gerard, J.-L., & Compère, V. (2018). Effectiveness of a blended learning course and flipped classroom in first year anaesthesia training. *Anaesthesia Critical Care & Pain Medicine*, *37*(5), 411–415. https://doi.org/https://doi.org/10.1016/j.accpm.2017.10.008.

Montolalu, C., & Langi, Y. (2018). Pengaruh Pelatihan Dasar Komputer dan Teknologi Informasi bagi Guru-Guru dengan Uji-T Berpasangan (Paired Sample T-Test). *D'CARTESIAN*, *7*(1), 44. https://doi.org/10.35799/dc.7.1.2018.20113.

Muliyani, T., & Huriaty, D. (2016). Pengembangan Instrumen Tes Geometri dan Pengukuran pada Jenjang SMP. *Math Didactic: Jurnal Pendidikan Matematika*, *2*(2), 91–98. https://doi.org/10.33654/math.v2i2.33.

Muttaqin, M. Z., & Kusaeri. (2017). Pengembangan Instrumen Penilaian Tes Tertulis Bentuk Uraian untuk Pembelajaran PAI Berbasis Masalah Materi Fiqh. *Jurnal Pemikiran Dan Penelitian Pendidikan*, *15*(1), 1–23. http://repository.uinsa.ac.id/id/eprint/315/.

Nengsi, A. R., & Efrina, G. (2019). Optimasi validitas dan reliabilitas tes pilihan ganda buatan guru mata pelajaran ips sd. *Journal Innovation in Islamic Education: Challenges and Readiness in Society 5.0, 4th International Conference on Education*, 43–48. https://ojs.iainbatusangkar.ac.id/ojs/index.php/proceedings/article/view/2138.

Nurfillaili, U., Yusuf, M., & Santih, A. (2016). Pengembangan Instrumen Tes Hasil Belajar Kognitif Mata Pelajaran Fisika pada Pokok Bahasan Usaha dan Energi SMA Negeri Khusus Jeneponto Kelas XI Semester I. *Jurnal Pendidikan Fisika*, *4*(2), 83. https://journal3.uin-alauddin.ac.id/index.php/PendidikanFisika/article/download/3709/3382.

Palimbong, S. M., Devi, O., Pompeng, Y., Ekonomi, F., & Kristen, U. (2022). *Pengaruh Penerapan Surat Pemberitahuan Elektronik (e-SPT) Masa Pajak Pertambahan Nilai (PPN) Terhadap Kepatuhan Wajib Pajak*. *2*(2), 475–481. https://doi.org/10.29264/jakt.v19i2.11169.

Perangin-angin, A. B., Sinar, T. S., & Zein, T. T. (2020). Interaksi Guru Dan Siswa Dalam Proses Belajar Mengajar Di SMA Negeri 1 Medan Perspektif Analisis Wacana Krisis van Dijk (1993). *Kode: Jurnal Bahasa*, *9*(2), 16–30. https://doi.org/10.24114/kjb.v9i2.18429.

Prameswari, D. P., & Rahayu, T. S. (2020). Efektivitas Model Pembelajaran Cooperative Learning Tipe Make a Match dan Numbered Head Together: Kajian Meta – Analisis. *Jurnal Ilmiah Pendidikan Profesi Guru*, *3*(1), 202–210. https://doi.org/10.23887/jippg.v3i1.28244.

Purnomo, P., & Maria Sekar Palupi. (2016). Pengembangan Tes Hasil Belajar Matematika Materi Menyelesaikan Masalah yang Berkaitan dengan Waktu, Jarak dan Kecepatan untuk Siswa Kelas V. *Jurnal Penelitian (Edisi Khusus PGSD).*, *20*(2), 151–157. http://e-journal.usd.ac.id/index.php/JP/article/view/872.

Puspasari, H., & Puspita, W. (2022). Uji Validitas dan Reliabilitas Instrumen Penelitian Tingkat Pengetahuan dan Sikap Mahasiswa terhadap Pemilihan Suplemen Kesehatan dalam Menghadapi Covid-19 Validity Test and Reliability Instrument Research Level Knowledge and Attitude of Students Towards. *Jurnal Kesehatan*, *13*, 65–71. https://doi.org/10.26630/jk.v13i1.2814.

Putri, D., & Nahadi. (2019). Perbandingan Reliabilitas Tes Hasil Belajar Matematika Sma Berdasarkan Teknik Penskoran Dan Ukuran Sampel. *Journal Education and Chemistry (JEDCHEM)*, *1*(1), 10–24. https://www.ejournal.uniks.ac.id/index.php/JEDCHEM/article/view/86.

Rahmawati, L., & Hardini, A. T. A. (2020). Pengaruh Model Pembelajaran Inquiry Berbasis Daring terhadap Hasil Belajar dan Keterampilan Berargumen Pada Muatan Pembelajaran IPS di Sekolah Dasar. *Jurnal Basicedu*, *4*(4), 1035–1043. https://doi.org/10.31004/basicedu.v4i4.496.

Riswanda, H., & Burhan, N. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, *8*(1), 160–180. https://doi.org/10.22219/kembara.v8i1.20530.

Sa'diyyah, F. N., Mania, S., & Suharti. (2021). Pengembangan Instrumen Tes untuk Mengukur Kemampuan Berpikir Komputasi Siswa. *Jurnal Pembelajaran Matematika Inovatif*, *4*(1), 17–26. https://doi.org/10.22460/jpmi.v4i1.17-26.

Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis Kualitas Instrumen untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring di Masa Pandemi. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika, Dan Statistika*, *1*(2), 130–139. https://doi.org/10.46306/lb.v1i2.

Setyawan, A., Syarifudin, A., & Akrom, R. (2021). Pengembangan Media Pembelajaran Teks Hikayat Berbasis Ispring untuk Siswa Kelas X SMA. Linguista: Jurnal Ilmiah Bahasa. *Sastra Dan Pembelajaranya*, *5*(2), 142–159. https://doi.org/10.25273/linguista.v5i2.11428

Sholihah, I., Fahrurrozi, & Abdul Gafur, H. (2017). Pengembangan Modul Pembelajaran Ekonomi Berbasis Guided Inquiry untuk Meningkatkan Hasil Belajar Siswa Kelas XI MA NW Sukamulia. *Pembelajaran Ekonomi*, *1*(2), 104–117. http://eprints.hamzanwadi.ac.id/4531/.

Srika Ningsih Pasi; Yusrizal. (2018). Analisis Butir Soal Ujian Bahasa Indonesia Buatan Guru MTSN di Kabupaten Aceh Besar. *Master Bahasa*, *6*(2), 195–202. https://doi.org/10.24173/mb.v6i2.11666

Sudaryono. (2013). *Pengembangan Instrumen Penelitian Pendidikan*. Graha Ilmu.

Sugiono, Noerdjanah, & Wahyu, A. (2020). Uji Validitas dan Reliabilitas alat Ukur SG Posture Evaluation. *Jurnal Keterapian Fisik*, *5* no *1*, 55–61. https://doi.org/10.37341/jkf.v5i1.167.

Sukardi. (2008). Metode Penelitian Pendidikan Kompetensi dan Praktiknya. In *Bumi Aksara*. Bumi Aksara.

Sulistianingsih. (2020). Pengetahuan Guru tentang Konstruksi Tes , Penguasaan Materi Pelajaran Sains dengan Reliabilitas Tes Buatan Guru. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *11*(2), 145–153. https://doi.org/10.37640/jip.v11i2.120.

Susilo, B., & Ernawati, A. (2018). *Pengaruh Model Pembelajaran Kooperatif Tipe Teams Games Tournament (TGT) Terhadap Persepsi Matematika Siswa*. *5*(2), 111–120. https://jurnal.uns.ac.id/jpm/article/view/26036.

Syachlani, A., & Setyorini, D. (2021). Pengembangan Instrumen Hasil Belajar Matematika Siswa (Tes Pilihan Ganda). *Jurnal Akrab Juara*, *6*(3). https://doi.org/10.58487/akrabjuara.v6i3.1523.

Tarigan, H. G. (2008). *Menulis Sebagai Suatu Keterampilan Berbahasa.* Angkasa.

Ulfah, A. A., Kartono, K., & Susilaningsih, E. (2020). Validity of Content and Reliability of Inter-Rater Instruments Assessing Ability of Problem Solving. *Journal of Educational Research and Evaluation*, *9*(1), 1–7. https://doi.org/10.15294/jere.v9i1.40423.

Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student asessment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika*, *7*(1), 57–68. https://doi.org/10.37058/jp3m.v7i1.2069.

Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe Hots Pada Kompetensi Sistem Rem Di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *17*(1), 95. https://doi.org/10.23887/jptk-undiksha.v17i1.22914.

Wenno, I. H., Tuhurima, D., & Manoppo, Y. (2021). How to Create a Good Test. *Jurnal Pendidikan Profesi Guru Indonesia*, *1*(1), 11–20. https://core.ac.uk/download/pdf/386366708.pdf.

Widayanti, W., Bistari, & Suparjan. (2021). Analisis Butir Soal Pilihan Ganda Penilaian Tengah Semester Pada Pembelajaran Tematik Kelas V Sekolah Dasar Negeri 39 Pontianak Kota. *Jurnal DIDIKA : Wahana Ilmiah Pendidikan Dasar*, *7*(2). https://doi.org/10.29408/didika.v7i2.4370.

Widiyawati, Y., Nurwahidah, I., & Sari, D. S. (2019). Pengembangan Instrumen Integrated Science Test Tipe Pilihan Ganda Beralasan Untuk Mengukur HOTS Peserta Didik. *Saintifika*, *21*(2), 1–14. https://core.ac.uk/download/pdf/297204372.pdf.

Wulan, E. R., & Rusdiana, H. A. (2015). *Evaluasi Pembelajaran*. Pustaka Setia.