Development Evaluation Tool *Two Tier Multiple Choice* Using Wondershare Quiz Creator to Identify Mathematical Connection

AUTHORS INFO

Sri Ayu Lestari
Universitas Muhammadiyah Gresik
sriayulestari75@gmail.com

Irwani Zawawi
Universitas Muhammadiyah Gresik
irwanizawawi@umg.ac.id

Fatimatul Khikmiyah
Universitas Muhammadiyah Gresik
fatim@umg.ac.id

Nur Fauziyah
Universitas Muhammadiyah Gresik
nurfauziyah@umg.ac.id

Suggetion for the Citation and Bibliography

Citation in Text:
Lestari, et al. (2021)

Bibliography:

Lestari, SA., Zawawi, I., Khikmiyah, F., & Fauziyah, N. (2021). Development Evaluation Tool Two Tier Multiple Choice Using Wondershare Quiz Creator to Identify Mathematical Connection. *Journal of Mathematics Education.* 6(2). 133-148. http://doi.org/10.31327/jme.v6i2.1607

Abstract

The research objective is to develop an evaluation tool two tier multiple choice using the wondershare quiz creator to identify students mathematical connection capabilities. The research model uses ADDIE. The test subjects in the study were class VII-D of Islamic Qon Middle School. The research instruments used were interview sheets, validation sheets, test instruments, and questionnaires. Data analysis techniques were carried out quantitatively and qualitatively. The results are: (1) the evaluation tool is feasible because it meets 2 valid and effective criteria. Valid criteria based on the assessment of media experts 94% (very feasible), and 80% material experts (feasible). Effective criteria are seen from the response of students and the quality of the items. The results of the students' responses were 82% (very interesting). While the quality of the items, the validity test contained 1 invalid question. Reliability test, there is 1 unreliable question. The difficulty level test has 7 moderate questions, 3 easy questions, and there are no difficult questions. The distinguishing power test contained 2 very good questions, 6 good questions, 2 enough questions, and not in the bad category. There were 11 distractors at the first level and 6 at the second level who were selected from all students. (2) The percentage of mathematical connections is very high 39.3%, high percentage 46.4%, moderate percentage 3.6%, low percentage 3.6%, and very low percentage 7.1%.

A. Introduction

To see the effectiveness of learning carried out by educators, it is necessary to evaluate learning. Evaluation according to Arifin (2013) is a systematic and continuous process, to determine the quality (value and meaning) of something, based on certain considerations and criteria in order to make a decision. In addition to the definition of evaluation, Gronlund in (Astiti, 2017) *"Evaluation is the systematic process of collecting, analyzing and interpreting information to determine the extent to which pupils are achieving instructional objectives".* Specifically, the purpose of evaluation is to provide clarification on the nature of learning outcomes, as well as provide input for learning progress (Ratnawulan, 2014). The evaluation function according to Sukardi (2015) is to find out aspects of the weaknesses of students in carrying out learning activities, to know the level of achievement of students in learning activities, as a means of feedback for an educator that comes from students, as a tool to determine the learning progress of students.

In essence, evaluation is carried out to measure and assess the ability of students, using an evaluation tool. Tools are often referred to as instruments, so evaluation tools are often referred to as evaluation instruments. The instrument is a tool, which is used to collect data, by means of measurement (Firdaos, 2016). According to (Purwanto, 2010), he explains the meaning of an instrument called a measuring instrument, which is used to measure in order to collect data. A good evaluation tool is an evaluation tool that is able to evaluate something, with results in accordance with the conditions being evaluated. In using this tool, the evaluator needs to use a method or technique, which is often called an evaluation technique. Evaluation techniques are divided into two types, namely test techniques and non-test techniques (Sudijono, 2016) . The test technique is a method used in the context of measuring and assessing in the field of education, in the form of giving assignments. While the non-test technique is an evaluation of student learning outcomes that is carried out without testing students, but by conducting systematic observations, conducting interviews, distributing questionnaires, and examining or researching documents. While the evaluation instrument is divided into two kinds, namely, objective tests and subjective tests. The objective test is a test that requires students to choose the answers that have been provided, while the subjective test is a test in the form of an essay that requires answers that are discussion and description of words (Arikunto, 2013).

Even though they have to be under the demands of the 21st century or the era of the industrial revolution 4.0, teachers should implement learning that focuses on 4C (Critical Thinking, Creative, Communicative and Collaborative) (Fauziyah et al., 2021). According to (Asep, 2008) in learning mathematics in America in 1989 there were five kinds of skills that must be developed including, mathematical communication, mathematical connection, mathematical thinking, problem solving, and understanding. concept. Mathematical Connection, is an ability that needs to be developed in studying mathematics. Connection comes from English which means relationship or connection. According to Utari Sumarno in (Listyotami, 2011), connections can be linked internally and externally, in the sense that internal linkages are associated with mathematical concepts, while external relationships are associated with mathematics in everyday life. Mathematical connection is one of the standards set by (NCTM, 2000) which includes problem-solving skills, reasoning, the ability of the connection, and the ability of representation..

While the objectives of learning mathematics contained in the Ministry of Education and Culture (2016) states that mathematics subjects aim to determine the ability of students to understand mathematical concepts, explain the interrelationships between concepts and apply concepts or algorithms, in a flexible, accurate, efficient, and precise manner, in problem solving. Based on TIMSS (Trend in International Mathematics and Science Study) study the 2015, Indonesia is ranked 44th out of 49 countries (Hadi, 2019). While the results of PISA study (Program for International Student Assessment) the 2018, Indonesia is ranked 6th from the bottom or 74th. With an average score of 371, it is below Panama which has a score of 377 from the highest average score of 555 obtained by China (Tohir & Ibrahimy, 2020). Based on these facts, it is evidence that learning mathematics has not emphasized the ability to connect mathematics, which is part of higher order thinking skills. This could be an indication of Indonesia's low ranking in the TIMSS study and the PISA study. If students have low mathematical connections, then students will have difficulty in relating mathematics to

everyday life. Therefore, the ability to connect mathematics needs to be trained from an early age, if students are able to relate mathematical ideas to their understanding, it will last longer because they are able to see the relationship between mathematical topics, with contexts other than mathematics in everyday life. In Standard National Council of Teachers of Mathematics (NCTM, 2000) defines mathematical connection as the relationship between mathematical topics, the relationship between mathematics and other disciplines and the relationship between mathematics and the real world or in everyday life.

One of the problem models that can be used to identify mathematical connection abilities is a multiple choice test equipped with reasons or known as a diagnostic test (Susanti, 2014). Two Tier Multiple Choice (TTMC) is a diagnostic test in the form of multiple choice questions, consisting of two levels, which was first developed by David F. Treagus in 1988. The first level contains questions that will be tested on students, while The second level contains the reasons students choose the first answer. Multiple Choice Tier Two is the development of the usual multiple-choice test, dith shaped test two levels, it is expected educators can measure and know the math skills tested.

The teaching process in the classroom carried out by the teacher is still conventional. There are still many teachers who are not familiar with Information Technology which is commonly used by teachers in urban areas to support the learning process. They cannot upgrade their knowledge where they may face so many obstacles which consequently cause their knowledge to become outdated (Fauziyah & Uchtiawati, 2017). Therefore, to give students a sense of enthusiasm, evaluation tools must be designed as attractive as possible to create effective, interactive, and efficient evaluation tools. In this study, the Wondershare Quiz Creator is used, which is software designed to create an IT-based assessment program or quiz (As'ari, 2017). This is very user friendly meaning, easy to use and does not require a complicated programming language. The results of the questions that have been made can be stored inform standalone flash or can stand alone on the website (Utomo, 2015). In addition, the results of the publication of the questions that have been made can be saved inform swf files, html files, and exe files (Astika, 2017; Farman et al., 2021). With Wondershare Quiz Creator it is easier for educators to make learning evaluations because they do not require a complicated programming language and there are many variations of questions such as true or false, multiple choice, fill in the blank, matching and short essays (Farman, 2020). There are several advantages of Wondershare quiz creator software including: (1) designing questions faster, because users are not required to master the action script; (2) there are many different types of questions; (3) can be changed background, color, font, and others; (4) questions can be published online; (5) questions can be made randomly; (6) examination of the results of student answers can be done quickly; (7) the result of student answers will automatically be sent to the email educator's; (8) users can also set KKM (Minimum Completeness Criteria) for students. In addition to the advantages mentioned above, Wondershare Quiz Creator also has weaknesses including: (1) limited templates; (2) writing cannot be animated with motion; (3) allow damage to occur hardware or software; (4) wondershare quiz creator software is not suitable for use on computers/ laptops/ PC that have low specifications.

There are several studies on two tier multiple choice, one of which is research conducted by (Shidiq, 2014). The difference from this study is that the instruments used are still in conventional form, while in this study using Wondershare quiz creator software. In addition, differences are also seen in the research subjects, in this study carried out at the SMP/MTs level, while Ari Syahidul Shidiq's research was conducted at the SMA/MA level. There is also research on wondershare quiz creators, one of which is research conducted by (Munscfatra, 2017) from that study, there is an increase in the percentage of student responses by 21% from the initial response of the *"interesting"* criteria to *"very interesting"*. Based on the description above, the researcher is interested in conducting research on *"Development of a Evaluation Tool Two Tier Multiple Choice Using Wondershare Quiz Creator Software to Identify Mathematical Connection"*

B. Methodology
1. Type of Research
Type of research used is researchand development. In this study using the ADDIE model with the stages of Analysis, Design, Development, Implementation and Evaluation.
2. Research Subjects
The subjects in this study were students of class VII D of SMP Islamic Qon which were used as development trials at thestage *development.* At the stage of *implementation* researchers took

VII C class subjec*t.* Class VII was chosen because the material that will be used as research is in class VII, and is based on the recommendation of the mathematics teacher.

3. Research Design

At the analysis stage there are several things that researchers do such as performance analysis, student analysis, analysis of facts, concepts & procedures of learning materials, and analysis of learning objectives. After the analysis stage, the next is the *design* stage that it goes through which includes preparing tests, designing media, and producing *draft 1.* After producing *draft 1,* validation is carried out by media and material experts at the development stage, it is said to be valid if the assessment shows an evaluation presentation $> 60\%$ which is classified as feasible or very worthy. If it is not valid, it will be revised and produce *draft 2.* If it is valid then a development trial is carried out, to find out whether it is effective or not, as seen from the quality of the questions and student responses. If it is still not effective, then another trial is carried out to produce effective criteria. After obtaining a valid and effective evaluation tool, the next step is the implementation stage by giving a mathematical connection test based on *Two Tier Multiple Choice,* after that a student response questionnaire is given and data analysis is carried out. So as to produce a valid and effective evaluation tool.

4. Research Procedure

This development model was carried out on social arithmetic material which refers to the ADDIE development model. The steps include: (1) *analysis,* relating to the analysis of the work situation and environment so that it can be found what products need to be developed; (2) *design,* is a product design activity as required; (3) *development,* is the activity of making and testing a product; (4) *implementation,* is the activity of using the product; (5) *evaluation, the* process of evaluating products that have been made according to specifications or not (Sugiyono, 2013).

5. Data Collection Method

In data using methods including: (1) interviews were conducted to conduct a preliminary study to two informants, the mathematics teacher and students; (2) validation questionnaires are given to material experts and media experts, with the aim that the evaluation tool that will be tested can measure what you want to measure, and can display what you want to display. Validation by material experts was filled out by validators of mathematics education lecturers at the University of Muhammadiyah Gresik and teachers of mathematics subjects. While the validation of the material is filled out by someone who is an expert in their field, in this study the researchers chose two lecturers of informatics engineering. The data collection method is done by showing the application wondershare quiz creator along with a validation sheet to the validator for assessment; 3) the evaluation test method two tier multiple choice was developed to identify mathematical connections. This is done by sharing the application inform *swf* through the group WhatsApp and asking to install and work on the evaluation test questions. After students answer all the questions asked by the researcher, the results will be automatically recorded and entered into the email researcher's. (4) the questionnaire method by distributing the link bit.ly google form about the student response questionnaire to the developed evaluation tool.

6. Instruments

There are several instruments used to obtain data in research, including: (1) interview sheets, questions to be asked to subject teachers about the habits of educators in evaluating learning, how the form of evaluation questions are given, applications that are often used in when the implementation is *online,* and of course the character or attitude of the students during the evaluation. While the questions given to students include students' interest in the developed evaluation tool, how to do it if it is in the form of two tier multiple choice; (2) validation sheet, consisting of 2 types of validation, material validation and media validation. Material validation is filled out by 2 validators, mathematics subject teachers and lecturers, the total statements in the material expert validation sheet are 20 statements, which consist of 4 assessment indicators, presentation, content quality, construction, and ease of use. While, media validation consists of 20 statements, which include presentation indicators, content design, display design, and ease of use; (3) the test instrument in this study was in the form of two-level multiple choice. The questions that are used as instruments related to social arithmetic material for class VII students are adjusted to the indicators of students' mathematical connection abilities; (4) questionnaire sheets, given to respondents in the form of questions or statements that must be filled out. Students fill out a questionnaire according to their opinion about the questions being worked on, the presentation of the questions, and the appearance of the evaluation tool.

7. Technique of Data Analysis

    Data analysis techniques were carried out quantitative and qualitative. Qualitative data comes from student interviews. Meanwhile, quantitative data analysis was carried out based on the presentation of the results of product development based on a evaluation tool two-tier multiple choice using wondershare quiz creator software and item analysis, and processed by using statistical tests. In this study, data analysis consists of several stages including:

a. Expert Validation Analysis

    Researcher used as a guide in revising each component in the preparation of an evaluation tool based on two tier multiple choice. There are several steps in the analysis of media validation instruments and expert validation, including:

1) Giving a score on each validation sheet with the following criteria:

Table 1. Eligibility Criteria Evaluation Tool for Validator

Score	Criteria
5	Very Good (SB)
4	Good (B)
3	Enough (C)
2	Less (K)
1	Very Less (SK)

(Sugiyono, 2013)

2) Then the results of the questionnaire will be analyzed by means of

$$P = \frac{f}{N} \times 100\%$$

(Sudijono, 2016)

Description :
P = Percentage
f = Raw score earned
N = Maximum score

3) The last step is to conclude the results of the analysis by looking at the table below:

Table 2. Feasibility Interpretation Percentage Range

Assessment	Criteria
$80\% < x \leq 100\%$	Very Eligible
$60\% < x \leq 80\%$	Eligible
$40\% < x \leq 60\%$	Fairly Eligible
$20\% < x \leq 40\%$	Not Eligible
$0\% \leq x \leq 20\%$	Very Inappropriate

(Purnamasari, 2015)

Based on the table above, the evaluation tool is said to be feasible if the percentage is ≥ 61

b. Mathematical Connection Ability

    According to (Jauhariansyah, 2014) the criteria for scoring on the two-level diagnostic test are as follows:

1) If the answer and reason are correct then the score = 1
2) If the answer is correct and the reason is wrong then the score = 0
3) If the answer is wrong and the reason correct then score = 0
4) If the answer and reason is wrong then score = 0

After being scored, the next step is to calculate the percentage score of each student using the formula:

$$\% \, Score = \frac{\sum score \; earned}{\sum maximum \; score} \times 100$$

(Zulkarnain, n.d.)

Next is to calculate the percentage of students' mathematical connection abilities using the formula:

$$\% \, Mathematical \; Connection = \frac{\sum students \; answer}{\sum all \; questions} \times 100$$

Percentage of students' mathematical connections for each item is qualified as follows:

Table 3. Qualifications of Students' Mathematical Connection

Percentage	Qualifications of Mathematical Connection
81 – 100	Very High
61 – 80,99	High
41 – 60,99	Enough
21 – 40,99	Low
0 – 20,99	Very Low

After knowing the percentage of students' mathematical connections, the average percentage calculation for each item is calculated.

c. Student Responses

To determine student responses to the evaluation tool, students are given a student response questionnaire which refers to the table below:

Table 4. Student Response Score

Score	Answer Choices
5	Strongly Agree
4	Agree
3	Don't Agree
2	Disagree
1	StronglyDisagree

Next, the calculation is carried out on each item using the formula :

$$P = \frac{f}{N} \times 100\%$$

(Sudijono, 2016)

Description :
P = Percentage
f = Raw score earned
N = Maximum score

After obtaining the percentage of student responses to the use of learning evaluation tools, they are grouped into several criteria as follows:

Table 5. Scale of Student Response Criteria

Average Score	Classification
$80\% < x \leq 100$	Very Interesting
$60\% < x \leq 80\%$	Interesting
$40\% < x \leq 60\%$	Less Attractive
$20\% < x \leq 40\%$	Not Interesting
$0\% \leq x \leq 20\%$	Very Unattractive

(Arikunto, 2013)

d. Item Analysis

There are several kinds of item analysis that are carried out including validity testing, reliability testing, difficulty level testing, discriminating power testing, and distractor effectiveness, as follows:

1) Validity Test

How to calculate item validity using the biserial correlation coefficient formula, namely:

$$r_{pbis} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Description :
r_{pbis} = biseral correlation coefficient
M_p = mean score of subjects who answered correctly for the items sought validity
M_t = mean total score
S_t = standard deviation of the total score
p = proportion of students who answered correctly

$$p = \frac{a\ lot\ \ of\ students\ who\ answered\ correctly}{a\ lot\ of\ students}$$

q = proportion of students who answered incorrectly

(Arikunto, 2013)

If $r_{pbis} > r_\alpha$ than the instrument is valid with $r_\alpha = 0{,}30$. The calculation of the validity test will use SPSS 16.

2) Reliability Test

In this study, the reliability test uses the K-R 20 formula, namely:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{s^2 \sum pq}{s^2}\right)$$

Description :
r_{11} = overall test reliability
p = proportion of subjects who answered each question correctly
q = proportion of subjects who answered each question incorrectly
$\sum pq$ = number of multiplication results between p and q
s = standard deviation from the test

If $r_{11} > r \propto$ then the instrument is reliable. With the tabe correlation coefficient $r_\alpha = 0,70$. It can be calculated using SPPSS 16. In addition, the reliability of each item can be said to be reliable, when viewed from *cronbach's alpha if item deleted* <value *cronbach's alpha* (r_{11}).

3) Level of Difficulty

Numbers that show difficult or easy called the index of difficulty. To find out the level of difficulty of each item, it can be calculated using the formula:

$$P = \frac{B}{J_S}$$

(Arikunto, 2013)
Description :
P = Difficulty index
B = Number of students who answered correctly
J_S = The number of all students

After the results are obtained, the next step is to categorize them in the index of difficulty criteria as shown in table 6

Table 6. Difficulty Index Creteria

Difficulty Index	Difficulty Category
0 − 0,30	Hard
0,31 − 0,70	Medium
0,71 − 1,00	Easy

(Arikunto, 2013)

4) Distinguishing Power Test

Distinguishing power is used to improve the quality of each item. To measure the distinguishing power of each item, the formula:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

(Arikunto, 2013)

Description :
D = distinguishing power
B_A = the number of participants in the upper group who answered correctly
B_B = the number of participants in the lower group who answered correctly
J_A = the number of participants in the upper group
J_B = the number of participants in the lower group.

After calculating using the above formula, the results of the distinguishing power will be obtained, and interpreted using the criteria in table 7.

Table 7. Classification of Distinguishing Power Interpretations

Distinguishing Power (D_p)	Criteria
$D_P \leq 0,00$	Very Bad
$0,00 < D_P \leq 0,20$	Bad
$0,20 < D\_P \leq 0,40$	Enough
$0,40 < D_P \leq 0,70$	Good
$0,70 < D_P \leq 1,00$	Very Good

(Hamzah, 2014)

5) Effectiveness of Distractors

Effectiveness of distractors is an answer option but is not an answer key. To calculate the distractor index, the following formula can be used:

$$IP = \frac{P}{(N-B)/(n-1)} \times 100\%$$

(Arifin, 2013)

Description :

IP = distractor index

P = number of students who chose distractors

N = number of students who took the test

B = number of students who answered correctly on each question

n = number of alternative answers (options)

1 = fixed number

After calculating then interpreted using the following criteria :

Table 8. Interpretation Classification of the Detractor Index

Distractor Index Value	Criteria
More than 200%	Very Bad
176% − 200%	Bad
151% − 175%	Enough
126% − 150%	Good
76% − 125%	Very Good

(Arifin, 2013)

If all students answer all correctly, then $IP = 0$ which means the distractor does not work. The distractor is said to be good if it is chosen by at least 5% of all students k.

8. Successful Development of Two Tier Multiple Choice Based TTMC

Evaluation tool is said to be feasible if it can meet 2 criteria, namely valid and effective. Here's an indicator of the validity and effectiveness can be seen in Table 9 :

Table 9. Indikator Validity and effectiveness

Criteria	Collection Method	Research Instrument	Success
Valid	Questionnaire validated by material experts and media experts	Material Expert and media expert validation sheet	It is said to be valid if the assessment of media experts and material experts shows an evaluation presentation $> 60\%$ classified as worthy or very worthy.
Efective	Student reponse questionnaires and the quality of the questions	Student reponse questionnaires and evaluation questions	1. Student responses to the evaluation tool are said to be interesting or very interesting if the results percentage show $> 60\%$. 2. Have goog quality from item, the question is based on the item indicator

The evaluation tool *two tier multiple choice* using the *wondershare quiz creator* can be said to be good, judging from the analysis of each item. Because grain analysis indicators can be seen in Table 10:

Table 10. Item Analysis Indicator

Indicator	Criteria
Validity Test	Have value $r_{hitung} > r_\alpha = 0{,}355$
Reliability Test	Have a degree of reliability more than $r_\alpha = 0{,}70$ or each item of value *cronbach's alpha if item deleted* < value *croncbach's alpha*
Level of Difficulty	Level of difficulty $0{,}31 − 0{,}70$
Distinguishing Power Test	Minimum distinguishing power is sufficient ($DP > 0{,}2$)
Effectiveness of Distractors	*Distractor selected minimum* 5% From alal students.

C. Findings and Discussion

This development research refers to the ADDIE model, which is the model developed by Dick and Carry in 1996. The stages are Analysis, Design, Development, Implementation and Evaluation. These stages are as follows: :

1. Analyze

 This stage begins with a performance analysis by conducting interviews with subject teachers about the problems faced in the implementation oflearning online. Interviews were conducted face-to-face with the mathematics teacher at the Qon Islamic Middle School. From the results of interviews and observations to schools, a basic problem was found, namely, an evaluation tool was needed that could identify students' mathematical connection abilities, which could be used inlearning online.

Next, analyze students with the aim of finding a suitable evaluation tool design, information obtained through interviews with 4 students online via video conference Google Meet with interview subjects taken randomly. From the results of interviews, students have never been given an evaluation of mathematics learning based on two tier multiple choice. Students want an evaluation tool that is interesting, interactive and not boring, which is equipped with pictures and audio.. Then an analysis of facts, concepts, principles and procedures is carried out to obtain relevant material in developing evaluation tools. The research was conducted in class VII even semester on social arithmetic material. In addition, the researcher also compiled the concept of the material used in the test, namely social arithmetic and the concept of application features in the evaluation tool.

The last step in the analysis stage is the analysis of learning objectives. Based on the results of the analysis and discussion with the mathematics teacher, the purpose of conducting a-based evaluation two-tier multiple choice using Wondershare quiz creator software is to identify students' mathematical connection abilities. Meanwhile, the mathematical connection indicators that will be used as a reference in preparing the test item instruments are as follows: (1) Connections among topics with in mathematics; (2) connections with other disciplines (Connection among topics across lessons); (3) Connection with the real world or knowledge of everyday life.

2. Design

At this stage, the first thing to do is the preparation of tests, which are based on indicators of connection ability, KI and KD. The test questions consist of 20 multiple choice questions with details of 10 questions at the first level, and 10 questions at the second level, with a duration of 90 minutes. Then the step of designing the media begins to design the features of the evaluationthat will be developed which consists of a login form, tool instructions for use and the final results of the evaluation tool when used. The following is a display of the evaluation tool used:
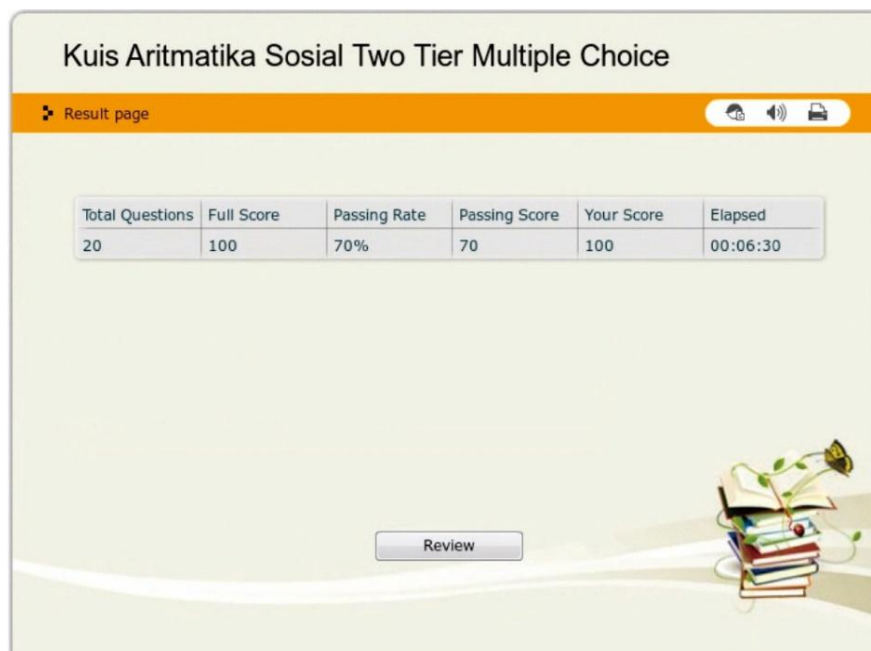

Picture 1. Login Form on the Evaluation Tool

In the login form there are several questions that must be filled out by students, such as full name, absentee number and class. There are several buttons on the login form such as thebutton start to start the quiz, and there are 3 buttons on the top right which include author information, sound, and print.
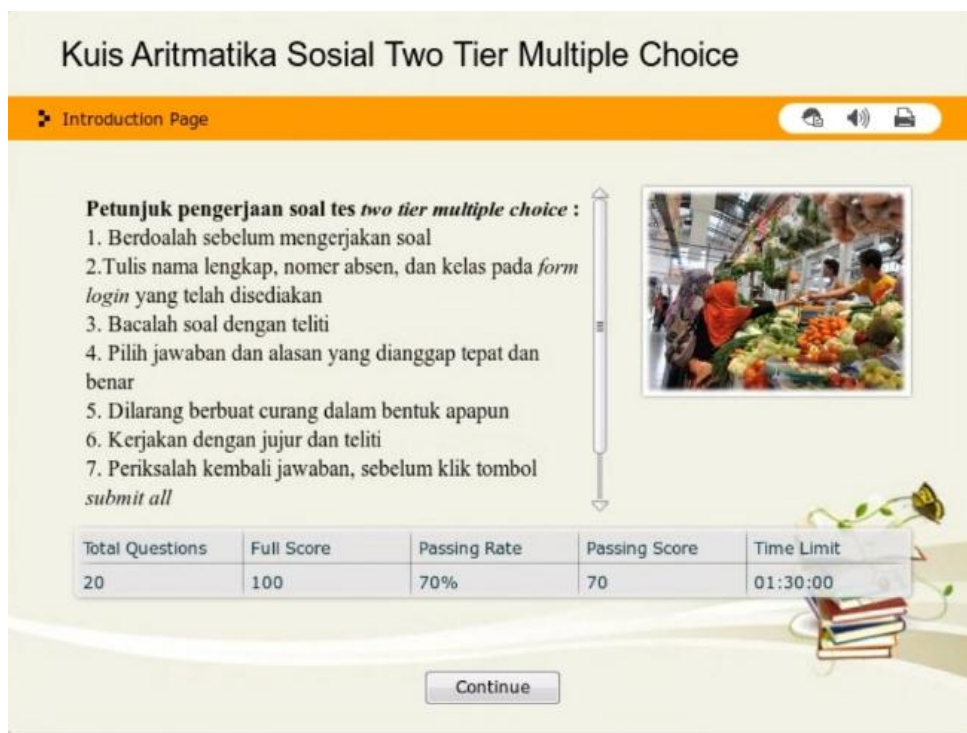
**Picture 2.** Instructions for Use of Evaluation Tools

In Picture 2 there is some information such as instructions for working on two-tier multiple choice, the questionstotal number of questions given is 20 questions with two levels, so a total score of 100 if you can answer all the questions correctly. In addition, there is a passing rate and a passing score, meaning that the minimum completeness limit is 70 and the working time is 1 hour 30 minutes.

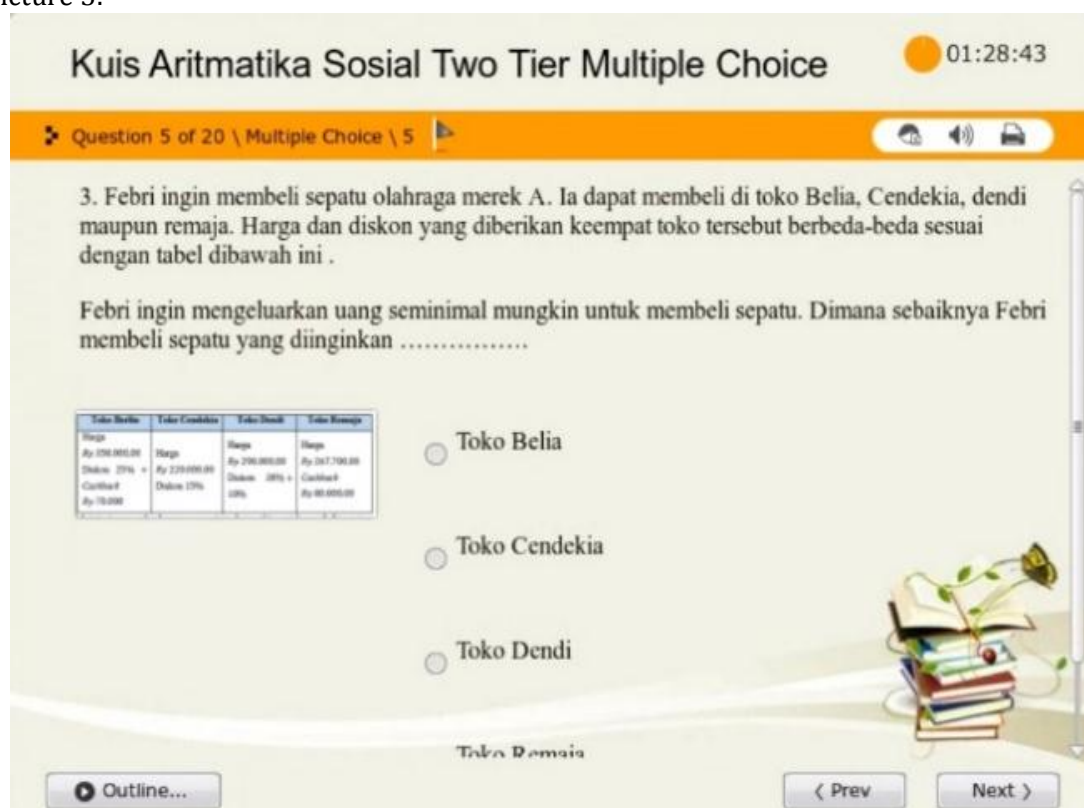**Picture 3.** Final Display of the Evaluation Tool

There is some information obtained in Picture 3 such as the number of questions, the score obtained, the passing rate, the passing score, the score obtained and the length of time students are working on. In addition, there is abutton review that serves to see the right and wrong questions after working.

The last step is the initial design, in this stepis produced thedraft initialof the learning evaluation tool. In making a-based evaluation tool two-tier multiple choice using the Wondershare Quiz Creator, the first step is to make evaluation test questions in Microsoft Word 2013. After the questions and the required pictures have been created, the researchers enter them in the Wondershare Quiz Creator.
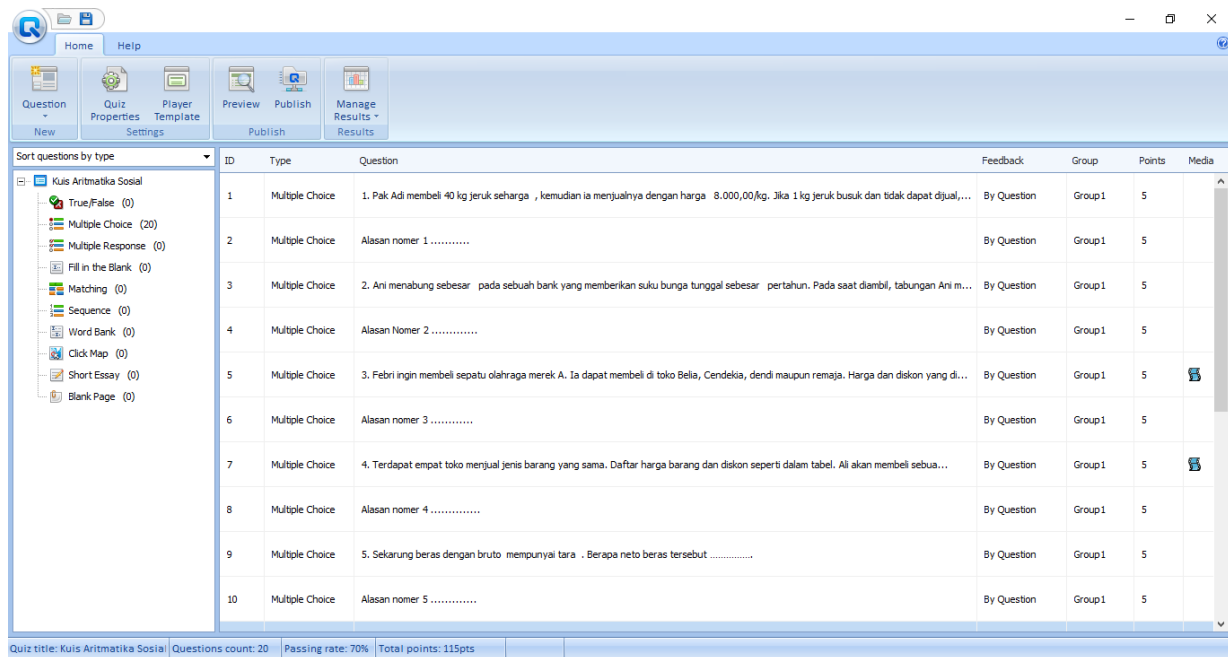
Picture 4. Making an Evaluation Tool Using *WQC*

In making the background design, the researcher did not make it, because the background was already provided in the application. The display of the questions at the first level can be seen in Picture 5:


Picture 5. Display of the First Level Questions for the TTMC Evaluation Tool

If you have finished choosing the answers at the first level, to move to the next question, students can click the button next.. The following view of the second level questions can be seen in Picture 6:

Picture 6. Display of the Second Level Questions TTMC Evaluation

Tool This evaluation tool cannot be opened if you do not use aweb browser supporting, because this file storage type is informat swf. In running the evaluation tool developed, students must download firsttheapplication puf in web browser in the play store as shown in Picture 7:
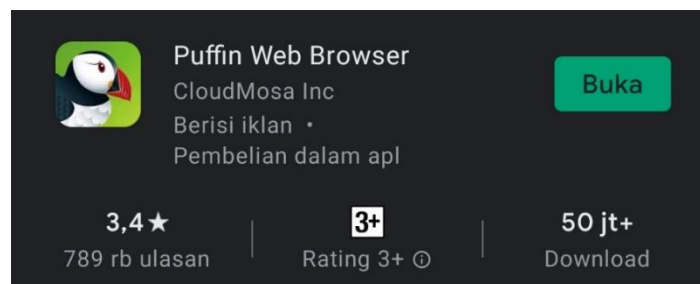
Picture 7. Icons Web Browser Supporting

3. Development

At this stage, expert validation is carried out on the evaluation tool developed. Expert validation consists of media experts and material experts. The name of the validator can be seen in table 1:

Table 11. Name of the expert validator

Validator	Expert Media	1. **Putri Aisyiyah R. D., S.Pd., M.Kom. (Lecturer of Informatics Engineering)**
		2. **Umi Chotijah, S.Kom., M.Kom. (Lecturer of Informatics Engineering)**
	Expert Materi	1. Sri Suryanti, M.Si. (Lecturer of Mathematic Education)
		2. Zaenal Abidin, S.Pd. (Mathematic Teacher in SMP Islamic Qon)

The results of the validation or assessment by experts can be concluded in table 2 below:

Table 12. Conclusion of the validation

	Media Expert	**Material Expert**
Total Score	187	160
Percentage	94%	80%
Qualification	Very Eligible	Very Eligible

From the results of the validation of the experts presented in table 2 above, it can be seen that the-based mathematics learning evaluation tool two-tier multiple choice using the wondershare quiz creator meets the valid criteria. This is because the results of the validation of media experts and material experts get a percentage of $\geq 61\%$, namely 94% and 80% .

Furthermore, a development trial was carried out to determine the quality of the items in the evaluation developed. There are several linkages between the answer criteria chosen by the

students, so that the researcher determines the number of correct answers to each question selected by the students. If students can answer correctly at the first level and answer correctly at the second level, they will get a score of 1. However, if the first level is correct and the second level is wrong, or vice versa then the score is 0. If students answer incorrectly at the first level and answer incorrectly at the first level the second level, the score obtained is 0. This can be seen in Picture 8:

	Nomer 7								Nomer 8							
	Tingkat Pertama				Tingkat Kedua				Tingkat Pertama				Tingkat Kedua			
Alternatif Jawaban	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Frekuensi	22	4	2	2	21	4	3	3	5	22	1	2	9	3	2	17

	Nomer 9								Nomer 10							
	Tingkat Pertama				Tingkat Kedua				Tingkat Pertama				Tingkat Kedua			
Alternatif Jawaban	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Frekuensi	1	28	1	0	25	2	1	2	13	14	2	2	9	14	3	4

Keterangan :

	Merupakan Kunci Jawaban

	Nomer 1								Nomer 2							
	Tingkat Pertama				Tingkat Kedua				Tingkat Pertama				Tingkat Kedua			
Alternatif Jawaban	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Frekuensi	2	1	1	26	27	1	2	0	1	26	2	1	21	1	2	7

	Nomer 3								Nomer 4							
	Tingkat Pertama				Tingkat Kedua				Tingkat Pertama				Tingkat Kedua			
Alternatif Jawaban	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Frekuensi	1	25	2	2	3	3	23	1	1	3	5	21	2	5	3	20

	Nomer 5								Nomer 6							
	Tingkat Pertama				Tingkat Kedua				Tingkat Pertama				Tingkat Kedua			
Alternatif Jawaban	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Frekuensi	19	4	2	5	16	2	6	6	24	4	2	1	3	24	2	1

Picture 8. Students' Answers to each item on the evaluation test

After knowing the students' answers at the first and second levels, the researcher gave a questionnaire for the students' responses to the evaluation tool used. di develop. The results of the student responses are in the appropriate criteria and can be seen in table 3:

Table 13. Student Response Questionnaire Results

No	Assessment Indicator	Percentage Score	Criterion
1	Attractiveness	84%	Very Acttractive
2	Content Quality	80%	Interesting
3	Language	80%	Interesting
4	Ease	84%	Very Acttractive
	Average Percentage	82%	Interesting

In addition, the researcher also analyzed the items in the evaluation test two-tier multiple choice when viewed from the validity, reliability, level of difficulty, discriminatory power and effectiveness of distractors. Where the results of the quality of the items are:

a. Item Validity Question

In the validity test, each question is said to be valid if the validity value is $r_{pbis} > r_\alpha = 0,355$. Based on the results of the validity test using IBM SPSS Stastistic 22, it was found that 1 item was invalid at number 10

b. Reliability

Quality of the item is said to be good if it has a high degree of reliability$(> r_\alpha)$. Based on the reliability test, it was obtained that $r_{hitung} = 0,790 > r_\alpha = 0,70$ then the question is included in the reliable criteria. Reliability test can also be seen on each item if cronbach's alpha if item deleted < cronbach's alpha value . Based on the reliability test on each item, it was found that 1 question was not reliable in item number 2. So that the questions that were used as the implementation stage only amounted to 8 questions.

c. Difficulty

Level In testing the level of difficulty, a good quality item produces a difficulty level of 0.31 0.70 or in the medium category. Based on the analysis of the difficulty level, it was found that the percentage of difficulty level was 70% said to be moderate, 30% was said to be easy, and there were no difficult questions.

d. Distinguishing Power

Quality of the items must have a minimum distinguishing power of sufficient criteria or ($DP > 0,2$). Based on the results of the discriminatory test carried out, it was found that the percentage of sol grains that had $DP > 0,2$ was 100% and there were no questions that had $DP < 0,2$.

e. Distractor Effectiveness

Distractor can function properly if a minimum ofselected 5% of all students are. Based on the results of the distractor effectiveness analysis, there are 11 distractors in the first level that are not functioning properly. While in the second level there are 6 distractors that do not function properly. Meanwhile, for each criterion, the percentage of distractors is very good at 35%, good criteria is 30%, quite good criteria is 26.7%, bad criteria is 5%, and very bad criteria is 3.3%.

4. Implementation

At this stage the researchers mengimplemntasikan evaluation tools VII C class with 28 learners. In qualifying the ability of mathematical connections on each item, the researcher groups each item in three criteria. BB answers mean that students can answer correctly at the first and second levels, BS/SB means that students can answer correctly at the first level and answer incorrectly. at the second level or vice versa. While SS means that students answer incorrectly at the first and second levels. For more details, it can be seen in table 4:

Table 14. Criteria for Student Answers in Each Item

Number Questions	Criteria Answer						Description
	BB		BS/SB		SS		
	Frek	% Koneksi	Frek	% Koneksi	Frek	% Koneksi	
1	25	89,29	2	7,14	1	3,57	Very High
2	24	85,71	2	7,14	2	7,14	Very High
3	22	78,57	2	7,14	4	14,29	High
4	20	71,43	8	28,57	0	0,00	High
5	21	75,00	2	7,14	5	17,86	High
6	14	50,00	6	21,43	8	28,57	Enough
7	17	60,71	7	25,00	4	14,29	Enough
8	18	64,29	6	21,43	4	14,29	High
Total	161	575	35	125	28	100	**Very High**
Average		71,88		15,63		12,50	

There are 224 frequencies spread across each answer criteria (BB, BS/SB, SS). On the BB criteria, 161 out of 224 were found spread across all answer criteria, so that an average percentage of 71.88% was found. In the BS/SB criteria, there are 35 out of 224 spread across all the answer criteria, so the average percentage is 15.63%. While the SS answer criteria amounted to 28 out of 224 spread across all answer criteria, so that an average of 12.50% was obtained. In addition, from the table above, information is obtained that the ability to connect mathematics is very high in questions number 1 and 2, the ability to connect mathematics is high in questions number 3,4,5,8 and the ability to connect mathematics is sufficient in questions number 6 and 7.

If the mathematical connection ability of each student is identified, it is found that the number of students who have very high mathematical connection abilities are 11 students, the high category is 13 students, the sufficient category is 1 student, the low category is 1 student and the very low category is 2 participants.

5. Evaluation

In this stage, a review of the evaluation tools are developed. The results of the evaluation are used to provide feedback on the product being developed. The research carried out is in accordance with the ADDIE development model procedure, which begins with the analysis stage and ends with the evaluation stage. In practice, there are several obstacles experienced by researchers. Like students who have difficulty downloading theapplication puf in web browser, due to signals or not enough space in their storage memory. But it does not reduce the enthusiasm of students in working, to get the best results.

D. Conclusion

From the research on the development of an evaluation tool for learning mathematics based on two tier multiple choice, it can be concluded that 1) Development of an evaluation tool for learning mathematics based on two tier multiple choice using the wondershare quiz creator was developed with the ADDIE stage, namely analysis (related to situation analysis activities). environment), design (product design as needed), development (product manufacturing and testing activities), implementation (product use activities) and evaluation (product assessment process). The evaluation tool for learning mathematics based on two tier multiple choice is said to be feasible because it has met 2 criteria, namely valid and effective. The valid criteria are based on the percentage of media experts' assessment of 94% (very feasible), and material experts of 80% (adequate). Effective criteria can be seen from the responses of students and the quality of the items. From the results of student responses obtained a percentage of 82% (very interesting). While the quality of the items, based on the validity test, there is 1 question that is not valid. The reliability test has 1 question that is not reliable, with a reliability value of 0.790. The difficulty level test consists of 7 questions in the medium category, 3 questions in the easy category, and there is no difficult question category. The discriminatory power test contained 2 questions that were included in the very good criteria, 6 questions in the good criteria, 2 questions in the sufficient criteria, and there was no discriminatory test in the bad category. While the effectiveness of distractors, there are 11 distractors at the first level and 6 distractors at the second level, which are chosen by 5% of all students. 2) The ability of students' mathematical connections identified through a-based evaluation tool two-tier multiple choice using the wondershare quiz creator, the results showed that, there were 11 students who had very high mathematical connection abilities with a percentage of 39.3%, 13 students who had high mathematical connection skills. high connection with a percentage of 46.4%, 1 student has sufficient connection ability with a percentage of 3.6%, 1 student in a low criterion with a percentage of 3.6%, and 2 students in a very low criterion with a percentage of 7.1% .

E. References

Arifin, Z. (2013). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Remaja Rosdakarya.

Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. PT. Bumi Aksara.

As'ari, D. R. (2017). Pemanfaatan Wondershare Quiz Creator Dalam Pembuatan Soal-Soal Bahasa Arab. *Arabi: Journal of Arabic Studies*, 2.

Asep, J. (2008). *Pengembangan Kurikulum Matematika (Tinjauan Teoritis dan Hidtoris)*. Multipressindo.

Astika, R. (2017). *Modul Tutorial Pembuatan Evaluasi Pembelajaran Menggunakan Software Wondershare Quiz Creator.*

Astiti, K. A. (2017). *Evaluasi Pembelajaran*. Penerbit ANDI.

Farman, F. (2020). Effectiveness Of Using The Wondershare Quiz Creator-Based Quiz On Understanding Concept In Calculus I. *Justek : Jurnal Sains dan Teknologi*, *3*(1), 58. https://doi.org/10.31764/justek.v3i1.3817

Farman, F., Sarimuddin, S., Chairuddin, C., & Hali, F. (2021). Computer based problem compilation training using wondershare quiz creator for junior high school teachers. *Community Empowerment*, *6*(5), 833–842. https://doi.org/10.31603/ce.4576

Fauziyah, N., Ma'rifah, U., Asari, S., & L. Maknun, C. (2021). Lesson Study for Learning Community to Support Creative Teachers in Designing Quality Learning: Lesson Study Practices on Bawean Island, Gresik Regency. *Kontribusia (Research Dissemination for Community Development)*, *4*(2), 443. https://doi.org/10.30587/kontribusia.v4i2.2663

Fauziyah, N., & Uchtiawati, S. (2017). Developing a Model of Educators' Professional Training Special for Remote Areas through the Implementation of Lesson Study. *International Education Studies*, *10*(8), 108. https://doi.org/10.5539/ies.v10n8p108

Firdaos, R. (2016). Metode Pengembangan Instrumen Pengukuran Kecerdasan Spiritual. *Edukasi: Jurnal Penelitian Pendidikan Islam*, *Vol.11 No.*

Hadi, S. (2019). *TIMSS INDONESIA (TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY)*. 562–569.

Hamzah, A. (2014). *Evaluasi Pembelajaran Matematika*. PT Raja Grafindo Persada.

Jauhariansyah, S. (2014). Pengembangan dan Penggunaan Tes Diagnostik Pilihan Gnada Dua Tingkat (Two Tier Multiple Choice) Untuk Menggungkap Pemahaman SiswaKelas X Pada Materi Konsep Redoks dan Larutan Elektrolit. *Program Sarjana (Strata 1) Di UN.*

Kemendikbud. (2016). *Permendikbud Nomor 22 Tahun 2016 Tentang Standar Proses Pendidikan Dan Menengah*. Kemendikbud.

Listyotami, M. K. (2011). Upaya Meningkatkan Kemampuan Koneksi Matematika Siswa Kelas VIII A SMPN 15 Yogyakarta Melalaui Model Pembelajaran Learning Cycle '5E'. *Jurnal Skripsi UNY*, *17*.

Munscfatra, N. (2017). Pengembangan Alat Evaluasi Berbentuk Test Online dengan Menggunakan Software Wondershare Quiz Creator Pada Mata Pelajaran Matematika Kelas X Materi Relasi dan Fungsi. *Jurnal Skripsi UIN Raden Intan Lampung*.

NCTM. (2000). *Principles and Standards for School Mathematics*. Tersedia di www.nctm.org.

Purnamasari, A. (2015). Pengembangan Alat Evaluasi Pembelajaran Berbasis Teknologi Informasi dan Komunikasi dengan Wondershare Quiz Creator Materi Sistem Penilaian Persediaan. *Jurnal Pendidikan, Vol 03. No. 01*.

Purwanto. (2010). *Evaluasi Hasil belajar*. Pustaka Pelajar.

Ratnawulan, E. (2014). *Evaluasi Pembelajaran*. Pustaka Setia.

Shidiq, A. S. (2014). Pengembangan Instrumen Penilaian Two-Tier Multiple Choice Untuk Mengukur Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa SMA/MA Kelas XI. *UNS-F. KIP Jur. Pendidikan Matematika Dan Ilmu Pengetahuan Alam-K.3310011-2014*.

Sudijono, A. (2016). *Pengantar Evaluasi Pendidikan*. PT. Raja Grafindo Persada.

Sugiyono. (2013). *Metodelogi Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabet.

Sukardi. (2015). *Evaluasi Pendidikan Prinsip dan Operasionalnnya*. Bumi Aksara.

Susanti, D. (2014). Penyusunan Instrumen Tes Diagnostik Miskonsepsi Fisika SMA Kelas XI Pada Materi Usaha dan Energi. *Jurnal Sarjana Program Pendidikan Fisika FMIPA Universitas Sebelas Maret*.

Tohir, M., & Ibrahimy, U. (2020). *Hasil PISA Indonesia Tahun 2018 Turun Dibanding Tahun 2015*. *January*, 10–12. https://doi.org/10.17605/OSF.IO/8Q9VY

Utomo, D. W. (2015). Pengembangan Sistem Ujian Online Soal Pilihan Ganda dengan Mneggunakan Software Wondershare Quiz Creator. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, *Vol. 04 No.*

Zulkarnain, I. dan N. A. S. (n.d.). Model Penemuan Terbimbing dengan Teknik Mind Mapping Untuk Meningkatkan Kemampuan Pemahaman Konsep Matematis Siswa SMP. *Jurnal Pendidikan Matematika FKIP Universitas Lambung Makurat*.