

## Development of Self-assessment Instruments to Measure Student Affective Domains on Online Biology Learning

Setiyo Prajoko<sup>1\*</sup>, Alfi Anjani<sup>2</sup>, Rezti Dwi Oktaviani<sup>3</sup>, Padmaning Fathimah<sup>4</sup>, Wimba Kamaludin<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup>, Universitas Tidar, Afiliation, Kapten Suparman Street, No 39, Magelang, Jawa Tengah, 56116, Indonesia

\*Correspondence: [setiyo prajoko@untidar.ac.id](mailto:setiyo prajoko@untidar.ac.id)

---

### Abstract

**Keywords:**  
instrument  
development,  
self-assessment,  
affective,  
biology learning,  
online learning

The purpose of this study was to determine the process of developing a self-assessment assessment instrument that measures the affective domain in online biology learning that meets the content and construct validity. The type of research and development that adopted the Plomp development model was used in this study. The subjects of this study were students of class XI MIPA at SMA Negeri Padamara, Purbalingga Regency, Central Java, totalling 40 students in May 2021. The instruments used in this study were self-assessment questionnaires and teacher questionnaires. The validity test results with the calculation of the validity coefficient value show a value of 0.45 so that the assessment instrument can be valid. The reliability test results obtained a coefficient value of 0.957 so that the assessment instrument can be said to be reliable. Furthermore, the results of expert validation got a score of B. These results met the criteria well and could be used with slight revisions. Thus, it can be concluded that the evaluation instrument produced in this development has met content and construct validity.

---

To cite this article:

Prajoko, S., etc. (2021). Development of Self-assessment Instruments to Measure Student Affective Domains on Online Biology Learning. *Thabiea : Journal of Natural Science Teaching*, Vol.4(2), 185-197.

### Introduction

According to Peraturan Pemerintah (Government Regulation) number 19 of 2005, assessment at the level of primary and secondary education consists of a) assessment of learning outcomes by educators; b) assessment of learning outcomes by educational units; c) assessment by the government. Assessment of learning outcomes by educators is carried out continuously to find out the success in the learning outcomes of learners and monitor the process of improving results in the form of daily repeats, midterm repeats, and final replays. Assessment is used to assess learners' competence as material for preparing a report on the progress of learning outcomes (Muslich, 2014). Assessment is an essential component in education. Efforts to improve the quality of education can be achieved through improving the quality of learning and the quality of assessment (Saputra, 2018). The two are interrelated; a sound learning system will result in good quality learning as well. Therefore, efforts to improve the quality of education will not be achieved without improving the quality of learning (Shofiyah, 2013).

Improving the quality of learning requires efforts to optimize processes and learning outcomes specifically overall because the quality of learning is the quality of implementation of previously designed programs (Shofiyah, 2013). The purpose of learning itself includes three domains, namely cognitive, affective and psychomotor. According to Benjamin S. Bloom's theory, these three domains are learning outcomes. The cognitive domain is concerned with knowledge, the affective domain is concerned with attitude, and the psychomotor domain is related to skills (Haris & Jihad, 2013).

The three types of learning outcomes that have been mentioned above, each domain has a definite and measurable measuring instrument or instrument. To measure the learning outcomes of the cognitive domain (knowledge) can use test techniques. As for measuring the learning outcomes of affective and psychomotor domains, we can use non-test techniques. Learning results and processes can be assessed using test techniques in the form of description tests and objective tests and can use non-test techniques. Test and non-test techniques are inseparable parts of measuring cognitive, affective and psychomotor learning outcomes (Sudjana, 2014).

Concerning using two types of learning measurement techniques, teachers in schools generally use more types using this type of test technique than non-test techniques. This happens because the test tool is easier to make, can be used more practically, and is considered only limited to cognitive aspects based on learning results obtained by learners after completing their learning experience in the classroom (Sudjana, 2014). In similar conditions, non-test assessments are still rarely used in assessing teaching and learning results, even though the data of assessment results through these tools is no less meaningful than assessment data through learning outcome tests (Sudaryono, 2012). Whereas teachers only use test techniques, the data collected becomes incomplete and meaningless, even harming certain parties. The use of non-test techniques is one form of criticism of the weaknesses of test techniques (Arifin, 2016). A non-test form or measuring tool is beneficial in evaluating learning outcomes related to personal quality or character development and skills (Sukardi, 2008). These personal qualities and skills can only be evaluated through appearance as the effect of mastering the skill domain in non-tests.

Educators implemented cognitive and psychomotor aspects of the three existing domains, while affective aspects are still not enough to gain attention as in both other aspects. Affective problems are fundamental, but the implementation is still lacking because to design the achievement of affective learning goals is not as easy as cognitive learning. The affective domain must appear in the process, and learning outcomes achieved by learners, therefore, must be assessed the results (Rahmawati et al., 2020; Sukanti, 2011).

The affective learning domain is related to students' feelings, emotions, or responses to their learning experience. Affective behaviour, among others, is indicated by attitude, interest, attention, and awareness (Yulinda, 2014). Therefore, it can be known that affective assessment means assessing attitudes and changes that occur in the behaviour of learners during learning (Sari, 2016). Two things related to affective assessment must be assessed (Arifin, 2016). First, practical competencies to be achieved in learning include response, appreciation, assessment and internalization. Second, the attitude and interest of learners

towards the subjects and learning process. There are four critical affective characteristics in the learning process: attitudes, interests, self-concepts, and values (Sukanti, 2011). It is used as a consideration of affective learning outcomes from learning learners have done that.

Affective learning outcomes consist of several levels, namely receiving, responding, valuing, organization, and characterization by a value complex (Krathwohl, 2002). Receiving is the willingness and sensitivity to pay attention to an activity or object in learning. Responding or responding is active participation to give a reaction to the material provided by the teacher. Valuing means giving value to an object so that there are actions carried out after learning. Organization means comparing the values of learning materials that will then connect and resolve a conflict. Characterization by a value or value complex is the integration of all the value systems that have been owned by learners, which affect personality patterns and behaviour.

In the assessment of affective aspects, some educators conduct assessments in different ways, namely observing learners during the teaching and learning process using observation sheets, observations made by educators by looking at the condition of learners during the lesson, if learners sit quietly and pay attention to be given good grades. At the same time, unsettled students are given less value, and some give affective values equated with positive grades. Notify. This condition indicates that affective assessments are carried out not using actual assessment procedures, so it is necessary to create an effective assessment model that follows the technical instructions of the curriculum (Muslich, 2014). Based on the 2013 curriculum, affective assessment can be done through observation, self-assessment, peer assessment by using instruments in the form of checklists or assessment scales equipped with rubrics, while journals are in the form of notes (Sabdinari, 2019).

The outbreak caused by Coronavirus Disease 2019 (Covid-19) caused many changes, especially in education, both in the world and in Indonesia. Learning that was initially done directly is now starting to adapt using online learning. Online learning is a learning system that is done by not face to face but using a platform that can help the teaching and learning process that is done even though it is remote. Online learning aims to provide quality learning services in a massive and open network to reach more and broader learning space enthusiasts (Sofyana & Rozaq, 2019).

The use of online learning systems is one of the efforts that can be done to overcome problems and make it easier for students to access the subject matter by communicating and discussing online using various platforms. Online learning systems can optimize the interaction between teachers and students through various platforms used in learning, such as Google Classroom, Zoom, WhatsApp, Zenius, and Quizizz. When the implementation of learning must be done online, affective domain assessment still needs to be considered. Affective domain assessment includes feelings, interests, attitudes, emotions or grades as students learn online (Gusti et al., 2020).

One of the affective domain assessments that can be used as an evaluation effort is self-assessment. Self-assessment is an assessment technique in which learners are asked to assess themselves concerning the status, process and level of achievement of competencies they learn in subjects based on specific criteria and references that have been prepared. The

primary purpose of self-assessment is to support or improve learning processes and outcomes. Nevertheless, the results of self-assessment can be used by teachers as a consideration to provide value. The role of self-assessment becomes important along with student-centred learning (Rusilowati, 2013). In addition, self-assessment is considered the easiest to implement as an evaluation instrument in online learning because teachers cannot directly monitor and assess student attitudes during learning.

The use of self-assessment instruments has advantages including 1) can foster the confidence of learners, because they are given the confidence to judge themselves; 2) the learner is aware of his strengths and weaknesses; 3) can encourage, familiarize and train learners to be honest (Rusilowati, 2013).

Biology is a science based on experimentation for development and application, so that it requires students to work to a high standard in conducting experimental. Biological learning requires experiments, descriptions and theories combined and interrelated. During this time, the assessment conducted by teachers is limited to measuring students' abilities on cognitive aspects only. It is still rare for teachers to measure a student's ability on an affective aspect. Cognitive and psychomotor learning outcomes will be optimal if learners have high affective abilities. Therefore, education should be organized by paying better attention to the affective domain. Based on the description above, the researcher intends to know whether the teacher during biology learning applies affective assessment instruments to students by using self-assessment instruments or not so that researchers can develop self-assessment assessment instruments that have been tested for feasibility. Thus, students can conduct self-assessment of disciplined attitudes, honesty, caring, confidence, polite and responsible.

## Method

This research is development research (*Research and Development*) that aims to obtain self-assessment assessment instruments to measure students' affective domain on biological learning. This development model is adapted from the Plomp development model consisting of five phases, namely (1) initial investigation phase, (2) design phase, (3) realization phase, (4) test phase, evaluation and (5) implementation phase. This research was conducted at one of the public high schools in Purbalingga Regency. The test subjects in the study were 40 students of class XI MIPA 2020/2021 Academic Year.

## Research Instruments

The research instruments used in this study are as follows. The first is Preliminary Investigation Data. Preliminary investigative data is obtained by providing Google form-based questionnaires to students guided by the statements provided and providing Google Form-based questions to biology teachers about how assessments have been applied in schools. The second is Validation Sheet instruments. Assessment device to obtain information about the quality of self-assessment by two expert validators. Information that has been obtained from validation results will be used as a reference in revising the assessment device to be developed so that it is suitable for use as an instrument of self-assessment-based affective domain assessment. The third is teacher response questionnaire. A teacher response

questionnaire obtains information or data about teacher opinions of self-assessment tools developed.

**Data Analysis**

The data analysis used in this study is quantitative analysis to explain the validity, reliability and practicality of the effectiveness of the developing affective area assessment device. Analysis of data from affective domain assessment instruments developed as follows.

**Content Validity Analysis**

The validity of the content points to the extent to which the test. It is judging by the contents that measure what is intended to be measured (Suryabrata, 2000). The validity of the content is determined through the opinion of experts, namely lecturers, in the process of reviewing instruments per statement item. The assessment instrument must measure the aspects set in terms of clarity of the measurement objectives formulated the conformity of the statement items for each aspect, the use of language and the clarity of instructions for the use of the instrument (Hobri, 2009). The results of expert validation were analyzed by considering assessments, inputs, comments and suggestions from validators that serve as guidelines for revising products that still get less assessment. The coefficient of validity of the contents is based on Gregory's formula (Ruslan et al., 2020). Here is how to determine the coefficient of content validity.

The formula of Coefficient content validity:

$$\frac{D}{(A + B + C + D)}$$

note:

- A: The number of statement items that derive the overlap value between the weak relevance (items are 1 and 2) of the first Validator to the weak relevance of the second Validator.
- B: The number of statement items that derive the overlap value between the strong relevance (items are 3 and 4) of the first Validator to the weak relevance of the second Validator.
- C: The number of statement items that derive overlap values between the first Validator's weak relevance and the Validator's strong relevance.
- D: The number of statement items that derive an overlap value between the vital relevance of the first Validator to the strong relevance of the second Validator.

**Table 1.** Model of agreement between appraisers for validation determination

<b>The 1<sup>st</sup> Validator</b>	
Less relevant score 1-2	Very relevant score 3-4

<b>The 2<sup>nd</sup> Validator</b>	Less relevant score 1-2	A	B
	Very relevant score 3-4	C	D

Source: (Westerterp, 2009)

To state that an instrument used has a high enough validity, the assessment results of both validators have strong relevance (3 and 4). If the coefficient of content validity is high (>75%), then it can be stated that the measurements are taken in the correct category. Conversely, if the value coefficient does not reach the specified standard, then it is declared invalid, and this needs to be revised again to obtain a correct category (Ruslan, 2009).

The validity of the construct relates to the extent to which the measurement scores with the instrument in question reflect the underlying theoretical construction of the instrument (Suryabrata, 2000). Reliability comes from the word reliable, which means trustworthy. Based on this meaning, the reliable instrument is an instrument whose measurement results can be trusted (Asrul *et al.*, 2019). Reliability is one of the main characteristics or characteristics of a suitable measurement instrument. A test is reliable if it always gives the same result when tested in the same group at different times or occasions (Cohen *et al.*, 2017; Heale & Twycross, 2015). Empirical reliability analysis uses reliability tests by looking at Cronbach Alpha coefficients of data obtained from the trial process using SPSS 22. If the Cronbach Alpha value is above 0.60 and below one, then the instrument has a high correlation or reliability, while if the Cronbach Alpha value is below 0.50, then the instrument has a low correlation or is not reliable (Xu & Lorber, 2014). Determination of reliability coefficients is used Guilford criteria as follows.

**Table 2.** Reliability Coefficient Criteria

Value	Information
0,00 – 0,20	Very Low
0,20 – 0,40	Low
0,40 – 0,60	Keep
0,60 – 0,80	High
0,80 – 1,00	Very High

Source: (Westerterp, 2009)

Differentiating Power, Questions/statements on non-test instruments should be able to distinguish positive and negative attitudes. Different indices are indicators of alignment or consistency between the function of item one and the overall scale in determining the differentiating power on this self-assessment instrument using interval scores.

**Table 3.** Interval Score Category

Answer Value	Scale
121 - 160	VG
91 – 120	G

61 - 90	E
31 - 60	B
0 - 30	VB

**Noted:**

VG: Very Good, G: Good, E: Enough, B: Bad, VB: Very Bad

**Results and Discussion**

Research development of self-assessment instruments to measure students' affective domains in online biology learning using Plomp development models that include 5 phases, namely: (1) initial investigation phase, (2) design phase, (3) realization/construction phase, (4) test phase, evaluation and revision, and (5) implementation phase. The first phase starts from the initial to the final investigation phase, which aims to discover and establish fundamental problems in learners' learning process and assessment. The initial investigation was conducted using instruments in the form of student questionnaires against respondents as many as 40 students of class XI and questionnaires against a biology teacher of class XI. The indicators of self assessment for affective domain presented in Table 4.

**Table 4.** The indicators of self assessment for affective domain

No.	Self-Assessment Indicators for Affective Domain
1	Be scientific, curious, thorough, diligent and honest.
2	Discipline and responsibility in doing assignments and conducting experiments.
3	Dare to ask questions, argue, care for the environment, work together, cooperate, be polite, friendly, critical, responsive and proactive in taking every action.

Based on the indicators, the questions is proposed. The are 20 items of questions of self assessment for affective domain. The next stage is the design phase (planning), a problem-solving activity based on the results of problem analysis in the initial investigation by producing a product to develop an affective domain assessment device based on self-assessment on biological learning. The next phase is the realization phase/ construction, which is the phase where two validators will validate the design until it produces a good design. The next phase is the test phase, evaluating and revising self-assessment instruments to produce valid and reliable instruments. The last stage is implementation. If the instrument is valid and reliable, then the assessment instrument can be used as a suitable assessment device.

Affective domain assessment is generally done by teachers using observation techniques for students; namely, teachers observe the attitude and behaviour of students directly when in the classroom when the lesson takes place. Supporting assessments in the form of self-assessment, friend assessment, or other affective assessments are rarely done by teachers because they will require more preparation. This is shown because the manufacture of affective assessment instruments requires systematic procedures ranging from preparation, implementation of assessments, instrument feasibility tests, and instrument repair (Jatmiko, 2018).

Self-assessment is one of the assessment techniques in which learners are asked to assess themselves related to the status, process and level of achievement of competencies

learned in certain subjects. By assessing themselves, learners can evaluate the advantages and disadvantages in themselves. Self-assessment instruments can measure all aspects so that in their manufacture adjusted to field conditions.

### Validity Analysis

Validity is a measure that indicates the degree of validity or validity of an instrument (Arikunto, 2021). Validity can be divided into three types. The first is the validity of the content, the second is the construction's validity, and the third is the validity of the criteria. This study only covered the validity of the contents and validity of the construct. In contrast, the validity of the criteria or empirical validity was not implemented because there are no instruments that can be used to compare intrusions developed by researchers. The validity of the contents and constructs in this study is as follows.

### Construct Validity

Construct validity is carried out by two validators who are lecturers of the Biology Education Study Program of The Universitas Tidar. The results of expert validation were analyzed by considering the assessment, input, comments and suggestions of the Validator. The results of the analysis are used as guidelines for revising products that still get less assessment. The coefficient of construct validity is calculated based on Gregory's formula.

**Table 5.** Expert validity results by two validators

		Validator 1	
		Less relevant Score 1-2	Very relevant Score 3-4
Validator 2	Less relevant Score 1-2	1	2
	Very relevant Score 3-4	8	9

The table shows the number of items with an irrelevant assessment by both presenters in column A is as much as one statement item. The number of items with irrelevant assessment by the second tester in column B is two statements. The number of items with irrelevant assessment by the first tester in column C is eight statements. At the same time, the number of items with relevant assessments by both presenters in column D is as many as nine statements. The results have been analyzed based on validation results by both examiners and subsequently calculated using Gregory's formula.

In general, the assessment of the self-assessment device on biology subjects from the calculation of the content validity coefficient shows that experts provide assessments that are already considered relevant. After doing the calculation, the value of the content validity coefficient for the first trial is 0.45. So, it can be concluded that the assessment device developed is declared valid with moderate validity criteria.

Self-assessment instruments to measure affective domains in developed biological learning have met the validity of content. This is because, based on the study of experts, this Self-assessment instrument follows affective domain assessment indicators on biological learning.



Based on the validation results from experts that two validators have done against 20 statement items in the instrument, the final result of validity by the first Validator, i.e. with a final value of C, which means the assessment instrument can be used with moderate revision. The value is set because nine statements are less valid, and there are eleven valid statement items. While the final result of validity by the second Validator is with a final value of B, which means the assessment instrument can be used with a slight revision. The value is given because there are three items of less valid statements and 17 valid statement items based on the second validator review. The results show that this assessment instrument is declared relevant and theoretically feasible based on the aspects that have been assessed on each statement item (Krisnawati, 2013).

**Content Validity**

Content validity in this self-assessment includes reliability, discriminatory, and difficulty index tests. The following is a discussion of each of these aspects.

**Reliability Analysis**

Reliability derived from the word reliability means the extent to which the results of a measurement can be trusted, the concept of reliability in the sense of the reliability of the measuring instrument is closely related to the problem of measurement error, the measurement error itself shows the extent to which inconsistencies in measurement results occur when re-measuring the same group of subjects. The quality of the instrument is obtained from the results of expert reviews and tests of question items, and reliability tests (Noviyanti et al., 2014).

Reliability is the consistency of one item with another. Reliability is obtained from the final value of learners from each indicator in the self-assessment-based affective domain assessment. The reliability analysis of the assessment instrument is conducted using SPSS 20 against 20 statement items resulting in the following coefficient values:

**Table 6.** Reliability Test Results

<b>Reliability Statistics</b>	
Cronbach's Alpha	N of Items
.957	20

Reliability is expressed by the reliability coefficient, whose numbers are in the range of 0 to 1.00. The higher the correlation coefficient of each item obtained, the higher the instrument's reliability level (Azwar, 1999). Based on the calculation of the self-assessment instrument coefficient in table 2.2 to 20 statement items in the instrument produces a Cronbach Alpha coefficient value of 0.957. So it can be concluded that the assessment developed can be declared reliable on very high-reliability criteria. This criterion is based on the reliability coefficient criteria according to Guilford, which is 0.80 - 1.00.

**Differentiating Power**

Determining the differentiating power in non-test instruments, not distinguishing between competent students (brilliant) and low-skilled students, but distinguishing positive

and negative students using interval score formulas. The use of interval scores is based on the data results of the trial of 40 students.

Based on calculations using interval scores against 20 statement items in the form of interval scales, the differentiating power of self-assessment *instruments* is 13 statement items (65%) of the difference in the Category of Excellent (SB). This category is obtained based on the calculation of other power and looks at the coefficient in table 1.3, which shows the Excellent category has a range of values of 121 - 160. In addition, there are 7 statement items (35%) of the difference in the Good category (B). This category is indicated in the range of values 91 – 120 to be expressed in the Good category. The acquisition of different power scores on these non-test instruments does not distinguish students who are able or clever and students who are unable or less clever (Darini et al., 2013). In this case, it means that the statement item has distinguished positive and negative attitudes in learners or respondents or other words, each question item has been known either based on the different power index. In addition, the different power index is also an indicator of alignment or consistency between item functions and the overall scale (Rusilowati, 2013).

### Difficulty Index

The difficulty level is the opportunity to answer a problem at a certain level of ability correctly, usually expressed in an index. This difficulty level index is generally expressed in proportions of magnitude 0.00-1.00 (Aiken, 2009). The self-assessment instrument is a non-test instrument that does not assess the right or wrong answer from respondents, so in the research and development of self-assessment instruments, there is no need to measure the level of difficulty in test instruments (Rusilowati, 2013). The difficulty level function of the problem item is usually associated with the test's purpose (Kadir, 2015). For example, for semester exams used problem items that have a moderate difficulty level, for selection purposes are used pesky items that have a high or difficult difficulty level, while for diagnostic purposes are usually used problem items that have a low or easy difficulty level.

Based on *the* self-assessment instrument developed, the teacher gave a positive response. Teachers find it easier to assess a student's affective domain as learning progresses. In this case, teachers get ease in assessing using self-assessment in online learning. This follows the principle of self-assessment that supports the teaching and learning process from various perspectives and can facilitate teachers in managing assessments (Barana et al., 2018). With the use of self-assessment *instruments*, teachers better understand the needs of students in learning, the shortcomings and advantages of students, the abilities of students and make it easier for teachers to recognize the character of students even though they do not directly observe.

Based on the research results, it can be concluded that the main thing related to the development of self-assessment instruments in class XI biology learning is that the instruments developed can be said to be valid and reliable so that the assessment instruments developed can be used by biology subject teachers. Test validity was used using Gregory's formula, which gains a value of 0.45. So that the self-assessment instrument can be valid, the reliability test results using SPSS 20 show a coefficient value of 0.957 so that the assessment instrument can be reliable. While the differentiating power by using interval scores, there are

13 statement items with excellent determination and 7 statement items with good categories. Thus, it can be concluded that the evaluation instruments produced in this development have met the validity of the contents and constructs. The development of self-assessment instruments makes it easier for teachers to assess students' affective domain in learning. This follows the principle of self-assessment that supports the teaching and learning process from various perspectives and can facilitate teachers in managing assessments.

### Acknowledgement

Special thanks to Mrs. Ika Sukmawati, M.Pd. and M. Radian N.A., M.Pd., who has agreed to become an expert validator. Their contribution is essential in this research.

### References

- Aiken, L. R. (2009). *Psychological testing and assessment*. Pearson Education India.
- Arifin, Z. (2016). Evaluasi Pembelajaran (Prinsip, Teknik, dan Prosedur), Cet. Ke-4, (Bandung: Remaja Rosdakarya, 2012).
- Arikunto, S. (2021). *Dasar-Dasar Evaluasi Pendidikan Edisi 3*. Bumi Aksara.
- Asrul, A., Nasution, I., Karima, M. K., & Juliani, J. (2019). *Evaluasi Program Pembinaan Guru*.
- Azwar, S. (1999). *Penyusunan skala psikologi*. Pustaka pelajar.
- Barana, A., Marchisio, M., & Sacchet, M. (2018). Advantages of using automatic formative assessment for learning mathematics. *International Conference on Technology Enhanced Assessment*, 180–198.
- Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (pp. 245–284). Routledge.
- Darini, N. M., Martha, N., & Artawan, G. (2013). Analisis Kualitas Tes Sumatif Bahasa Indonesia Kelas VIII Smpn8 Denpasar Semester Genap Tahun 2012. *Jurnal Pendidikan Dan Pembelajaran Bahasa Indonesia*, 2.
- Gusti, A. R., Afriansari, Y., & Walid, A. (2020). Penilaian Afektif Pembelajaran Daring IPA Terpadu dengan Menggunakan Media Whatsapp. *Diffraction*, 2(2), 65–73.
- Haris, A., & Jihad, A. (2013). *Evaluasi pembelajaran: Yogyakarta: Multi Pressindo. Achmad Rifa'I Dan Chatarina Tri Anni. 2009, Psikologi*.
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66–67.
- Hobri, H. (2009). Metode Penelitian Pengembangan (Developmental Research) Aplikasi pada Penelitian Pendidikan Matematika. *Jember: Word Editor*.
- Jatmiko, A. (2018). Pengembangan Instrumen Penilaian Autentik Kurikulum 2013 Aspek Afektif dalam Mata Pelajaran PAI Kelas VII di SMPN 3 Kalasan. *Edulab: Majalah Ilmiah Laboratorium Pendidikan*, 3(2), 73–92.
- Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *Al-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, 8(2), 70–81.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218.
- Krisnawati, Y. D. (2013). *Pengembangan Instrumen Penilaian Domain Afektif yang*

*Berkualitas pada Mata Pelajaran Geografi Kelas X di SMA N 1 Boja Kabupaten Kendal Tahun Ajaran 2012/2013.* Universitas Negeri Semarang.

- Muslich, K. M. (2014). Pengembangan Model Assessment Afektif Berbasis Self Assessment dan Peer Assessment di SMA Negeri 1. *Jurnal Kebijakan Dan Pengembangan Pendidikan*, 2(2).
- Noviyanti, L., Indriyanti, D. R., & Ngabekti, S. (2014). Pengembangan instrumen self dan peer assessment berbasis literasi sains di tingkat SMA. *Lembaran Ilmu Kependidikan*, 43(1), 32–39.
- Rahmawati, R. F., Imaduddin, M., Haqiqi, A. K., Fikri, A. A., Fawaida, U., Prasetyo, D. R., & Faikhamta, C. (2020). Assessing Psychosocial Outdoor Learning Environment of Pre-service Science Teachers through The Field Trip Experiences. *Participatory Educational Research*, 7(2), 135–150.
- Rusilowati, A. (2013). Pengembangan Instrumen Nontes. *Makalah. Seminar Nasional Evaluasi Pendidikan Di Universitas Negeri Semarang*, 7–21.
- Ruslan, R., HR, I. S., & Hindi, A. N. A. M. (2020). The Development of Reasonable Mathematical Problems Based on Content Validity in Mathematics Teachers Conference (MGMP) of Bone Regency, Indonesia. *Mattawang: Jurnal Pengabdian Masyarakat*, 1(1), 15–19.
- Sabdinar, T. (2019). *Pengembangan Instrumen Penilaian Domain Afektif Berbasis Self Assesment Pada Mata Pelajaran Biologi Kelas X Di SMA Negeri 5 Bulukumba.* Universitas Negeri Makassar.
- Saputra, A. A. (2018). *Pengembangan Instrumen Penilaian Domain Afektif Berbasis Self Assessment pada Mata Pelajaran Biologi Kelas X di SMA Negeri 16 Bulukumba.* Universitas Islam Negeri Alauddin Makassar.
- Sari, N. I. (2016). *Penilaian Afektif dan Psikomotorik dalam Pembelajaran Sejarah di SMA Negeri Se Kabupaten Kendal.* Universitas Negeri Semarang.
- Shofiyah, H. (2013). Penerapan Self Assesment (Penilaian Diri) Pada Kegiatan Praktikum Untuk Meningkatkan Hasil Belajar Siswa Kelas X Sman 1 Sidayu. *Inovasi Pendidikan Fisika*, 2(3).
- Sofyana, L., & Rozaq, A. (2019). Pembelajaran Daring Kombinasi Berbasis Whatsapp Pada Kelas Karyawan Prodi Teknik Informatika Universitas PGRI Madiun. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 8(1), 81–86.
- Sudaryono, D. E. P. (2012). Cet. I. Yogyakarta: Graha Ilmu.
- Sudjana, N. (2014). *Penilaian hasil proses belajar mengajar.*
- Sugiyono, S. (1999). (1999). *Statistika Untuk Penelitian.* Bandung: Alfabeta. Bandung: Alfabeta.
- Sukanti, S. (2011). Penilaian Afektif Dalam Pembelajaran Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 9(1).
- Sukardi, H. M. (2008). Evaluasi pendidikan prinsip dan operasionalnya. Jakarta: Bumi Aksara.
- Suryabrata, S. (2000). Pengembangan alat ukur psikologis. Yogyakarta: Penerbit Andi.
- Westerterp, K. R. (2009). Assessment of physical activity: a critical appraisal. *European Journal of Applied Physiology*, 105(6), 823–828.
- Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation

of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology*, 82(6), 1219.

Yulinda, R. (2014). Kemampuan Mahasiswa dalam Menganalisis Materi Ipa SMP Berdasarkan Taksonomi Kognitif, Psikomotor Dan AfektIF. *Vidya Karya*, 33(2), 181–187.