ANALYSIS OF INSTRUMENT VALIDITY ASSESSMENT ITEMS FOR STUDENT LEARNING ACHIEVEMENT

Muh Syauqi Malik
Fakultas Ilmu Tarbiyah dan Keguruan, UIN Sunan Kalijaga Yogyakarta
19204080006@student.uin-suka.ac.id

Maya Setia Priyadi
Fakultas Ilmu Tarbiyah dan Keguruan, UIN Sunan Kalijaga Yogyakarta
19204080016@student.uin-suka.ac.id

Abstract

This article aims to determine the analysis of instrument validity assessment items. The formulation in this article is to explain the concept of validity, instruments, as well as a detailed discussion of the feasibility analysis of the valuation instruments. The notion of validity is a concept related to the extent to which the test has measured what must be measured. To analyze instruments to evaluate learning outcomes in the form of test items, it can be done in 2 ways, namely qualitative analysis of questions and quantitative analysis of questions. This instrument is a benchmark for student learning achievement. The assessment instruments used by educational institutions in the form of school exams or madrassas must meet the requirements of substance, construction, language, and have evidence of empirical validity. While the analysis of the validity of the quantitative test includes analysis of totality validity which includes content validity, construct validity, forecast validity, and comparison validity, and item validity analysis. Researchers used data analysis techniques using instrument analysis sheets, manual calculations, Microsoft Excel applications, and SPSS applications (Statistical Product and Service Solution). This article is presented systematically and is accompanied by supporting instructions and images so that readers can understand it more quickly and more easily.
Keywords: *Validity, Instruments, Assessment*

INTRODUCTION

The basic principle of the problem in assessment is to conclude whether the test can measure what should be measured or not, as well as conclusions about whether this test can be used correctly for decision making or not. This requires in-depth testing to know the results. Instrument testing begins through the process of testing the suitability of the respondent's instruments. Assessment is not only the collection of student data but also processing it to obtain an overview of the process and learning outcomes of students. Assessment does not just give students questions and then finish, but the teacher must follow up on it for the sake of learning. To

carry out the assessment, the teacher needs an assessment instrument in the form of questions both to test cognitive, affective, and psychomotor abilities.[1]

In addition to improving the learning process that is being implemented, improving the quality of learning can be done through improvements to the assessment used to measure student learning outcomes. Improvement in management can be done by providing assistance and training in performing services. Assessments are used to obtain, synthesize, and interpret information obtained from students to make decisions about students in the classroom.[2]

The assessment has an important role in learning activities. Assessment can help students, the extent to which they have succeeded in following the lessons given by the teacher, and assessment helps teachers to be able to find out which students are having difficulty and successfully mastering the material. Also, with the right and efficient instrument assessment, it will be easy to find out the causes of difficulties experienced by students and easier to find ways to overcome these difficulties. There are 4 roles as an organization, namely, assessment is a tool for diagnosing student strengths and weaknesses, the assessment that is useful for monitoring student progress, the assessment helps determine the level of students, can also determine the effectiveness of learning that has been designed. The more quality the learning assessment is, the better the teacher's understanding of the students' weaknesses and strengths in learning material.[3]

The results obtained can be used to consider whether the instrument used is good or not. An instrument must be measured whether the instrument can measure accurately what will be measured, it is necessary to test the validity of the assessment instrument. Therefore, this article explains in detail the concept of validity as well as the steps in analyzing the validity of valuation instruments.

---

[1] Agus Budiman dan Jailani Jailani, "Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1," *Jurnal Riset Pendidikan Matematika* 1, no. 2 (2014): 139–151.

[2] Khoirul Bashooir dan Supahar Supahar, "Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM," *Jurnal Penelitian dan Evaluasi Pendidikan* 22, no. 2 (2018): 219–230.

[3] Aloysius Rabata Nova TFT dan Parno Parno, "Pengembangan Instrumen Asesmen Penguasaan Konsep Tes Testlet Pada Materi Suhu dan Kalor," *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan* 1, no. 6 (2016): 1197–1203.

METHODOLOGY

The methodology used in this research is the literature study. The author summarizes solutions from various sources to serve as a clearer and more comprehensive reference. Researchers used data analysis techniques using instrument analysis sheets, manual calculations, Microsoft Excel applications, and SPSS applications (Statistical Product and Service Solution). This article is presented systematically and is accompanied by supporting instructions and images so that readers can understand it more quickly and more easily.

DISCUSSION

Educational evaluation is the control, determination, and guarantee of the quality of education of the education system in every path, level, and type of accountability for the administration of education. [4] One form of evaluation is the test technique. Permendikbud No. 66/2013 challenges the Standard of Assessment conducted comprehensively to assess, starting from the process to the learning outcomes. An authentic assessment includes the realm of attitudes, skills, and knowledge.[5] The instruments commonly used in assessing learning outcomes on cognitive aspects with test techniques are items.

The test is one of the tools to take the measurement, namely a tool to collect information on the characteristics of an object.[6] A test used in the assessment process must be of good quality and the test can measure the student's actual ability[7]. Analysis of test quality is a step that must be taken to determine the degree of quality of a test[8]. According to Anastasi, the main purpose of analyzing the items

---

[4] "Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional," 2003.

[5] Umi Salamah. *Penjaminan Mutu Penilaian Pendidikan.* Jurnal Managemen Pendidikan Islam, Vol 2. No.1. Maret 2013

[6] Pardimin. "Analisis Butir Soal Tes Pemecahan Masalah Matematika". Jurnal Wacana Akademika, Vol 1, No 1. 2017.

[7] Nila Hayati dan Djemari Mardapi. "Pengembangan Butir Soal Matematika di SD di Kabupaten Lombok Timur sebagai Upaya dalam Pengadaan Bank Soal". *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran, Vol 4 No 1.* 2014.

[8] Ata Nayla Amalia. "Analisis Butir Soal Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia Vol X No 1. 2012.*

in a test Made by the teacher is to identify deficiencies in the test or learning.[9] According to Sudijono, tests in measuring and assessing the field of Education are in the form of giving assignments that must be done by students so that they can find out student achievement.[10] However, before the instrument is used, the evaluator must know the validity of the instrument. This is done to see whether the instrument is truly capable of measuring student learning outcomes. Or instruments used only to measure a student's memory or language skills. So it can be said that the instrument is the basis for decision making that illustrates student learning outcomes.

The instrument to be used was first analyzed. The analysis is carried out to describe the information whether the goods meet the criteria or requirements as a high-quality evaluation tool, both in terms of substance, language, or construction requirements. As stated in Permendiknas Number 20 of 2007 concerning Assessment Standards:

> *"The learning outcome assessment instrument used by educators meets the requirements (a) substance, is to represent the competence being assessed, (b) construction, is to meet the technical requirements following the form of the instrument used, (c) language, is to use language that is good and correct and communicative according to the level of development of students. The assessment instrument used by education in the form of school exams meets the requirements of substance, construction, language, and has evidence of empirical validity. The assessment instrument used by the government in the form of the UN meets the requirements of substance, construction, language, and has evidence of empirical validity and produces scores that can be compared between schools, between regions, and between years".*[11]

Based on Permendiknas No. 20/2007 on Assessment Standards it can be concluded that it is important to pay attention to the requirements of learning outcomes assessment instruments including materials, construction, and language. To analyze instruments to evaluate learning outcomes in the form of test items to find out how far the test has measured what needs to be measured is called

---

[9] Riza Yonisa Kurniawan, dkk. "Pemberian Pelatihan Analisis Butir Soal Bagi Guru di Kabupaten Jombang: Efektif?. *Jurnal Pemberdayaan Masyarakat Madani (JPMM), Vol 1 No 2. 2017.*

[10] Tutut Kurniawan. "Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar. Journal of Elementary Education, Vol 4 No 1 2015.

[11] Menteri Pendidikan Nasional, "Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian" (Menteri Pendidikan Nasional, 2007).

validity[12], it can be done in 2 ways, namely qualitative analysis of questions, and quantitative analysis of questions. A test can be said to be good as a measuring tool if it meets 3 requirements, namely valid, reliable, has good distinguishing power, and level of difficulty[13]. Test analysis is one of the activities in the context of building a test to get an idea of the quality of the test, both the overall quality of the test and the quality of each item. The analysis is done after the test is compiled and tested on all subjects and the results become feedback to improve the quality of the test concerned.[14]

1. Qualitative Test Analysis

    The test is a systematic procedure that is made in the form of standardized tasks and given to individuals or groups to be done, answered, or responded to, both in written, oral, and action forms.[15] Analysis of qualitative tests can be done before or after the trial. The steps are to look at the items that have been determined by paying attention to aspects of substance, construction, and language.

2. Quantitative Problem Analysis

    Empirical validity means loading experience. Validity is said to be empirical if it has been tested in terms of experience. For example, a student is said to be creative when viewed from his experiences while in college or outside of college if students always produce new ideas that have never existed before. Empirical test analysis or also called quantitatively is intended to investigate the level of validity, reliability, distinguishing features, and the level of difficulty of questions.[16]

    a. Validity test

---

[12] Sumarna Surapranata, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004* (Bandung: PT. Remaja Rosdakarya Offset, 2009).page. 165.

[13] Murbangun Nuswowati. "Pengaruh Validitas dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Kimia terhadap Pencapaian Kompetensi". *Jurnal Inovasi Pendidikan Kimia, Vol 4 No 1. 2010*

[14] Mahmud Alpusari. *Analisis Butir Soal Konsep Dasar IPA 1 Melalui Penggunaan Program Komputer Anates Versi 4.0 For Windows.* Jurnal Primary Program Studi PGSD FKIP Universitas Riau. Vol 3s. No 2. Oktober 2014

[15] Zulkifli matondang. "Validitas dan Reliabilitas suatu Instrumen Penelitian". Jurnal Tabularasa. Vol 6. No 1. Juni 2009.page.87

[16] Sukiman, *Pengembangan Sistem Evaluasi* (Yogyakarta: Insan Madani, 2011).page. 175.

Validity is the level of confidence and validity of the measuring instrument used to measure the test, the instrument can be labeled valid if it shows reliable data so that validity must be able to measure what must be measured.[17] The validity of the test must always be related to specific goals or decision making.[18]

1) Analyze overall test validity

Analysis of the validity of the test in totality is a test analysis as a whole, for example, there is a test consisting of 30 questions, then 30 questions that need to be analyzed as a whole. From this statement, it can be concluded that the level of the totality of test validity can be divided into two categories, namely the validity of considerations through rational analysis and empirical data analysis. The type of validity of rational analysis consists of content validity and construct validity. While the validity of empirical data consists of the validity of predictions and the validity of comparisons.

a) Content validity

According to Sekaran, content Validity is a function of how well the dimensions and elements of a concept have been described[19]. Content validity means that the content assessment instrument is following the competencies taught both in terms of facts, concepts, principles, and procedures contained in KD learning.[20] In other words, content validity states whether the test includes a representative sample of the domains of behavior

---

[17] Idrus Alwi, "Kriteria Empirik dalam Menentukan Ukuran Sampel Pada Pengujian Hipotesis Statistika dan Analisis Butir," *Formatif: Jurnal Ilmiah Pendidikan MIPA* 2, no. 2 (5 Agustus 2015), https://doi.org/10.30998/formatif.v2i2.95.

[18] Zulkifli matondang. "Validitas dan Reliabilitas suatu Instrumen Penelitian". Jurnal Tabularasa. Vol 6. No 1. Juni 2009.page.89

[19] Hendryadi. "Validitas Isi: Tahap Awal Pengembangan Kuisioner. *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT Vol 2 No 2. 2017.*

[20] Asep Ediana, *Evaluasi Pembelajaran di SD dan MI* (Bandung: PT. Remaja Rosdakarya Offset, 2018).page. 150.

measured. Before compiling the test, compile the grid of questions first, then write the question item. The minimum question box contains the aspects of Basic Competence, description of materials, indicators, and the number of questions from each indicator. Lattice is examined in more detail before being used as a guideline for the preparation of goods, after being declared good it can only be used.

b) Build validity

The construct of validity refers to the extent to which the instrument can measure the ideas contained in the material to be measured.[21] The construct of validity discusses the points in the test whether it is following the achievement of the competencies needed in the curriculum or not. For example, "students can explain the meaning of doomsday", then the formulation of the questions contained in the test must be able to demonstrate the ability of their level of understanding as well.

Analysis of the construct validity of the test was carried out by matching the thinking skills contained in each item with the thinking skills listed in each indicator formulation to be measured. In other words, construct validity can be done rationally, by thinking critically or using logic. The construct validity can be done with a computer program, using factor analysis. If using this method, the instrument must be tested first, then the data will be analyzed using a computer.[22]

Construct validity can be divided into two, namely convergent validity, which refers to the level of concordance between the measurement results of measurement instruments and theoretical concepts that explain the existence of attributes of these

[21] Hamzah Uno dan Satria Koni, *Assessment Pembelajaran* (Jakarta: Bumi Aksara, 2013).page. 81-82.

[22] Yusrizal Yusrizal, "Pengembangan Instrumen Penilaian Kinerja Dosen Di Perguruan Tinggi," *PARAMETER: Jurnal Pendidikan Universitas Negeri Jakarta* 29, no. 1 (30 Mei 2017): 108–20, https://doi.org/10.21009/parameter.291.10.

variables, and discriminant validity refers to mismatches between attributes that should not be measured by measuring instruments and theoretical concept about this variable.[23]

c) Prediction validity

The term "estimate" if associated with validity, what is meant by predictive validity is a condition that indicates the extent to which a test has accurately demonstrated its ability to predict what will happen in the future.[24] For example, the entrance test for junior high school (SMP) is a test that is expected to predict the success of test-takers in taking lessons in the future. Candidates who are screened in exams are expected to reflect high and low abilities to take lessons in junior high school. If the test scores are high, it will guarantee its success later. And vice versa. So it can be concluded that the junior high school entrance test serves to predict the ability of prospective students.

Test the validity of the forecast can be done by connecting test scores with predetermined criteria. For example the value of learning outcomes in the first semester. It is said that a junior high school entrance test has good predictability validity if it has a directional alignment between the test whose validity is tested and the criteria for a direct relationship between the test tested for validity and the specified criteria. The simplest method most often used is applying the product-moment correlation analysis technique from Karl Pearson, namely: (a) the null hypothesis (H0) to be tested; (B) there is no significant positive correlation between

---

[23] Bhisma Murti. *Validita*s *dan Reliabilita*s *Pengukuran.* Fakultas Kedokteran Universitas sebelas Maret Surakarta, 2011.

[24] Mujianto Solichin, "Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan," *Dirasat: Jurnal Manajemen dan Pendidikan Islam* 2, no. 2 (30 Juni 2017): 192–213.
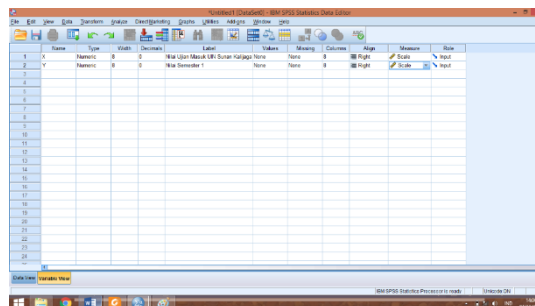
the learning outcomes test tested for predictive validity (variable X); (c) predetermined criteria (variable Y).[25]
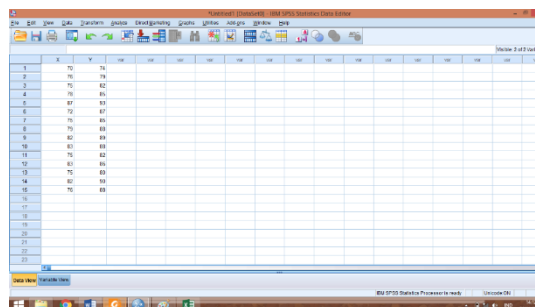
Using the SPSS Application

(1) Open the SPSS application
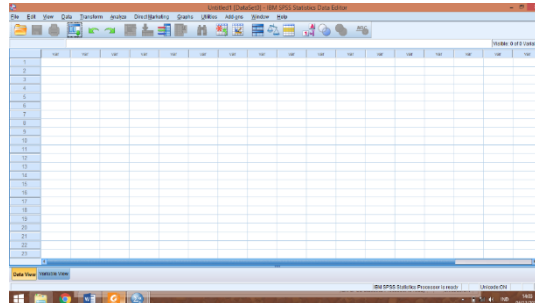
(2) Select Variable Display, enter the Variable description

(3) Select Data Display, enter the value of the Sunan Kalijaga UIN Entrance Exam into column X and Semester 1 Value into column Y
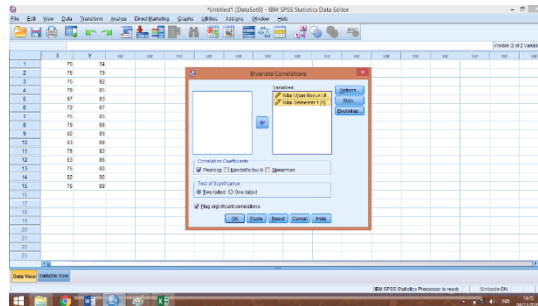

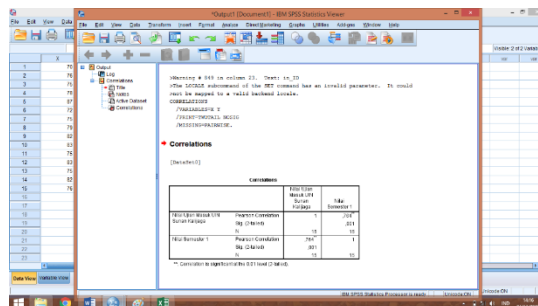(4) Click Analysis, Correlate, Bivariate

[25] Mujianto sholicin. *Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan.* Jurnal Managemen dan Pendidikan Islam. Vol.2 No.2. Juni 2017
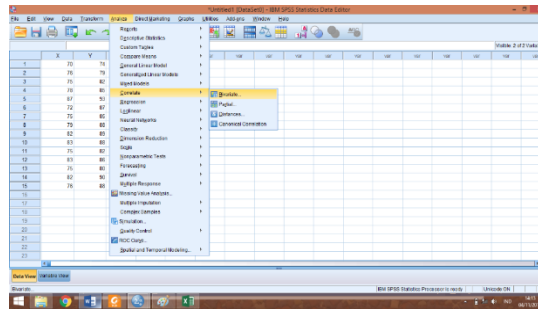
(5) Enter the X and Y variables into the right column

(6) Click OK, the results are as follows.

(7) Decision rules are as follows:[26]

Valid: If the r-value is greater than the r-value (r-value> table) or sig value. 2-tailed is smaller than the value of α = 0.05 (p-value <α)

Invalid: if the calculated value is smaller than the measured value (calculated value <r table value) or sig value. 2-tailed is greater than the value of α = 0.05 (p-value α)

(8) Therefore, a decision can be made that there is a correlation between variable X and variable Y, which means the UIN entrance test has a good estimate validity because r count is

---

[26] Imam Machali, *Statistik itu Mudah: Menggunakan SPSS Sebagai Alat Bantu Statistik* (Yogyakarta: MPI FITK UIN Sunan Kalijaga Yogyakarta, 2015).page. 74.

0.764> r table 0.553 and sig. 2-tailed is smaller than the value of α = 0.05.

d) Comparable Validity

During this time the test used as a measurement can be said to have peer validity if it has been able to show a direct relationship, between the first and last tests. Comparable validity is also called the validity of experiences that already exist today. We compare the results of the test obtained now with data that reflect experiences gained in the past. If the current test has a direct relationship with the results of the test based on experience, then the test can have validity or comparative experience, and vice versa.[27]

Examples of comparative validity analyzes are the same steps as estimated validity, the difference between estimated validity measures the extent to which test instruments demonstrate future capabilities, while comparative validity shows a comparison between current tests and tests that have been done in the past.

2) Analysis of item validity

The accuracy of the measurement held by the problem in the measurement that must be measured is an understanding of the validity of the item. The test that is usually made by the teacher is a set of various items used by the test compiler to measure student learning outcomes after they have followed the learning process for a certain time.[28]

There is a close relationship between items and learning achievement tests as a totality. It can be seen that the more items answered by students, the higher the score obtained, and vice versa. This shows that the greater the support provided by items on the test of learning outcomes, it will show stability and vice versa. This shows that the validity of the test is influenced by the validity of the items.

[27] Sukiman, *Pengembangan Sistem Evaluasi*.page. 180.

[28] Sukiman.page. 182

The importance of an educator first tracing the learning outcomes test which has been used as a measurement tool for student learning outcomes, usually, the correlation technique used to analyze items uses the product-moment correlation technique.

3. Analysis of Non-Test Validity
   a. a list of questions

   The questionnaire comes from the Latin Questionnaire which means a series of questions about the object being assessed to obtain data or information. Through questionnaires, relevant information can be collected following the purpose of the assessment and valid and reliable information.[29]

   Based on the type, the questionnaire was divided into three, closed questionnaires, open questionnaires, and closed and open questionnaires. The explanation is as follows.[30]

   1) Closed questionnaire

   Alternative answers in the closed questionnaire have been predetermined, respondents only choose the available alternatives.

   Example: Do you like the Learning Assessment Development course?

   a) Like it

   b) Do not like

   Some of the advantages of a closed questionnaire are:

   a) Easy to process

   b) Alternative answers are the same, it is not possible to give rise to different answers

   c) Respondents more easily answer

   d) Easy to do

   e) Easy to make code

   Some shortcomings in the closed questionnaire are:

---

[29] Muri Yusuf, *Asesmen dan Evaluasi Pendidikan: Pilar Penyedia Informasi dan Kegiatan Pengendalian Mutu Pendidikan* (Jakarta: Kencana, 2015).page. 104.

[30] Yusuf.page. 105.

a) Limit respondents from expressing their opinions

b) Easy to guess

c) Sometimes the alternative answers do not fit the situation

2) Open the questionnaire

Alternative answers in the open questionnaire are not provided, open questionnaires allow respondents to express answers in more depth.

Example: In your opinion, what factors make students lazy to learn?

....

Some of the advantages of open questionnaires are:

a) Preparation for closed questions

b) Respondents can answer according to conscience

c) Give opportunities to develop thinking skills

d) Allow being creative

Some of the shortcomings in the open questionnaire are:

a) Difficult in coding

b) Difficult to analyze

c) Not uniform

d) The time needed is longer

3) The questionnaire is closed and open

Combined between these two forms of the questionnaire in addition to providing alternative answers, there is also an opportunity to express respondents' answers according to their circumstances.

Example: How long do you read the Koran in a day?

a) 5 minutes

b) 10 minutes

c) 15 minutes

d) 20 minutes

e) .... (write it down)

Important things to consider in preparing the questionnaire instrument grid are as follows:[31]

1) The goal to be achieved

2) Types of data needed

3) Data elements

4) Respondents

b. Observation

Observation is a careful and systematic observation of objects.[32] Through observation, an educator knows the student's nonverbal behavior or other educational program activities. Observation serves as an evaluation tool in the world of education and teaching that is used to assess the learning activities of that skill.[33] According to Morries, observation is the activity of recording a symptom with the help of an instrument and recording it for scientific or other purposes[34]. So it can be concluded that the observation is not only done in the classroom but also outside the classroom. Observation of students when throwing trash, practicum, or picket classes is part of the observation. Observations can be divided into two, namely:

1) Participant Observation, observers are regularly involved directly in the program or activity being observed. In this way, the observer understands and understands the incident. One thing to note is that the student or individual being observed does not know that the observer is conducting an assessment. For example, to find out students' habits in studying at home, the observer lives together with students, if necessary in one room. The thing to note in this observation is that the observation material must be adjusted to the purpose of the observation, the time and form of the recording is done immediately after the event with

[31] Susilo Rahardjo dan Gudnanto, *Pemahaman Individu: Teknik Nontes* (Jakarta: Kencana, 2016).page. 107.

[32] Muri Yusuf, *Asesmen dan Evaluasi Pendidikan* (Jakarta: Prenada Media Group, 2015).page. 102.

[33] Siti Mania. *Observasi Sebagai Alat Evaluasi Dalam Dunia Pendidikan Dan Pengajaran.* Dirasat: jurnal tarbiyah dan keguruan. Vol 11, No 2. (2017)

[34] Hasyim Hasanah. *Teknik-Teknik Observasi.* Jurnal : at-Taqadum, Vol 8, No 1, Juli 2016.

keywords, systematic chronological order, building relationships to prevent suspicion, using good items. approach, maintaining the situation makes sense.[35]

2) Non-participant observation, the observer is not directly involved in the observed activity. For example, to find out how students behave in learning at school, observers observe student behavior while at school.

The key to successful observation in the educational process is largely determined by educators. If it is not possible to carry out continuous observations, the "time sampling schedule" approach can be used, ie the educator makes a detailed list so that the observation unit is selected systematically, which represents the observed behavior with the period that has been applied. For example, making 10-minute observations for each unit taken at random, for repeated observations. According to Nana, steps that need to be taken before making an observation instrument, 1) make an initial observation about the behavior to be observed; 2) sort the observer behavior according to what it should be; 3) determine the shape of the instrument; 4) notify the prospective observer (students) about the observation instrument; 5) give special notes at the end of the observation guide to note interesting thing[36]. If viewed from the side of whether observations are controlled, then observations can be divided into two types, namely structured observations, and unstructured observations.

In structured observations, teacher educators determine what will be observed, how to observe it, and when it is observed, whereas, in unstructured observations, greater flexibility is given to educators. This can be seen when the conditions are observed. Records of what was observed must be done as quickly as possible after the observations were made, and perfected at a later time, while what was observed was still fresh in the mind of the observer. Observations at school can be made using special notebooks about activities related to students while in the school environment.[37]

---

[35] Hasyim Hasanah. *Teknik-Teknik Observasi.* Jurnal : at-Taqadum, Vol 8, No 1, Juli 2016.

[36] Siti Mania. *Observasi Sebagai Alat Evaluasi Dalam Dunia Pendidikan Dan Pengajaran.* Dirasat: jurnal tarbiyah dan keguruan. Vol 11, No 2. (2017)

[37] Satria Koni, *Assessment Pembelajaran*.page. 31.

CONCLUSION

Based on the results of the analysis above, it can be concluded that validity is a concept related to the extent to which the test has measured what must be measured. Analysis of the validity of the instrument can be done in 2 ways, namely qualitative and quantitative. Qualitatively, the analysis must meet the requirements of the realm of substance, the realm of construction, and the realm of language. Whereas in quantitative terms include analysis of the total validity of the test (content validity, construct validity, estimated validity, comparative validity) and item analysis. This article also discusses how to analyze the validity of the test and non-test.

BIBLIOGRAPHY

Alwi, Idrus. "Kriteria Empirik dalam Menentukan Ukuran Sampel Pada Pengujian Hipotesis Statistika dan Analisis Butir." *Formatif: Jurnal Ilmiah Pendidikan MIPA* 2, no. 2 (5 Agustus 2015). https://doi.org/10.30998/formatif.v2i2.95.

Bashooir, Khoirul, dan Supahar Supahar. "Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM." *Jurnal Penelitian dan Evaluasi Pendidikan* 22, no. 2 (2018): 219–230.

Budiman, Agus, dan Jailani Jailani. "Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1." *Jurnal Riset Pendidikan Matematika* 1, no. 2 (2014): 139–151.

Ediana, Asep. *Evaluasi Pembelajaran di SD dan MI.* Bandung: PT. Remaja Rosdakarya Offset, 2018.

Hamzah Uno, dan Satria Koni. *Assessment Pembelajaran.* Jakarta: Bumi Aksara, 2013.

Hasanah, Hasyim. "Teknik-teknik Observasi. (Sebuah Alternatif Metode Pengumpulan Data Kualitatif Ilmu-ilmu sosial). *Jurnal At Taqaddum Vol 9 No 1. 2017*

Hayati, Nila, dan Djemari Mardapi. "Pengembangan Butir Soal Matematika di SD di Kabupaten Lombok Timur sebagai Upaya dalam Pengadaan Bank Soal". *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran Vol 11 No 1. 2014*

Hendryadi. "Validitas Isi: Tahap Awal Pengembangan Kuisioner". *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT, Vol 2 No 2. 2017*

Kurniawan, Tutut. "Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar". *Journal of Elementary Education, Vol 4 No 1.* 2015.

Mania, Sitti. "Observasi sebagai Alat Evaluasi dalam Dunia Pendidikan dan Pengajaran". *Lentera Pendidikan: Jurnal Ilmu Tarbiyah dan Keguruan, Vol 11 No 2. 2017*

Machali, Imam. *Statistik itu Mudah: Menggunakan SPSS Sebagai Alat Bantu Statistik*. Yogyakarta: MPI FITK UIN Sunan Kalijaga Yogyakarta, 2015.

———. *Statistik Manajemen Pendidikan Teori dan Praktik Statistik dalam bidang Pendidikan, Penelitian, Ekonomi, Bisnis, dan Ilmu-Ilmu Sosial Lainnya*. Vol. Vol. 1. Yogyakarta: Kaukaba Dipantara, 2015.

Menteri Pendidikan Nasional. "Peraturan Menteri Pendidikan Nasional Nomor 20 Tahun 2007 tentang Standar Penilaian." Menteri Pendidikan Nasional, 2007.

Nayla, Ata Amalia, dan Ani Widayanti. "Analisis Butir Soal Tes Kendali Mutu Kelas XII Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012". *Jurnal Pendidikan Akuntansi Indonesia, Vol X No 1.* 2015.

Nuswowati, Murbangun, dkk. "Pengaruh Validitas dan Reliabilitas Butir Soal Ulangan Semester Bidang Studi Kimia terhadap Pencapaian Kompetensi". *Jurnal Inovasi Pendidikan Kimia. Vol 4 No 2. 2010*

Pardimin. "Analisis Butir Soal Tes Pemecahan Masalah Matematika". *Jurnal Wacana Akademika, Vo 1, No 1. 2017.*

Rahardjo, Susilo, dan Gudnanto. *Pemahaman Individu: Teknik Nontes*. Jakarta: Kencana, 2016.

Solichin, Mujianto. "Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan." *Dirasat: Jurnal Manajemen dan Pendidikan Islam* 2, no. 2 (30 Juni 2017): 192–213.

Sudjana, Nana. *Penilaian Hasil Belajar Mengajar*. Bandung: PT. Remaja Rosdakarya Offset, 2015.

Sukiman. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani, 2011.

Surapranata, Sumarna. *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*. Bandung: PT. Remaja Rosdakarya Offset, 2009.

TFT, Aloysius Rabata Nova, dan Parno Parno. "Pengembangan Instrumen Asesmen Penguasaan Konsep Tes Testlet Pada Materi Suhu dan Kalor." *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan* 1, no. 6 (2016): 1197–1203.

"Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional," 2003.

Yonisa, Riza Kurniawan, dkk. "Pemberian Pelatihan Analisis Butir Soal Bagi Guru di Kabupaten Jombang: Efektif?. *Jurnal Pemberdayaan Masyarakat Madani (JPMM), Vol 1, No 2. Desember 2017.*

Yusrizal, Yusrizal. "Pengembangan Instrumen Penilaian Kinerja Dosen Di Perguruan Tinggi." *PARAMETER: Jurnal Pendidikan Universitas Negeri*

Jakarta 29, no. 1 (30 Mei 2017): 108–20. https://doi.org/10.21009/parameter.291.10.

Yusuf, Muri. *Asesmen dan Evaluasi Pendidikan*. Jakarta: Prenada Media Group, 2015.

———. *Asesmen dan Evaluasi Pendidikan: Pilar Penyedia Informasi dan Kegiatan Pengendalian Mutu Pendidikan*. Jakarta: Kencana, 2015.