perbaikan cek 3

by Muhammad Fajrul Bahri

Submission date: 18-Jul-2019 09:24AM (UTC+0700)
Submission ID: 1152810094
File name: 4._Fajrul_versi_proofreader_editor_1.doc (791K)
Word count: 4300
Character count: 22985

Content Validity and Reliability Analysis of Integrated Islamic-Science Test Instrument to Measure the Student's Critical Thinking Ability

Muhammad Fajrul Bahri
Program Pascasarjana, Yogyakarta State University
email: fajrulbahri17@gmail.com

Supahar
F MIPA, Yogyakarta State University
email: pahar.fis@gmaail.com

Diterima: 20 Maret 2019	Direvisi : 24 Juni 2019	Diterbitkan:tanggal bulan tahun

Abstract

This research is aimed to asses validity and reliability of Islamic-science integrated test instrument. The instrument was formed in multiple choice test instrument which is regarded appropriate to assest student critical thinking ability. The validity of instrument and its reability were assessed quantitively. While Aiken's V was used for assessing content validity, and Alpa Cronbach was used for assessing content reability. In terms of Item Response Theory, it was assessed through test I Function information and Standard Error of Measurement (SEM). The result of developing Islamic-science integrated test instrument to measure critical thinking ability was formed in form of closed-minded multiple choice that consisting of two test packages. The total test items on the instrument are 45 items consisting of 25 items with 5 shared items (anchor). Estimation of the validity test instrument obtained an index ranging between 0.75 to 1. according to classical test theory, the coefficient of reliability was 0.70 and according to the total function information and Standart Error of Measurement, the test instrument was suitable for range ability from 1.75 to +3. Thus, the test instrument is proved reliable and valid. A total of seven teachers using the decelared that the test instrument is appropriate to use in measuring learning outcomes, especially critical thinking abilities of student.

Keywords: *validity, reliability, Islamic, science.*

Abstrak

Penelitian ini bertujuan untuk menganalisis konten validitas dan reliabilitas dari intrumen tes terintegrasi agama dan sains. Konten validitas dan reliabilitas dianalisis secara kuantitatif. Pendekatan kuantitatif untuk konten validitas menggunakan Aiken's V, sedangkan untuk konten reliabilitas menggunakan Alpha Cronbach dan secara teori respon butir menggunakan fungsi informasi dan kesala17 pengukuran. Hasil pengembangan berupa instrumen tes terintegrasi agama dan sains untuk mengukur kemampuan berpikir kritis berupa tes pilihan ganda beralasan tertutup yang terdiri atas dua paket tes. Total item tes pada instrumen sebanyak 45 item yang terdiri dari 25 item dengan 5 item bersama (*anchor*). Estimasi validitas instrumen tes menggunakan formula Aiken diperoleh rentang indeks antara 0,75 sampai 1. Estimasi reliabilitas tes menurut teori tes klasik ditunjukkan oleh koefisien *Alpha Cronbach* sebesar 0,70 dan menurut kurva total fungsi informasi dan SEM yang berdasar pada teori respon butir (IRT) instrumen tes cocok untuk rentang kemampuan -1,75 sampai +3. Dengan demikian instrumen tes terbukti valid dan reliabel. Sejumlah tujuh guru pengguna instrumen tes menyatakan bahwa instrumen tes layak digunakan dalam pembelajaran. Dengan demikian instrumen tes cocok dan layak digunakan untuk mengukur hasil pembelajaran khususnya pada kemampuan berpikir kritis peserta didik.

Kata Kunci: validitas, reliabilitas, agama, sains.

Background

At present, the dynamics of learning education always changes in line with the development times. Ideal learning requires teacher to develop learning in a more conventional direction. The Constitution Number 20th on 2003 about the national education system defined an ideal learning activity involves interaction between student,

teacher, and learning resources in a learning education.[1] This idea leads to an understanding that learning should be regarded as a process in which transferring knowledge from teacher to student is happened. This urges a teacher to creative preparing an appropriate instrument for measuring learning outcomes. Generally, Anderson and Krathwohl divided the cognitive abilities of students into six parts, namely remembering, understanding, applying, analyzing, evaluating, and creating. The six levels of cognitive ability then divided into two levels of thinking skills, namely lower order thinking skills (that are remembering, understanding, and applying), and high order thinking skills (that are analyzing, evaluating, and creating).[2]

The differences of cognitive abilities among student encourage teachers to develop learning at high level abilities. It is mostly important for student to acquire high level ability especially in critical thinking skill. Quitadamo et. al defined Critical thinking is a purposeful self-regulation assessment process that encourages problem solving and decision making, or the engine that encourages how we decide what to do or believe in a particular context.[3] More research defined about the critical thinking ability, Yusri (2015) imply the critical thinking skills are important for student especially in natural science, even from basic education, which can be acquired through the implementation of scientific learning model.[4] Similarly with that statement, Istianah explain that critical thinking skills are tended in mathematics and natural science, where, using a particular model can improve the critical thinking skills in Mathematics learning[5]. Unfortunately, approaches that impelemented in Islamic education subjects do not encourage and train student to use critical thinking skills. Mumtahanah explain, on the reality, general teachers still using conventional methods in his leaning, which is make the teacher always dominates in teaching and learning process. Learning situations where the teacher is the central point making the role of student becomes very small. They were just sitting, listening to information provided by the teacher, recording what the teacher delivered, and memorizing what he recorded.[6]

Kunter et al. state that teacher education is an important variable as a quality of control that has a contribution in achieving the success of learning to achieve the goals, the teacher who teaches the subject must be in line with his field.[7] It is important to needed critical thinking for making sure the learning has getting the best achievement. Pradana et. al imply one of the demands that must be fulfilled on current learning is critical thinking ability .[8] The learning that directs student to develop critical thinking skills will make student compete in the global arena.

---

[1] Salinan Undang- Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional. p. 3

[2] Anderson, L. W., & Krathwohl, D. R. *Kerangka landasan untuk pembelajaran, pengajaran, dan asesmen (revisi taksonomi pendidikan bloom), cetakan kedua.* (Terjemahan Agung Prihantoro). (Yogyakarta. Pustaka Pelajar, 2017) (Edisi asli diterbitkan tahun 2001 oleh Pearson Education, Inc.). p. 105

[3] Quitadamo, Ian J., et al. Community-based inquiry improves critical thinking in general education biology. *CBE—Life Sciences Education*, Vol. 7 No.3, 2008. p. 328

[4] Yusri, Keterkaitan Scientific Learning dengan Kemampuan Berpikir Kritis (Kajian Proses Pembelajaran di Sekolah Dasar Budi Mulia Dua Sedayu Bantul). *Tesis*. Yogyakarta: Program Pascasarjana. UIN Sunan Kalijaga, 2015 (unpublished).

[5] Istianah, Euis. (Meningkatkan Kemampuan Berpikir Kritis Dan Kreatif Matematik Dengan Pendekatan Model Eliciting Activities (Meas) Pada Siswa Sma. *Infinity Journal*, Vol. 2, No.1, p. 93

[6] Mumtahanah, Nurotun. Meningkatkan Kemampuan Berpikir Kritis Siswa Melalui Metode Cooperative Learning dalam Pembelajaran PAI. *Jurnal* AL-HIKMAH Jurnal Studi Keislaman. Vol. 3, No. 1, 2013. p. 49

[7] Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, Vol. *105 No.*3, 2013. p. 206

[8] Pradana, Shan Dut Sukma, Parno; Handayanto, Supriyono Koes. Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, , Vol. 21, No. 1, 2017. p. 52

Integrated of Islamic-science requires student to develop knowledge to higher-order thinking skills. The integration is supposed to train student for increasing higher thinking skills especially in critical thinking ability, since it leads them to engage in learning activities in which they should aware of links between science or a scientifical phenomenon and religious values. Hove imply the using of critical thinking strategies in the high school classroom can improve the student appearance.[9] The combination of Islamic and science based learning will create a critical thinking ability, because student are expected to be able to find links between Islamic material learning and the state of science learning that surrounds them as well as those that have been proven.

This high level of ability is then expressed through the domain of critical thinking skills in order to understand learning material that are integrated in Islamic-science. in Islamic Islamic learning. It is necessary to have an instrument model developed in compiling an appropriate test instrument to measure the critical thinking ability. Wagner imply one of the the skill that needed by student in order to survive on 21st century is critical thinking skills and solve the problem.[10]

The achievement of students' competence on the cognitive dimension is generally assessed by the test.[11] Integrated of Islamic-science in Islamic education learning to measure critical thinking skills can develop with test instrument. According to Mardapi, the test is an instrument that used to make the measurement.[12] Azwar describes the test as a question set devised to uncover certain attributes through response to the question[13] Similarly, Gronlund & Linn defined test was an instrument or systematycal procedures that making to measuring a sample of behavior. Similarly with that, Cronbach imply test is a systematic procedure for viewing out a person's behavior and explain it with the aid of a numerical scale or a category-system.[14] The purpose of the test is to measure the achievement of learning outcomes or competencies to be achieved by student. It is necesarry to make a standard test to assess critical thinking ability based on integrated Islamic-science test.

The Study About Reliability and Validity

Reliability

The results of test can be trusted if the measuring instrument used reliably, such as able to produce a careful score with minor measurement errors. Reliability of test measurement is estimated through the computation of reliability coefficients and standard error measurement in accordance with specific analytical procedures.[15] Reliability similarly with the consistency of student scores that approve alternative forms of the same test. Due to differences in the exact content are assessed on alternative types, environmental variables such as fatigue or lighting, or student errors in responding, there are no two tests that will consistently produce identical results. This is true regardless of how similar the two are. In fact, even the same test given to the same student category the day after will produce two sets of scores that don't coincide perfectly. Obviously, when we conducted two tests that included the same material, we preferred similar student scores.

[9] Hove, G. (2011). *Developing Critical Thinking Skills in The High School English Classroom.* (Unpublished master's thesis). University of Wisconsin-Stout, WI.

[10] Wagner, T, *The global achievment gap*.New York: Basic Book, 2008. p. 14

[11] Ikhsanuddin et. al, Estimation of Cronbach Reliability based on Sample Size, Gender, and the Grades. Article. p. 1. (unpublished)

[12] Mardapi, Djemari. *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta. Parama Publishing, 2016. p. 94

[13] Azwar, S. *Constructing Psychological Scale*. Yogyakarta: Pustaka Pelajar, 2017. p. 6

[14] Gronlund, Lin and Croncbach in, Suwarto. Pengembangan tes ilmu pengetahuan alam terkomputerisasi. *Jurnal Penelitian dan Evaluasi Pendidikan*, 2017, Vol. 21.No. 2. p. 154

[15] Azwar, S. *Constructing* ..., 2017. p. 14

The more the value is, the more reliable the test score are[16]

In order to properly use, a good test instrument should be valid and reliable. A set of instrument of measuring must having high consistency, if the instrument of measure is carried out repeatedly and giving the same results. Reliability applies to the level of a test device, can't valid for each test item composing a measurement instrument. [17] It is widely known that reability is one of the basic criteria for a research instrument can be accepted. Reliability supports the validity of a test instrument. Particularly, validity was related to the accuracy of the result measurement in the ability of the test that measured. Therefore, it is important to assest reability of an instrument. steps for resolving reliability can be finished manually or assisted by a computer program. There are several techniuqes for measuring the reability include test repetition techniques, parallel test, halves, Rulon formulas, Hoyt formulas, Kuder-Richardson, Kappa coefficients, Cronbach Alpha, etc.[18]

Cronbach's alpha is a popular estimator which underestimates the reliability of a test.[19] In line with this assertion, Mardapi [20] and Azwar [21] describe through the reliability estimation using the Cronbach formula, the actual reliability is greater than or equal to the calculated α- coefficient. If the α-coefficient is of high value, then there is a real possibility of reliability even higher, but if it is in low-value there is a probability of unreliability or simply unfulfilled assumptions of parallel tests (τ-equivalent) because α-coefficient ignore that one. The value of the recommended construct reliability coefficient is above 0.7.[22] Researchers who get reliability coefficient score below 0.7 are expected to modify the measurement model they developed.

Reliability on Item Resphonse Theory (IRT) are expressed by the index of separation of items (item separation) and index of person separation (case / person separation). The item separation shows how the sample is scattered along the linear scale, while the person separations show how trusted the arranged test reach from high ability to low ability of individuals. In addition, the reliability in the Item Resphone Theory was expressed by the information function, which is relation among the test function information with standard error of measurement (SEM). Based on graph of the test information fuction and standard error of measurement, it is known that the developed test is fit with certain ability. [23]

Validity

Aiken proposed the concept of content validity more detailed. The detail was seen from the standard validity which is influenced by the number of rater and the rating scale[24]. Retnawati [25] and Widoyoko [26] also explain Validity is a degree to which an interpretation is properly supported by empirical facts and theories. An instrument is considered as valid as it can be used in order to accurately measure something that is supposed to measure. In other words, validity is related to "accuracy" with measuring instrument. A test as a tool for measuring learning outcome is considered as valid when it can be used to measure what is supposed to measure. In connection with

[16] WELLS, Craig S.; WOLLACK, James A. An instructor's guide to understanding test reliability. *Testing & Evaluation Services. University of Wisconsin*, 2003. p. 2

[17] Subali, B. *Pengembangan tes: beserta penyelidikan validitas dan reliabilitas secara empiris*. Yogyakarta, UNY Press. 2016. p. 60

[18] Suryabrata, S. *Pengembangan alat ukur psikologi (edisi ketiga)*. Yogyakarta, ANDI. 2005. p. 30

[19] Subali, B. *Pengembangan* ..., 2016. p. 65

[20] Mardapi, Djemari. *Pengukuran*, ..., 2016. p. 73

[21] Azwar, S. *Constructing* ..., 2017. p. 26

[22] Hair, J.F. et.al. Multivariate data analysis 7th edition. New York, Pearson Prentice Hall. 2010. p. 125

[23] Mardapi, Djemari. *Pengukuran, Penilaian*... p. 22-23.

[24] Aiken, L. R. Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, Vol. *45, No.* 1, 1985, p. 135

[25] Retnawati, H. *Validitas, reliabilitas, dan karakteristik butir: panduan untuk peneliti, mahasiswa, dan psikometri*. Yogyakarta, Parama Publishing. 2016. p. 16

[26] Widoyoko, E.P. *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta, Pustaka Pelajar. 2016. p. 141-142

educational test, the validity make it available for measuring learning outcomes.

Azwar[27] imply there are three types of validity, namely content validy, construct validity and empirical validity: Content validation is the extent to which the elements in a measuring instrument are truly relevant and represent representations of constructs that are in accordance with the measurement objectives. 2) Validity's construct: validity's construct proves whether the measurement results obtained through the test items correlate highly with the theoretical constructs underlying the arrangement. Acording with this state, Huda & Mardapi stateConstruct validity explains the extent to which performance on the test is consistent with the constructs in a particular theoretical consideration [28] 3) Validity based on criteria: the validity of the criteria is also called empirical validity interpreted as validity determined based on criteria, both internal criteria and external criteria. Internal criteria are test that are themselves criteria, while external criteria are the results of measuring instrument or other test outside the instrument itself which are the criteria.

Reynold et. al explained that quality testing is one step to show that evaluation instruments have been optimally developed. The main proof of the quality of the evaluation instrument is its validity. Messick defines validity as "one of integrative on how far the theoretical and rational empirical evidence support the feasibility of interpretation and action based on the results of the measurement process".[29] Validity is an assessment based on various categories of evidence. The relevant proof includes the measure's reliability, whether it includes interest constructs, and whether the score produced correlates with other variables that are expected to correlate with and not correlate with variables that are conceptually different.[30]

According to Yaghmaie[31] validity is used to measure related variables. Validity was significant factor to identifying the concept of measurement, but it is not an assumption whether the instrument is good to measure what it will measured. Yaghmaie also explained that for a test to be considered a good index validity, the value should be 0.75 or more. A single item with a value lower than 0.75 must be discarded..[32]

Method

Content reliability in this research can be determined with using classical test theory that indicated by Cronbach's Alpha coefficient, in addition, based on item response theory determined by the total curve of information function and standard error of measurement (SEM). Cronbach alpha provides an estimate of the internal consistency of the test instrument, so that (a) alpha does not show the stability or consistency of the test instrument over time, which would be estimated using either the retest-reliability strategy, and (b) alpha does not show stability or consistency in the test all forms of tests, which would be better estimated using the equivalent form a reliability strategy. [33]

The test was conducted on 563 student from 7 high schools. Either reliability or test information fuction proved the consistency of

[27] Azwar, S. *Konstruksi tes kemampuan kognitif (edisi kesatu)*. Yogyakarta, Pustaka Pelajar. 2016. p. 5

[28] Hudha, S.A & Mardapi, D, et al. Developing an instrument for measuring the spiritual attitude of high school students. *REiD (Research and Evaluation in Education)*, V. 4 No.1, 2018. p. 38.

[29] Reynold, C. R., Livingstone, R. B., & Wilson, V. *Measurement and assessment in education*. Upper Saddle River: Pearson Education Inc. 2009.

[30] https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/. Accesed on 21 June 2019.

[31] Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education*, Vol. 3 No. 1. p. 24

[32] *Ibid.* p. 26

[33] Brown, James Dean. The Cronbach alpha reliability estimate. *JALT Testing & Evaluation SIG Newsletter*, 2002, 6.1. p. 18

the test instrument, so that the test results can be trusted.

Table 1. Formula calculation of the content reliability quantitative

Alpha Cronbach (Clasical test theory)	Information function and SEM (Item Response theory)
$r_i = \frac{k}{(k-1)}\left\{1 - \frac{\sum s_i^2}{s_t^2}\right\}$ r_i = koefisien reliabilitas Alfa Cronbach K = jumlah item soal $\sum s_i^2$ = jumlah varians skor tiap item s_t^2 = varians total	to get the standart error of measurement, first, we must find the reliability of the test with this formula: $r_{xx'} = 1 - S_E^2 / S_{x'}^2$ $r_{xx'}$ = reliability S_E^2 = variance of error score $S_{x'}^2$ = variance of observe score And this is the formula for the standard error of measurement $SEM = \sqrt{S_x^2(1 - r_{xx'})}$ SEM = Standard error of measurement $r_{xx'}$ = reliability S_x^2 = variance of observe score

Whereas, in order to make an assessment on content validiy a expert judgment in form of quantitative analysis is applied. Quantitative analysis on the content validity using Aiken's V. Data was obtained from four expert judgment. The four expert judgments consisted of measurement expert, expert in the Islamic-science, and two of Islamic education teachers.

Table 2. Formula calculation of the content validity quantitatively

Aiken's V
$\sum s/[n(c-1)]$ s = r – lo r = the value given by expert lo = lowest validity score c = highest validity score n = number of expert who gave the score

The test instrument used the Partial Credit Model (PCM) with four scales. PCM was developed to analyzing items that needed some settlement step. According to Istiyono et al., the assumption of using PCM is based on each item have the same different power.[34] Similarly, according to Widiarso, PCM is one of the Rasch models that focuses on the location of the item in the modeling.[35] The model of PCM can accurate measuring the results of a test by combining dichotomous and polytomus models into one scoring.

Reliability Estimate; a Result

Estimation reliability with classical test theory of integrated Islamic-science test instrument using the Quest program was shown in Figure 1 where the estimated value was obtained 0.70.

```
Tes terintegrasi
------------------------------------------------------------------------------
Case Estimates                                        30/ 8/18 22: 9
all on all (N = 563 L = 45 Probability Level= .50)
------------------------------------------------------------------------------

Summary of case Estimates
=========================

Mean                              -.15
SD                                 .36
SD (adjusted)                      .30
Reliability of estimate            .70


 Fit Statistics
 ==============

 Infit Mean Square          Outfit Mean Square

    Mean     1.01             Mean     1.00
    SD        .29             SD        .39
```

Figure 1. Score reliability's estimate with Quest program

According to Hair, these results show that the test is reliable.[36] If the Alpha Cronbach reliability coefficient is less than 0.70 (ri < 0.70), Tavakol & Dennick Travakol and Dennick suggest to revise the items or to elimate them. The simple method to establish

[34] Istiyono, E., Mardapi, D. & Suparno. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) peserta didik SMA . *Jurnal* Penelitian dan evaluasi pendidikan. 2014. p. 4

[35] Widiarso, W. Aplikasi Teori Respon Item untuk Pemodelan Respons Menipu Pada Skala Kepribadian, *Laporan hasil penelitian*, Fakultas Psikologi UGM, Yogyakarta. 2010. p. 6

[36] Hair, J.F. et.al. Multivariate ....p. 125

that items is using computer program. the Alpha Cronbach reliability coefficient is more than 0.90 (ri > 0.90), they suggest to reducing the number of questions with the same criteria even though in different sentence forms.[37]

Similarly with the result of Alpha Cronbach reliability, depend on the test information function curve and SEM, the coefficient of integrated religious-science test instrument is in the ability range of -1.75 to +3. These results showed that the test instrument is fit for student with that range of abilities (between 1.75 to +3). The the test information function curve was presented in figure 2.

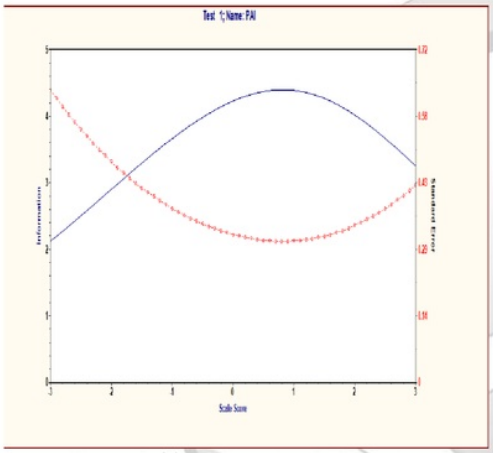

Figure 2. Total Function Information Curve and SEM.

Validity Estimate

Some improvements on the instrument construction was made based on expert judgment. According to judgment from the expert about the result of test instrument. Then, the researcher revision the construction of the test. The experts said, focused on the test instrument should be on the aspects universal of subject matter in the Islamic and science, not only at the structure of content material of Islamic-science obtained on the curriculum.

The results of the improvement instrument based on expert judgment are then re-analyzed the content of validity with Aiken' V index.. The results of analysis validity according on the Aiken's V index are presented in table 3.

Table 3. Result of validity instrument according at Aiken's V

No. Item	V1		V2		V3		V4		Total score		Aiken's V	
	A	B	A	B	A	B	A	B	A	B	A	B
01	4	4	5	5	4	5	4	5	17	19	0,81	0,94
02	3	3	4	3	5	5	3	3	15	14	0,75	0,75
03	3	4	5	5	4	4	5	5	17	18	0,81	0,88
04	4	3	5	5	5	4	4	4	18	16	0,88	0,75
07	5	4	5	4	4	4	4	4	18	16	0,88	0,75
08	5	5	5	4	4	4	5	5	19	18	0.94	0,88
09	5	4	5	5	5	4	3	5	18	18	0,88	0,88
10	5	4	5	5	4	5	4	4	18	18	0,88	0,88
12	5	4	5	5	5	5	5	5	20	19	1	0,94
13	4	4	5	5	4	4	4	5	17	18	0,81	0,88
14	3	5	4	5	5	3	5	3	17	16	0,81	0,88
15	5	5	5	4	5	4	4	4	19	17	0,94	0,81
17	3	3	5	5	4	4	3	3	15	15	0,81	0,81
18	3	4	4	4	5	4	5	4	17	16	0,81	0,75
19	3	3	5	5	4	4	4	4	16	16	0,81	0,75
20	3	3	5	5	5	4	3	4	16	16	0,88	0,75
21	3	3	4	5	5	4	5	5	17	17	0,81	0,81
22	3	3	5	4	4	4	4	5	16	16	0,75	0,75
24	5	3	4	5	3	4	3	4	15	16	0,88	0,75
25	3	2	4	4	4	4	4	3	15	13	0,75	0,75
5	5	5	5	5	5	5	3	3	18	18	0,88	0,88
6	5	5	5	5	4	4	4	4	18	18	0,88	0,88
11	5	5	5	5	5	5	5	5	20	20	1	1

[37] Lavakol & Dennick dalam Febrianawati Yusup, "Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif", *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, Vol. 7 No. 1, Juni 2018, p. 22

16	3	3	4	4	4	4	5	5	16	16	0,75	0,75
23	5	5	5	5	5	5	4	4	19	19	0,94	0,94
Content Validity Test Instrument												

Based on Aiken's V index, the value of item validity is ranging from 0.75 to 1. According to Yaghmaie, the value indicates that the instrument is proper to use for measuring and assessing student ability in understanding Islamic-science integrated learning material. Therefore, based on the appropriateness of the content, it can be said that the instrument is ready for use.

The quality of test instrument can be determine from the criteria of validity and reliability described previously. The developed test instrument have been proved to e valid and reliable so the test instrument is in good quality. In addition, the feasibility of the test instrument is supported by the statement of the Islamic Education teacher as the user and assessor of the instrument that being developed. Seven islamic education teacher from different school stated the test instrument is feasible to use in the learning processes toasses and support the achievement of learning outcomes especially in Islamic Education. Result of statement Islamic education teacher shown in table 4

Table 4. Results of the Statement of of Islamic Education Teachers as a Users of Test Instrument

No.	Guru PAI	Nama Sekolah	Pernyataan Kelayakan Instrumen	
			Layak	Tidak Layak
1.	Guru PAI 1	SMAN A	√	-
2.	Guru PAI 2	SMAN B	√	-
3.	Guru PAI 3	SMAN C	√	-
4.	Guru PAI 4	SMAN D	√	-
5	Guru PAI 5	SMAN E	√	-
6	Guru PAI 6	SMAN F	√	-
7.	Guru PAI 7	SMAN G	√	-

Conclusion

Content validity of the test instrument is classified good according to the Aiken index amount from the assess of four experts, which is in the ranging of 0.75 to 1. All items in the test instrument developed proved to be fit. Reliability of the test also classified to good test instrument expressed by Cronbach Alpha coefficient of 0.70. Furthermore, based on the total item function information and standard error of measurement, it is known that the test instrument is fit (suitable) for student with ability ranging from -1.75 to +3. In addition, Seven islamic education teacher from different school stated that the test instrument is feasible use in the learning processes toasses and support the achievement of learning outcomes especially in Islamic Education

Bibliography

Text Books

Anderson, L. W., & Krathwohl, D. R. (2017). *Kerangka landasan untuk pembelajaran, pengajaran, dan asesmen (revisi taksonomi pendidikan bloom), cetakan kedua.* (Terjemahan Agung Prihantoro). (Yogyakarta. Pustaka Pelajar, 2017) (Edisi asli diterbitkan tahun 2001 oleh Pearson Education, Inc.)

Azwar, S. *Constructing psychological scale.* Yogyakarta: Pustaka Pelajar, 2017

Azwar, S. *Konstruksi tes kemampuan kognitif (edisi kesatu).* Yogyakarta: Pustaka Pelajar. 2016.

Brown, James Dean. The Cronbach alpha reliability estimate. *JALT Testing & Evaluation SIG Newsletter*, 2002, 6.1. p. 18

Hair, J.F. et.al.. *Multivariate data analysis 7th edition.* New York, Pearson Prentice Hall. 2010.

Mardapi, Djemari. *Pengukuran, Penilaian dan Evaluasi Pendidikan.* Yogyakarta. Parama Publishing, 2016.

Retnawati, H. *Validitas, reliabilitas, dan karakteristik butir: panduan untuk peneliti, mahasiswa, dan psikometrian.* Yogyakarta, Parama Publishing. 2016.

Reynold, C. R., Livingstone, R. B., & Wilson, V. *Measurement and assessment in education.* Upper Saddle River: Pearson Education Inc. 2009.

Subali, B. *Pengembangan tes: beserta penyelidikan validitas dan reliabititas secara empiris.* Yogyakarta, UNY Press. 2016.

Sumintono, B., & Widhiarso, W. *Aplikasi pemodelan rasch pada assessment pendidikan.* Tim Komunikata. 2015.

Suryabrata, S. *Pengembangan alat ukur psikologi (edisi ketiga).* Yogyakarta, ANDI. 2005.

Wagner, T, *The global achievment gap.*New York: Basic Book, 2008. p. 14

Wells, Craig S.; Wollack, James A. An instructor's guide to understanding test reliability. *Testing & Evaluation Services. University of Wisconsin*, 2003.

Widoyoko, E.P. *Teknik Penyusunan Instrumen Penelitian.* Yogyakarta, Pustaka Pelajar. 2016.

Journal

Aiken, L. R. Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, Vol. *45, No.* 1, 1985, p. 135

Hudha, S.A & Mardapi, D, et al. Developing an instrument for measuring the spiritual attitude of high school students. *REiD (Research and Evaluation in Education)*, Vol. 4 No.1, 2018. p. 38.

Istianah, Euis, "Meningkatkan Kemampuan Berpikir Kritis Dan Kreatif Matematik Dengan Pendekatan Model Eliciting Activities (Meas) Pada Siswa SMA", *Infinity Journal*, Vol. 2, No.1, p. 43.

Istiyono, E., Mardapi, D. & Suparno. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) peserta didik SMA . *Jurnal Penelitian dan evaluasi Pendidikan.* 2014. p. 4

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, Vol. *105 No.*3, 2013. p. 206

Mumtahanah, Nurotun. Meningkatkan Kemampuan Berpikir Kritis Siswa Melalui Metode Cooperative Learning dalam Pembelajaran PAI. *Jurnal* AL-HIKMAH Jurnal Studi Keislaman. Vol. 3, No. 1, 2013. p. 49

Pradana, Shan Duta Sukma, Parno; [illegible]ndayanto, Supriyono Koes. Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, , Vol. 21, No. 1, 2017. p. 52

Quitadamo, Ian J., et al. Community-based inquiry improves critical thinking in general education biology. *CBE—Life Sciences Education*, Vol. 7 No.3, 2008. p. 328

Widiarso, W. Aplikasi Teori Respon Item untuk Pemodelan Respons Menipu Pada Skala Kepribadian, *Laporan hasil penelitian*, Fakultas Psikologi UGM, Yogyakarta. 2010. p. 6

Yusup, Febrianawati, "Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif", *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, Vol. 7 No. 1, Juni 2018, p. 22

Papers and Articles

Ikhsanuddin, mardiyanti, A., Subali, B, Estimation of Cronbach Reliability based on Sample Size, Gender, and the Grades. Article.. (unpublished)

Yusri, Keterkaitan Scientific Learning dengan Kemampuan Berpikir Kritis (Kajian Proses Pembelajaran di Sekolah Dasar Budi Mulia Dua Sedayu Bantul). *Tesis*. Yogyakarta: Program Pascasarjana. UIN Sunan Kalijaga, 2015 (unpublished)

Hove, G. (2011). *Developing critical thinking skills in the high school English classroom.* (Unpublished master's thesis).

University of Wisconsin-Stout, WI. Retrieved from https://tinyurl.com/y7dm8blh.

Website

https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/. Accesed on 21 June 2019.

perbaikan cek 3

ORIGINALITY REPORT

23%	19%	10%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journal.uny.ac.id Internet Source	2%
2	Submitted to Syiah Kuala University Student Paper	1%
3	nuspaces.nu.edu Internet Source	1%
4	jois.eu Internet Source	1%
5	www.mecd.gob.es Internet Source	1%
6	Submitted to London School of Marketing Student Paper	1%
7	Ikhsanudin, B Subali. "Content validity analysis of first semester formative test on biology subject for senior high school", Journal of Physics: Conference Series, 2018 Publication	1%
8	Submitted to Leeds Trinity and All Saints Student Paper	1%

9	uad.portalgaruda.org Internet Source	1%
10	www.researchgate.net Internet Source	1%
11	International Journal of Contemporary Hospitality Management, Volume 25, Issue 7 (2013-09-21) Publication	1%
12	fkip.ummetro.ac.id Internet Source	1%
13	digilib.uin-suka.ac.id Internet Source	1%
14	psasir.upm.edu.my Internet Source	1%
15	ejournal.radenintan.ac.id Internet Source	1%
16	Submitted to Florida Atlantic University Student Paper	1%
17	Submitted to Program Pascasarjana Universitas Negeri Yogyakarta Student Paper	1%
18	aip.scitation.org Internet Source	1%
	www.vsb.cz	

19	Internet Source	1%
20	www.ijcmph.com Internet Source	<1%
21	jurnal.uin-antasari.ac.id Internet Source	<1%
22	www.ncbi.nlm.nih.gov Internet Source	<1%
23	uhra.herts.ac.uk Internet Source	<1%
24	Submitted to State Islamic University of Alauddin Makassar Student Paper	<1%
25	candrajiwa.psikologi.fk.uns.ac.id Internet Source	<1%
26	eprints.uny.ac.id Internet Source	<1%
27	Submitted to University of Newcastle Student Paper	<1%
28	S R Maulita, Sukarmin, A Marzuki. "The Content Validity: Two-Tier Multiple Choices Instrument to Measure Higher-Order Thinking Skills", Journal of Physics: Conference Series, 2019 Publication	<1%

29	repository.uinjkt.ac.id Internet Source	<1%
30	Submitted to Sriwijaya University Student Paper	<1%
31	es.scribd.com Internet Source	<1%
32	docplayer.info Internet Source	<1%
33	fr.scribd.com Internet Source	<1%
34	matematika.fmipa.unp.ac.id Internet Source	<1%
35	Rosi Angelia, Ahmad Fauzi, Yohandri. "Validity and practicality of IPA textbook integrated the theme of a hurricane with the type of shared inquiry-based training", Journal of Physics: Conference Series, 2019 Publication	<1%
36	www.scribd.com Internet Source	<1%
37	academic.oup.com Internet Source	<1%
38	ejer.com.tr Internet Source	<1%

39	"Second Handbook of Information Technology in Primary and Secondary Education", Springer Science and Business Media LLC, 2018 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		