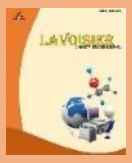
Available online at
https://jurnal.uinsyahada.ac.id/index.php/Lavoisier/index

LAVOISIER: CHEMISTRY EDUCATION JOURNAL

Vol. 3 (1), 2024



Validity and Reliability of Test and Non-Test Research Instruments on Chemical Bond Materials

Syah Nurul Riswana[1*], Haqqi Annazili Nasution[2]

[1,2] Postgraduate Chemistry Education, Faculty of Mathematics and Natural Sciences, Medan State University, Jalan Willem Iskandar Pasar V Medan Estate, 20221, North Sumatra, Indonesia.

*Correspondence: syahnurulriswana1999@gmail.com

Article History

Received 06 12th 2024
Revised 06 27th 2024
Accepted 06 27th 2024
Available Online 06 30th 2024

Keywords:
Chemical Bonding
Research Instrument
Reliability
Test and Non-test
Validity

Abstract
This research aims to determine the feasibility of research instruments through the validity and reliability of test and non-test instruments on chemical bonding materials. The instruments tested were in the form of multiple-choice questions, and the non-test instruments were in the form of questionnaires. Validity and reliability testing aims to determine the suitability of the test instrument. The test instruments aim to measure learning outcomes, and the non-test instruments aim to measure students' motivation for chemical bonding material. Validity and reliability testing of the test instrument was carried out by testing content validity with expert validators and small-scale field trials with students in class XI Science 1 at SMA Negeri 14 Medan. Non-test instrument testing was carried out through content validation tests on three expert validators. The research results showed that of the 40 questions, there were 28 valid questions, and the reliability test obtained r-11 = 0.85 and was included in the high reliability category. At the difficulty level of the questions, the number of questions in the difficult category was 5; 24 in the medium category; and 11 in the easy category. Then, regarding the differentiating power of questions in the good category, the number obtained was 25, and in the bad category, the number obtained was 15. The percentage of eligibility for non-test instruments obtained was 90%, which is a feasible criterion.

1. Introduction

Test and non-test instruments are two types of instruments that can generally be used in chemistry education research. Research instruments are tools used to collect data in research. The measurement objectives of the research instrument are adjusted to the theory used as a basis (Purwanto, 2018). The steps that need to be taken in developing test results or learning achievements are: compiling test specifications, writing test questions, reviewing test questions, conducting test trials, analyzing question items, improving tests and assembling tests (Mardapi, 2008).

The test instrument in this research is in the form of multiple-choice questions on chemical bond material to measure learning outcomes, while the non-test instrument in this research is a questionnaire to measure student motivation. Several things need to be considered to make a good instrument. The steps for making instruments for both tests and non-tests are as follows: determine the purpose of making the instrument; find relevant theories or coverage of material; create indicators or instrument details; content validation; revision based on validator input; conduct trials with appropriate respondents to obtain

participant response data; analysis (validity, reliability, difficulty, and discriminating power); and then compose the instrument (Retnawati, 2016).

This research aims to determine the feasibility of this research instrument through the validity and reliability of test and non-test instruments on chemical bonding material. To answer research objectives, data is needed. Conclusions that are appropriate to the actual situation will be made with correct data. The quality of data collection instruments or measuring research variable objects determines the validity of the data (Sembiring & Nasution, 2021).

A good instrument has certain criteria in research, thereby producing good quality research data too. Likewise, instruments that do not have good criteria for research will produce poor-quality research data (Sukendra & Atmaja, 2020).

Whether a research instrument is good or not is determined by its validity and reliability. Validity is a test that functions to see whether a measuring instrument is valid or invalid, while reliability is an index that shows the extent to which a measuring instrument can be trusted or relied upon. Then the reliability test can be used to determine the consistency of a measuring instrument or its consistency even after repeated measurements. A measuring instrument is considered reliable if it produces constant results even though it is used repeatedly (Janna & Herianto, 2021).

Validity testing is one of the steps taken to test the content of an instrument. The purpose of the validity test is to measure the accuracy of the instrument that will be used in research. Meanwhile, reliability testing is a process of measuring the accuracy (consistency) of an instrument. Reliability testing is a condition that can be trusted or something that can be trusted. The purpose of the validity test is to find out how consistent the instruments used by researchers are (Hakim, Mustika, & Yuliani, 2021).

2. Materials and Methods

This test was carried out at SMA Negeri 14 Medan. Validity and reliability testing of the test instrument was carried out by testing content validity on three expert validators and small-scale field trials on class XI IPA 1 students at SMA Negeri 14 Medan. Non-test instrument testing was carried out through content validation tests on three expert validators. Several aspects that will be analyzed from test instruments that are tested on a limited scale are validity, reliability, level of difficulty of questions and distinguishing power of questions.

Validity test

A valid instrument means that the measuring instrument used is appropriate for measuring what is to be measured. To test the validity of the instrument items, the instrument must be tested and analyzed using item analysis. Item validity is carried out by calculating the correlation between each instrument item score and the total score, using the "product moment" correlation formula as follows (Silitonga, 2014):

Information:
r-xy : correlation coefficient between variables x and y
x : Score for each question item
y : Total score
∑xy : Number of multiplications between x and y
x2 : Square of
xy2 : Square of y
N : Number of samples

Based on the validity requirements of the test instrument, namely the validity coefficient obtained (r-xy) is compared with the r-table product moment value with degrees of freedom (db = N-2) at α = 0.05 with the criteria: If r-xy > r-table then the test instrument is said to be valid (Silitonga, 2014).

Reliability

Reliability is the stability/reliability/consistency of a measuring instrument so that if the instrument is used it always produces consistent results. A reliable instrument means an instrument that, when used several times to measure the same object and under the same conditions, will produce the same data (Silitonga, 2014). Reliability can be found using the Kuder & Richardson formula (K-R.20):

With the formula: $q = 1 - P$

Information:

r11 : test reliability coefficient
K : number of test items
: score variance
p : Proportion of subjects who answered correctly
q : Proportion of subjects who answered incorrectly

To interpret the reliability value of the questions, if r11 (calculated) > r-table for the $\alpha = 0.05$ level then the test is declared reliable (Silitonga, 2014).

Question Difficulty Level

A good test item is a test item that is neither too easy nor too difficult. The number that shows the level of difficulty of a test item is called the Item Difficulty Index (P) which can be calculated using the formula (Silitonga, 2014):

Information:

P: Item difficulty index
B: Number of students who answered the item correctly
Q: Total number of students

The larger the price P, the easier the item, conversely the smaller then P, the more difficult the item. A test item is said to meet the requirements if the P value ranges between: 0.20 – 0.80. If $P < 0.20$ it means the test item is too difficult, and if $P > 0.80$ it means the test item is too easy. The test item difficulty level categories can be seen in Table 1:

Table 1. Test Item Difficulty Level Category

Level of Difficulty	Description
$P < 0,20$	Hard
$0,20 \leq P \leq 0,80$	Currently
$P > 0,80$	Easy

Question Differentiating Power

Discriminating power is the ability of a question item to differentiate between smart students (high ability) and stupid students (low ability) (Silitonga, 2014). The differentiating power formula is:

Information:
D: Differentiating power
JA: number of upper-group test students
JB: number of lower group test students

BA: Number of upper-group who answered correctly
BB: Number of lower-group who answered correctly

The number that shows the magnitude of the different power of an item is called the Different Power Index (Discrimination Index) symbolized by "D" where the value of D ranges from -1 to +1. An item is declared to meet the requirements if D ranges between: + 0.20 to +1.0.

Feasibility of the questionnaire

Qualitative analysis was carried out by testing non-test instruments through content validation tests on three expert validators. Calculation of results from expert validators on non-test instruments is as follows:

$$P = f/n\ 100\%$$

After getting the results from the three validators, the average of the validator results is calculated. Validation in research needs to be carried out to determine the suitability of the instrument used. The categorization of the eligibility criteria can be seen in Table 2.

Table 2. Categories of eligibility criteria for non-test instruments

Score Interval % Criteria	Score Interval % Criteria
81% - 100%	Very worthy
61% - 80%	Worthy
41% - 60%	Decent enough
21% - 40%	Not worth it
<21%	Not feasible

Quantitative analysis tests the validity of the questions, reliability, level of difficulty of the questions and the distinguishing power of the questions. The following is a presentation of the results of quantitative analysis calculations.

3. Results and Discussion

Quantitative analysis tests the validity of the questions, reliability, level of difficulty of the questions and the distinguishing power of the questions. The following is a presentation of the results of quantitative analysis calculations.

Validity test

To find out the results of the validation of the question instrument, the test instrument is tested first on students. There are 40 questions used in the test instrument. Determining decisions regarding valid or invalid instruments is seen based on the r value obtained. If the value of r-count> r-table then the instrument is said to be valid. The r-table value is 0.344 with N=33 and a significance level of 0.05 (5%). The following are the results of calculating the validity of the questions in Table 3.

Table 3. Question Validity Test

Number of Questions	Numbers	Description
28	1,3,4,6,7,9,10,11,12,13,14,15,16,17,18,21,22,24,25,26, 27,29,30,32,35,36,38,39	Valid
12	2,5,8,19,20,23,28,31,33,34,37,40	Invalid
40	Amount	

Based on the calculations, 28 questions have r-count > 0.404 and the remaining 12 questions have r-count < 0.404.

Reliability

Reliability is an index that shows the extent to which a measuring instrument can be trusted or relied upon. The reliability value obtained based on test instrument data is 0.85 and includes high reliability. The presentation of reliability criteria can be seen in Table 4.

Table 4. Reliability exposure

Quantity	Criteria
Number of items	40
Number of students	33
R Table	0,344
Valid Amount	28
Invalid Amount	12
Score variance	42,484
KR-20	0,85

Question Difficulty Level

Good questions are questions that are not too easy and not too difficult. A good instrument is an instrument that has a moderate level of difficulty, namely with a difficulty level index value of 0.2–0.8. If an index <0.2 is obtained then the instrument is said to be difficult. And if the instrument is>0.8 then the instrument is said to be easy. The following are the test calculation results. The level of difficulty of the trial questions is in Table 5.

Table 5. Question Difficulty Level

Number of Questions	Numbers	Category
5	2,8,31,33,37	Hard
24	1,4,5,6,10,13,15,18,19,20,21,22,23,24,26 ,28,29,30,32,34,35,38,39,40	Currently
11	3,7,9,11,12,14,16,17,25,27,36	Easy
40	Amount	

Based on the difficulty level test, 24 questions are good to use because they are instruments that have a medium level of difficulty.

Question Differentiating Power

Based on the test instrument data, the distinguishing power was obtained for questions with a good category of 25 and a bad category of 15. The following are the results of the calculation of the Differentiating Power test for trial questions in Table 6.

Table 6. Test the Discriminating Power of Questions

Number of Questions	Numbers	Category
25	1,3,4,6,7,10,12,13,14,15,20,21,22,24,25,2 6,27,29,30,32,34,35,36,38,39	Good
15	2,5,8,9,11,16,17,18,19,23,28,31,33, 33,40	Not good
40	Amount	

Based on the different power tests, questions that may be used are questions that have a different power category of 0.20-1.00. This means 25 instrument questions are good to use.

Expert Validator Eligibility Percentage

Qualitative analysis was carried out by testing non-test instruments through content validation tests on three expert validators. The percentage results for each of the three expert validators are 89%, 89% and 92% and the average percentage result for the three expert validators is 90%. Based on these percentage results, it can be said that the non-test instruments used in this research fall into the very appropriate criteria category. The results of the presentation of the percentage of expert validators can be seen in Table 7.

Table 7. Percentage of Expert Validators

Expert Validator	Total Score	Percentage Results
Lecturer I	25/28 × 100%	89%
Lecturer II	25/28 × 100%	89%
Chemistry teacher	26/28 × 100%	92%
	Average percentage	90%

Based on Table 7, the total scores for each expert validator are 25, 25, and 26. The percentage results for each expert validator are obtained from the total score using the formula explained in the explanation of the method section.

5. Conclusions

Based on the preparation of the validity and reliability of test and non-test instruments on chemical bond material, it can be concluded that. In testing the validity of the questions on the test instrument, it was found that of the 40 questions, there were 28 valid questions and 12 invalid questions. In the reliability test, r11 = 0.85 was obtained and was included in the high-reliability category. At the level of difficulty of the questions, the number of questions in the difficult category was 5 questions, 24 questions in the medium category and 11 questions in the easy category. In the distinguishing power of questions in the good category, the number obtained was 25 questions and in the bad category, the number obtained was 15 questions. In terms of the feasibility percentage for non-test instruments, the percentage results for each of the three expert validators are 89%, 89% and 92% and the average percentage result for the three expert validators is 90%. Based on the average percentage results, the non-test instrument used falls into the very appropriate criteria category.

References

Hakim, R. Al, Mustika, I., & Yuliani, W. (2021). Validitas Dan Reliabilitas Angket Motivasi Berprestasi. *FOKUS (Kajian Bimbingan & Konseling Dalam Pendidikan)*, *4*(4), 263. https://doi.org/10.22460/fokus.v4i4.7249

Janna, N. M., & Herianto. (2021). Konsep Uji Validitas Dan Reliabilitas dengan Menggunakan Spss. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, *18210047*, 1–12.

Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes Dan Non-Tes*. Yogyakarta: Mitra Cendekia.

Purwanto. (2018). *Teknik Penyusunan Instrumen Uji Validitas Dan Reliabilitas Penelitian Ekonomi Syariah*. Magelang: Staial Press.

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa Dan Psikometrian).* Yogyakarta: Parama Publishing.

Sembiring, F. S., & Nasution, H. A. (2021). Validitas Dan Reliabilitas Instrumen Penelitian Tes Dan Non-Tes Pada Materi Laju Reaksi. *Prosiding Seminar Nasional Kimia & Pendidikan Kimia#2 - 2021 Jurusan Kimia FMIPA UNIMED*, *2*, 32–38.

Silitonga, P. M. (2014). *Metodologi Penelitian Pendidikan*. FMIPA UNIMED.

Sukendra, K., & Atmaja, K. S. (2020). Instrumen Penelitian. In *Mahameru Press*.