International Journal of Educational Research Excellence (IJERE)
https://ejournal.ipinternasional.com/index.php/ijere
Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

Development of a Virtual Test to Measure Mastery of Green Chemistry Material

Dewi Sanusi Noor [1*)], Nahadi [1)], Hernani [1)]

[1)] Departemen Pendidikan Kimia, Universitas Pendidikan Indonesia, Indonesia
[*)] Corresponding Author: dewinoor21@gmail.com

Article history: received July 17, 2023; revised August 12, 2023; accepted September 13, 2023

Abstract. This research aims to obtain a virtual test that measures proficiency in green chemistry with good instrument quality, meeting the criteria of validity, reliability, readability value, difficulty level, and good discrimination power. The virtual test in this research consists of text, images, animations, graphs, and videos. The method used is the development and validation method, which involved developing twenty items for the virtual test that have been deemed valid. First, a readability test was conducted on the items with five students, followed by a real-class trial involving 101 tenth-grade students from three high schools that already use the independent curriculum. The findings of this research show that out of the 36 developed test items, 20 items have a CVR value above 0.99, a reliability value of 0.702, categorized as high. The difficulty level test resulted in a value of 0.39, categorized as moderate, and the discrimination power test yielded a value of 0.41, categorized as good.

Keywords: Assesment, Green Chemistry, Virtual Test

I. INTRODUCTION

Green chemistry is the latest subject in the independent curriculum, which is currently being implemented by the government to realize better education and a better future. Green chemistry itself is an important topic in chemistry because in the past 10 years from 2011 to 2021, the trend of green chemistry has been closely related to organic chemistry, polymers, toxicology, industrial chemistry, general chemistry, and environmental chemistry, which address catalysts, energy, chemical synthesis, environmental pollution, and even chemical products such as plastics, oils, and so on. The emergence of green chemistry in high schools is partly due to environmental issues such as pollution in the surrounding areas. Previous research has shown that by applying the principles of green chemistry, environmental pollution can be prevented (Inayah et al., 2022). The application of these green chemistry principles can be done by all sectors, including the government, society, and students, so that green chemistry can be included in the independent curriculum to enable students to understand both the theory and practice in efforts to reduce, eliminate, and replace the use of toxic and hazardous chemicals in the learning process, even in daily activities.

Every learning process implemented in schools and higher education institutions is inseparable from instruments. These instruments are used to assess students' competency achievements, compile progress reports on learning outcomes, and improve the learning process (Firman, 2018). Various forms of instruments have been developed to measure students' learning outcomes in general, whether they are cognitive, psychomotor, or affective, but there is still a need for instruments to measure the mastery of concepts in green chemistry as included in the independent curriculum.

Research on the development of instruments to measure proficiency in green chemistry has not been widely conducted, and the forms of these instruments are still limited. This creates challenges for educators in implementing the teaching and learning process. However, there are some similar studies, such as the research by Grieger, K., Schiro, A., and Leontyev, A (2022), which proves that the developed instrument is efficient and accurate in measuring students' knowledge of the principles of green chemistry. This study was conducted with college students using a true-false format instrument. Further research can be conducted to explore the development of these instruments.

To be more innovative, the development of the test is conducted using a virtual test with the advantage of utilizing multimedia such as videos, animations, graphs, or experimental data, which have been proven to help students better understand the material (Valantika et al., 2016) and receive positive responses in terms of ease of use and user satisfaction (Mayangsari, 2022). Moreover, in both online (distance learning) and offline (face-to-face) learning conditions, the use of technology facilities remains prominent. This proves that online testing, whether using mobile phones or computers, becomes an alternative and solution in the teaching and learning process and learning outcomes.

This research aims to develop a virtual test that measures proficiency in green chemistry in schools that implement the independent curriculum. The development of this virtual test will result in a high-quality instrument that meets the criteria of validity, reliability, readability value, difficulty level, and good discrimination power to measure proficiency in green chemistry.

Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

II. METHODS

The research design used in this study is the development and validation design, consisting of four stages: (1) planning stage, which includes problem identification, determining the objectives of instrument development, and identifying the aspects to be measured through the developed instrument; (2) instrument development stage; (3) development testing stage; and (4) evaluation stage, specifically the real-class testing (Adams & Wieman, 2010). Data were obtained from the validation of the test items by five experts, the readability test conducted by five students, and the students' responses to the developed virtual test. These data were used to analyze reliability, difficulty level, and discrimination power.

This virtual test is based on the learning outcomes specified in the independent curriculum (Kemendikbudristek, 2022). The test blueprint for the virtual test can be seen in Table 1.

Table 1. Test blueprint for virtual test on proficiency in green chemistry

Learning Outcomes	Indicators of Concept Achievement	The number of test items
Applying Chemical Concepts in Environmental Management, including Explaining the Phenomenon of Global Warming	Determining evidence of environmental issues	3
	Explaining environmental issues through chemical reactions	3
	Providing examples of environmental management solutions for local issues	3
	Explaining the importance of green chemistry in environmental management	3
	Defining the concept of green chemistry	3
	Analyzing the principles of green chemistry in daily life	3
	Analyzing environmental issues and solutions in support of environmental conservation efforts	3
	Identifying evidence of issues contributing to global warming	3
	Explaining the causes of global warming	3
	Analyzing the impacts of global warming	3
	Identifying impacts contrary to the principles of green chemistry	3
	Analyzing activities that support green chemistry	3

The virtual test was conducted online on the website www.classmarker.com, with each student having a different account. The participants in the real-class trial consisted of 101 tenth-grade students from three high schools (Sekolah Menengah Atas/SMA) that implemented the independent curriculum and had studied green chemistry.

III. RESULTS AND DISCUSSION

A. Content Validation

Based on the critical values table proposed by Lawshe (1975), if there are five validators, the critical value for CVR is 0.99. A test item can be considered valid if the determined CVR value is greater than 0.99, indicating that all validators rated it as "Yes". Conversely, if a test item is deemed invalid and cannot proceed to the field testing stage, the CVR value obtained will be lower than 0.99. The results of the CVR analysis for each test item in the virtual test in this study are shown in Figure 1.

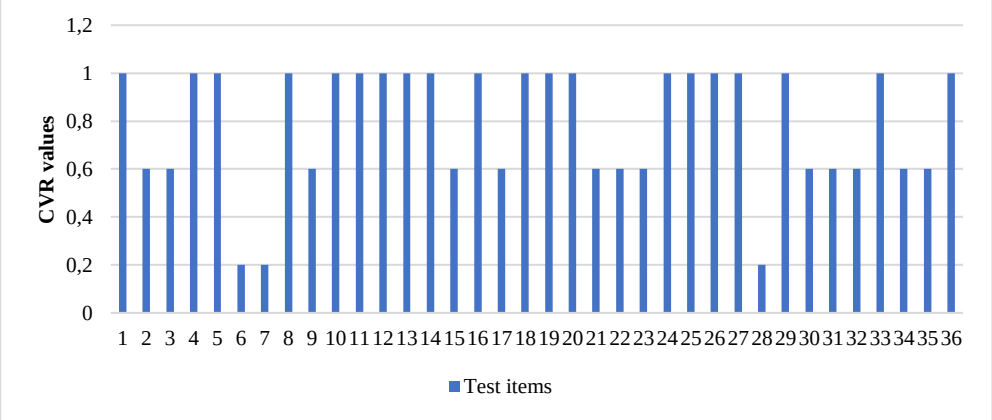


Figure 1. CVR Analysis Results

Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

Based on Figure 1, it can be observed that 20 test items obtained a CVR value of 1, indicating a valid category. On the other hand, there are 16 test items with CVR values below 0.99, specifically 3 test items with a CVR value of 0.2 and 13 test items with a CVR value of 0.6, indicating an invalid category. The test items categorized as invalid were subsequently revised or excluded from use. The test items categorized as valid required revisions based on expert feedback to improve the developed virtual test. These revisions included language improvements, illustration or image enhancements. After obtaining the CVR values, the CVI (Content Validity Index) was calculated to determine the overall content validity of the test. It is a simple average of the CVR values for all the items (Gilbert & Prion, 2016). The calculated CVI value was 0.79, indicating that there is consistency between the test items in the virtual test based on the indicators of concept achievement developed to measure students' level of concept mastery.

B. Validity of the Virtual Test

The readability test was conducted to assess students' readability level of the developed virtual test using a questionnaire. Four aspects were observed, including 1) comprehension of the presented article/image/video/animation, 2) understanding of language usage, 3) understanding of the problem in the stem, and 4) understanding of the answer choices. The average readability test results for all aspects for each test item are presented in the form of a diagram, which can be seen in Figure 2.

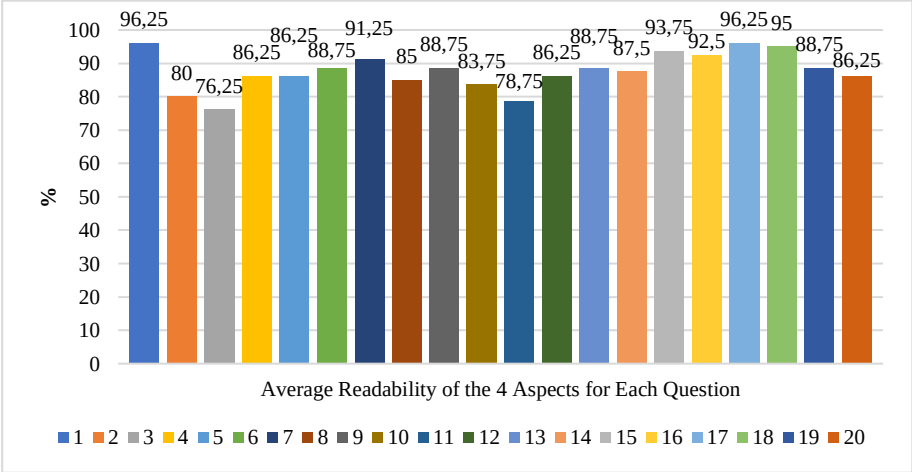


Figure 2. Recapitulation of (%) Readability of Each Test Item

Based on Figure 2, the average readability of each test item, according to Koentjaraningrat (1997), is categorized as mostly readable with a percentage of 87.81%. Although the test items received the category of mostly readable, the researcher conducted a review before administering this virtual test to students. Test items number 2, 3, and 11 were revised based on the readability test results conducted by the respondents, as these items were categorized as mostly readable. The revisions for these three items were made in terms of language usage, problem stem, and answer choice selection. However, no revisions were made for the comprehension of articles, images, graphs, videos, and animations as a whole, as the test items were categorized as mostly or completely readable.

Therefore, the high readability score indicates that the developed virtual test in this study is easy to read and comprehend. This ensures that students can accurately answer the questions, as the issue of unreadable test items is avoided.

C. Reliability

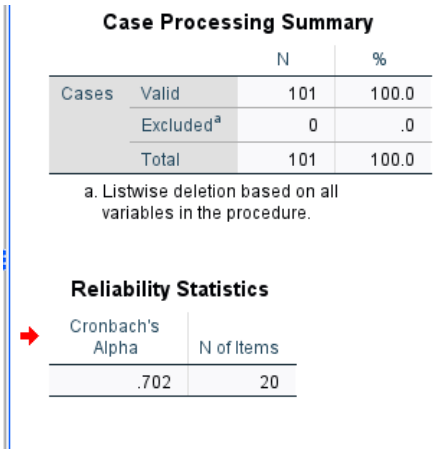
Reliability testing was conducted to determine the extent to which a measurement tool provides consistent and trustworthy information about an individual's ability (Firman, 2018). Reliability tests the level of consistency of a used instrument. The reliability of the virtual test was obtained from the field testing conducted in one of the tenth-grade classes in a high school that implemented the independent curriculum. Prior to the field testing, the virtual test had undergone validation by experts and a readability test. The reliability test results in this study referred to Jacobs and Chase, and the results were categorized according to the clarification in Table 2.

Table 2. Classification of reliability coefficients

Correlation Coefficient	Specification
0.00-0.19	Very low
0.20-0.39	Low
0.40-0.59	Moderate
0.60-0.79	High
0.80-1.00	Very high

Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

Reliability testing in this study was conducted using the SPSS (Statistical Package for Social Sciences) software, version 27.0, with the Cronbach's Alpha method. The obtained reliability coefficient was 0.702, as shown in Figure 3.



Case Processing Summary

		N	%
Cases	Valid	101	100.0
	Excluded[a]	0	.0
	Total	101	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.702	20

Figure 3. Reliability Results using SPSS

Based on the interpretation of Table 2 developed by Jacobs and Chase, the instrument used in this study is considered acceptable and falls into the category of high reliability. This indicates that the virtual test designed to measure proficiency in green chemistry exhibits good consistency. In other words, the virtual test with green chemistry content developed in this study can provide consistent results when repeated with the same subjects under the same conditions (Firman, 2013).

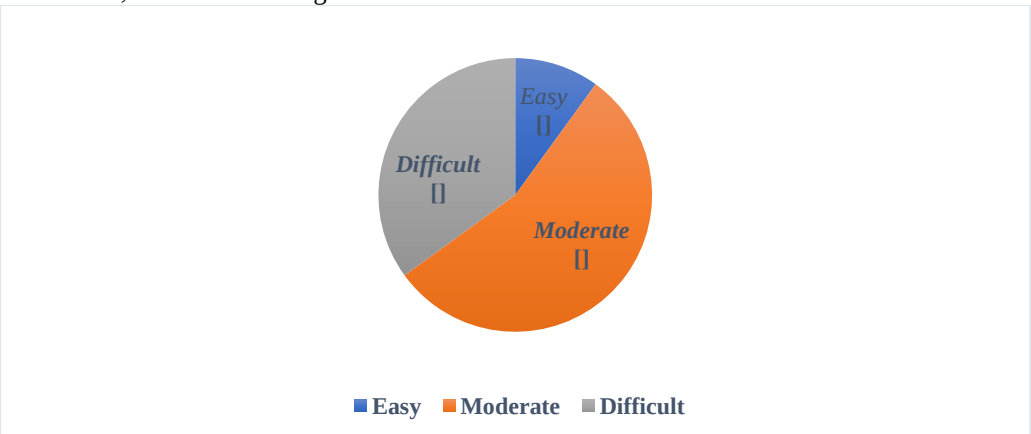
D. Quality of the Virtual Test

The quality of the virtual test in this study is assessed through two measures: the difficulty level and the discrimination power. The difficulty level of the multiple-choice questions in this study is categorized according to Farida (2014), using the following formula:

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

Explanation:
P : difficulty index
B : number of students who answered the question correctly
JS: total number of students participating in the test

Referring to the difficulty level categories proposed by Farida (2014), the average difficulty level in this study is 0.39, which falls into the moderate category. The percentage distribution of the difficulty level for the total of 20 test items, categorized as easy, moderate, and difficult, can be seen in Figure 4.



Interpretation: 0.00-0.29: Difficult; 0.30-0.69: Moderate; 0.70-1.00: Easy (Farida & Nuryantini, 2014)
Figure 4. Percentage Diagram of the Difficulty Level of the Virtual Test

Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

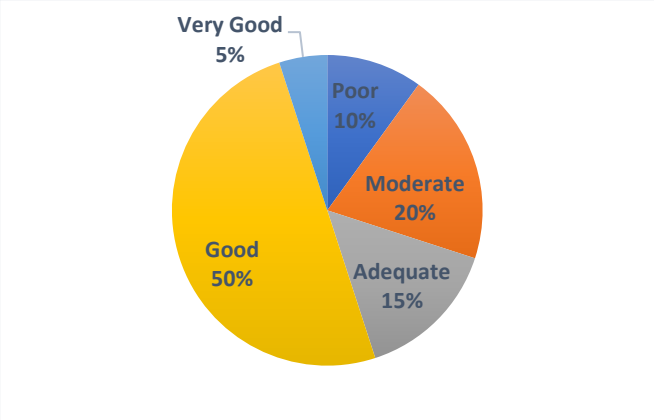
In detail, the results of the pilot test of the developed virtual test yielded 2 items (10%) categorized as "easy," 11 items (55%) categorized as "moderate," and 7 items (35%) categorized as "difficult." A good test item is one that is neither too easy nor too difficult. In this case, a good test item has a moderate difficulty level on average (Arikunto, 2015). Therefore, the virtual test in this study has met this criterion overall.

Furthermore, the discrimination power of the test items aims to assess the ability of each item to differentiate between students who have and have not mastered the tested material. The discrimination power (D) tested in this study was determined according to Farida (2014), and each test item was categorized based on its discrimination power, using the following formula:

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

Explanation:
D : Discrimination Power
B_A : number of students in the upper group who answered the item correctly
B_B : number of students in the lower group who answered the item correctly
J_A : number of students in the upper group
J_B : number of students in the lower group

The discrimination power of the virtual test, consisting of a total of 20 multiple-choice items, was assessed on 101 students in grade X of schools using the independent curriculum. The students were divided into upper and lower groups. The student scores were ranked from highest to lowest, and the top 27% of scores were allocated to the upper group, while the bottom 27% were allocated to the lower group (Farida & Nuryantini, 2014).
The results of the discrimination power test in this study are presented in the form of percentages, which can be seen in Fig 5.



Interpretation: 0.00-0.19: Poor; 0.20-0.29: Moderate; 0.30-0.39: Adequate; 0.40-0.69: Good; 0.70-1.00: Very Good (Farida & Nuryantini, 2014)
Figure 5. Percentage of Average Item Discrimination Power

In detail, based on the interpretation of discrimination power according to Farida (2014), there are 2 items (10%) categorized as poor, 4 items (20%) categorized as moderate, 3 items (15%) categorized as adequate, 10 items (50%) categorized as good, and 1 item (5%) categorized as very good. The overall average discrimination power for this virtual test is 0.42, which falls into the good category. Based on these results, it can be concluded that the developed virtual test for green chemistry can differentiate between students who have mastered the material and those who have not (Firman, 2013).

IV. CONCLUSIONS

The developed virtual test has met the criteria for validity and reliability with high or good categories. The virtual test is also considered suitable for use as it meets the criterion of readability, with almost all test items being easily readable. Additionally, the quality of the virtual test is considered good, with a moderate difficulty level and good discrimination power. Overall, the average virtual test can be considered suitable for measuring students' mastery of concepts in green chemistry.

ACKNOWLEDGEMENTS

Author thanks to the Supervisor Prof. Dr. Nahadi, M.Pd., M.Si. and Dr. Hernani, M.Si. who have provided support and guidance.

Volume 02, Number 02, December 2023
e-ISSN: 2830-7933
Doi: https://doi.org/10.55299/ijere.v2i2.489

REFERENCES

Adams, W. ., & Wieman, C. . (2010). Development and Validation of Instruments to Measure Learning of Expert-Like Thinking. *International Journal of Science Education*, *1*(1), 1–24.

Arikunto, S. (2015). *Dasar-Dasar Evaluasi Pendidikan* (2nd ed.). Jakarta: Bumi Aksara.

Farida, I., & Nuryantini, A. (2014). *Evaluasi Pembelajaran*. Bandung: Patragading.

Firman, H. (2013). *Evaluasi Pembelajaran Kimia*. Bandung: Jurusan Pendidikan Kimia FPMIPA UPI.

Firman, H. (2018). *Evaluasi Pembelajaran Kimia*. Bandung: Jurusan Pendidikan Kimia FPMIPA UPI.

Gilbert, E. ., & Prion, S. (2016). Making sense of methodes and measurenment: lawshe's content validity index. *Clinical Simulation in Nursing*, *12*, 530–531.

Grieger, K., Schiro, A., & Leontyev, A. (2022). Development of the Assessment of Student Knowledge of Green Chemistry Principles (ASK-GCP). *Chemistry Education Research and Practice Journal*.

Inayah, S., Dasna, I. W., & Habiddin, H. (2022). Implementasi Green Chemistry Dalam Pembelajaran Kimia: Literatur Review. *Hydrogen: Jurnal Kependidikan Kimia*, *10*(1), 42. https://doi.org/10.33394/hjkk.v10i1.4611

Kemendikbudristek. (2022). *Keputusan Nomor 008/H/Kr/2022 Tentang Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah Pada Kurikulum Merdeka*. Jakarta.

Koentjaraningrat. (1997). *Metode-Metode Penelitian Masyarakat - Metode Wawancara*. Jakarta: Gramedia Pustaka Utama.

Lawshe, C.H. (1975). *A Quantitative Approach to Content Validity*. Purdue University

Mayangsari, D. (2022). *Pengembangan dan Validasi Tes Virtual Berbasis Tiga Level Representasi Kimia untuk Mengukur Penguasaan Konsep Hidrolisis Garam*. Universitas Pendidikan Indonesia.

Valantika, L., Firman, H., & Nahadi. (2016). Development virtual test of reaction rate based visual perceptual skills to measure students' mastery concept. *Journal of Physics: Conference Series*, *755*(1). https://doi.org/10.1088/1742-6596/755/1/011001