❒ 44

Development and Validation of Chemistry Virtual Test Based Multiple Representations

Nahadi[1]*, Harry Firman[2], Hendri Kurniadi[3]
[1,2]Department of Chemistry Education, Universitas Pendidikan Indonesia, Indonesia
[3]SMK Terpadu Al-Itihad Cianjur, Indonesia

Article Info	ABSTRACT
	The purpose of this study was developing and validation virtual test based multiple representations to facilitate students to understand the questions that assess students' decision-making competence of buffer solution. This research was conducted using the method development and validation. The participants were 136 high school students who take science class. Data obtained were content validity, reliability, item difficulty index, discrimination index and readability of virtual test that has developed. This study also compared the item difficulty and readability between virtual test and paper and pencil test. The results showed that the virtual test had CVI's value was 0.71, the Cronbach's Alpha value was 0.925 which was showed that the reliability with included in "very good" category, the item difficulty index was moderate categories and the readability of virtual test was higher than paper and pencil test. It could be concluded that the virtual test that had developed was feasible to use and could facilitate students in understanding the questions that assess students' decision-making competence of buffer solution.

Corresponding Author:

Nahadi,
Department of Chemistry Education,
Universitas Pendidikan Indonesia,
Jl. Dr. Setiabudi 229 Bandung, Indonesia.
Email: nahadi@upi.edu

1. INTRODUCTION

National Exam is standard evaluation of primary and secondary education nationally and this is to equal the quality of education levels among regions that conducted by the Centre for Educational Assessment, Ministry of National Education in Indonesia, with the aim to provide guidance and assistance to the education unit in its efforts to improve the quality of education. One of which must be evaluated from the National Exam is that the quality of questions used, which the quality of the National Exam have to meet validity and reliability.

The chemistry item test on National Exam has not been able to measure the ability of high order thinking skills (HOTS). Preliminary study results showed that only six items (15%) were categorized at analysis level (C4), 22.5% on memorizing (C1), understanding (C2), and 35% on application level (C3) [1]. In 2015, 555 schools carried out National Exam with Computer-Based, consisting of 42 Junior High Schools, 135 Senior High Schools and 378 Vocational Schools in 29 provinces and Foreign Affairs. Based on the phenomenon that has been described, the test only measure the ability of low order thinking skills (LOTS) and multiple chemical representation media used in Paper Based Test (PBT) can not be shown in their entirety [2].

Study results also showed that the characteristic of the HOTS item test on National Exam for High School Rayon B Year 2012/2013 was the stimulus, with the form of the stimulus in National Exam were

pictures/graphs/charts (15%), table (15%), symbols/formula/chemical equations (47.5%), samples (22.5%), and a fragment of cases (32.5%). The research results prove that item test on National Exam need to be revisited in order to measure the HOTS. The item test that measure the high order thinking skills make participants to be able to think logically, critically and analytically. One high order thinking skills is the ability of decision-making [3].

Decision making is the process of thinking to identify and decide the choice of various options. The ability of decision-making is one aspect of Graduate Competency Standards on the curriculum in 2013. Making the right decision can be made if the student is able to understand the concept as a whole. Students are said to understand the chemical concepts thoroughly and deeply if the student is able to connect the three levels of the chemical representation (macroscopic, sub-microscopic and symbolic) [4], [5]. Representation level macroscopically obtained through observation using five senses, symbolic knowledge denotes symbols and equations of reactions and submicroscopic knowledge includes the representation of atoms, molecules, and ions using image or molecular model [6]. However, in fact student can't connect chemical representation. Buffer solution is concepts that need to be observed. The concept can also be explained in various forms of representation that can visualize these materials so that students are expected to observe the symptoms that occur, analyze and make the conclusions [7].

On the other hand, science and technology has brought its own impact to the world of education and also support the success of education. In the field of education, the use of computers is necessary to facilitate complex educational activities as well as in assessing students. Competency assessment has shifted toward the use of computer-based procedures [8]-[10]. The use of computer technology to assess students provides some benefits, In general, the implementation of the CBT system can saving time, reducing the burden of teachers, the assessment results are also faster and can be reducing the possible fraud of students during the exam [11]-[15]. The results of its assessment is more valid, more interactive and its assessment can be more contextual [16]. In addition, the results of test can be directly known and may indicate a student's weaknesses, so as to provide immediate feedback [17]. The computer is a medium of instruction to visualize different facts, skills, concepts and computer also displays images that move in accordance with its requirements [18].

Van Merrienboer said that importance to use a computer as an assessment and instructional tool, since the tool can simulate real-world problems, which are ill-structured and complex in nature [19]. Besides assessment by Computer Based Test, has advantages which the tests can explore the computer so that the test form with multimedia can be images, text, graphics, animation, video and others. One of the computer based test that has been and should be developed further is a virtual test. Virtual test is a test form using software that can be carried out both online and offline. From the above definition, it can be concluded that the virtual test is a form of testing by using multimedia computer such as images, text, graphics, diagrams, animations, and video.

Mushonev states that the use of visual forms in the test questions will help evaluators to measure students' cognitive abilities were higher compared to just use a statement or just questions and can train simultaneously measuring the ability of the students' science process [20]. The use of media in virtual test such as images, audio, animation, and video can help student understand the statement items are delivered. Multimedia effects provide sufficient retention great for students. This is so that make the question be more interactive. Type this assessment also received a good response from the students [21], [14].

Virtual test is developed by eight visual perceptual skills are developed Rocford and Archer with the aim to measure the level of student mastery concepts on the reaction rate material [22]. There are several research of virtual tests have been developed, but the virtual development of visual perceptual skills test based on specific chemical content has never been developed. Research that have been developed include Firman and Rusyati develop a virtual test used to measure critical thinking skills of junior high school students on the theme of human diseases, Saukani in his research to develop a virtual test used to measure the ability of decision-making high school students on the material solution acid-base, Anggarjati in his research to develop a virtual test used to measure critical thinking skills of high school students on the material Chemical Equilibrium, and Christian has developed VSCs (Visual Spatial Chemistry Specific) [23]-[25]. VSCs are assessment tools developed to assess visual perceptual skills of the students, where these skills are involved in the formation of a mental image of visual objects. Christian VSCs focuses on the representation of the molecule, and have not focused on a specific chemical content [22].

Based on phenomena that have been described, assessment only measure low-level thinking skills (low order thingking skills-LOTS) and media used in the form PBT so multiple representations chemistry can't be shown in full. Issues such as basic statements tests containing abstract concepts that are difficult to describe with words causing students to have difficulty understanding statement of principal test.Use virtual test can create types more interactive questions. With use virtual test allows for problem creators using images, graphics, animations, and deep videos problem-solving so as to clarify the intent of the underlying

statement test. Therefore, it needs to be developed virtual test based multiple representation which can measure decision-making ability in competence buffer.

2. RESEARCH METHOD

The method used was the Development and Validation to produce a virtual test that measured the ability of decision making on buffer solution [26]. The research included analysis of buffer solution concept and analysis indicators student's decision making capability; construct virtual test; validation of contents based on judgment of experts; revise draft of the product; trials; data analysis and conclusion. Virtual test was published in *exe* file format, can be played on any computer using the operating system based on windows without having to install software. Subjects in this study were students of class XI in Al-Ittihad Senior High School in Cianjur, academic year 2015/2016. The research instruments were a virtual test that provides a simple multiple-choice questions, written test and interview guide sheets. Valuation techniques were expert judgment, decision-making ability tests and interviews. The data obtained were the result of the students' answers to the decision-making ability tests and the results of interviews with teachers and students. Data processing was performed using the Microsoft Excel program. The data in this study include determining the validity of the instrument, the analysis of items which include difficulty index discrimination index, instrument reliability and concurrent validity and analysis of the results of interviews with teachers and students. Validation results were calculated using the CVR (Content Validity Ratio) and averaged using CVI (Content Validity Index). The validity test using Pearson Product Moment analysis and reliability test using Cronbach's Alpha formula.

3. RESULTS AND ANALYSIS

3.1. Development of Pencil and Paper Test and Virtual Test

Pencil and paper test and virtual test based on multiple representations of buffer solutions were developed based on curriculum applied in schools. The developed item problem refers to competence related to the real of knowledge. The material of the buffer solution lies in the basic competence 3.13, which is to analyze the role of the buffer solution in the body of a living being. Based on these competencies, then it was revealed to be integrated indicators of decision-making capabilities developed by Bavalon. The decision-making indicators developed by Bavalon are: (1) Analyze possible alternative answers and possible risks; (2) Be able to analyze the relevance of existing rules or concepts; (3) Detects incorrect answer errors; (4) Understanding irrelevant decision-making basics; (5) Integrate related values; (6) Evaluate yourself. The six indicators were then reduced to 21 sub-indicators, then from 21 sub-indicators were produced 31 items about buffer solutions based on multiple representations that refer to the decision-making indicator. Pencil and paper test and virtual test which was developed in the form of multiple choice test.

3.2. Quality of Pencil and Paper Test and Virtual Test

Content validity to test decision-making ability is used by expert judgment. There are two aspects that are assessed by the expert, were the suitability of the sub-indicator of decision-making with the question of pencil and paper test and the conformity of the sub-indicator of decision-making with the virtual test question. Content validation data about pencil and paper test and virtual test by experts are presented in table 1.

Based on CVR value calculation, 31 questions which were developed for paper and pencil test, 26 items of that were valid consist of 16 items with CVR value 1.00 and 10 items with CVR 0.71. While the other 5 questions were invalid because of the CVR value lower than 0.62 (2 items with CVR value 0.43 and 3 items with CVR value 0.14). 31 questions were developed for a virtual test, 22 items of that were valid consist of 11 items with CVR value 1.00 and 11 items with CVR value 0.71. While the other 9 questions were invalid because of the CVR value lower than 0.62 (6 items with CVR value 0.43 and 3 items with CVR value 0.14). Each 31 questions from paper and pencil test and virtual test were compared, and 22 questions were equal. In addition also was determined CVI value of paper and pencil test and virtual test. Based on the calculations, CVI value of paper and pencil test was 0.79 and a virtual test was 0.71.

The paper and pencil questions and virtual tests used in the trial test after validated content were valid and equivalent. The Trial was conducted on 136 students of class XI. The result of the validity of item about paper and pencil test and virtual test are presented in table 2.

Table 1. Recapitulation Table of Expert Assessment Pencil and Paper Test and Virtual Test For each Decision Making Indicator

The decision-making indicators	Problem No.	Pencil and Paper Test		Virtual Test	
		CVR	Conclusions	CVR	Conclusions
1. Analyze possible alternative answers and possible risks.	1	1,00	Valid	0,71	Valid
	2	0,71	Valid	0,71	
	3	1,00	Valid	0,71	Valid
2. Be able to analyze the relevance of existing rules or concepts.	4	0,71	Valid	0,71	Valid
	5	0,71	Valid	0,43	Not Valid
	27	1,00	Valid	1,00	Valid
	28	1,00	Valid	1,00	Valid
	6	1,00	Valid	1,00	Valid
	7	0,14	Not Valid	0,14	Not Valid
	8	0,14	Not Valid	0,14	Not Valid
	9	0,43	Not Valid	0,43	Not Valid
	10	0,14	Not Valid	0,14	Not Valid
3. Detects incorrect answer errors.	11	0,71	Valid	0,71	Valid
	12	0,71	Valid	0,71	Valid
4. Understanding irrelevant decision making basics.	13	0,71	Valid	0,71	Valid
	14	1,00	Valid	1,00	Valid
	15	0,43	Not Valid	0,43	Not Valid
	16	0,71	Valid	0,43	Not Valid
	17	0,71	Valid	0,71	Valid
	18	1,00	Valid	0,71	Valid
	29	0,71	Valid	0,71	Valid
	30	1,00	Valid	1,00	Valid
	31	1,00	Valid	1,00	Valid
5. Integrate related values.	19	1,00	Valid	0,71	Valid
	20	1,00	Valid	1,00	Valid
	21	1,00	Valid	1,00	Valid
	22	1,00	Valid	1,00	Valid
	23	1,00	Valid	1,00	Valid
	24	0,71	Valid	0,43	Not Valid
	25	1,00	Valid	0,43	Not Valid
6. Evaluate yourself.	26	1,00	Valid	1,00	Valid

Table 2 Recapitulation Validity of Item about Paper and Pencil Test and Virtual Test

The decision-making indicators	Problem No.	Validity Item Problem			
		Paper and Pencil Test	Category	Virtual Test	Category
1. Analyze possible alternative answers and possible risks.	1	0,542	Enough	0,657	High
	2	0,596	Enough	0,711	High
	3	0,679	High	0,786	High
2. Be able to analyze the relevance of existing rules or concepts.	4	0,525	Enough	0,772	High
	5	0,502	Enough	0,633	High
	6	0,728	High	0,543	Enough
	7	0,533	Enough	0,732	High
	8	0,541	Enough	0,772	High
3. Detects incorrect answer errors	9	0,574	Enough	0,683	High
	10	0,502	Enough	0,557	Enough
4. Understanding irrelevant decision making basics.	11	0,707	High	0,677	High
	12	0,544	Enough	0,786	High
	13	0,500	Enough	0,786	High
	14	0,630	High	0,772	High
	15	0,728	High	0,732	High
	16	0,728	High	0,671	High
	17	0,728	High	0,655	High
5. Integrate related values.	18	0,552	Enough	0,659	High
	19	0,709	High	0,632	High
	20	0,729	High	0,657	High
	21	0,711	High	0,633	High
6. Evaluate yourself.	22	0,535	Enough	0,557	High

Based on the Table 2., 22 items developed, there were 10 items of paper and pencil test that the category is high and 12 items enough category, whereas in a virtual test 19 items were a high category and 3 items enough category. Figure 1 shows the percentage of the validity of paper and pencil test items and virtual test
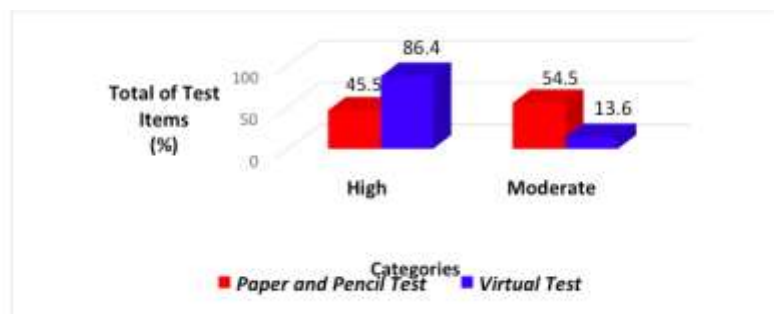
Figure 1. Comparison of the validity of the items of paper and pencil test and virtual test

Trials conducted to obtain empirical validity, the results were analyzed with SPSS 23, the results of data analysis showed that 86.4% of items test in virtual test was in the high category and 13.6% was at enough category, while for paper and pencil test was 45.5% in the high category and 54.5% was at enough category. From the data, the validity of items test in a virtual test that was at the high category was greater than paper and pencil test. Test had high validity if the test was to measure the ability of students in chemistry material that had been taught, so it could be concluded that the items test of virtual test that had developed was capable to measure the decision-making ability of buffer solution.

Reliability with very good category means that the test had been developed had a very good consistency in measuring the ability of student decision-making so that when tested to other students with the same conditions and circumstances will produce the same information or close to the same. Based on the analysis of the trial results by using SPSS 23 obtained Cronbach's Alpha value for paper and pencil test was 0.925 and for a virtual test was 0.947. Based on the value of Cronbach's Alpha was known that the reliability of the two types of tests was "very good". It was based on the categories of interpretation performed by George and Mallery that the reliability with Cronbach's Alpha value 0.910 to 1.000, the reliability included in the category of "very good" [27]. Reliability with the category of "very good" means that the items test had been developed had excellent consistency in measuring decision-making ability of students so that when tested to other students with the same conditions and state it would produce the same information or close to the same.

The difficulty index indicates the relationship between the student's probability of answering the question correctly or incorrectly. Recapitulation of difficulty index about paper and pencil test and virtual test Decision-making capabilities are represented in table 3.

Table 3. Recapitulation of Difficulty Index about Paper and Pencil Test and Virtual Test

The decision-making indicators	Problem No.	Difficulty Index			
		Paper and Pencil Test	Category	Virtual Test	Category
1. Analyze possible alternative answers and possible risks.	1	0,30	Medium	0,35	Medium
	2	0,56	Medium	0,58	Medium
	3	0,52	Medium	0,57	Medium
	4	0,51	Medium	0,53	Medium
2. Be able to analyze the relevance of existing rules or concepts.	5	0,29	Medium	0,35	Medium
	6	0,38	Medium	0,36	Medium
	7	0,4	Medium	0,41	Medium
	8	0,49	Medium	0,53	Medium
3. Detects incorrect answer errors	9	0,58	Medium	0,61	Medium
	10	0,29	Difficult	0,38	Medium
	11	0,38	Medium	0,39	Medium
	12	0,43	Medium	0,57	Medium
	13	0,49	Medium	0,57	Medium
4. Understanding irrelevant decision making basics.	14	0,51	Medium	0,53	Medium
	15	0,38	Medium	0,41	Medium
	16	0,38	Medium	0,40	Medium
	17	0,38	Medium	0,41	Medium
	18	0,57	Medium	0,60	Medium
Integrate related values.	19	0,32	Medium	0,34	Medium
	20	0,32	Medium	0,35	Medium
	21	0,32	Medium	0,35	Medium
Evaluate yourself.	22	0,34	Medium	0,38	Medium

     Based on the data, difficulty index for paper and pencil test varies ranged from 0.29 to 0.58 with an average of 0.42. From 22 questions of paper and pencil test that was developed, difficulty index for 2 questions was "difficult" and 20 questions were "moderate". While the difficulty index for virtual test ranged from 0.34 to 0.61 with an average of 0.45. Difficulty index of 22 questions of virtual test that was developed were "moderate". Based on these results it could be seen that the virtual test had the moderate items test more than the paper and pencil test and in the virtual test, there was no difficult item test if it's compared with paper and pencil test. This was shown that the items test of the virtual test had difficulty index higher than paper and pencil test, it was also indicated that the items test of a virtual test was easier to understand than paper and pencil test [27].

     Discrimination index is the ability of items to distinguish between students who have mastered the material in question and the students who have not mastered the material being tested. Recapitulation of the differentiating index of paper and pencil test and virtual test are represented in table 4. Percentage differentiating index for paper and pencil test and virtual test are presented in Figure 2.

Table 4 Recapitulation of The Differentiating Index of Paper and Pencil Test and Virtual Test

| The decision-making indicators		Problem No.	differentiating index			
			Paper and Pencil Test	Category	Virtual Test	Category
1.	Analyze possible alternative answers and possible risks.	1	0,44	Good	0,64	Good
		2	0,55	Good	0,656	Good
		3	0,61	Good	0,67	Good
		4	0,40	Good	0,67	Good
2.	Be able to analyze the relevance of existing rules or concepts.	5	0,39	Good	0,62	Good
		6	0,66	Good	0,44	Good
		7	0,51	Good	0,58	Good
		8	0,56	Good	0,67	Good
3.	Detects incorrect answer errors	9	0,52	Good	0,60	Good
		10	0,39	Good	0,45	Good
		11	0,64	Good	0,55	Good
		12	0,36	Enough	0,67	Good
4.	Understanding irrelevant decision making basics.	13	0,51	Good	0,67	Good
		14	0,57	Good	0,67	Good
		15	0,66	Good	0,58	Good
		16	0,66	Good	0,56	Good
		17	0,66	Good	0,55	Good
		18	0,5	Good	0,62	Good
5.	Integrate related values.	19	0,67	Good	0,60	Good
		20	0,69	Very Good	0,64	Good
		21	0,67	Good	0,62	Good
6.	Evaluate yourself.	22	0,48	Good	0,45	Good
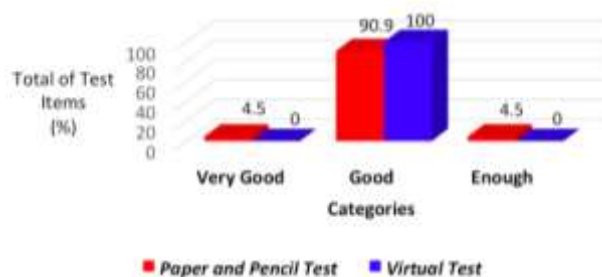
Figure 2. Percentage of discrimination index for paper and pencil test and virtual test

     Based on data, items test of the virtual test had discrimination index values higher than paper and pencil test. Virtual test items had discrimination index ranges from 0.44 to 0.67 with an average 0.60 (good categories), while paper and pencil test had discrimination index ranges from 0.36 to 0.69 with an average of 0.55 (good categories). The greater of discrimination index of the question the more obvious differences between the high group and the low group [28]. From 22 questions of paper and pencil test that was developed, 1 question was in "very good" category and 20 questions were in "good" category and the other 1

was in "enough" category. The trial was conducted to determine the readability to the items test that was developed.

In the matter of developed virtual test is expected students can easily answer the problem compared with the matter of paper and pencil test. With the easier the students answer the question correctly will certainly cause differentiation of the matter will be lower because of the upper group and the lower group value is not too different. Differentiating power obtained on the virtual test shows that the number of low and moderate groups are correct in answering the problem so that the differentiation of virtual test is lower than the paper and pencil test. This certainly shows that the virtual test allows students to answer questions.

The Trial test question is done using a questionnaire that contains four aspects of assessment, clarity on articles, pictures, videos, tables or animations; use of language that is easy to understand; understanding of problems in question; and an understanding of the choice of answers. Based on the test legibility, legibility of paper and pencil test is 88.8%, while the virtual test legibility 95.6%. This shows that the readability of the question in the form of virtual test is greater than the matter of paper and pencil test. Based on the T-test with a significance value of 0.001 (<0.05), this means that there is a significant difference to the level of legibility of paper and pencil test with a virtual test [19].

In addition to validity test of the content, it did also correlation test between virtual test and paper and pencil test. Data related to the analysis of virtual tests and paper and pencil test, which the value of the correlation between virtual test and a paper and pencil test was 0.921 included in "very high" category, so it could be considered that the items test in virtual tests and paper and pencil test had a high relationship.

The existence of paper and pencil test is still widely used in schools, but the problem in this form has many weaknesses because it is ineffective and efficient. Therefore the question in the form of virtual test can reduce the weaknesses that exist in the matter of paper and pencil test. This is in line with the application of computer based test can save time, reduce the burden of teachers, the assessment results are also faster and can reduce the possible fraud of students, and make it easier for students to understand the problem because the use of technology in computer based test can accommodate complex cognitive appraisal processes that can't be accommodated with paper based test [11], [13], [15].

This virtual test which development had a positive response from both students and teachers, it could be seen from the students' responses, that the virtual test very interesting. Implementation of a virtual test at school is still rarely is done, this is because the school infrastructure is not yet complete. The Virtual test can be used as a solution for education, for improvement and measurement of high-order thinking skills, one of that is to measure the decision-making ability [12].

4.   CONCLUSION

Based on research of the development and validity of virtual test to measure the decision-making ability at buffer solution can be concluded that: multiple representations-based virtual test that have been developed had CVI value of 0.71 and included "high" category. It also had reliability with "very good" category. The virtual test based multiple representations that had been developed had difficulty index in "moderate" category. In general, the readability of items test in virtual test was more than in paper and pencil test. Interview result showed that virtual test based multiple representations that had been developed had a positive response from teachers and students in terms of appearance, grammar of the questions, ease of access and operation, multimedia contained in test, benefits, and time.

REFERENCES

[1]   A. Satrisman, "Analisis Soal Ujian Nasional Kimia SMA Tahun 2013 Berdasarkan Taksonomi Bloom Dua Dimensi," Proseeding UNJ Jakarta. pp. 48-56, 2014.

[2]   W. Sarjono, "Ujian Nasional Berbasis Komputer (UN BK/Computer Based Test) di SMK Negeri 1 Jakarta Tahun Pelajaran 2014/2015." Mercubuana Jakarta. 2015.

[3]   Lailly, et al.. "Analisis Soal Tipe Higher Order Thinking Skill (HOTS) Dalam Soal UN Kimia SMA Rayon B Tahun 2012/2013." *Kaunia*. 9 (4) pp. 27-39, 2015.

[4]   Sunyono, "Validitas Model Pembelajaran Kimia Berbasis Multipel Representasi Untuk Meningkatkan Model Mental Siswa Pada Topik Struktur Atom. Jurnal Pendidikan Kimia FKIP Universitas Lampung," 3 (1) pp. 1 – 9, 2012.

[5]   I. W. Suja, "Strategi "Ermo" Dalam Pengajaran Konsep-Konsep Kimia Abstrak-Teoritis," Doctoral Thesis, Universitas Negeri Surabaya, 2014.

[6]   A.L., Chandrasegaran, D.F. Treagust, & M. Mocerino,. "The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation". *Chemistry Education Research and Practice*. 8(3), pp. 293–307, 2007.

[7]   J.W. Russell, et al., "Use of Simultaneous-Synchronized Macroscopic, Microscopic, and Symbolic Representations to Enhance the Teaching and Learning of Chemical Concepts" *Journal of Chemical Education*. 74.(3). pp. 330-334, 1997.

[8]   A.Jurecka, "Introduction to the Computer-Based Assessment of Competences. Cambridge/Gottingen; Hogrefe & Huber Publishers. pp. 193-214, 2008.

[9]   D.H. Jonassen, "Learning to Solve Problems; An Instructional design guide". San Fransisco; John Wiley & Sons. 2004.

[10] J., Trindade, et al. "Science Learning in Virtual Environments; a descriptive study," *British Journal of Educational Technology*. 33 (4) pp. 45-55, 2002.

[11] S. M. Bodmann, and D. H. Robinson, "Speed and Performance Differences among Computer-Based and Paper-Pencil Tests" *Journal of Educational Computing Research*. 31(1). pp. 51-60, 2004.

[12] M., Jamil, K. J., Topping, R. H. Tariq, "Perceptions of university students regarding computer assisted assessment," *The Turkish Online Journal of Educational Technology*. 11 (3). pp. 267-277, 2012.

[13] B. J., Mason, M., Patry, D. J. Bernstein,). "An Examination of The Equivalence Between Non-Adaptive Computer-Based and Traditional Testing," *Journal of Educational Computing Research*. 24(1). pp. 29 – 39, 2001.

[14] V. Terzis, and A. A. Economides, "The acceptance and use of computer based assessment," *Computers and Education*. 56. pp. 1032-1044, 2011.

[15] M.Thelwall, "Computer-based assessment: a versatile educational tool," *Computers and Education*. 34. pp. 37-49. 2000.

[16] K.Scalise, and B.Gifford, "Computer-Based Assessment in E-Learning: A Framework for Constructing Intermediate Constraint Questions and Tasks for Technology Platforms," *The Journal of Technology, Learning, and Assessment* 4(6). 2016.

[17] M., Krause, et al. "On the development and assessment of a computer based learning and assessment environment for the transition from lower to upper secondary chemistry education," *Journal Royal Society of Chemistry.Chemical Education Research & Practice.* 14. pp. 345-353, 2013.

[18] P.Coburn, "Practical Guide to Computer in Education 2nd. California" Addison-Wesley Publication Company Inc. 1985.

[19] R., Scherer, et al. "Developing a computer-based assessment of complex problem solving in Chemistry," *International journal of STEM Education*, 2014.

[20] R. E. Mayer, "Multimedia learning. New York" Cambridge University Press. 2001.

[21] M, Jawaid, F. A., Moosa, F., Jaleel, and J. "Ashraf Computer Based Assessment (CBA): perception of residents at Dow University of health sciences". *Pakistan Journal of Medical Sciences*. 30 (4). pp. 688-691, 2014.

[22] C. M. Christian, "The Development and Validation of a Visual-Spatial Chemistry Specific (VSCS) Assessment Tool," *Journal of Research in Science Teaching*. 51(8). 963–981, 2014.

[23] H. Firman, and L. Rusyati, "Virtual Test: Sebuah Student Center Software Sebagai Alat Ukur Berpikir Kritis Siswa SMP Pada Tema Penyakit Manusia (Laporan PPKBK)," Bandung: Universitas Pendidikan Indonesia. 2014.

[24] M. Saukani, "Pengembangan dan Validasi Virtual Test untuk Mengukur Kemampuan Pembuatan Keputusan (Decision Making) Siswa pada Materi Larutan Asam Basa," Tesis, Universitas Pendidikan Bandung, 2015.

[25] A.D. Anggarjati, "Pengembangan dan Validasi Virtual Test untuk Mengukur Kemampuan Berpikir Kritis Siswa Pada Materi Kesetimbangan Kimia". Tesis, Universitas Pendidikan Bandung. 2015.

[26] W. K., Adam, C.E. Wieman, "Development and Validation of Instruments to Measure Learning of Expert-Like Thinking," *International Journal of Science Education*. 33(9) pp. 1-24, 2010.

[27] J. A Gliem, and R. R. Gliem, "Calculating, Interpreting, And Reporting Cronbach's Alpha Reliability Coefficient for Likert-types Scales," Columbus: Ohio State University. 2003.

[28] H. Firman, "Evaluasi Pembelajaran Kimia. Bandung" Jurusan Pendidikan Kimia FPMIPA Universitas Pendidikan Indonesia. 2013.