Development of Evaluation tool using Exam view Test Generator to measure students' critical thinking skills at SMAN 4 Padangsidimpuan

R Ariaji[1*], Abubakar[1], S A Siregar[1], A F D Harahap[2], S Siregar[3], and F H Marpaung[4]

[1]Chemistry Education Study Program, Faculty of Teacher Training Education,Universitas Muhammadiyah Tapanuli Selatan, Padangsidimpuan, Indonesia.
[2]Economic Education Study Program, Faculty of Teacher Training Education, Universitas Muhammadiyah Tapanuli Selatan, Padangsidimpuan, Indonesia.
[3]Program Studi Administrasi Publik, Faculty of Teacher Training Education, Universitas Muhammadiyah Tapanuli Selatan, Padangsidimpuan, Indonesia
[4]Pancasila and Civic Education Study Program, Faculty of Teacher Training Education, Universitas Muhammadiyah Tapanuli Selatan, Padangsidimpuan, Indonesia.

* rizki.ariazi@um-tapsel.ac.id

Abstract. A learning strategy that is quite varied so that it needs a more practical and creative evaluation tool. Evaluation is still manual, giving students an opportunity to share their answer keys during the exam. This research aims to implement an evaluation tool using electronic media or commonly called electronic test (E-Exam) of chemical material in grade X SMAN 4 Padangsidimpuan. In order to change the teacher's behavior of the learning pattern/assessment traditionally moved/migrated towards communicative system by utilizing information technology both offline and online. The study aims to develop a Exam View Test Generator Evaluation tool in accordance with BSNP standards, due diligence on expert validator lecturer and chemical teachers, as well as to the critical thinking ability of students of chemistry lessons in X-grade odd semester SMAN 4 Padangsidimpuan. Research and Development (R&D) methods use the Model DDD-E (Decide, Design, Develop, Evaluate). The results of the research with the average of expert validator feasibility test scores of 3.59 are Valid and do not need revision (worth). Based on the results of the hypothesis testing on the critical thinking ability of the students acquired SIG value. < 0.05 ie = 0.46 < 0.05 so that Ha was accepted and Ho was rejected.

1. Introduction

Education is a conscious and well-planned endeavor to create a learning atmosphere and learning process so that learners actively develop their potential to possess religious spiritual strength, self-control, personality, intelligence, and skills required by him and his community [1]. Learning is a system consisting of various components interconnected with each other, such components include: purpose, material, method, and evaluation of learning [2].

The term critical thinking is often likened to mean by thinking convergent, logical thinking and reasoning. As the formulation of the essence of critical thinking also varies. Critical thinking is the ability to solve problems by thinking seriously, actively, thoroughly in analyzing all the information

they receive by including rational reasons so that every action is done properly [3]. One of the most important skills to have and develop students is the ability to think critically with the critical thinking process of students will analyze, rethink, or raise new ideas [4]. Critical thinking is a very essential ability, and works effectively in the optimal learning process and requires critical thinking from the learner [5]. Students ' experience in critical thinking during the science learning and topic characteristics process appears as a reason behind students ' critical thinking skills level on specific science topics [6].

A training generator software developed with a virtual platform in enhancing the learning process and evaluation of subjects based on problems. For students as training, with content on a specific subject, it can monitor their performance and as an effective feedback. This tool is used in classroom learning on face-to-face, mixed or online learning [7]. Electronic test (E-Exam) is a form of test by utilizing electronic media especially using computer or commonly called Computer Based Test (CBT). CBT can be used offline as well as online. With such innovations should be able to change the teacher behaviour of the learning pattern and assessment traditionally moved/migrated towards communicative system by utilizing information and communication technology Computer Based Test (CBT) One solution of the assessment problem is in Exam View Test Generator software [8]. Material understanding has a positive effect on the skills to develop learning instruments. Assessment as a process of collecting information on students ' performance to make decisions in knowing the success of student learning activities Program [9].

This Software can also help students in terms of evaluation issues, in accordance with the regulation of the Minister of Education and Culture of the Republic of Indonesia No. 3 year 2017 article 9 paragraph 1 on the implementation of national examinations conducted through national-based examinations Computer. Muhadjir Effendy's Kemendikbud target, in the year 2019 all HIGH School/SMK will apply UNBK (100 percent). SMA and SMK next year can be implemented UNBK, which is still constrained by SMP, because there are many in the remote area [10]. SMAN 4 Padangsidimpuan will implement UNBK, with the software Exam View Test Generator will then be able to train students in the implementation of UNBK for the following years. The study aims to develop a Exam View Test Generator Evaluation tool in accordance with BSNP standards, due diligence on expert validator lecturer and chemical teachers, as well as to the critical thinking ability of students of chemistry lessons in X-grade odd semester SMAN 4 Padangsidimpuan.

2. Methods

This research was conducted at SMAN 4 Padangsidimpuan which is located at Jl. Sutan Soripada Mulia No. 38 Village Sadabuan subdistrict Padangsidimpuan Utara City Padangsidimpuan North Sumatra province. Research and Development (R&D) methods by combining the model of ADDIE (Analyze, Design, Development, Implementation, and Evaluation) and Model 4-D include (define, design, develope, and disseminate) so that the research uses the model DDD-E (Decide, Design, Develop, Evaluate) [11] [12].

The data collection techniques in this research are the data used in the development of the Exam View Test Generator Evaluation tool is a qualitative and quantitative data. Qualitative Data is obtained from the results of the interviews and questionnaire of feasibility test on expert validators and chemical teachers to analyze student needs, suggestions and comments. Quantitative Data is obtained from the test results (Pretest and Postest), and the test results of the students ' critical thinking ability. Classter sampling technique random sampling, i.e. take 2 (two) classes randomly from 7 classes. The first class, referred to as the experimental class, is a class that is given treatment using the Exam View Test Generator Evaluation Tool and the second class is called the control class i.e. the class that is given treatment using the usual evaluation tools applied to teachers.

Data analysis technique is product feasibility in the form of qualitative data analysis obtained from the needs analysis, questions in polls, criticism and advice from the expert validator. Data from polls is a qualitative data that is quoted through the scale of Likert score 1 and the highest score is 5. Determination of range can be known through the highest score range minus the lowest scoring range divided by the highest score. Then, the effectiveness of the product is a test of normality, homogeneity test, and hypothesis testing.

3. Results and Discussion

1) Decide

At this stage have 3 phases, namely:

a) Setting Goal evaluation Tool

Students still use manual evaluation systems (using paper) that cause students to share their answer keys with each other. Therefore, in the presence of evaluation tool ExamView Test Generator is expected that students can study independently and integrated.

b) Determining the materials used in the evaluation tool

Select material in the odd semester X class.

c) Assessing resources

Based on observations at SMAN 4 Padangsidimpuan, SMA has adequate computer laboratory facilities, and can be used to the fullest.

2) Design

There are 3 phases in this stage, namely: (1) Install the Exam View Test Generator program on the laptop, (2) Make a question using the Internet network (online), (3) Make a question without using the Internet (offline).

Figure 1. Exam view test generator program online versions

Figure 2. Exam view test generator program offline versions

3) Develop

a) Assessment Tool Evaluation Exam View Test Generator by expert Validator

Assessment of Evaluation tool by 4 people of expert validator consisting of 2 lecturers and 2 teachers of chemistry. The expert tried to run the program then asked for input and criticism through the poll given. This expert assessment is done to know the eligibility of the Exam View Test Generator Evaluation tool can be seen in the table 1.

Table 1. Assessment of the overall evaluation tool of experts

No	Validator	Standard Rating				
		Content Eligibility	Language Eligibility	Presentability	Grafapia Eligibility	Total Amount
1	Lecturer	3,79	3,45	4,25	3,82	
2	Teacher	3,35	3,22	3,50	3,36	
Amount		7,14	6,67	7,75	7,18	
Average		3,57	3,33	3,87	3,59	3,59
Criteria		Valid	Valid	Valid	Valid	Valid

In table 1 It appears that the average value of due diligence based on the BSNP poll is: (1) The standard of content eligibility is 3.79 by a lecturer validator and 3.35 by a chemical teacher validator. (2) The language eligibility standard is 3.45 by the faculty validator and 3.22 by the chemical teacher Validator. (3) The standard presentation eligibility is 4.25 by a lecturer's validator and 3.50 by a chemical teacher validator. (4) The standard of graphic worthiness is 3.82 by a lecturer's validator and 3.36 by the chemical teacher Validator.

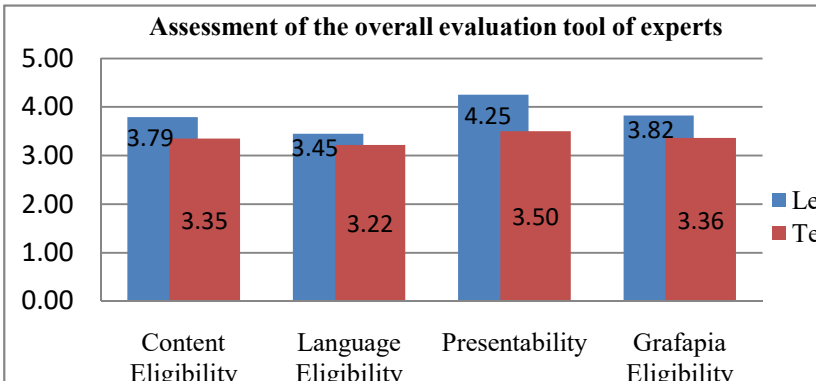

Figure 3. Expert validator assessment chart

For the average value of all aspects of the eligibility standards based on BSNP is 3.59 validation criteria {Valid and no need for revision}.

b) Testing Requirements Analysis

Before conducting the test statistic the initial steps needed to be performed is screening the data to be processed. One assumption of using parametric statistics is that each variable is normally distributed and homogeneous. It aims to reduce barriers in planned analysis. The prerequisite tests conducted include normality testing, homogeneity testing.

Table 2. Summary of normality and homogeneity test results

Variable	The value of Sig.	Significance level	Conclusion
Tool evaluation Exam	0,16	>0,05	Normal distribution
View Test Generator	0,78	> 0,05	Homogeneous
Student Critical	0,22	>0,05	Normal distribution
Thinking ability	1,00	> 0,05	Homogeneous

After fulfilling the requirements of normality and homogeneity, a hypothesis test is performed. The hypothesis testing was done with a simple linear regression test using SPSS. The hypothesis test compares with the significance value (Sig) by 0.05. Based on the above output is known significance value (Sig.) of 0.46 < 0.05, so it can be concluded that H0 rejected and Ha accepted, which means that there is an impact tool evaluation using the Exam View Test Generator against the ability of thinking Critical to students.

Table 3. Hypothesis test results

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig
		B	Std. Error	Beta		
1	(Constant)	64.893	8.304		7.815	.000
	Tool evaluation Exam View Test Generator	.085	.115	.091	.743	.460

a. Dependent Variable: Student Critical Thinking ability

4) Evaluate

Based on the test results and the assessment discussion given by experts on the evaluation tool using Exam View Test Generator gained an average total of 3.59, according to expert lecturers gained an average of 3.82 and the results are Subjects with an average of 3.35. The overall average amount gained from both expert lecturers and teacher subjects indicated the evaluation tool using the Exam View Test Generator which was developed worthy of use as an evaluation tool in class X SMAN 4 Padangsidimpuan. After doing the data analysis calculation is obtained test result normality indicating that the data obtained from the experimental class and from the control class is normal distribution. Whereas homogeneity test results indicate that the data obtained from the experimental and control classes comes from a homogenized variance. Based on the hypothesis testing results, it is derived that the value of sig. < 0.05 is = 0.46 < 0.05 so that Ha is accepted and Ho is rejected. This means that the evaluation tool using the Exam View Test Generator has an influence on the ability of critical thinking than the teacher using a manual evaluation tool (using paper) that the teacher or evaluate is checking Throughout the design and development process.

The willingness of teachers on computer-based assessments provides relevant insights to improve computer-based exam implementation, computer-based exams benefit young growing generation with computer and digital technology [13]. Computerised Adaptive Beat Alignment Test (CA-BAT), a variant of the Beat Alignment Test (BAT) that utilizes the latest advancements in psychometric theories, including item response theory, adaptive testing, and automatic item creation [14]. Online services generate new requests for test and shift activity between services, requiring system-level analysis to effectively predict cost effectiveness [15]. The distance education/course offered online is growing rapidly, to measure the level of learning the survey is used to evaluate the level of taste and meaning that is owned by each lecture, assignment, and exam [16]. A computerized cognitive assessment is an essential tool in behavioral sciences, testing three gamified variants of the Go-No-Go task, delivered both in the lab and online [17]. An integrated eClass evaluation experiment that considers multiple dimension LMS trial focus using online platform [18]. The development of the CAT-NE (Clinical Assessment Tool for Nursing Education) explains learning objectives, improves process assessment focus, and makes evaluation more objective [19].

4. Conclusions

The development process of the Exam View Test Generator Evaluation Tool uses DDD-E model (Decide, Design, Develop, Evaluate). The average value of all aspects of eligibility standards based on BSNP on criteria is Valid and does not need revision (feasible). There is a influence of evaluation tool Exam View Test Generator against the students ' critical thinking ability in SMAN 4 Padangsidimpuan.

Acknowledgments

The author thanked the Director of research and community service, Ministry of Research, Technology, and Higher Education Directorate General of strengthening Research and development.

References

[1] Musriadi, *Profesi Kependidikan Secara Teoritis dan Aplikatif*. Yogyakarta: Deepublish, 2016.
[2] Rusman, *Belajar dan Pembelajaran Berorientasi Standar Proses Pendidikan*. Jakarta: Kencana, 2017.

[3] H. Liberna, "Peningkatan Kemampuan Berpikir Kritis Matematis Siswa Melalui Penggunaan Metode Improve pada Materi Sistem Persamaan Linear Dua Variabel," *Form. J. Ilm. Pendidik. MIPA*, vol. 2, no. 23, pp. 190–197, 2015.
[4] E. Prihartini, P. Lestari, and S. A. Saputri, "Meningkatkan Kemampuan Berpikir Kritis Matematis Menggunakan Pendekatan Open Ended," *Pros. Semin. Nas. Mat. IX 2015*, pp. 58–64, 2016.
[5] D. Ahmatika, "Peningkatan Kemampuan Berfikir Kritis Siswa Dengan Pendekatan Inquiry/Discovery," *J. Euclid*, vol. 3, no. 1, pp. 394–403, 2016.
[6] N. I. Maulida, H Firman, and L. Rusyati, "Profile of Students' Critical Thinking Skill Measured by Science Virtual Test on Living Things and Environmental Sustainability Theme," *J. Phys. Conf. Ser.*, vol. 812, pp. 1–6, 2017.
[7] O. A. González-Estrada, J. F. Cáceres-Ramírez, and W. Pinto-Hernández, "Virtual tool for training and evaluation of problem-based engineering subjects," *J. Phys. Conf. Ser.*, vol. 1161, pp. 1–8, 2019.
[8] S. Siregar, *Sukses Mebuat Bank Soal dengan ExamView*. Yogyakarta: Deepublish, 2017.
[9] T. I. Prasetya, "Meningkatkan Keterampilan Menyusun Instrumen Hasil Belajar Berbasis Modul Interaktif Bagi Guru-Guru IPA SMP N Kota Magelang," *J. Educ. Res. Eval.*, vol. 1, no. 2, pp. 106–112, 2012.
[10] Y. Medistiara, "Mendikbud Targetkan SMA/SMK 100% UNBK Tahun 2019," *news.detik.com*, Kebayoran Baru, Jakarta Selatan, p. 2/4/2018, Apr-2018.
[11] M. Tegeh and M. Kirna, "Pengembangan Bahan Ajar Metode Penelitian Pendidikan dengan ADDIE Model," *J. IKA*, vol. 11, no. 1, pp. 12–26, 2013.
[12] Ayang Kinasih, W. Sunarno, and Sukarmin, "Pengembangan Modul Fisika Dengan Pendekatan Keterampilan Proses Sains Pada Materi Listrik Dinamis Untuk Meningkatkan Kemampuan Berpikir Kreatif Siswa Kelas X SMA," *J. INKUIRI*, vol. 7, no. 1, pp. 29–38, 2018.
[13] A. J. Boevé, R. R. Meijer, C. J. Albers, Y. Beetsma, and R. J. Bosker, "Introducing Computer-Based Testing in HighStakes Exams in Higher Education: Results of a Field Experiment," *PLoS One*, vol. 10, no. 12, pp. 1–13, 2015.
[14] P. M. C. Harrison and D. Müllensiefen, "Development and Validation of the Computerised Adaptive Beat Alignment Test (CA-BAT)," *Sci. Rep.*, vol. 8, no. 12395, pp. 1–20, 2018.
[15] K. M. E. Turner, K. J. Looker, J. Syred, A. Zienkiewicz, and P. Baraitser, "Online testing for sexually transmitted infections: A whole systems approach to predicting value," *PLoS One*, vol. 14, no. 2, pp. 1–14, 2019.
[16] J. Domenghini, D. Bremer, S. Keeley, J. Fry, C. Lavis, and S. Thien, "Assessing Student Learning with Surveys and a Pre-Test/Post-Test in an Online Course," *Nat. Sci. Educ.*, vol. 43, no. 1, pp. 109–116, 2014.
[17] J. Lumsden, A. Skinner, A. T. Woods, N. S. Lawrence, and M. Munafò, "The effects of gamelike features and test location on cognitive test performance and participant enjoyment," *PeerJ*, pp. 1–19, 2016.
[18] K. Kabassi, I. Dragonas, A. Ntouzevits, T. Pomonis, G. Papastathopoulos, and Y. Vozaitis, "Evaluating a learning management system for blended learning in Greek higher education," *Springerplus*, vol. 5, no. 101, pp. 1–12, 2016.
[19] H. Skúladóttir and M. H. Svavarsdóttir, "Development and validation of a Clinical Assessment Tool for Nursing Education (CAT-NE)," *Nurse Educ. Pract.*, vol. 20, pp. 31–38, 2016.