Biosfer: Jurnal Pendidikan Biologi, 16 (2), 220-231 (2023)

Biosfer: Jurnal Pendidikan Biologi

Journal homepage: http://journal.unj.ac.id/unj/index.php/biosfer

Developing critical thinking skills assessment of excretory system

Yenny Anwar*, Adeng Slamet, Ayu Tania

Biology Education, Faculty of Mathematics and Natural Science, Universitas Sriwijaya, Indonesia

*Corresponding author: yenny_anwar@unsri.ac.id

ARTICLE INFO

Article history
Received: 22 September 2021
Revision: 15 June 2023
Published: 29 June 2023

Keywords:
Critical thinking skill
Excretory system
Question development

ABSTRACT

This study aims to developing valid, reliable and practical Assessment to measure or to treat the critical thinkings skills of senior high school students on excretory system. The method used on this study is the development research (DR) that consists of synthesising theory and needs analyzes, step design (variable construction, indicator development, questions arrangement, making the instrumen, scoring and evaluation). Evaluation steps divided into three parts, namely: validation, reliability, and items analyzes. The quality of the test questions was determined by theoretical validation and empirical validation. The quality of the item was analised using the results of a field trial included the difficulty level, discriminatory power, and distractor. This study produced a valid questions in terms of content, construct, and language, with a final koefisien kappa above 0,61. The reliability of the test was 0,84. The result of students questionnaire analyzes in the phase of field trial gained an average of 3,60 (scale of 5). The average of value obtained from the student questionnaire indicates that the Critical Thinking Skills questions of exretory system is easy to use (practical). Thus, the development research produced 32 questions of critical thinking are valid, applicable, and easy to use.

Anwar, Y., Slamet, A., & Tania, A. (2023). Developing critical thinking skills assessment of excretory system. *Biosfer: Jurnal Pendidikan Biologi, 16*(2), 220-231. https://doi.org/10.21009/biosferjpb.23171

INTRODUCTION

Science and technology are developing so rapidly and been giving rise to new demands in all aspects of life, including in the education system. Education has a very important role in increasing quality human resources so that they can develop thinking skills and personal skills in the era of globalization. The ability to think is needed to analyze a problem and find a solution. Education should be one of the vehicles in a reliable thought formation process. One effort that can be done to prepare a learning process that trains students to develop high-level thinking skills (Arifin, 2018) . This is supported by the aim of the curriculum in 2013 and merdeka curriculum that emphasizes the skills of learners are equipped early with skill 4C namely critical thinking and problem solving skills, collaboration skills, creativities skills and communication skills (Permendikbud No. 36 2018). The provision of 2013 curriculum objectives, namely 4C skills, requires students to hone higher order thinking skills. one of the high-level thinking skills is the ability to think critically. The fundamental task of the modern education system is to develop students' critical thinking (Gilmanshina et al., 2021). Critical thinking is a particular generic skill (Liang & Fung, 2021). Critical thinking is a skill that can be developed in the learning process. In the learning process, the role of the teacher is needed to develop the mindset of students to explore their ideas on the learning activities and supporting to learning that makes it more easier for students to acquire knowledge (Adiyatmaningsih, 2014).

However, in the fact results of TIMSS (Trends in International Mathematics and Science Study 2015), Indonesia has ranked 45 out of 48 countries participating on the study in the administration of the study, while according to the OECD in 2016 for PISA (Program for International Student Assessment) 2015, Indonesia has been ranked 64th out of 72 participating countries. Testing the ability of students to complete high-order thinking questions has been carried out by inserting higher-order thinking questions in the national exam, it is known that students have difficulty working on these UN questions (Sani, 2019) . This proves that the thinking skills of students are still in the realm of low-level thinking. In line with the results of research conducted by Perdana which states that students' critical thinking skills are still lacking, one of which is the ability to provide basic explanations (Pradana, 2018; Rahdar et al., 2018). An appropriate subject in training students' critical thinking skills is learning biology, this is supported by Anwar 2020, which revealed its insights into the Indonesian National Framework (KKNI) curriculum which emphasizes that biology learning provides direct experience for students to be able to develop, explore the environment scientifically. The emphasis is expected to improve the ability of students in the learning process, namely fostering critical thinking skills of students. Based on an interview conducted with one of the biology subject teachers at SMA Negeri 2 Palembang, it was stated that biology learning is a subject that has so much material that students have difficulty remembering, analyzing, and even solving problems based on critical thinking skills so that it needs to be trained. One of the biological materials that are interconnected with one another is the excretion system. Material about the excretory system is material that is concrete in nature, but the process cannot be sensed, because the study includes physiological processes that occur in the human body (Ibrahim, 2014).

The importance of critical thinking skills had given to the necessary develop questions that can be used as a means of training students 'thinking skills so that they can improve students' critical thinking skills. Formative Assessment as a reflective thinking instruments is needed by students' perceptions of something, this perception will emerge if we direct it and we facilitate it continuously (Anwar, 2021). Researchers try to develop critical thinking questions on the excretory system material in basic competency 3.9, namely analyzing the relationship between the tissue structure of the organs in the excretory system in relation to bioprocesses and functional disorders that can occur in the human excretion system, with the research problem formulation, namely how far the question is feasible. - about the developed critical thinking skills that meet the valid, reliable, and practical categories? This research on developing questions based on critical thinking skills aims to assess the feasibility of the developed questions in the hope that the questions that are developed can be used as a reference in measuring the thinking abilities of students at a higher level.

An appropriate subject in training students critical thinking skills is learning biology, this is supported by Anwar 2020 that revealed his insight into the Indonesian National Framework (KKNI) curriculum which emphasizes that biology learning provides direct experience for students to be able to develop, explore the environment scientific. The emphasis is expected to improve the ability of

students in the learning process, namely growing critical thinking skills. Based on an interview conducted with one of the biology subject teachers at SMA Negeri 2 Palembang, it was stated that biology learning is a subject that has so much material that students have difficulty remembering, analyzing, and even solving problems based on critical thinking skills so that it needs to be trained. One of the biological materials that are interconnected with one another is the excretion system. The excretory system is material that is concrete in nature, but the process cannot be sensed, because the study includes physiological processes that occur in the human body (Ibrahim, 2014).

Given the importance of critical thinking skills, it is necessary to develop questions that can be used as a means of training students 'thinking skills so that they can improve students' critical thinking skills. Researchers try to develop critical thinking questions of the excretion system in basic competency 3.9, namely analyzing the relationship between the tissue structure of the organs in the excretory system in relation to bioprocesses and functional disorders that can occur in the human excretory system. The research problem formulation is how far the feasibility of the developed critical thingking skills questions has valid, reliable, and practical categories? Research development of matter based on the skills to think critically is aimed at assessing the feasibility of the product matter that developed with the hope of questions developed later will be the information measure thinking skills of students at a highers level.

METHODS

Research Design

This development research which refers to the design flow of the question instrument development according to Djaali and Mulyono (2008). The development flow is tailored to the needs. The procedure of research follows the steps the instrument development, among others: (1) Research study tand analyzes supporting theories (2) Set Indicators (3) Create problem grids (4) Preparation of problem and scoring instruments product (5) Validation of theoretical (6) Revision (7) Empirical validation (Figure 1)

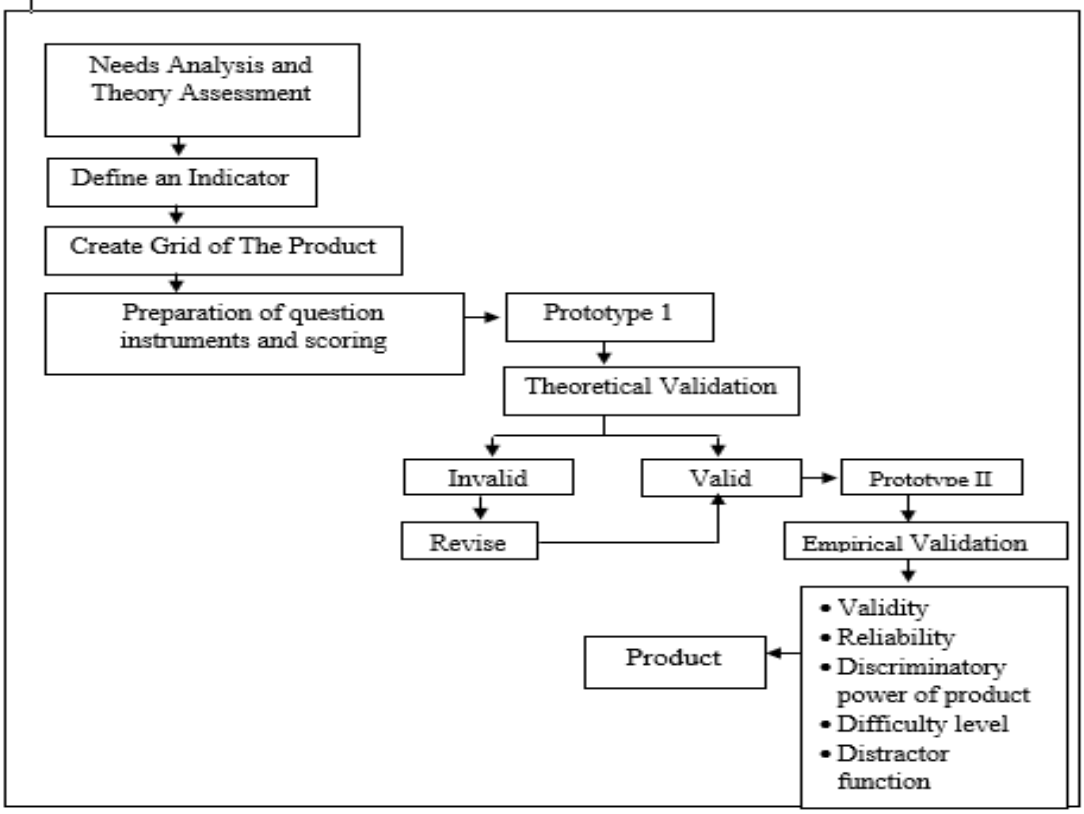

Figure 1. The assessment development stages

This development research aims to produce a product that is a matter of which feasibility to tested. The products produced in this study were items in the form of test objective multiple choice with

five possible answers, questions that develop amounted to 40 grains to the basic competence 3.9 to analyze the relationship between network structure constituent organs of the excretory system in relation to the bioprocess and malfunctioning that can occur in the human excretory system.

Instrument

This research was conducted in the second semester of the academic year 2019/2020 on 10 June 2020 in the SMA Negeri 2 Palembang, Indonesia to sample as many as 72 participants students are in class XI IPA 5 and XI 6 randomly selected or *random sampling* . Indicator of the product was using indicators of critical thinking skills by Ennis (1985) .

Table 1

Indicator critical thinking skills of ennis (1985).

No	Indicator	Sub Indicators
1	Elementary clarification	Focusing question
		Analyze arguments
		Ask and answer questions
2	Basic support	Consider whether the source can be trusted or not
		Observe and consider the observation report
3	Inference	Deducing and considering the results of the deduction
		Induce and consider the results of induction
		Make and determine the results of considerations
4	Advance clarification	Define terms and consider a definition
		Identify assumptions
5	Strategy and tactics	Specifies an action
		Interact with other people

Procedure

The Techniques of collecting data on research study are through interviews, documentation, questionnaire practicality, and the data tes. The results on this research is the answer of participant students as samples in the study and will had been analyzes to obtained a conclusion of the object that is being investigated. In this study preliminary research, interviews had doing with two biology teacher.

Data Analysis Techniques

Analyzes results of this research used test validity qualitative and quantitative. Qualitative analyzes is the validator's response on a form validation theoretical. These are suggestions and any comments. The analyzes quantitative on this study is marks checklist on a statement in the form and carculating using the formula coefficient kappa. Questions that can be used by researchers if the value of the Kappa coefficient is ≥ 0.61.(Viera & Garrett, 2005).

Kappa Coefficient Calculation:

$K = \frac{Po-Pe}{1-Pe}$

In which :

K = Kappa coefficient

Po = Proportion of observed agreement

Pe = Proportion of expected agreement

The Analyzes validity of the questions was calculating the correlation between the score of the items and the total score. In this study, the analyzes of item validity, reliability, and item analyzes (degree of difficulty , discrimination power and distractor) was carried out using the *Anates V4* program.

RESULTS AND DISCUSSION

Needs Analysis and Theory Assessment

Research on the development of Assessmet based on critical thinking skills uses the development research method that follows the development model steps according to Djaali & Mulyono (2008), this

step begins with the analyzes of supporting theories in developing critical thinking skills-based questions, one of the supporting theories in conducting this research is research on critical thinking skills that has been conducted by Anwar (2020) which states that 21st century education requires the emergence of a superior generation who is able to think creatively and critically, the emphasis is expected to increase their ability in the learning process that can foster and encourage critical thinking skills. According to Aripin (2018) emphasis The critical thinking ability of students can be trained through learning by giving critical thinking questions. The theory above is in line with the results of interviews with biology teacher at SMA Negeri 2 Palembang which stated that biology has a broad and complex material so that students thinking skills can be trained through question development.

Setting Indicators

The indicators this research are used critical thinking indicators according to Ennis in Costa which consist of five indicators and are further grouped into twelve sub indicators . Can be seen on table 1.

Creating Problem Grids

The grid is a description of the questions to be developed, in the form of a table of specifications**.** The question grid is made based on the analyzes of the learning implementation plan (RPP) which contains learning objectives, competency achievement indicators that cover the subject matter to be achieved in basic competencies. 3.9. Question making needs to adjust to the indicators of critical thinking skills. The critical thinking skills questions were will be developed as much 40 items questions with five answer choices, the number of questions is determined based on the analyzes of the lesson plan.

Preparation of Problem and Scoring Instruments

Preparation of question instruments based the initial draft of the question grid on critical thinking skills which is called the initial prototype. The initial prototype was then consulted with the supervisor to make improvements and set a score of 1 for correct answers and 0 for wrong answers. After the initial prototype is consulted and goes through the revision stage it will become a prototype I. The stages of determining the question score 1 for the correct answers and 0 for wrong answers. After the initial prototype was consulted and went through the revision stage, it would become prototype I.

Theoretical Validation

Theoretical validity is the evaluation of prototype I, which is tested for validity by the expert, which includes: content validity, construct validity, and language validity by using the kappa coefficient calculation. Based on the calculations obtained, the question instrument is declared valid by calculating the kappa coefficient for each item having a calculation result of more than 0.61 which states that the calculation of the question is feasible to be tested on condition that it revises the suggestions that have been given by the expert.

Revision

The researcher makes revisions based on suggestions and comments that have been given by validators or experts. After passing the expert validation stage and being revised, the instrument is called prototype II and is ready to conduct trials.

Empirical Validation

Apart from evidence for the convergent and discriminant validity, another fundamental analysis associated with CFA is the model-fit evaluation. In brief, when the fit indices are within the acceptable values, it indicates that the model is acceptable (Noh & Khairani, 2020). Empirical validity consists of analyzing the validity of the practicality of the questions, the reliability of the questions, and the analyzes of the items that have been developed. Prototype II was tested on a research sample of 72 students using online-based electronic media, namely google forms. Students' answers will be calculated using the Anates V4 application program. The results of the analyzes of the calculation of the empirical validity of the items. Can be seen in Table 2.

Table 2
Empirical Validation Result.

Item	Correlation coefficient	Interpretation	Item	Correlation coefficient	Interpretation
1	0,187	Invalid	**21**	0,477	Valid
2	0,325	Valid	**22**	0,188	Invalid
3	0,071	Invalid	**23**	0,455	Valid
4	0,325	Valid	**24**	0,398	Valid
5	0,351	Valid	**25**	0,453	Valid
6	0,356	Valid	**26**	0,411	Valid
7	0,32	Valid	**27**	0,528	Valid
8	0,215	Invalid	**28**	0,463	Valid
9	0,157	Invalid	**29**	0,496	Valid
10	0,332	Valid	**30**	0,317	Valid
11	0,089	Invalid	**31**	0,186	Invalid
12	0,34	Valid	**32**	0,465	Valid
13	0,392	Valid	**33**	0,354	Valid
14	0,327	Valid	**34**	0,335	Valid
15	0,335	Valid	**35**	0,428	Valid
16	0,393	Valid	**36**	0,367	Valid
17	0,212	Invalid	**37**	0,488	Valid
18	0,459	Valid	**38**	0,357	Valid
19	0,334	Valid	**39**	0,467	Valid
20	0,32	Valid	**40**	0,315	Valid

Table 2 shows the results of the validity of product developed critical thinking skills as many as 32 valid items and 8 invalid items. According to Arifin, (2016) invalid questions because many various factors, namely (1) the instrument factor used for the test, namely the preparation of instruments such as syllabus, question grids, instructions for solving questions, detailed answers, use of effective sentences, alternative answers , level of difficulty, discernment and so on. (2) administrative and scoring factors, namely disproportionate allocation of time for handling the questions, scoring errors, and the test taker's physical condition. (3) the factor of the students' answers, namely the tendency of students to answer quickly but not correctly so they did trial and error. Factors originating from the test, namely: (1) unclear instructions; (2) use of difficult vocabulary and sentence structures; (3) ambiguity the possibility of multiple interpretations in understanding and solving test questions; (4) excessive emphasis on certain levels, so it is too easy to predict the tendency of the answers to questions; (5) the quality of the test items is not coherent. Based on the researchers observations on questions number 1, 3, 8, 9, 11, 17, 22, and 31 which are categorized as invalid, it can be influenced by the use of biological vocabulary that seems new to students, sentence structures that are difficult for students to understand,

functional problem distractors , as well as questions that are too long. Through the analyzes of the achievement of the critical thinking ability indicators, it was found that the sub-indicators of interacting with other people had the largest percentage value.

The results calculation the achievement of indicators critical thinking skills that were answered most correctly by students had the highest percentage of 52%, that's the fifth indicator managing strategies and tactics, sub-indicators interacting with others, these questions are found at numbers 23, 33, and 40. In question number 23 there were 41 students who answered correctly, question number 33 there were 33 students who answered correctly, and in question number 40 there were 38 students who answered correctly, so the total number of students who answered correctly on the interacting sub indicator with 112 other people, based on these results it has been proven that half of the test takers have answered correctly, which means that the ability of students to set strategies and tactics, especially interacting to find answers using logic, has been good. In previous research conducted by Sobiatin, 2016, the results of the same analyzes showed that 48% of students had good high-order thinking skills through empirical validation trials about high order thinking skills on the circulation system material, although it showed good results but in reality students are more accustomed to working on lower order thinking questions that only prioritize aspects of the ability to recall existing information.

Practicality questionnaire

The practicality of the Assessment was tested by using a questionnaire to analyze the practicality of the product and obtain the practicality value. Practicality means the ease of a test both in preparing, using, processing, and interpreting. The practicality questionnaire was tested during the research process, namely after testing the question instrument, an assessment of the practicality of the product was obtained by exploring the opinions of 72 students through a practicality questionnaire consisting of 18 positive statements. Calculation results Questionnaire practicality of product based on critical thinking skills can be seen in table 3.

Table 3
Student's Assessment of the practicality questions

No.	Indicators	Students responses value (N=72)
1.	question instructions	3.70
2.	Words or sentences in the question	3.45
3.	Characteristics of Critical Thinking	3.62
4.	Language	3.63
	Final Score	3.60
	Categories	Practical

The results student's questionnaire assessment of the practicality had been final score 3.60, that proved the questions developed had good quality, because the statement instrument in the questionnaire rubik led to an assessment of the practicality of the questions. Good quality means that the product is practical to use. Question products are said to be practical or good if the results of the questionnaire calculations are in the value interval from 3.41 to 4.20 (Sugiyono, 2016)

Reliability

Reliability aims to see the stability of the questions based on the developed critical thinking skills, the high reliability of the test is one of the requirements for a learning outcome test that can be said to be a good test. Reliability is a consistency or similarity in the results of measuring objects carried out many times at different times. Analysis of the items in terms of reliability was carried out to measure the consistency of the test. A test is said to be reliable if the test's mean value gives the same results if it is given to the same object at different times. The reliability value on the question instrument based on critical thinking skills is 0.84, if the reliability value is above 0.70 it has a high test reliability and the product moment correlation value is 0.72 so it can be stated that the question of critical thinking skills has good reliability.

Analyze of items

Analyze of items was carried out after the prototype II questions were tested and then obtained the calculation of the validity and reliability values. This evaluation is carried out to determine the feasibility of each item as a functioning measuring tool.to carry out its function as a measuring tool.

Degree of Difficulty

The results of the difficulty index calculation will be interpreted in three criteria, namely difficult, medium, and easy. Following are the results of the item analyzes based on the degree of difficulty.

Table 4

Distribution Degree of Difficulty Based on Critical Thinking Skills.

No.	Difficulty index	Number Questions	Items
1.	< 0,30 (difficult)	1,8,14,21,35,37	6
2.	0,30-0,70 (moderate)	2,3,4,5,6,7,9,10, 11,12,15,16,17, 18,19,20,22,23, 24,25,26,27,28, 29,30,31,32,33, 34,36,38, 39,40	33
3.	>0,70 (easy)	13	1
	Total items		40

Table 4 shows the degree of difficulty of the questions based on critical thinking skills. Based on this table, a percentage degree of difficulty the product had been made as shown in Figure 1.

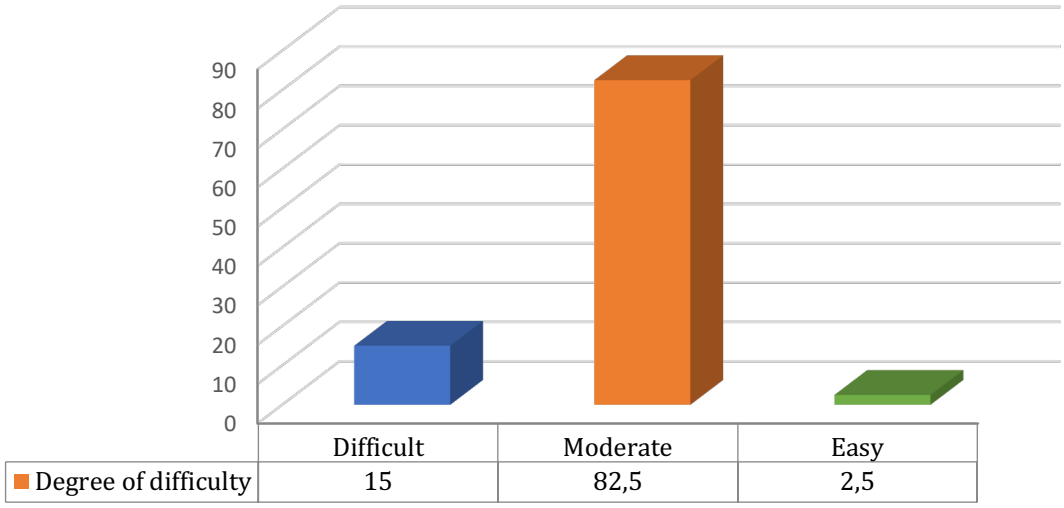

Figure 2. Percentage Degree of Difficulty Index based on Critical Thinking Skills

Based on Figure 1, questions in the difficult category have a percentage of 15%, questions in the medium category have a percentage of 82.5, questions in the easy category have a percentage of 2.5%. An item is a good answer if it has a difficulty level of 0.25-0.75 with a medium question category (Sudijono, 2016) so it can be denied that instruments based on critical thinking skills include questions that have a good level of difficulty because some items are or equal to 82.5 % have medium difficulty.

Discrimination Power

Discrimination Power is one basis in compiling test items, test questions must also be able to reflect the differences in abilities that exist among students. The results of the discriminate power distribution of the questions are shown on Table 5.

Table 5

Discriminatie Power distribution Based on Critical Thinking Skills Questions

No.	Discernment	Number Questions	Items
1.	<20 (not good)	3,9,11,31	4
2.	0.20-0.40 (moderate)	1,2,5,7,8,14, 17,20,22,40	10
3.	0.40-0.70 (good)	4,6,10,12,13,15, 16,18,19,21,23, 24,25,26,27,28, 29,30,32,33,34, 35,36,37,38,39	26
4.	0.70-1.00 (very good)	0	0
5.	Signed negative (-)	0	0
	total items		40

Table 5 shows the results discrimination power based on the answers students during the field test. The percentage distribution of discriminate power had been obtained as shown in Figure 2.

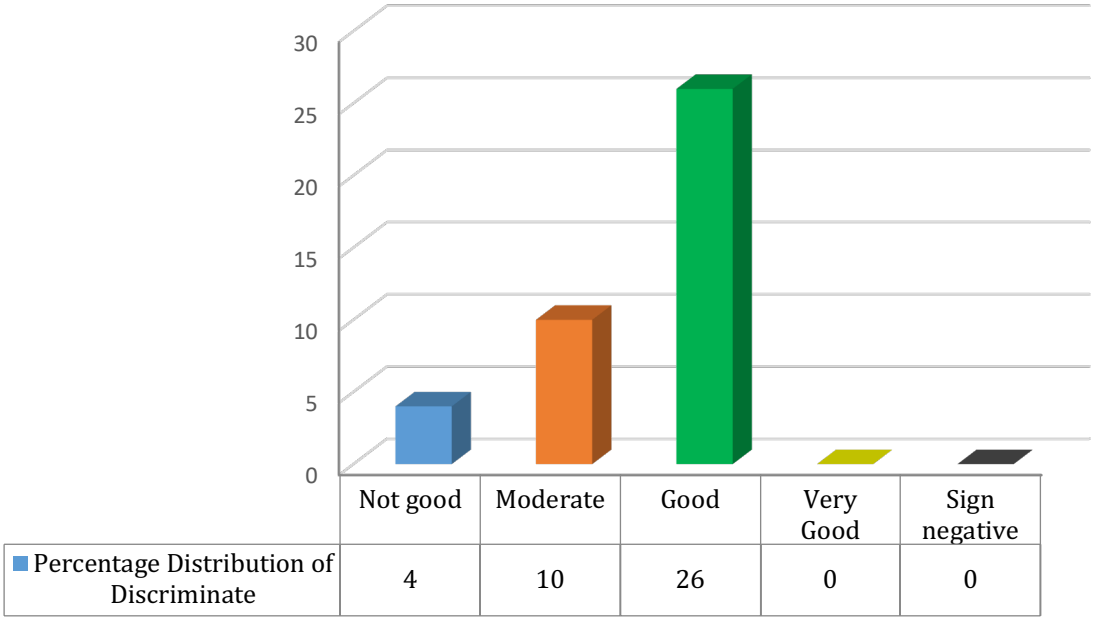

Figure 3. Percentage Distribution Discriminate Power of Critical Thinking Skills Question

The results of the calculation of discrimination power with the *Anates V4* program show that 26 items or 65% have good discrimination power, 10 questions or 25% have moderate discrimination power, 4 items or 10% have bad discrimination power. The results of the analysis indicate that the instrument of critical thinking skills that the researcher developed has discrimination power with good quality.

Distractor

The main purpose of the distractor on each item is so that some of the students who take the test are interested in choosing the alternative answer as the correct answer. The more students who are fooled, it can be stated that the distractor has been able to carry out its function properly.

The calculation of the effectiveness of the distractor (distractor index) on questions based on critical thinking skills on the excretion system material using the help of the Anates V4 program. The results of the analyzes show that on average the answers of the deceivers have performed their function well. At least 5% of all test takers have chosen the cheating answer. Of the 40 test items, there were 26

questions that had good distracting quality, namely questions on numbers 4, 5, 6, 8, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 25, 26, 27 , 29, 30,31, 33, 34, 35, 37, 38, 39, the analyzes results were obtained from the draft Anates V4 results based on the analyzes of students 'answers showing the spread of students' answers in each item caused the results of the questioners to be good, if calculated and analyzed questions that have good distractors chosen by at least 3-4 students so that it can be concluded that the questions were selected at least 5% of all students who participated in the trial.

The results collected are used as material for researchers to analyze both the value of validity, reliability, analyzes of the items, and the statistical test which shows that there is a relationship. Based on the researcher's analyzes, the value of practicality shows that the questions can be read and considered practical by the participants students through the practicality questionnaire statement obtained a value of 3.60 which is categorized as practical. If the question is practical it can be interpreted that the question is easy for students to understand. Some students will answer correctly the items that cause the product moment correlation value is high and the questions become valid. The more valid items will increase the reliability value of the question product. A good measuring instrument must be valid, reliable and practical. In addition to validity and reliability, an instrument in the form of test questions must also be analyzed by the compiler items. From the evaluation analyzes, it can prove the feasibility of the product that the researcher developed is measured based on how far the total questions are valid, the reliability value of the question products, the value of the item analyzes obtained, and the results of the practicality calculation of the questions based on the interpretation of the interval set in the study. Evaluation is also carried out to find out whether the items developed can carry out their function properly or not. So, in the future the compiled test questions had been a function as a measuring device. Based on the results of this research on the development of critical thinking skills based on the index difficulty of the items, the discrimination power, and distractor items. Interval category was according to Sudijono, 2016 the product was has functioned well if the range or interval category showed interpretation good category.

These critical thinking questions can be used as a treatment when used as a formative assessment so that it is expected to improve students' critical thinking skills. As some experts claim that: critical thinking skills can be created by giving students treatments. The treatments used to adjust to students' potential, such as student's critical thinking skills (Ong et al., 2020; Tarch et al., 2020; Susetyarini et al., 2020; Chusni, M et al, 2022). Apart from being familiar with critical thinking-based assessments, their critical thinking skills can be developed through the learning experiences provided. Critical thinking skills are developed through inquiry in exploring, discovering, and interpreting (Ernita et al., 2021; Firiyani et al 2019).

CONCLUSION

The research of developing questions critical thinking skills on the excretion system material produces question that are categorized as valid, reliable and practical. There were has 32 items categorized as valid and 8 items were categorized as invalid. The reseaech study has a high reliability rate avarage is 0.84. The problem is categorized as practical because the practicality test has been carried out and the calculation of the practicality average is 3.60 so that it is declared practical. This research produces questions of critical thinking skills that is feasible to use.

REFERENCES

Adiyatmaningsih, H., Suara., Rini, K. (2014). Mind mapping-assisted problem-based learning models have an effect on the science learning outcomes of class V SD cluster III Gianyar. *Journal of the PGSD University of Ganesha Education Forum 2 (1).* Retrieved from https://ejournal.undiksha.ac.id/index.php/JJPGSD/article/view/3057/2531

Anwar, Y., & Permata, S. (2020, March). Measuring biology educations students' critical thinking skill using online systems. In *Journal of Physics: Conference Series* (Vol. 1480, No. 1, p. 012068). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1480/1/012068/meta

Anwar, Y., Zen, D., & Permata, S. (2021, January). Enhancing critical thinking skills of biology education students using online formative assessment. In *4th Sriwijaya University Learning and Education*

International Conference (SULE-IC 2020) (pp. 14-17). Atlantis Press. https://doi.org/10.2991/assehr.k.201230.076

Arifin, Ipin. (2018). Development of Multiple Choice Questions to Measure the Critical Thinking Ability of Siswap on the Concept of the Human Regulatory System for High School Level. *Manggifera Journal edu*. 3 (1): 13-25. Retrieved from https://jurnal.biounwir.ac.id/index.php/mangiferaedu/article/view/10

Center for Educational Assessment of the Research and Development Agency (2015). Regarding the results of TIMSS (Trends in International Mathematics and Science Study. Retrieved from https://nces.ed.gov/timss/timss2015/

Chusni, M. M., Saputro, S., Suranto, S., & Rahardjo, S. B. (2022). Empowering critical thinking skills on different academic levels through discovery-based multiple representation learning. *Cakrawala Pendidikan*, *41*(2), 330-339. https://etheses.uinsgd.ac.id/id/eprint/67142

Directorate General of Primary and Secondary Education. (2018). 2013 Curriculum Implementation Module SMA Revision 2018. Jakarta: Kemendikbud.

Djaali., M, P. (2008). Measurement in the field of education. Jakarta: Grasindo.

Eka Sobiatin. (2016). Development of Higher Level Thinking Questions (Hots) For High School Human Circulation System Material. Essay. Sriwijaya University: Palembang.

Ennis, R. H. (1985). Goal critical thinking curriculum. Virginia: Association for Supervision and Curriculum Development (ASCD).

Ernita, N., Muin, A., Verawati, N. N. S. P., & Prayogi, S. (2021). The effect of inquiry learning model based on laboratory and achievement motivation toward students' physics learning outcomes. *Journal of Physics: Conference Series*, 1816(1), 012090. https://iopscience.iop.org/article/10.1088/1742-6596/1816/1/012090/meta

Fitriani, H., Asy'ari, M., Zubaidah, S., & Mahanal, S. (2019). Exploring the Prospective Teachersâ€TM Critical Thinking and Critical Analysis Skills. Jurnal Pendidikan IPA Indonesia, 8(3), 379–390. https://doi.org/10.15294/jpii.v8i3.19434

Gilmanshina, S., Smirnov, S., Ibatova, A., & Berechikidze, I. (2021). The assessment of critical thinking skills of gifted children before and after taking a critical thinking development course. *Thinking Skills and Creativity,* 39, 100780. https://doi.org/10.1016/j.tsc.2020.100780

Ibrahim, A., Sariwulan D., & Ana R. W. (2014). Application of learning log class to diagnose learning difficulties of high school students on the human excretion system material. Formica Education Online. 1 (1). Retrieved from Retrieved from https://onesearch.id/Record/IOS1006.article-108

Liang, W., & Fung, D. (2021). Fostering critical thinking in English-as-a-second-language classrooms: Challenges and opportunities. *Thinking Skills and Creativity*, *39*, 100769. https://doi.org/10.1016/j.tsc.2020.100769

Noh, A. M., & Khairani, A. Z. (2020). Validating the S-stem among malaysian pre-university students. *Jurnal Pendidikan IPA Indonesia*, *9*(3), 421-429. https://doi.org/10.15294/jpii.v9i3.24109

OECD. (2016). PISA Result From PISA 2015. Paris: OECD. Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Ong, K. J., Chou, Y. C., Yang, D. Y., & Lin, C. C. (2020). Creative drama in science education: The effects on situational interest, career interest, and science-related attitudes of science majors and non-science majors. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(4), em1831. https://doi.org/10.29333/ejmste/115296

Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 36 of 2018 concerning Amendments to the Regulation of the Minister of Education and Culture Number 59 of 2014 concerning the 2013 Curriculum for Senior High Schools / Madrasah Aliyah. Jakarta: Ministry of Education and Culture.

Pradana, S. (2018). Development of Critical Thinking Ability Tests in Optical Geometry Materials for Physics Students. *Journal of Educational Research and Evaluation*. 21 (1): 51. Retrieved from https://journal.uny.ac.id/index.php/jpep/article/view/13139

Rahdar, A., Pourghaz, A., & Marziyeh, A. (2018). The Impact of Teaching Philosophy for Children on Critical Openness and Reflective Skepticism in Developing Critical Thinking and Self-Efficacy. *International Journal of Instruction*, 11(3), 539-556. https://eric.ed.gov/?id=EJ1183430

Sani, Ridwan A. (2019). HOTS Based Learning. Jakarta: Tiara Smart.

Sudijono, A. (2016). Introduction to Educational Evaluation. Jakarta: Rajawali Press.

Sugiyono. (2016). Educational Research Methods (quantitative, qualitative, and R&D approaches). Bandung: Alfabeta.

Susetyarini, E., & Fauzi, A. (2020). Trend of Critical Thinking Skill Researches in Biology Education Journals across Indonesia: From Research Design to Data Analysis. *International Journal of Instruction*, 13(1), 535-550. http://eprints.umm.ac.id/id/eprint/71287

Tarchi, C., & Mason, L. (2020). Effects of critical thinking on multiple-document comprehension. *European Journal of Psychology of Education*, 35(2), 289-313. https://doi.org/10.1007/s10212-019-00426-8

Viera and Garrett (2005), Understanding Interobserver Agreement: The Kappa Statistic. Fam Med 2005; 37 (5): 360-363. Retrieved from https://pubmed.ncbi.nlm.nih.gov/15883903/

Zainal Arifin. (2016). Learning Evaluation. Bandung: PT Remaja Rosdakarya.