Unnes.J.Biol.Educ. 11 (3) (2022)

Journal of Biology Education

http://journal.unnes.ac.id/sju/index.php/ujbe

The Development of Critical Thinking Skills Test Instrument on Human Excretory System

Siti Nur Alimah, Nugroho Edi Kartijono[1]✉

[1]Biology Department, FMIPA, Semarang State University, Indonesia

Article Info

Article History:

Received : August 2022

Accepted : August 2022

Published : December 2022

Keywords:

Human excretion system, Critical Thinking Skills, Test Instrument

Abstract

The purpose of this study was to develop a suitable test instrument to measure the critical thinking skills of high school students on the material of the human excretory system. This study is Research and Development (R&D) adapted from Sugiyono (2015). The test subjects in this study were students of class XI MIPA SMA Negeri 1 Juwana. Data was collected by tests referred to the Facione indicators with 18 items to measure students' critical thinking skills, and questionnaires to collect data on students' responses on the instruments. Student questionnaires are in Likert scale. The data then was analyzed by tests of validity, reliability, level of difficulty, and distinguishing power by SPSS version 20. The small-scale test showed that 3 out of 18 questions were not valid and needed improvements. Students' critical thinking skills in the small-scale test were in high category (48.5%). In the large-scale test, 18 items were valid and feasible to use and 52.3% of students' critical thinking skills were in high category. It is concluded that the instrument used is valid and feasible based on expert validation, item analysis and field trials. The critical thinking skills of SMA Negeri 1 Juwana students were in the high category.

✉Correspondence Address:
D6 Building 1st Floor Jl Raya Sekaran Gunungpati Semarang
E-mail: nugrohoedik@mail.unnes.ac.id

p-ISSN 2252-6579

e-ISSN 2540-833X

INTRODUCTION

21st Century Education requires students to have knowledge in various fields. Knowledge is important for someone to have because it can develop thinking skills. The ability to think can determine a person's success in life (Zubaidah, 2010). 21st century learning is known as 4C learning skills. 4C skills that is Critical Thinking and Problem Solving (critical thinking and problem solving), Communication, Creativity and Innovation and Collaboration. Critical Thinking is an activity that aims to solve problems logically, can understand a problem and find relevant answers. Communication namely using the ability to express their ideas, both during discussions and in writing. Creativity and Innovation is the ability to develop and convey new ideas and to be open and responsive to new perspectives. Collaboration namely the ability to work in groups (Kemendikbud, 2017). One of the subject matter that can train students' 4C thinking skills is biology which covers all aspects of living things as well as matters relating to these living things. Examples of aspects discussed include morphology, anatomy, ecology, and physiology. Based on these four aspects, the material on physiology is quite difficult. This is in line with previous research by Michael (2007) which stated that the most complex biological material to study is physiology.

One of the materials related to daily life in terms of physiology is the excretory system. Excretory system material requires students to have learning competencies at a high level of understanding. These competencies include several abilities, namely thinking critically in receiving various information, thinking creatively in solving problems from the knowledge possessed and making decisions in complex situations (Saputra, 2016). In fact, the learning objectives to achieve a high level of understanding have not been fully achieved. Evidence that students' science performance abilities are low is based on the results of the 2018 PISA report which shows that in the science performance category, Indonesia ranks 71 out of 79 PISA participating countries. Regarding this matter, it shows that the scientific ability of students in Indonesia has not been superior at the global level so that there needs to be an improvement in the vision and mission in advancing education. Improving scientific competence can be done by improving the assessment system. The Minister of education and culture stated that the assessment needs to be improved so that it refers more to basic competencies that are useful for improving the quality of education (OECD, 2019).

Written test is more commonly used by teachers as a tool for assessing student learning outcomes. A good test instrument is one that can measure students' higher order thinking skills. This is in accordance with the program developed by the Ministry of Education and Culture through the Directorate General of Teachers and Education Personnel (Ditjen GTK), namely the development of learning oriented to higher order thinking skills (HOTS) to improve the quality of learning and improve the quality of graduates (Ariyana, 2018). The type of higher order thinking skill that is the basis and initial to be trained is critical thinking. Sani (2019).

Based on the results of an unstructured interview with one of the biology teachers at SMA Negeri 1 Juwana in May 2021, it was stated that students actually had critical thinking skills, but these abilities were not honed. This was because the evaluation process done by biology teachers did not seem to be able to improve students' critical thinking skills. The absence of development of questions based on critical thinking skills made students' critical thinking skills less honed. In this regard, the teacher assumed that if students were hard to solve critical thinking questions and affected the learning outcomes. Therefore, it is necessary to develop test instruments that can train students' thinking skills so that they are more accustomed to working on them and finally improve the quality of education.

METHODS

This study is research and development (R&D). The development model adapts the stages of Sugiyono (2015) involves 10 stages, namely potential and problems, data collection, product design, design validation,

design revision, small-scale trials, product revisions, large-scale trials, product revisions, and product reviews.. The product resulting from this development is a critical thinking skill test instrument in the form of descriptive questions.

The activity at the potential and problem stage was a needs analysis carried out by identifying problems from the results of an interview with a biology teacher at SMA Negeri 1 Juwana. Furthermore, the method of data collection was to collect information and data as material and reference for developing critical thinking skills test instruments. Then the product design was planning for the preparation of critical thinking questions instruments which include the formulation of learning indicators and critical thinking indicators that were adapted to basic competencies in the human excretory system concept, formulating questions framework and its answer based on the material. Furthermore, the test instrument was tested for validity to two validators, namely a lecturer of a biologist at the State University of Semarang and a biology teacher at SMA Negeri 1 Juwana. After the validation process, product revisions were done based on the suggestions from experts to produce a valid and suitable draft for use to the trial stage. The next stage was conducting product trials on a limited scale to 33 students. The aim was to determine the validity, reliability, level of difficulty and distinguishing power. The questions were analyzed using the SPSS 20. Subsequently, a product revision was made from a small-scale trial to produce 2 drafts of instruments to be tested on a wider scale. The next step was to conduct largescale trials or field tests on 107 students to determine the quality of the products developed based on students' comments as improvement materials as well as to determine the level of critical thinking skills of each student.

RESULTS AND DISCUSSION

1. Potential and Problems

Based on the results of an interview with one of the biology teachers at SMAN 1 Juwana, the tests used were multiple choices and descriptions that only measured learning outcomes but did not measure students' critical thinking skills. The teacher had not fully made questions based on critical thinking skills so that students were less familiar with these questions and tended to use LOTS type questions (Lower Order Thinking Skills). As a result these skills had not been measured properly. Whereas instruments based on critical thinking skills are needed to train students' thinking skills**.**

2. Data Collection

The next activity was collecting data to support the development of test instruments. Some of the data collected included Student Activity Sheets (LKPD) used during learning, Learning Implementation Plans (RPP) and materials on the human excretory system obtained from various sources.

3. Product Design

The designed questions consisted of 18 questions with each critical thinking indicator containing 3 questions. The indicators used were critical thinking indicators according to Facione which consist of 6 indicators, namely interpretation, analysis, inference, evaluation, explanation and self- regulation . Questions were made according to basic competencies and learning indicators. Before compiling the questions, first a questions framework was made and then the test questions, answer line and assessment guidelines were made.

4. Design Validation

After making the initial product, the product was then validated by experts. The expert team consisted of 2 validators, they were UNNES Biology Lecturer and the Biology Teacher at SMAN 1 Juwana. The instrument used was a Likert scale assessment questionnaire. The assessment was carried out after going through several revision processes. The results of the validation by experts are presented in the Figure 1.

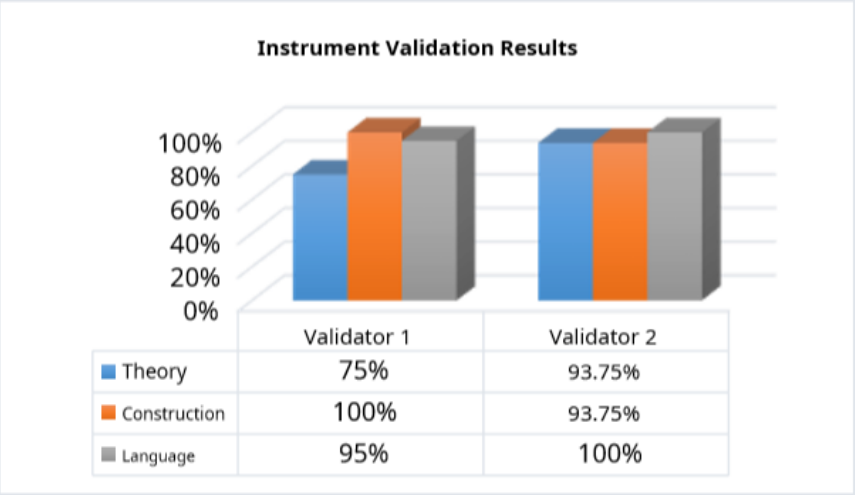

Figure 1. Expert Validation Results

Based on Figure 1, the highest assessment results on the test instrument product are in the language aspect, namely the value of validator 1 is 95% and validator 2 is 100%. This means that in this aspect, the sentence formulations presented are communicative and effective, using language according to PUEBI rules, and using words that do not cause misunderstandings. Meanwhile, the material and construction aspects also received good ratings. The results of this assessment are the results of the assessment for the test instrument that has been revised so that it is suitable for use in small-scale tests.

5. Design Revision

Based on the results of instrument validation, there are several inputs for each item. Suggestions for improvement given by the validator are as follows.

Table 1. Expert Validator Improvement Suggestions

No.	Aspect	Revision Suggestions	Revision Results
1.	Theory	The questions are adjusted to the learning indicators	The questions have been adjusted to the learning indicators
2.	Construction	Presentation table and picture zoomed in	Enlarged tables and figures
3.	Language	Sentences are corrected according to PUEBI rules	Sentences in the questions have been adjusted according to PUEBI

Other inputs from the validator are related to the suitability of the questions with the editorial answers. Based on some comments and suggestions from the validator, it is necessary to consider as material for instrument revision.

6. Small Scale Test

After a limited trial was carried out with 33 students, the data obtained were then analyzed using the SPSS version 20 to determine the quality of the developed test instrument. The quality in question is validity, reliability, level of difficulty and distinguishing power.

Table 2. Questions Validity

Category	Question number	Amount	Percentage
Valid	1, 2, 3, 4, 5, 7, 8, 9, 11,12,13, 14, 15, 17, 18	15	83.3%
Invalid	6, 10, 16	3	16.7%

The table shows that there are 3 items that are not valid, namely questions number 6, 10, and 16. This is because r_{Count} is smaller than r_{Table}. So that the 3 items must be revised before being used in a large-scale test.

Table 3. Questions Reliability

Cronbach's Alpha	N of Items	Correlation	Interpretation	Note
0.862	15	High	Well	Reliable

The reliability index ranges from 0-1. The higher the reliability coefficient of the question, the higher the accuracy (Kadir, 2015). In this reliability test, the result is 0.862, which means it has a high level of accuracy/consistency.

Table 4. Level of Problem Difficulty

Category	Question number	Amount	Percentage
Very difficult	-	-	-
Hard	8	1	5.6%
Medium	3, 5, 7, 9, 11,12,13, 14, 15, 17, 18	11	61%
Easy	1, 2, 4, 6, 10	5	27.8%
Very easy	16	1	5.6%

In the table above, questions that were too easy, namely number 16, needed to be corrected because questions that were too easy would not train students' critical thinking skills. So, it could be concluded that the items included in the easy, medium, and difficult categories would be accepted and used as a test for the purpose of formative tests while the questions that were included in the too easy category would be revised.

Table 5. Distinguishing Power of Questions

Category	Question number	Amount	Percentage
Very Well	1,2,4,5,7,8,9,11,12,13, 14, 15, 18	13	72.2%
Well	3, 17	2	11%
Bad	10, 16	2	11%
Very bad	6	1	5.6%

The table above shows that the overall test items compiled are in the good and quite good categories. However, questions number 6, 10 and 16 included in bad and very bad criteria. It could be concluded that the three numbers could not distinguish well between smart and less intelligent students. So, these three questions should be revised. From this analysis, it could be concluded that questions number 6, 10, and 16 were revised. The three items that had been revised were then combined with questions that were not revised to be used in field tests on a wider scale. Next, analyze the results of student responses. The following are the results of the responses to the developed instrument.

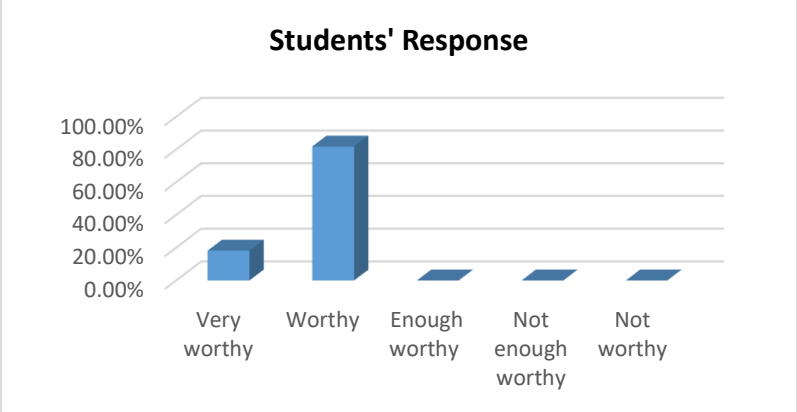

Figure 2. Students' Response

Based on Figure 2, students gave a positive response to the developed test instrument. A total of 27 respondents with a percentage of 81.8% stated that the instrument was suitable for use, while 6 other respondents with a percentage of 18.2% stated that the instrument prepared was very suitable to be used and no one gave an assessment that was quite appropriate to not feasible. In addition to the assessment in accordance with the 8 statements in the questionnaire given, students commented that the time given to work on the questions was too short so they lacked concentration and were in a hurry to write the answers. Suggestions from students so that the time given could be extended again so that the critical thinking skills of students could be optimally measured. Next was to analyze the test results of critical thinking skills. Student test results are presented in Table 6.

Table 6. Test Results of Students' Critical Thinking Skills

Ability interval (%)	Category	Amount	Percentage (%)
80 < X ≤ 100	Very High	2	6
60 < X ≤ 80	High	16	48.5
40 < X ≤ 60	Moderate	13	39.4
20 < X ≤ 40	Low	2	6
0 < X ≤ 20	Very Low	0	0

Based on the Table 2, the critical thinking skills of class XI MIPA 2 students are mostly in the high category. This means that in solving problems, there were already many students who were able to fulfill or answer all critical thinking indicators on the material of the human excretory system, as many as 48.5% (16 students from 33 students.). Meanwhile, there were 13 students were in the medium category with a percentage of 39.4% and there were even 2 students who reached the low category. Students who were in the moderate category were because of the fact that most students misunderstood the questions and did not provide explanations regarding their answers by linking concepts to the human excretory system.

Critical thinking skills were also analyzed based on 6 indicators of students' critical thinking skills, namely interpretation, analysis, inference, evaluation, explanation and self- regulation. The results of the analysis per indicator are presented in the Figure 3.

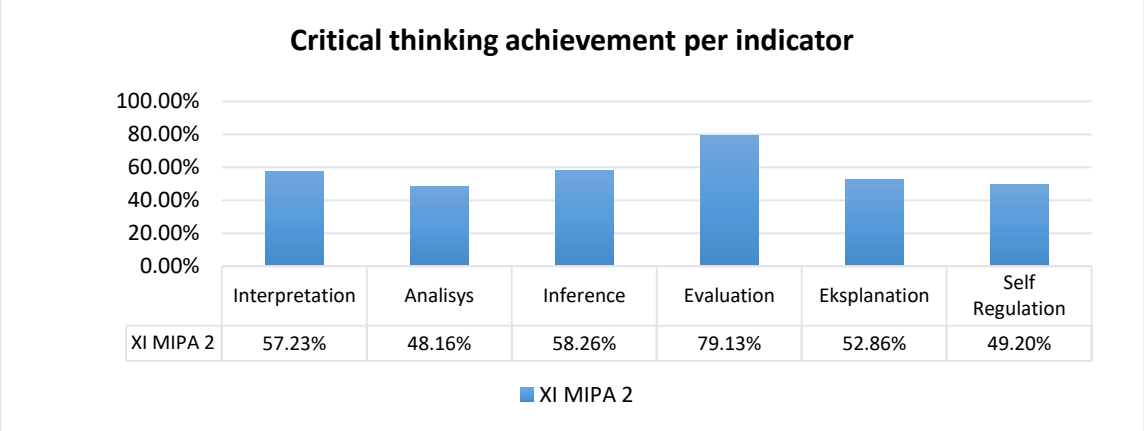

Figure 3. Critical Thinking Achievement Per Indicator

Based on diagram 3, the interpretation indicator obtained 57.23% in the medium category, 48.16% analysis indicators are in the very moderate category, 58.26% inference indicators are in the medium category, evaluation indicators 79.13% are in the high category, explanation indicators are 52.86%, and indicators self-regulation 49.20% which means both are in the very moderate category. So, it can be concluded that the evaluation indicator has the highest percentage compared to other indicators. This means that in this indicator, students were able to provide solutions or write solutions to problems that were known to be true and appropriate. Meanwhile, the other five indicators are in the medium category. This happened because most of the students were able to answer the questions, but they were not quite right. They did not understand the meaning of the information provided, had not been able to describe and provide further explanations so that the responses given from the assumptions were incomplete and inaccurate in drawing a conclusion.

7. Large-Scale Test

The following are the results of responses from 107 respondents to the developed instrument.

Table 7. Student Response Results

Criteria	Amount	Percentage (%)
Very worthy	29	27.10
Worthy	78	72.90
Enough	-	-
Not Worthy	-	-
Not feasible	-	-

Based on the Table 7, the three classes of XI MIPA gave a positive response. A total of 29 respondents stated that the product developed was very feasible to use and 78 other respondents stated that it was feasible to use. So, it can be concluded that 100% of students stated that the instrument developed was feasible and even very feasible to be used as an evaluation tool to measure students' critical thinking skills.

Table 8. Critical Thinking Skills Test Results

Ability interval (%)	Category	Amount	Percentage (%)
$80 < X \leq 100$	Very High	18	16.8
$60 < X \leq 80$	High	56	52.3
$40 < X \leq 60$	Moderate	25	23.3
$20 < X \leq 40$	Low	8	18.6
$0 < X \leq 20$	Very low	-	-

Students who have critical thinking skills in the high to very high category mean that these students have met the six critical thinking indicators and are able to answer questions well. Judging from the test results, students were able to identify known problems in the questions, could describe and link concepts with questions from the questions given correctly and provided complete explanations or reasons so as to obtain the correct answers. In addition, they were also able to draw conclusions well.

Furthermore, students whose critical thinking skills are in the medium category means that most of them have been able to answer questions but there are some of the six critical thinking indicators that have not been met or have not been answered correctly. Another reason was because most of the students had met all the indicators but there were some questions that scored not maximal. Meanwhile, critical thinking skills with low or very low categories were caused by students not being able to solve problems from the questions given and not giving proper explanations. Students who are in the low category only meet 2 or 3 of these critical thinking indicators and the scores obtained from all questions are relatively minimal.

Table 9. Critical Thinking Achievement Per Indicator

Indicator	Percentage per indicator (%)	Category
Interpretation	73.8	High
Analysis	69	High
Inference	60.6	High
Evaluation	72	High
Explanation	65.4	High
Self regulation	52.5	Currently

Based on Table 9, it can be seen that the three classes on indicators of interpretation, analysis, inference, evaluation, explanation are in the high category. Meanwhile, for self-regulation indicators, the three classes are in the medium category. Based on the description above, it could be seen that the critical thinking skills of each student are different. These differences are certainly influenced by many factors, including physical conditions, anxiety, intellectual development, motivation, habits, experience, consistency and feelings. These factors can increase or decrease a person's critical thinking ability (Sutriyanti & Mulyadi 2019).

Research by Samadya & Dahlan (2020) states that learning motivation has a very strong relationship with critical thinking skills. Students with high learning motivation have high critical thinking skills, students with moderate learning motivation, have moderate critical thinking skills, as well as students with low learning motivation, their critical thinking skills are also low. Students who have low learning motivation and low critical thinking skills have less interest in questions based on problem solving, do not like challenges so that their insight is limited, and they do not understand what is being asked in the questions as a result their answers are incomplete and illogical.

CONCLUSIONS

The research shows that the critical thinking skills test instrument developed on the human excretory system concept for high school students is feasible to use based on the results of expert validation tests, small-scale tests and field tests. The critical thinking skills of SMA Negeri 1 Juwana students measured by the developed instrument are in the high category.

Based on the results obtained from this research, some suggestions can be put forward as follows: based on the results obtained from this study, it is suggested that teachers should develop test instruments in order to train students' critical thinking skills with appropriate methods in ongoing learning activities to improve students' critical thinking skills. In addition, teachers could also provide practice questions more often, such as this developed test instrument so that students' critical thinking skills could be measured properly.

REFERENCES

Ariyana, Y. (2018). *Buku Pegangan Pembelajaran Berorientasi Kemampuan Berpikir Tingkat Tinggi*. 6.

Kadir, A. (2015). Menyusun Dan Menganalisis Tes Hasil Belajar. *Al-Ta'dib*, *8*(2), 70–81.

Kemendikbud. (2017). Panduan Implementasi Kecakapan Abad 21 Kurikulum 2013 Di Sekolah Menengah Atas. In *Jakarta: Direktorat Pembinaan Sekolah Menengah Atas. Direktorat Jenderal Pendidikan Dasar Dan Menengah Kementerian Pendidikan Dan Kebudayaan*.

Michael, J. (2007). What makes physiology hard for students to learn? Results of a faculty survey. *American Journal of Physiology - Advances in Physiology Education*, *31*(1), 34–40. https://doi.org/10.1152/advan.00057.2006

OECD. (2019). Programme for international student assessment (PISA) results from PISA 2018. *Oecd*, 1–10.

https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-iii_bd69f805-en%0Ahttps://www.oecd-ilibrary.org//sites/bd69f805-en/index.html?itemId=/content/component/bd69f805-en#fig86

Purnama, S. (2018). Metode Penelitian Dan Pengembangan (Pengenalan Untuk Mengembangkan Produk Pembelajaran Bahasa Arab). *Indonesian Journal of Economics Education*, *1*(1), 1–8. https://doi.org/10.17509/jurnal

Samadya, M., & Dahlan, D. (2020). Identifikasi Perbedaan Keterampilan Berpikir Kritis Ditinjau dari Motivasi Belajar Peserta Didik Identify Differences of Critical Thinking Skill Based on The Student ' s Learning Motivation in Economic Learning. *Jurnal Ilmu Sosial*, *17*(1), 91–98.

Sani, R. A. (2019). *Pembelajaran Berbasis HOTS (Higher Order Thinking Skills)*.

Saputra, H. (2016). *Pengembangan Mutu Pendidikan Menuju Era Global : Penguatan Mutu Pembelajaran dengan Penerapan HOTS (High Order Thinking Skills)*. SMILE's Publishing.

Sugiyono. (2015). Metode Penelitian Pendidikan. Bandung. *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, Dan R&D)*, 308.

Sutriyanti, Y., & Mulyadi, M. (2019). Analisis Faktor-Faktor yang Mempengaruhi Penerapan Berpikir Kritis Perawat dalam Melaksanakan Asuhan Keperawatan di Rumah Sakit. *Jurnal Keperawatan Raflesia*, *1*(1), 21–32. https://doi.org/10.33088/jkr.v1i1.394

Zubaidah, S. (2010). Berfikir Kritis : Kemampuan Berpikir Tingkat Tinggi Yang dapat Dikembangkan Melalui Pembelajaran Sains. *Seminar Nasional Sains 2010 Dengan Tema "Optimalisasi Sains Untuk Memberdayakan Manusia," January 2010*, 11.