Four-Tier Multiple Choice Test Characterized by Local Wisdom Values for Analyzing Critical Thinking Skills

Zuafatun Ni'mah[1*], Agus Sudarmanto[2], Qisthi Fariyani[3]

[1,2,3] Physics Education UIN Walisongo Semarang
*Correspondence: zuafatunnikmah123@gmail.com

Abstract	
Keywords: Four-Tier Multiple Choice Test Local Wisdom Critical Thinking Skill	This study aims to develop a four-tier multiple-choice test instrument characterized by the values of local wisdom to describe the characteristics of the instrument, determine the validity and reliability of the instrument, determine the characteristics of items, and analyze the critical thinking skills of class X students on Momentum and Impulse material. This research is Research and Development (R&D) and uses the Borg and Gall development model. The research methods used included interviews, tests, questionnaires, and documentation. The four levels of the four-tier multiple-choice test instrument consist of multiple-choice questions, answer confidence level, reasonable choice, and reason confidence level. Validity testing by the three validators obtained the result that the instrument developed was the very valid category. Reliability testing shows that the developed question instrument is reliable with the acquisition of a value of 0.946. The test questions consist of 14 questions in the medium category and 7 questions in the difficult category. The difference in power index is 0.22 to 0.53 which consists of 11 good enough questions, 6 good questions, and 4 very good questions. Overall results of critical thinking skills of students 11.93% in the very critical category, 5.93% in the critical category, 12.2% in the less critical category type 1A (LC1A), 10.87% in the less critical category type 1B (LC1B), 8.56% in the less critical category type 2A (LC2A), 9% in the less critical category type 2B (LC2B), and 41.36% in the non-critical category.

To cite this article:

Introduction

The low ability of students in science is one of the problems of education in Indonesia. This is based on the results of studies of international institutions TIMSS and PISA related to the cognitive abilities of students. The results of the TIMSS study in 2015 showed that in the field of science from 49 participating countries Indonesia was ranked 44th (Hadi & Novaliyosi, 2019). PISA results in 2018 also showed that out of 79 the number of participating countries Indonesia was ranked 71st (Harususilo, 2019). PISA and TIMSS study subjects are students at elementary and junior high school levels. Critical thinking is one of the aspects tested in the PISA (Pratiwi, 2019) and TIMSS (Hadi & Novaliyosi, 2019).

Critical Thinking Skill is a way of thinking that is reflective, reasonable, or based on the reason that is focused on determining what is believed and will be done. The 2013 curriculum

aims to improve the quality of learning by implementing the 4C characteristic learning program, one of which is critical thinking skills (Arifin, 2018). Many students still experience difficulties in physics. The cause of the difficulties is because the physics learning process that has been implemented has not yet integrated formal knowledge of physics with daily experiences that are experienced and occur in the environment of students (Oktaviani, Gunawan, & Sutrio, 2017). Innovative and critical thinking from students can be raised through the integration of local wisdom in educational and learning activities (Satriawan, Subhan, & Fatimah, 2017). The integration of local wisdom values   can also be done by teachers in assessment activities.

Essay questions are a form of questions that are often used to analyze students' critical thinking skills, such as research conducted by Putri, Nevrita, & Hindrasti (2017). The emergence of the subjectivity effect in the assessment process is a weakness of the essay problem (Mukti & Istiyono, 2018). Multiple choice questions instruments are rarely used to measure students' critical thinking skills because their assessment is only on the content of knowledge without considering the reasons behind the choice. New test model solutions are needed to avoid weaknesses in essay and multiple-choice questions used to analyze students' critical thinking skills. The form of four-tier multiple choice test questions is a form of questions that provides answer choices, choice of reasons, and the level of confidence in choosing answers and reasons.

The four-tier multiple-choice test instrument consists of four tiers. The first tier is a multiple choice question with five answer choices. The second tier is the level of confidence in the answer to the first tier. The third tier is the reason for the answer at the first tier with four choices and one choice of open reasons that can be filled by yourself. According to Suparno (2013), the purpose of providing open reasons is for students to have freedom in giving reasons in choosing the answers to be conveyed. The fourth tier is the level of confidence in the reason for the answer at the third tier. The level of confidence developed is in the range of one to six.

One of the advantages of the four-tier multiple-choice test compared to conventional multiple-choice is that it reduces errors in measurement (Ramadhan, Dwijananti, & Wahyuni, 2018). The advantages of a four-tier multiple-choice test include reducing errors in measurement and based on differences in the level of confidence in the answers and the level of confidence the reason has chosen, the teacher can dig deeper about the power of students' critical thinking and can plan better learning strategies to improve critical thinking skills learners.

The writing of the four-tier multiple-choice test items developed in addition to referring to the cognitive domain is also associated with examples of the value of local wisdom that has developed in society. One of the physics material which is closely related to local wisdom is the material of Momentum and Impulse. Examples of local wisdom related to the material Momentum and Impulse include the process of making a keris related to the concept of impulses, "*bola kasti*" games related to the concept of momentum, and "*kentongan"* musical instruments that apply the concept of collision. Students are required to be able to explore and construct critical thinking skills through the material Momentum and Impulses that are learned. Research conducted by Agustin, Yuliati, & Zulaikah (2016) shows that many

students experience errors in solving Momentum and Impulse problems, including errors in presenting questions, concepts, and problem-solving strategies. This study aims to develop a four-tier multiple-choice test instrument characterized by the values   of local wisdom to describe the characteristics of the instrument, determine the validity and reliability of the instrument, determine the characteristics of items, and analyze the critical thinking skills of class X students on Momentum and Impulse material.

Method

This research is categorized as research and development (R&D). The development procedure follows the development procedures of Borg and Gall (1979), which include needs analysis and information gathering, research objectives, product development, small-scale trials, product revisions I, wide-scale trials, product revisions II, field tests, revision of final products, dissemination and implementation. The development procedure used in this study only reached the second stage of the revision, because at this stage the validity, reliability, difficulty level, and differentiating power of the questions were known, and the final product of the four-tier multiple-choice test was characterized by local wisdom values can be used to analyze students' critical thinking skills.

The study was conducted at MAN Demak. The research subjects were students of class X MAN Demak. Small-scale trial subjects were 32 students from class X IPA 6 and wide-scale trial subjects were 68 students from class X IPA 1, X IPA 2, and X IPA 5.

Data collection techniques are done through tests, questionnaires, interviews, and documentation. The test is carried out by testing the four-tier multiple-choice test instrument on a small-scale trial and a wide-scale trial. An example of a four-level multiple-choice test is shown in Figure 1.

There are two types of questionnaires given to students, namely the response questionnaire given on the small scale trial and the questionnaire on the wide-scale trial response. Indicators for assessing student response questionnaires on a small-scale test include the suitability of the material, use of sentences, readability of questions, readability of images, and suitability of time to work on questions. Indicators for assessing student response questionnaires on a wide-scale test include the suitability of questions with learning materials, clarity of questions, readability of questions, number and timing of questions, indicators of critical thinking skills, the content of local wisdom, and the benefits of the four-tier multiple-choice test.

Interviews with teachers were conducted for the analysis of initial needs, gathering information, and knowing the teacher's assessment of the four-tier multiple-choice test characterized by the values   of local wisdom developed. The assessment indicators used as an interview guide for teachers include the suitability of the questions with basic competencies and learning indicators, the suitability between the content of local wisdom and the material that has been taught, the readability of the questions, the suitability between the number of questions and the processing time of the questions, the benefits of the four-tier multiple-choice test in learning, and teacher interest in the four-tier multiple-choice test characterized by local wisdom values. Documentation techniques are used to obtain data from secondary sources. Documentation techniques in this study were used to collect the names of students who were used as research subjects.

A group of children are playing fireball. If the ball is kicked hard then the ball is difficult to stop, or it can be said that the momentum of the ball is large. Meanwhile, if the ball is kicked slowly then the ball is easily stopped or it can be said that the momentum of the ball is small. Based on this description, what is the definition of momentum in your opinion?

A. Momentum is a measure of the average force received by an object
B. momentum is a measure of the average speed of a moving object
C. momentum is a measure of the change in impulse of a moving object
D. Momentum is a measure of the force used to make an object move
E. Momentum is a measure of the difficulty of stopping a moving object

Are you sure about your answer?

1	2	3	4	5	6
just guessing	very unsure	not sure	sure	very sure	very very sure

Reason

A. The magnitude of the momentum is directly proportional to the mass and velocity of the object
B. The magnitude of the momentum is inversely proportional to the mass and velocity of the object
C. The magnitude of the momentum is inversely proportional to the force received by the object
D. The magnitude of the momentum is directly proportional to the change in the object's impulse
E. …………………………………………………………………………………..

Are you sure of your reasoning?

1	2	3	4	5	6
just guessing	very unsure	not sure	sure	very sure	very very sure

Figure 1. An example of a four-level multiple-choice test

Data analysis techniques included validity, reliability, level of difficulty, different power, and interpretation of students' critical thinking skills. Validity testing is carried out by material expert lecturers, evaluation expert lecturers, and physics teachers. The assessment criteria for product validity include the readability of the questions, the material of the questions, the language used, the construction of the questions, and the processing time of the questions Reliability testing uses the Alpha Cronbach formula. Analysis of the interpretation of the results of students' critical thinking skills using the results interpretation table can be seen in Table 1. A score of '1' is given if the answer or reason given by students is correct and a score of '0' is given if the answer or reason given by students is wrong. If the chosen scale is 4, 5, or 6 then the level of confidence is high and if the scale chosen is 1, 2, or 3 then the level of confidence is classified as low (Fariyani, Rusilowati, & Sugianto, 2015).

Table 1. Interpretation of Four-Tier Multiple Choice Test Results
Characterized by Local Wisdom Values

Answer	Answer's Confidence Level	Reason	Reason's Confidence Level	Category Critical Thinking Skill
correct	High	correct	High	very critical
correct	High	correct	Low	critical
correct	Low	correct	High	
wrong	High	correct	High	less critical type 1A
wrong	Low	correct	High	
correct	Low	correct	Low	less critical type 1B
wrong	High	correct	Low	
wrong	Low	correct	Low	
correct	High	wrong	High	less critical type 2A
correct	Low	wrong	High	
correct	High	wrong	Low	less critical type 2B
correct	Low	wrong	Low	
wrong	High	wrong	High	non-critical
wrong	High	wrong	Low	
wrong	Low	wrong	High	
wrong	Low	wrong	Low	

Results and Discussion

The following stages of research:
1. **Needs Analysis and Information Gathering**

    The information obtained based on the results of the TIMSS and PISA studies shows that the critical thinking skills of students at the elementary and junior high school levels are still low. Based on research conducted by Sari, Parno, & Taufiq (2016) also shows that the critical thinking ability of students in class XI science on Newton's Law material is still low.

    In addition to the literature review, needs analysis and information collection were also conducted through interviews with physics teachers at MAN Demak. The information obtained shows that in the evaluation activities the teacher has never conducted an evaluation related to the critical thinking skills of students.

2. **Research Objectives**

    Based on the data and information obtained from the needs analysis and information collection stages, it was determined that the purpose of this study was to develop a four-tier multiple-choice test instrument characterized by local wisdom values in the Momentum and Impulse material to analyze the critical thinking skills of class X students.

3. **Product Development**

The results of the development of the four-tier multiple-choice test instrument are characterized by local wisdom values in the form of 35 questions obtained from the translation of 7 question indicators from the description of 2 Basic Competencies in Momentum and Impulse Materials.

4. Instrument Validity

Validity analysis is used to determine the validity of the instruments developed. The instrument is said to be valid if it can measure what is to be measured (Suparno, 2010), in this study is the critical thinking skills of students. Validity test is carried out by evaluation expert lecturers, material expert lecturers, and physics teachers.

Validity testing is carried out on each item, to find out that each item developed has been feasible and can measure what is being measured (Fariyani et al., 2015). If the assessment is carried out on each item, then identifying which parts need improvement, both the suitability of the material, sentence questions, choice of answers, and choice of reasons will also be easier. If the assessment is done as a whole, then the location of weaknesses in each item will be difficult to find. The results of the validation of the three validators received an average rating of 24.4 or 93.9% and included in the very valid category. The validity of the instrument is based on the validity criteria for PKGS instruments in Table 2.

Table 2. Validity Criteria for PKGS Instruments

Percentage (%)	Interpretation
81 – 100	Very Valid
61 – 80	Valid
41 – 60	Less Valid
21 – 40	Invalid
0,0 – 20	Very Invalid

(Akbar, 2013)

5. Small-Scale Trials

The results of the small-scale test are used to determine the reliability of the questions, the level of difficulty of the questions, and the different power of the questions. Reliability is a level of consistency or constancy. The instrument is said to be reliable if it has consistent results in measuring what it wants to measure (Sukardi, 2011). Based on the reliability analysis results show that $r_{11} > r_{table}$, with $r_{11} = 0.946$ and $r_{table} = 0.349$, this shows that the four-tier multiple-choice test instrument characterized by the values of local wisdom developed is reliable. The instrument developed can analyze students' critical thinking skills in the Momentum and Impulse material.

A question is said to be good if the problem is valid, reliable, has a good level of difficulty, and different power. Characteristics of items include the level of difficulty and power of different questions. The difficulty level of the items is a measurement of the level of ease or difficulty of an item for students. Questions that are tested on a wide-scale test and serve as the final product of development are questions that are currently being selected with a greater number than questions with difficult categories. A good problem is

a question with a medium category, which is not too easy and not too difficult (Arikunto, 2007). The difficulty level of 21 items consists of 14 questions in the medium category and 7 questions in the difficult category.

The difference in power is the ability of a question to be able to distinguish between high-capacity students and low-ability students (Arikunto, 2007). The questions used for wide-scale testing and used as the final product of development are questions with quite good, good, and very good categories. The power difference of 21 items has a different power index from 0.22 to 0.53 with the category 11 good enough, 6 good questions, and 4 very good questions.

6. Product Revisions I

The results of the revision after the small-scale trial resulted in 21 multiple choice questions with four levels characterized by local wisdom values. The characteristics of the final product developed in the form of a four-tier multiple-choice test instrument characterized by local wisdom values are used to analyze students' critical thinking skills on the Momentum and Impulse material. Each item consists of four tiers, namely questions with answer choices, choice of the confidence level of answers, choice of reasons, and choice of reasons for confidence level. The level of confidence in the reasons provided ranges from one to six. The confidence level of scale one is just guessing, scale two is very unsure, scale three is unsure, scale four is sure, scale five is very sure, and scale six is very very sure.

7. Wide-Scale Trials

Test results on a wide-scale test are used to analyze students' critical thinking skills.

8. Product Revisions II

The final question of the results of the development is 21 items, consisting of seven problem indicators, five indicators of critical thinking skills, and three levels of the cognitive realm. The values of local wisdom contained in the matter include traditional musical instruments, traditional games, traditional means of transportation, and traditions of local communities.

**Student Critical Thinking Skill Results**

Data obtained from the wide-scale test are then interpreted to classify students into very critical (VC), critical (C), less critical type 1A (LC1A), less critical type 1B (LC1B), less critical type 2A (LC2A), less critical type 2B (LC2B), and not critical (NC).

First Indicator: Elementary Clarification

The indicator elementary clarification contained five items of the total number of items developed. The summary of students' critical thinking skills research results on indicators providing elementary clarification is presented in Figure 2.

Based on the results obtained from the study, the uncritical category gets the highest percentage, this condition is caused by students' lack of understanding of the concepts given or learned. The uncritical condition is also caused because in the learning activities the concepts taught are not following the mental development of students (Mubarak, Susilaningsih, & Cahyono, 2016). Categories that have a high percentage under the uncritical

category are less critical categories of type 1A. Less critical type 1A occurs when students give the wrong answer but the reason given is right with a high level of confidence, this condition occurs when students understand the material provided with an incorrect concept, so students can give the right analogy but with the concept incorrect (Mubarak et al., 2016).
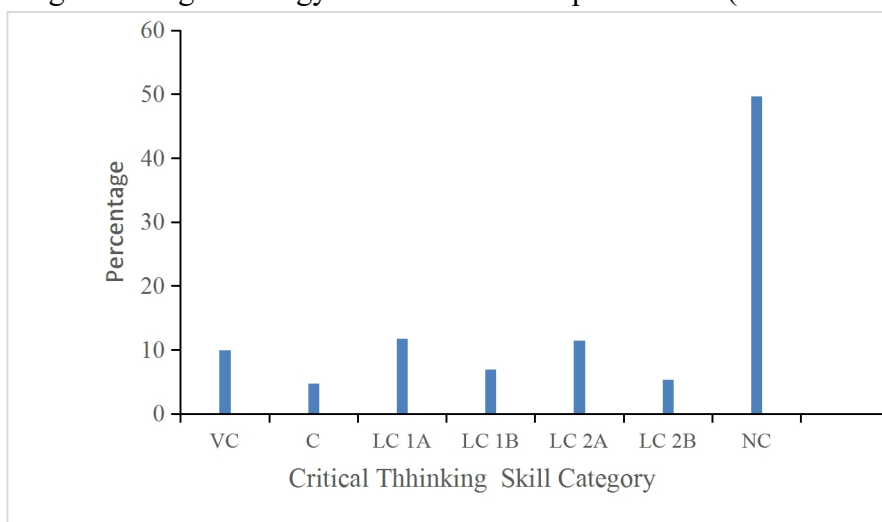

Figure 2. Grouping Diagrams on Elementary Clarification Indicators

According to Nurhayati, Rahayu, & Yahmin (2016), the cause of the low ability of students in making elementary clarifications is because students have not been able to properly analyze the problems contained in the questions. The flow of thinking of students is still not good in providing answers and reasons.

The Second Indicator: Basic Support

Indicators basic support contained in five items from the total number of items developed. A summary of the results of students' critical thinking skills on the indicators of building basic support is presented in Figure 3.
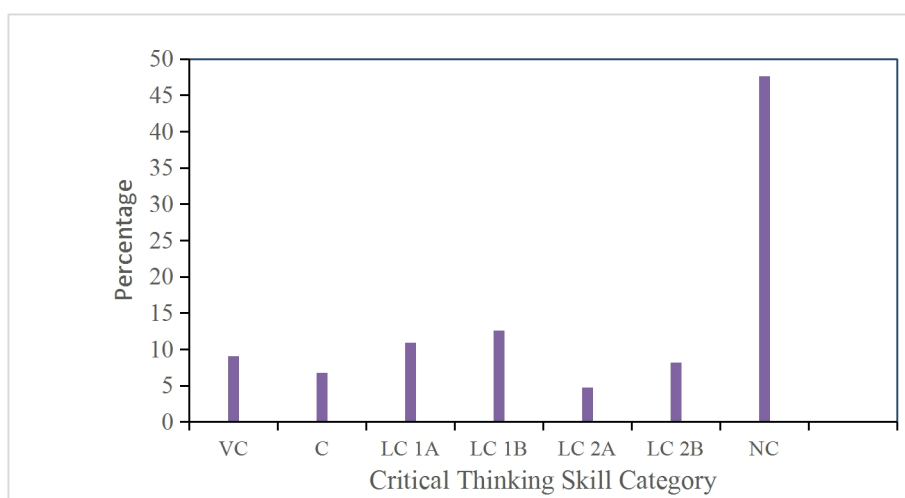
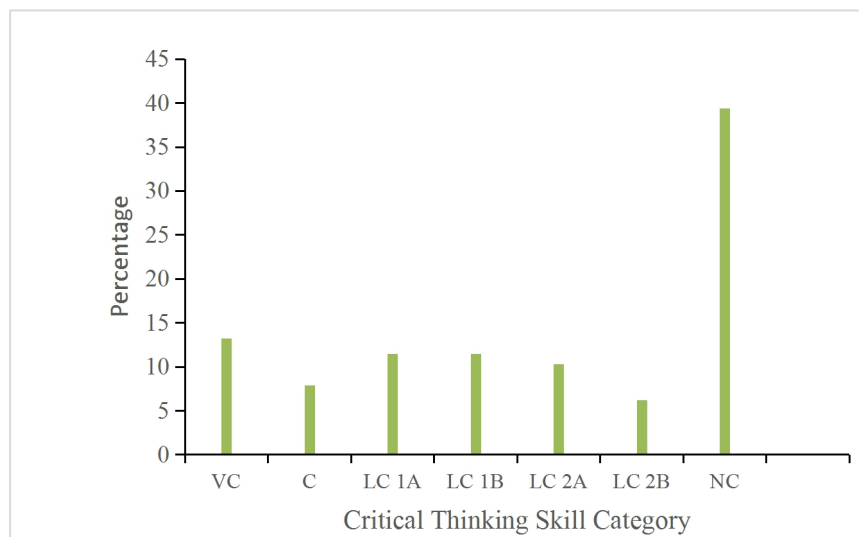**Figure 3.** Grouping Diagrams on Basic Support Indicators

Based on the results obtained, the uncritical category has the highest percentage of indicators building basic support. Uncritical conditions occur when students provide answers and reasons that are not appropriate to the problems presented. Categories that have a high percentage under the uncritical category are less critical categories of type 1B. The condition of LC1B occurs when students are unsure of the reasons they give. This is because students are between understanding and not understanding the problems given (Bandura, 1993).

Poor results in the aspect of basic support skills were also obtained in research conducted by Rahmawati, Hidayat, & Rahayu (2016) on junior high school students. Lack of practice questions that train students to observe cases is the cause of students' difficulties in observing a case in indicators of building basic skills. Students tend to more easily solve problems if the problem is known and refers to the formula used (Pradana, Parno, & Handayanto, 2017).

Third Indicator: Inference

The indicator makes an inference contained in five items from the total number of items developed. The summary of students' critical thinking skills research results on the indicators making an inference is presented in Figure 4.

Figure 4. Grouping Diagrams on Inference Indicators

Based on the results obtained, the uncritical category has the largest percentage of indicators making inferences. Uncritical conditions occur when the responses given by students are unclear and the information provided is also illogical. The cause of the low ability of students to make inferences is the lack of student practice to make inferences in the learning process and the students' lack of understanding of the material presented.

The very critical category gets the biggest percentage under the uncritical category. When students have mastered the material provided and can distinguish between something they already know and don't know, students are classified as very critical (Mubarak et al., 2016).

Fourth Indicator: Advanced Clarification

The indicator advanced clarification contained in three items of the total number of items developed. The summary of students' critical thinking skills research results on indicators advanced clarification is presented in Figure 5.
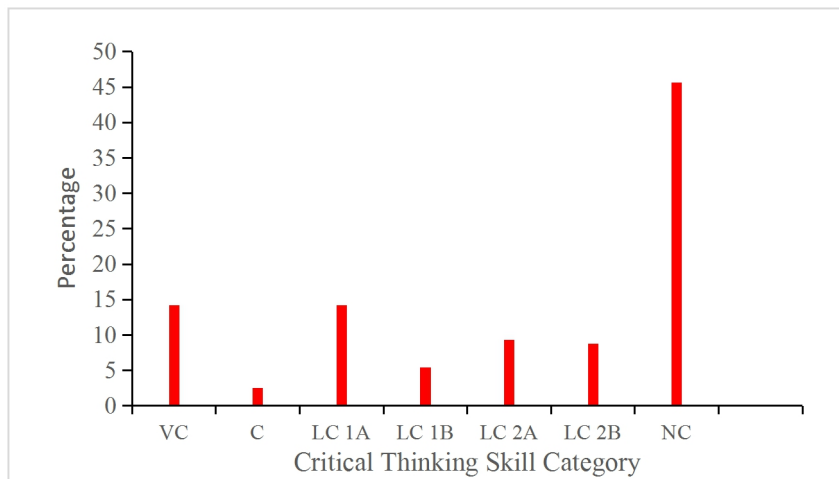

Figure 5. Grouping Diagrams on Advanced Clarification Indicators

Based on the results obtained, the uncritical category has the largest percentage of indicators providing advanced clarification. Categories that have a high percentage under the uncritical category are the very critical category and the less critical category type 1A. Very critical conditions occur when students can provide answers and the right reasons and with a high level of confidence in the questions given. While less critical type 1A, namely the condition of students giving the wrong answer but the reason given is right with a high level of confidence. The cause of this condition is the carelessness of students in choosing answers, students will not give the right reasons if the answers given are wrong (Syahrul & Setyarsih, 2015).

A learning method is needed that can bring up critical thinking skills of students, such as research conducted by Nurhayati et al., (2016) on skills to make advanced clarification. The results of this study indicate that the ability to make advanced clarification of students in the control class who get conventional learning methods is low, in contrast to the experimental class that uses the Learning Cycle method contextual Socio Scientific Issues has moderate criteria.

Fifth Indicator: Strategy and Tactics

The indicators strategy and tactics contained in three items of the total number of items developed. The summary of students' critical thinking skills research results on indicators compiling strategies and tactics is presented in Figure 6.

Based on the results obtained, the uncritical category has the highest percentage of indicators compiling strategies and tactics. The category that has a high percentage under the uncritical category is the less critical category type 1B, namely when students provide the correct answer and reason but with a low level of confidence. Students who give wrong

answers, but the reasons are correct with a low level of confidence included in this category. The cause of this condition is that students are in between understanding and not understanding of the problem given so that they become less sure of the answers they give (Bandura, 1993).
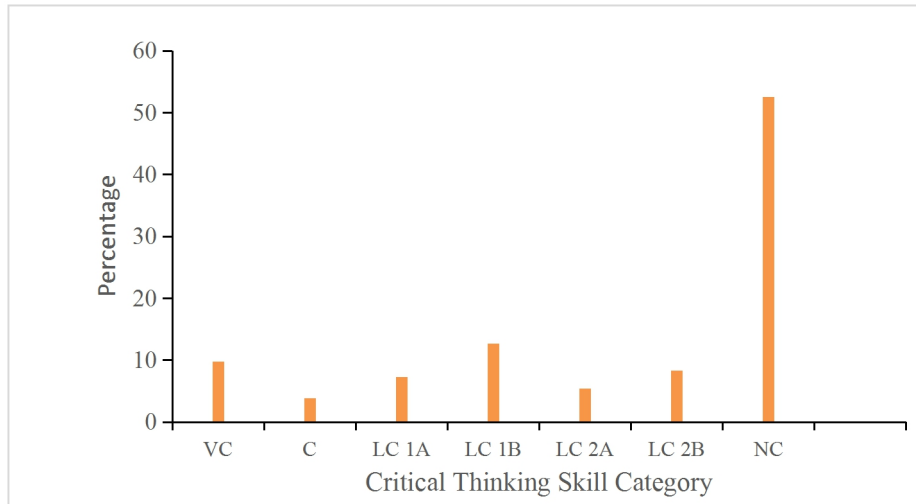

Figure 6. Grouping Diagrams on Indicators Strategy and Tactics

Research conducted by Pradana *et al.*, (2017) also shows that the ability of students in managing strategies and tactics is still lacking. Students must be trained to get used to linking the knowledge they learn with real events that are experienced, such as local wisdom that is around the environment of students. It aims if students face problems in real life, students can analyze and determine an action using the sciences they have learned.

The indicator that gets the highest percentage in the very critical category is the advanced clarification indicator. These results are following Pradana et al. (2017) who obtained the results that the critical thinking skills of physics students in providing further clarification had the highest average value of 79.92.

The indicators that get the highest percentage in the non-critical category are indicators of building strategies and tactics. These results are consistent with the research conducted by Rahmawati et al. (2016) towards students at the junior high school level, which shows that students' answers to the questions given are still not good. The flow of thinking of students and the concepts they have is not related. Interpretation of students' critical thinking skills analysis results for all indicators of critical thinking skills is presented in Figure 7.
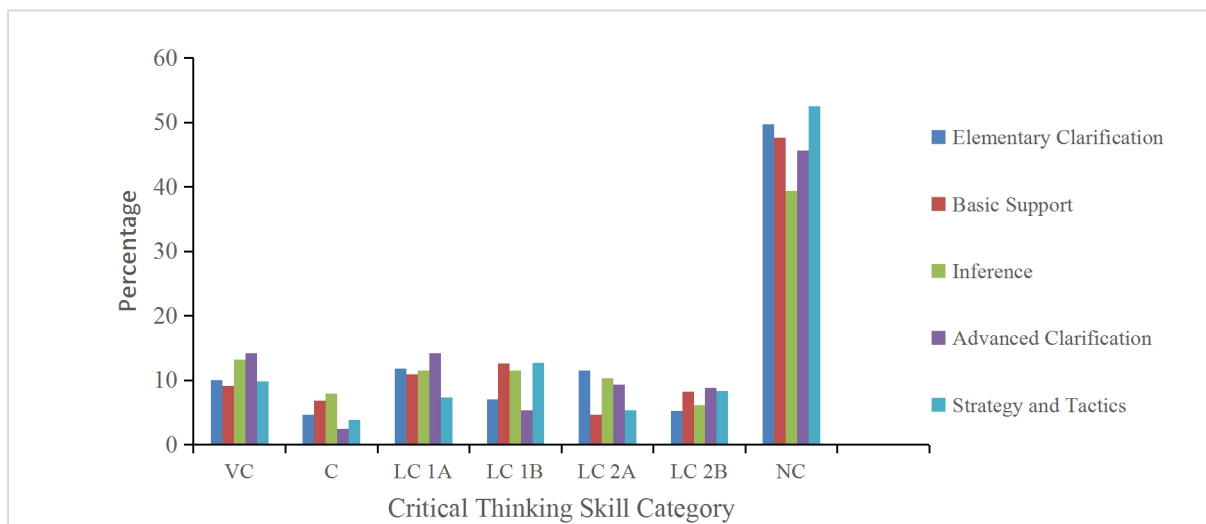
Figure 7. Interpretation of critical thinking skill

The category that gets the highest percentage on each indicator of critical thinking skills is the non-critical category. This shows that students' critical thinking skills at the high school level as a whole are still low. This result is following several other studies that analyzed the critical thinking skills of students at the junior high and tertiary education levels. Students' critical thinking skills at each level are still low.

The cause of the low critical thinking skills of students is that students are not accustomed to working on questions that require critical thinking aspects. Students should be given learning that trains students' abilities in solving problems that require critical thinking analysis. In addition to learning activities, teachers can also integrate the values of local wisdom in assessment activities. The integration of local wisdom in educational and learning activities has the potential to generate innovative and critical thinking from students (Satriawan, Subhan, & Fatimah, 2017). Improving critical thinking power can be done through the provision of test instruments that are well designed and by the level of students' thinking abilities.

Conclusion

The final product of the development is a four-level multiple-choice test instrument characterized by local wisdom values that serve to analyze students' critical thinking skills on the Momentum and Impus material. Each item consists of four levels. The final product produced consisted of 21 items, consisting of 7 question indicators, 5 indicators of critical thinking skills, and 3 levels of cognitive domains. The values of local wisdom listed in the matter include traditional musical instruments, traditional games, traditional means of transportation, and local community traditions. Based on the validation analysis of the three validators, it was shown that the question instruments developed were included in the excellent category by getting an average rating of 24.4 or 93.85% validation. The reliability test analysis of the developed test instruments obtained the results of $r_{11} > r_{table}$, with $r_{11} = 0.946$ and $r_{table} =. 0.349$, it shows that the instrument is a matter of reliability. Questions consisted of 14 questions with moderate difficulty and 7 questions with difficult difficulty. The difference in power includes 11 good enough questions, 6 good questions, and 4 very good questions. Analysis of critical thinking skills on the Momentum and Impulse material

shows that students are in the very critical category at 11.93%, the critical category at 5.93%, the less critical category type 1A (LC1A) at 12.2%, the less critical category type 1B (LC1B) by 10.87%, less critical category type 2A (LC2A) by 8.56%, less critical category type 2B (LC2B) by 9%, and uncritical category by 41.36%.

References

Agustin, D. K., Yuliati, L., & Zulaikah, S. (2016). Kesalahan Siswa SMA dalam Memecahkan Masalah Momentum-Impuls. *Pros. Semnas Pend. IPA Pascasarjana UM*, 174–183. Malang: Universitas Negeri Malang.

Akbar, S. (2013). *Instrumen Perangkat Pembelajaran*. Bandung: PT Remaja Rosdakarya Offset.

Arifin, Z. (2018). Pengembangan Instrumen Analisis Kesalahan untuk Mengukur Critical Thinking Skill Berdasarkan Kompetensi Inti Kurikulum. *EduMa*, *7*(1), 53-62.

Arikunto, S. (2007). *Dasar-Dasar Evaluasi Pendidikan* (Revisi). Jakarta: Bumi Aksara.

Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, *28*(2), 117–148.

Fariyani, Q., Rusilowati, A., & Sugianto. (2015). Pengembangan Four-Tier Diagnostic Test untuk Mengungkap Miskonsepsi Fisika Siswa SMA Kelas X. *Journal of Innovative Science Education*, *4*(2), 41–49.

Hadi, S., & Novaliyosi. (2019). TIMSS Indonesia (Trends In International Mathematics And Science Study). *Prosiding Seminar Nasional & Call For Papers*, 562–569.

Harususilo, Y. E. (2019, December 4). Skor PISA Terbaru Indonesia. *Kompas.Com*.

Mubarak, S., Susilaningsih, E., & Cahyono, E. (2016). Pengembangan Tes Diagnostik Three Tier Multiple Choice. *Journal of Innovative Science Education*, *5*(2), 101–110.

Mukti, T. S., & Istiyono, E. D. I. (2018). Instrumen Penilaian Kemampuan Berpikir Kritis Peserta Didik SMA Negeri Mata Pelajaran Biologi Kelas X Instrument for Assessing the Critical Thinking Ability of High School Students on Biology Subject X Grade. *Jurnal Pendidikan Biolog*, *11*(2), 105–110.

Oktaviani, W., Gunawan, & Sutrio. (2017). Pengembangan Bahan Ajar Fisika Kontekstual untuk Meningkatkan Penguasaan Konsep Siswa. *Jurnal Pendidikan Fisika Dan Teknologi*, *III*(1), 1–7.

Pradana, S. D. S., Parno, & Handayanto, S. K. (2017). Pengembangan Tes Kemampuan Berpikir Kritis Pada Materi Optik Geometri Untuk Mahasiswa Fisika. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *21*(1), 51–64. https://doi.org/DOI: http://dx.doi.org/10.21831/pep.v21i1.13139

Pratiwi, I. (2019). Efek Program Pisa Terhadap Kurikulum di Indonesia. *Jurnal Pendidikan Dan Kebudayaan*, *4*, 51–71.

Putri, O. D., Nevrita, & Hindrasti, N. E. K. (2017). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Siswa SMA pada Materi Sistem Pencernaan. *Jurnal Pendidikan Biologi*, 14–27.

Ramadhan, G., Dwijananti, P., & Wahyuni, S. (2018). Analisis Kemampuan Berpikir Tingkat Tinggi (High Order Thinking Skills) Menggunakan Instrumen Two Tier Multiple Choice Materi Konsep dan Fenomena. *Unnes Physics Education Journal 7*, *7*(3), 85–90.

Sari, A. L. R., Parno, & Taufiq, A. (2016). Kemampuan Berpikir Kritis dan Pemahaman Konsep Fisika Siswa SMA pada Materi Hukum Newton. *Pros. Seminar Pend. IPA Pascasarjana UM*, *1*, 88–99.

Satriawan, M., Subhan, M., & Fatimah. (2017). Pembelajaran Fisika Berbantuan Bahan Ajar Berbasis Kontekstual dengan Mengintegrasi Kearifan Lokal untuk Meningkatkan Penguasaan Konsep Fisika. *Jurnal Penelitian Pembelajaran Fisika*, *8*(2), 115–120.

Sukardi. (2011). *Evaluasi Pendidikan (Prinsip & Operasionalnya)* (1st ed.; F. Yustianti, ed.). Jakarta: Bumi Aksara.

Suparno. (2013). *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. Jakarta: PT. Grasindo.

Suparno, P. (2010). *Metode Penelitian Pendidikan Fisika (Buku Kuliah Mahasiswa)*. Yogyakarta: Universitas Sanata Dharma.

Syahrul, D. A., & Setyarsih, W. (2015). Identifikasi Miskonsepsi dan Penyebab Miskonsepsi Siswa dengan Three-tier Diagnostic Test pada Materi Dinamika Rotasi. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, *04*(03), 67–70.