The Validity of the Internalized Argumentation Skills Test for Chemistry Students

Suyono, Suyono^{1*}, Harun, Nasrurdin¹, Bertha, Yonata¹, Duhita², Savira Wardani², and Anggita, Intan Kumala Dewi¹

¹ Chemistry Department, Universitas Negeri Surabaya, Indonesia

² IKIP Siliwangi Bandung, Bandung, Indonesia

Abstract. This study aims to produce a package of questions measuring internalized argumentation skills (IAS) that meet the requirements of logical validity (content and construct) and empirical validity. The research follows Fenrich's development model. The implementation of the test package to test the empirical validity involved 30 participants, students of the Chemistry Education Study Program. The research data were analyzed descriptively. The results of the research, a package of IAS measuring questions for chemistry students has been successfully developed. The IAS measuring test package consists of five questions with details of one easy level question, three medium level questions, and one difficult level question. The package of questions developed has met the requirements of logical validity (content and construct) and empirical validity.

Keywords: IAS, Logical validity, Empirical validity

1 Introduction

Argumentation skills (AS) need to be trained to students by the parties for at least four reasons. First, AS is a moderating skill over critical thinking skills [1,2]. Second, AS together with critical skills and problem-solving skills are one unit [3]. Third, problem solving skills are one of the thinking skills demanded by the 21st century [4,5]. Fourth, AS is not an ability that can develop by itself along with human physical development [6,7].

Researchers try to ratify the recommendations of the experts above. In 2019 through Physical Chemistry 3 lectures, researchers trained AS to chemistry students at FMIPA Unesa. The exercises are managed through structured assignments assisted by SLM-PC3_AS. SLM-PC3_AS is an acronym for Structured Lecture Materials- Physical Chemistry 3 to train AS. The SLM-PC3_AS used has met the requirements of validity (consistency and relevance), practicality, and effectiveness [8,9]. Researchers have measured the student's AS after the exercise and concluded that the student's AS achievement is above a score of 70 [9].

Students' AS due to the training they undergo is expected to be integrated or internalized in their cognitive structure. This expectation is based on a statement that learning science through scientific argumentation has a positive and significant impact on the ability to argue and understand students' concepts [3]. If this internalization occurs, students will be able to

assess chemical claims outside of PC3 content. This hypothetical sentence can be based on a framework of thinking based on Piaget's assimilation theory [10].

According to assimilation theory, every person has a schema that allows him to remember and respond to stimuli that come from the environment. When AS has been internalized in the student's scheme, it allows students to use their experiences in argumentation when presented with a new claim? When presented with a new claim and asked to rate, students will make a decision to accept or reject the claim, collect data to support the decisions taken, and write a narrative that explains the relationship between the data held and the decisions made. When the argumentation practice is rehearsed repeatedly, the AS will be integrated or internalized in the student's scheme.

If the internalization of AS has occurred in the student's scheme after being trained through PC3 lectures, then students will be able to assess chemistry claims outside the PC3 topic. This hypothetical sentence needs to be tested through research activities. The ability of post-training students to use SLM-PC3_AS in this article is named Internalized Argumentation Ability (IAS). IAS needs to be tested scientifically through an assessment process. The assessment process begins with measurement. This study aims to produce a package of questions measuring IAS that meets the requirements of logical validity and empirical validity.

Chemical claims outside the topic of PC3 have colored video content that has gone viral in the

* Corresponding author : duhita@ikipsiliwangi.ac.id

community. The chemical claims circulating through social media are both true and false. Not a few people were then confused because they did not get an explanation from the parties about the truth of the claims. The presence of students in the midst of society like this is very much needed. The qualifications of students who will participate in solving problems in society must be guaranteed by higher education institutions. The availability of a learning device in the form of a student IAS measuring test package will provide benefits for efforts to guarantee the quality of graduates by higher education institutions. This is the benefit of development research conducted by researchers.

Each assessment requires an assessment method according to its respective characteristics [11,12]. The assessment process begins with measurement. Various student activities in learning often cannot be measured by traditional assessment [13]. This is in line with the shift in learning from teacher centered to student centered. The assessment approach shifts not only prioritizing assessment of learning but more towards assessment for learning and assessment as learning. Assessment as learning requires teachers to switch from knowledge bearers to knowledge guides, namely guiding students through the process of understanding "cognitive processes" so that students learn to monitor their learning and make adjustments [14].

Assessment as learning refers to assessment as a student's metacognitive skill and as a literacy [15]. Comparison between students with each other is almost non-existent [16]. Assessment as learning places more emphasis on involving students in routinely reflecting on their work and making decisions about how students can play a major role in what has been done [16]. In more detail, it is stated that in assessment as learning, what is assessed is each student's thoughts about their learning, what strategies are used to support or improve their learning, and the mechanism by which students make adjustments to help their learning [17].

If you pay attention to the description in the paragraph above, the IAS assessment can be classified into assessment as learning. IAS assessment that involves students actively in assessing claims, collecting supporting data, and creating narratives that connect data with decisions made. Such a series of activities requires students to learn about how they should learn (metacognitive abilities) and need to have the ability to search data to support decisions made (literacy skills). IAS assessment conditions students to assess themselves. This conditioning refers to the opinion that the best judge for a person is himself. Students are their own best assessors [16].

The assessment process begins with measurement. To measure IAS, a question package that can be accounted for is needed. The package of questions must be accountable in terms of appropriateness, validity, reliability, interpretability, and usability. Through this research, a package of questions measuring IAS was developed that can be justified in terms of validity.

A test is said to be valid if it measures what it is intended to measure. The validity of a test can be known

from the results of thought and from the results of experience. The first thing will get logical validity and the second thing will get empirical validity. Logical validity contains the word "logical" derived from the word "logic" which means reasoning, with such meaning then logical validity for an evaluation instrument indicates the conditions for an instrument that meets valid requirements based on the results of reasoning. There are two kinds of logical validity that can be achieved by an instrument, namely content validity and construction validity. A test is said to have content validity if it measures certain specific objectives that are parallel to the material or content of the lesson given. A test is said to have construction validity if the items that make up the package measure every aspect of thinking that is the instructional goal. Empirical validity contains the word "empirical" which means "experience" [18].

Empirical validity is based on test results, not just a test tool. This data testing guide is based on the statement that empirical validity refers to the appropriateness of interpretations made based on test scores that are associated with the specific use of the test instrument and not the instrument itself [19]. So, testing the validity of the IAS measuring test package is seen from two sides, namely on the items and on the test results using those items.

The steps taken in developing a question package include setting clear goals, formulating goal-directed specifications, making grids, compiling instruments, reviewing instruments, analyzing test results, and revising or perfecting question packages [20]. The IAS measuring test package developed has the aim of describing argumentation skills that have been internalized in students' cognitive structures. The formulation of IAS specifications is in accordance with the components that represent argumentation skills (AS). AS is divided into four components, namely (1) compiling claims, (2) showing evidence, (3) compiling reasons, and (4) compiling counterarguments [21].

The formulation of the first, second, third, and fourth AS component indicators, respectively, are: (1) students can write a statement which is a claim that is compiled in providing an assessment of a phenomenon, (2) students can write down some appropriate evidence to strengthen it. claims that have been prepared, (3) students can formulate statements to explain the suitability of the evidence submitted in strengthening the claims that have been prepared, and (4) students can write a statement and include the reasons put forward to say that a statement is false [21]. The fourth component indicator is only active when a student has to reject a claim.

The package of measuring questions developed is limited to the second, third, and fourth AS component indicators. The first AS indicator, namely the preparation of claims is intentionally not measured because in the IAS measuring instrument students are tasked with assessing existing chemical claims, which are contained in viral videos on social media. In 2020, researchers have collected several viral videos containing chemical claims and analyzed 11 videos. The

results of the analysis found five videos containing true claims and six videos containing scientifically false claims [22]. The title of the video and the chemical claims contained therein are presented in Table 1.

Table 1. The title of the video and the chemical claims

No.	Video Title	Claim Statement
1.	Dish soap	The amount of foam is related to the washing power of the soap.
2.	Fizzy Drinks and Mentos Candy	There will be a dangerous chemical reaction if you consume a soda with mentos.
3.	Coconut Shell Smoke Becomes Covid 19 Medicine	The coconut couch contains acid.
4.	Test Iron on Various Quality Drinking Water Products	The water that is drunk from one of the products is a boiled nail.
5.	Benefits of Moringa Leaves	The vermicelli that had been white turned into pitch black is evidence of poison contamination in the vermicelli.
6.	Fake the Look of a Chicken Egg	Citrus vinegar can soften egg shells.
7.	Foods That Shouldn't be Consumed Together	Do not consume at the same time between milk and eggs.
8.	Joint Pain	Betel leaf is rich in antioxidants
9.	Drain the Dirt and Mucus in the Lungs	Guava leaves are rich in vitamin C and contain high levels of iron.
10.	Motorcycle Burns When Sprayed with Disinfectant.	The motor caught fire due to being sprayed with disinfectant
11.	Making Mouthwash from ORS and Honey	The real honey when put in the water won't mix.

In video number 5 entitled "The Benefits of Moringa Leaves" there are false chemical claims. In this video, the presenter makes a statement (claim) that "The vermicelli that was white turned black is proof that there is poison contamination in the vermicelli." This claim is false because the change in the color of the vermicelli from white to deep black (actually blue-black) is not evidence of any toxic contaminants in the vermicelli. The color change is the result of the iodine test which proves the presence of starch (polysaccharide) in the test material. It is known that the main content of vermicelli is starch [23,24].

False claims as in the paragraph above can be used as content in the stem of a IAS measuring question. In this case, at the initial stage students are asked to make a decision to accept or reject the claim. Furthermore, students are asked to find and present supporting data for the decisions taken and provide explanations that connect the data with the decisions taken. The answers of students who accept claims are certainly different from the answers of students who reject claims. Thinking skills like this are what in this study is called IAS. The stages of answering questions (assessing a claim) like this are in accordance with the indicators of skill or argumentation skill (AS).

2 Method

This research is a development research, referring to the development stages of Fenrich (2004) [25] as shown in Figure 1.

Fig. 1. The Fenrich Model Development Cycle (2004)

The analysis was carried out in order to (1) identify, collect, and select viral videos containing chemical knowledge content and (2) content analysis to identify and formulate the claims stated in the selected videos. The results of activities one and two in this analysis stage can be seen in Table 1.

At the planning stage, it begins with determining the IAS indicators to be measured. As already mentioned, the IAS indicators that will be measured are the second, third, and fourth indicators. At this stage, the researcher also compiled a specification table that describes the relationship between IAS indicators and chemical claims contained in each video. This table is then used by researchers for evaluation and decision making that the chemical claims exposed in each video deserve to be used as stems for each IAS item that will be made.

At the design stage, the researcher wrote 11 drafts of IAS questions as a follow-up to the planning stage. Furthermore, a joint review was carried out by the research team for evaluation and revision of the 11 questions made. The revised IAS question package is ready to be validated by an expert. To support the validation process (determination of logical validity), the researcher prepared a validation format to guide the expert in providing validity judgments on the 11 IAS questions that had been made. The researcher hopes that the 11 questions made are valid as instruments for generating argumentation processes in students' cognitive structures. The eleven questions made are expected to be valid or valid to "force" students to write answers to questions that can be used to determine their

IAS scores. The validation table design can be seen in Table 2.

Table 2. Validation Format for IAS

No	Video Title	Claim	Validation														
			IAS Indicator 2					IAS Indicator 3					IAS Indicator 4				
1.			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
2.																	
dst																	
11.																	

Note: Validity scores 1, 2, 3, 4, and 5 represent very invalid, invalid, moderately valid, valid, and very valid, respectively.

To maintain the consistency of the assessment given by the validator, the researcher includes an assessment rubric as follows :

- 1) If the validator has a prediction that students will work on this question 100% based on feelings alone ; not based on search results and reflection on information from related libraries, then the validator can give a very invalid assessment. That is, the question is very invalid in measuring the IAS indicator.
- 2) If the validator has a prediction that students will work on this question 75% based on feelings alone ; 25% based on search results and reflection on information from related libraries, the validator can give an invalid assessment. That is, the question is not valid in measuring the IAS indicator.
- 3) If the validator has a prediction that students will work on this problem 50% based on feelings alone ; 50% based on search results and reflection on information from related libraries, the validator can provide a fairly valid assessment. That is, the question is quite valid in measuring the IAS indicator.
- 4) If the validator has a prediction that students will work on this question 75% based on search results and reflection on information from related libraries, 25% based on mere feelings, then the validator can provide a valid assessment. That is, the question is valid in measuring the IAS indicator.
- 5) If the validator has a prediction that students will work on this problem 100% based on search results and reflection on information from related libraries, then the validator can provide a very valid assessment. That is, the question is very valid in measuring the IAS indicator.

This rubric was formulated by the researcher by adapting the individual confidence level measurement pattern developed by Hassan et al [26]. The use of the phrase “have a prediction” is based on the researcher's consideration that the validators understand the content being assessed.

To determine whether or not there is agreement between validators in providing an assessment, the percentage of agreement value is used using the formula developed by Borich (1994) [27]. The formula is as follows :

$$\text{Percentage of agreement (R)} = \left[1 - \frac{A - B}{A + B} \right] \times 100\%$$

Information :

R : Coefficient of percentage of agreement (R)

A : The highest score from the validator and B is the lowest score from the validator. The conclusion criterion for the R value is that the validators agree on the assessment given if the

R value is 75% [27].

The IAS measuring question is declared to meet the requirements of construct validity, if each indicator gets an assessment with a mode (Mo) of at least 4 and there is no disagreement between the validators. The IAS measuring test package is declared to meet the content validity requirements, if there is not a single negative statement from the validator. The negative statement is that the questions developed do not measure the argumentation thinking process being studied which is one of the principles in assessment as learning.

At the design stage, a format is also prepared for determining the classification of questions into easy, medium, and difficult levels. The difference in the level of questions is based on the level of student difficulty in fulfilling the three IAS indicators. The design of the format is presented in Table 3.

Table 3. Formats for Determining the Level of IAS Questions

No.	Video Title	Claim statement	Rated Question Level		
			Easy	Medium	Difficult
1.					
2.					
etc					
11.					

The IAS question package consisting of 11 questions accompanied by a validation format and a question level determination format was given to three experts for assessment in order to determine logical validity. After the validation stage by experts and the determination of the level of the questions, the researcher revised according to the validator's input and carried out an evaluation to determine the five questions that would be implemented. Student IAS will not be measured using 11 items that have been validated. Measurements using too many essay items are feared to be invalid because of the boredom factor of the tested party. Researchers set five IAS items to be tested for empirical validity ; consisting of one easy question, three medium questions, and one difficult question. The comparison of the number of items at the level of easy : medium : difficult to adapt Silverius' opinion, which is 1 :2 :1 [28].

IAS assessment can be classified into assessment as learning. In assessment as learning, what is assessed is the mind of each student [17]. Comparison between students with each other is almost non-existent [16]. Referring to these two opinions, the researcher does not use the results of quantitative tests such as the different test items and others in determining the validity of the questions. The empirical validity of the student IAS measuring item package is determined by referring to the opinion which states that empirical validity refers to the feasibility of interpretations made based on test scores related to certain uses [19]. The package of IAS measuring questions that were tested on students consisted of questions with easy, medium, and difficult levels. If the student's IAS score has a trend according to the level of difficulty of the questions, then the package of questions developed is declared to meet the requirements of empirical validity.

The scoring of the student's IAS quality is guided by the rubric as shown in Table 4.

Table 4 Rubric for Scoring Student Answers on the IAS Test Sheet

Score	Rubric
20	Students only write statements of acceptance (agree) or reject (disagree) to the claim, but do not write down supporting data.
40	Students write statements of acceptance (agree) or reject (disagree) to claims and write down supporting data for acceptance or rejection of claims but are less precise.
60	Students write statements accepting (agree) or reject (disagree) with claims and write down supporting data for acceptance or rejection of claims that are appropriate, but have not written narratives connecting the data submitted in strengthening or rejecting claims.
80	Students write statements of acceptance (agree) or rejection (disagree) of claims and write down supporting data for acceptance or rejection of claims that are correct. Students have tried to write a narrative of the relationship between the data submitted and the decision to accept or reject the claim, but the interconnection is not visible (not connected).
100	Students write statements of acceptance (agree) or rejection (disagree) of claims and write down supporting data for acceptance or rejection of claims that are correct. Students have tried to write a narrative of the relationship between the data submitted and the decision to accept or reject the claim, the interconnection has also appeared (connected).

This rubric was developed based on the IAS indicator. Scores of 20, 40, 60, 80, and 100 adapt the argumentation quality analysis framework developed by Dawson & Venville (2009) [29]. There is no score 0 because in the IAS question, the claim component must exist because it occupies the stem position of the question. The lowest IAS score is 20 as is the level 1 score in the framework developed by Dawson & Venville. The scoring rubric can be operated by the scorer (corrector) when a key is included with several alternative answers or argumentative statements for each IAS question (available to the researcher).

The target of the implementation of the questions are students of the Chemistry Education Study Program, FMIPA Unesa class of 2017. The research target students are 30 people, from PKA, PKB, and PKU classes, 10 people each taken randomly. The research data were analyzed descriptively.

3 Result and Discussion

The results of the three validators' assessment of logical validity, especially constructs are presented in Table 5.

Table 5. Validity Scores Given by Three Validators to 11 IAS Questions

No	Indicators Measured by Questions	Skor dari Validator			M ₀	Percent of Agreement (%)			Criteria
		V ₁	V ₂	V ₃		R ₁₂	R ₁₃	R _{2,3}	
	Question 1								
1	IAS Indicator 2	4	5	4	4	90	100	90	Valid
2	IAS Indicator 3	4	4	5	4	100	90	90	Valid
3	IAS Indicator 4	4	4	4	4	100	100	100	Valid
	Question 2								
1	IAS Indicator 2	4	5	5	5	90	90	100	Very Valid
2	IAS Indicator 3	4	4	4	4	100	100	100	Valid
3	IAS Indicator 4	4	4	4	4	100	100	100	Valid
	Question 3								
1	IAS Indicator 2	4	5	5	5	90	90	100	Very Valid
2	IAS Indicator 3	4	4	5	4	100	90	90	Valid
3	IAS Indicator 4	4	3	5	4	86	90	75	Valid
	Question 4								
1	IAS Indicator 2	4	4	5	4	100	90	90	Valid
2	IAS Indicator 3	4	3	5	4	86	90	75	Valid
3	IAS Indicator 4	4	3	5	4	86	90	75	Valid
	Question 5								
1	IAS Indicator 2	4	5	5	5	90	90	100	Very Valid
2	IAS Indicator 3	4	4	5	4	100	90	90	Valid
3	IAS Indicator 4	4	4	5	4	100	90	90	Valid
	Question 6								
1	IAS Indicator 2	4	5	5	5	90	90	100	Very Valid
2	IAS Indicator 3	4	5	5	5	90	90	100	Very Valid
3	IAS Indicator 4	4	5	5	5	90	90	100	Very Valid
	Question 7								
1	IAS Indicator 2	4	4	4	4	100	100	100	Valid
2	IAS Indicator 3	4	3	4	4	86	100	86	Valid
3	IAS Indicator 4	4	3	4	4	86	100	86	Valid
	Question 8								
1	IAS Indicator 2	4	4	4	4	100	100	100	Valid
2	IAS Indicator 3	4	4	4	4	100	100	100	Valid
3	IAS Indicator 4	4	4	5	4	100	90	90	Valid
	Question 9								
1	IAS Indicator 2	4	3	5	4	86	90	75	Valid
2	IAS Indicator 3	4	3	5	4	86	90	75	Valid
3	IAS Indicator 4	4	3	5	4	86	90	75	Valid
	Question 10								
1	IAS Indicator 2	4	5	5	5	90	90	100	Very Valid
2	IAS Indicator 3	4	5	5	5	90	90	100	Very Valid
3	IAS Indicator 4	4	5	5	5	90	90	100	Very Valid
	Question 11								
1	IAS Indicator 2	4	4	5	4	100	90	90	Valid
2	IAS Indicator 3	4	4	5	4	100	90	90	Valid
3	IAS Indicator 4	4	4	5	4	100	90	90	Valid

Note: IAS Indicator 2: Students can write down evidence/data that can be used to substantiate claims. IAS Indicator 3: Students can write a narrative that connects the evidence/data submitted in strengthening the claim. IAS Indicator 4: Students can write a statement and include reasons and evidence/data submitted to say that a statement (claim) is false.

The three validators agreed with each other in providing a validity score for the 11 questions made by the researchers, both for the two, three, and four IAS indicators. The percentage of agreement value for each question and/or each indicator is not below 75 [27]. When the construct validity scores given by the first, second, and third validators are 4, 5, and 5 respectively, the decision to take a score of 5 (mode value, Mo) as a representation of the assessment is very valid for the items being assessed (see data on question number 2, indicator 1) cannot be interpreted that actually the first validator does not agree with the decision. The same applies to question number 1, indicator 1. When it is decided that this item is valid (score 4 as the mode), it cannot be interpreted that actually the second validator does not agree with the decision. When the construct validity scores given by the first, second, and third validators are 4, 3, and 5 respectively on question number 4 indicator 2, then the decision to take a score of 4 (median value, Me) as a representation of valid assessment for the item is already in accordance with the principles of descriptive statistics used for data analysis. The eleven AS questions that were developed received a valid (majority) and very valid assessment. This means that the 11 AS items developed can be trusted to measure students' IAS. This means that the 11 questions developed are declared to meet the requirements of logical validity, especially in terms of constructs.

There was not a single statement given by the validator that the questions developed did not measure certain specific objectives that were parallel to the chemical content in the argumentation thinking process being studied. This data is evidence that the package of questions developed meets the requirements of content validity [18]. Thus, it can be concluded that the 11 IAS measuring items developed have met the requirements of logical validity, both construct (consistency) and content (relevance).

Of the 11 items measuring IAS that have met logical validity, five items will be tested for empirical validity. The selection of these five questions begins with determining the level of the questions into three classes, namely easy, medium, and difficult. The ranking of questions is based on the results of the assessment of three experts in the field of chemistry learning. The ranking of items by the three experts is presented in Table 6.

Table 6. Easy, Medium, and Difficult Ranking Results for 11 IAS Questions

No	Question	Difficulty Level			Results
		Easy	Med	Difficult	
1					
1.	Validator 1	√			Easy
2.	Validator 2			√	
3.	Validator 3	√			
2					

No	Question	Difficulty Level			Results
		Easy	Med	Difficult	
1	Validator 1		√		Medium
2	Validator 2		√		
3	Validator 3		√		
3					
1	Validator 1	√			Medium
2	Validator 2		√		
3	Validator 3		√		
4					
1	Validator 1			√	Medium
2	Validator 2		√		
3	Validator 3	√			
5					
1	Validator 1		√		Easy
2	Validator 2	√			
3	Validator 3	√			
6					
1	Validator 1	√			Easy
2	Validator 2	√			
3	Validator 3	√			
7					
1	Validator 1			√	Difficult
2	Validator 2			√	
3	Validator 3			√	
8					
1	Validator 1		√		Medium
2	Validator 2	√			
3	Validator 3		√		
9					
1	Validator 1			√	Difficult
2	Validator 2			√	
3	Validator 3		√		
10					
1	Validator 1		√		Easy
2	Validator 2	√			

No	Question	Difficulty Level			Results
		Easy	Med	Difficult	
3	Validator 3	√			
11					
1	Validator 1		√		
2	Validator 2	√			Easy
3	Validator 3	√			

Eleven IAS measuring items have been classified into easy level questions (questions number 1, 5, 6, 10, and 11), medium level questions (questions number 2, 3, 4, and 8), and difficult level questions (questions number 7 and 9). To fulfill the comparison of the proportions of easy: medium: difficult questions = 1: 2: 1 with a total of five questions, then from the five easy questions one item is taken, namely question number 5, from four easy questions, three questions are taken, namely questions number 3, 4, and 8, and from the two difficult questions, one question was taken, namely question number 7. These five questions were combined into the Chemistry Student IAS Measuring Question Package.

The number or percentage of students who obtained IAS scores of 100, 80, 60, 40, and 20 on the five items are presented in Table 7.

Table 7. Percentage of Students with IAS Scores

Question Class	IAS Student Score Percentage				
	100	80	60	40	20
Easy	60	33	7	-	-
Medium 1	20	56	17	7	-
Medium 2	27	40	27	6	-
Medium 3	30	40	27	3	-
Difficult	7	20	30	30	13

IAS students in solving argumentation questions with low complexity are in the distribution of scores of 60, 80, and 100. The percentage of students with a IAS score of 100 (the highest score) reaches 60%, and only 7% of students get a score of 60. Students' IAS scores on three level questions are having the same pattern. Students' IAS in solving medium-level questions was in the distribution of scores of 40, 60, 80, and 100. There were no students with a IAS score of 20 (lowest score). If averaged, then the percentage of students with IAS scores of 100, 80, 60, and 40 are 26, 45, 24, and 5%, respectively.

The percentage of students with an IAS score of 100 when working on medium level questions is smaller than the same IAS score when students work on easy level argumentation questions. As in the case of working on the easy-level argumentation questions, in the work on the moderate-level argumentation questions, there were no students with a IAS score of 20 (the lowest score).

Students' IAS scores in solving argumentative questions with high complexity are in the distribution of scores of 20, 40, 60, 80, and 100. The percentages of students with IAS scores of 100, 80, 60, 40, and 20 are 7, 20, 30, respectively. 30, and 13%. The percentage of students with a IAS score of 100 is very small, only 7%. A condition of IAS achievement that is reversed when compared to IAS's achievement when students work on easy level argumentation questions.

The percentage of students who get a IAS score of 100 decreases as the level of complexity in assessing claims or the level of complexity of the questions increases. Easy level questions; which in accepting or rejecting student claims, it is easy to find supporting data and does not require complicated explanations, none of the students got a score of 40. IAS with the lowest score only appears when students have to undergo a more complex argumentation process. Students with the lowest IAS scores were found when students completed difficult level questions, reaching 13% of students.

Empirical validity refers to the appropriateness of interpretations made based on test scores related to a particular use [19]. The package of IAS measuring questions that were tested on students consisted of questions with easy, medium, and difficult levels. If the student's IAS score has a trend according to the level of difficulty of the questions, then the package of questions developed is declared to meet the requirements of empirical validity. If you pay attention to the data in Table 7 and the explanatory description of the data, it appears that the students' IAS scores have a parallel trend with the level of difficulty of the questions. Because students' IAS scores have a trend according to the level of difficulty of the questions, the package of questions developed is declared to meet the requirements of empirical validity.

4 Conclusion

Researchers have succeeded in developing a package of chemistry students' IAS measuring questions. The IAS measuring test package consists of five questions with details of one easy level question, three medium level questions, and one difficult level question. The package of questions developed has met the requirements of logical validity (content and construct) and empirical validity.

This article was developed based on research data using Unesa PNPB funding sources for the 2021 fiscal year. The researcher expresses his gratitude to the Unesa Chancellor who has allocated funds for this research. The highest appreciation is conveyed to all participating students in this study.

References

1. Liliarsari. The use of interactive multimedia to enhance students' generic science skills. *International Conference on Rural Information and Communication Technology*, ITB, Bandung, June 2009. (2009)

2. Marttunen, M., Leena, L., Litosseliti, L., & Lund, K.. Argumentation skills as prerequisites for collaborative learning among finish, french, and english secondary school students. *Educational Research and Evaluation*, **11(4)**, 365-384. (2005)
3. Chen, C.H., & She, H. C. The impact of recurrent on-line synchronous scientific argumentation on students' argumentation and conceptual change. *Educational Technology & Society*, **15(1)**, 197-210. (2012)
4. Trilling, Bernie & Fadel, Charles. *21ST Century Skills, Larning for Life in Our Times*. San Fransisco CA: John Wiley & Sons. (2009)
5. Suto, Irenka. *21st Century Skills: Ancient, ubiquitous, enigmatic?* Paper Published in January 2013 in Research Matters: Cambridge Assessment Publication. (2013)
6. Muslim. *Pengembangan program perkuliahan fisika sekolah berorientasi kemampuan berargumentasi calon guru fisika* (Disertasi tidak dipublikasikan). Universitas Pendidikan Indonesia, Bandung. (2014)
7. Sadler, T. D. Informal reasoning regarding socioscientific issues: a critical review of research. *Journal of Research in Science Teaching*, **41(5)**, 513-536. (2006)
8. Suyono; Nasrudin, Harun; & Yonata, Bertha. Consistency and Relevance of Structured Lecture Materials in Physical Chemistry 3 Subjects. Series *Advances in Social Science, Education and Humanities Research*. (2019)
9. Suyono; Nasrudin, Harun; & Yonata, Bertha. *Pengembangan Bahan Kuliah Terstruktur Pada Mata Kuliah Kimia Fisika 3 untuk Memfasilitasi Keterampilan Proses Sains, Berargumentasi, dan Pemecahan Masalah*. Laporan Penelitian. LPPM Unesa. (2019)
10. Suyono & Hariyanto. *Belajar dan Pembelajaran, Teori dan Konsep Dasar*. Bandung: PT Remaja Rosdakarya. (2016)
11. Routman, R. *Invitations changing as teachers and learners*. Toronto: Irwin Publishing. (1991)
12. Masters, G. N., & Forster, M. Developmental assessment. Victoria: ACER. (2008)
13. Stears, M., & Gopal, N. Exploring alternative assessment strategies in science classrooms. *South African Journal of Education*. **Vol 30**, pp: 591-604. (2010)
14. Rowe, J. *Assessmen as Learning- ETEC 510*, Diakses Oktober 2016 dari www.etcet.ubc.ca/510/wiki/Assessment_as_Learning#Assessment_for_Learning. (2012)
15. Earl, K. & Giles, D. An-other Look at Assessment: Assessment in Learning. *New Zealand Journal of Teachers'Work*, **8(1)**, 11-20. (2011)
16. Earl, L. *Assessmen as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA, Corwin Press. (2003)
17. Earl, L. & Katz, S. *Rethinking classroom assessment with purpose in mind. Assessment for learning, assessment as learning, assessment of learning*. Western and Northern Canadian Protocol for Collaboration in Education (WNCP). <http://www.wncp.ca/media/40539/rethink.pdf>. (2006)
18. Arikunto, Suharsimi. *Dasar-dasar Evaluasi Pendidikan Edisi Revisi*, h. 64-79 9 (2005)
19. Gronlund, N. E. *Assessment of student achievement (7th ed.)*. Boston: Allyn and Bacon. (2003)
20. Yusuf, Muri. *Asesmen dan Evaluasi Pendidikan*, (Jakarta: PT Fajar Interpretama Mandiri, 2015), h. 64. (2015)
21. Chin, M., & Osborne, J. Supporting argumentation through student's question in science classroom. *Journal of The Learning Science*, **19(2)**, 230 – 284. (2010)
22. Suyono; Nasrudin, H.; Yonata, B.; & Sabtiawan, W. The Claims Statements from Viral Videos for Instrument Development to Assess Argumentation Thinking Skills. *Journal of Physics: Conference Series*, Volume 1899, 2nd Workshop on Engineering, Education, Applied Sciences and Technology (WEAST) 2020 5 Octpber 2020, Makasar Indonesia. (2020)
23. Febrianto, Yulizar Verda Febrianto; Palupi, Nurheni Sri; & Kusnandar, Feri. Pengujian Mutu Bihun Instan sebagai Produk dalam Program Pemberian Makanan Tambahan (PMT) untuk Ibu Hamil. *Jurnal Mutu Pangan Vol. 4(1)*: 44-50, 2017 ISSN 2355-5017 44 @JMP2017. (2017)
24. Fu, Bin Xiao. Asian noodles: History, classification, raw materials, and processing. *Food Research International* **41** (2008) 888–902. (2008)
25. Fenrich, P. Instructional design tips for virtually teaching practical skills. In *Proceedings of the 2004 Informing Science and IT Education Joint Conference*. Rockhampton, Australia June 25–28, 2004. (2004)
26. Hasan, Saleem, Diola Bagayoko, dan Ella L. Kelly. Misconceptions and The Certainty of Response Index (CRI). *Journal: Physics Education*, **Vol. 34**, No. 5, hal. 294-299, <http://iopscience.iop.org/0031-9120/34/5/304>, diakses tanggal 13 Januari 2011. (1999)
27. Borich, G. D. *Observation skills for effective teaching*. Texas: The University of Texas. (1994)
28. Silverius, Suke. *Evaluasi Hasil Belajar dan Umpan Balik*. Cetakan Ke-1. Jakarta: PT Grasindo. (1991)
29. Dawson, Vaille & Venville, Grady Jane. High-School Students' Informal Reasoning and Argumentation about Biotechnology: An Indicator of Science Literacy? *International Journal of Science Education*. **Vol. 31**, No. 11, 15 July 2009, pp 1421-1445. (2009)