1. Introduction
Evaluation is one of the processes to measure success in learning. Evaluation can be used as feedback for teachers to see the level of success in the learning process. According to [1] evaluation plays an important role in the learning process as it can present the level of student mastery of the material, give an overview of learning difficulties and the position of students among their peers. In the 2013 curriculum, students’ development is evaluated thoroughly, both in the domain of knowledge, attitudes, and skills. The learning process with an emphasis on mastering knowledge and skills begins to be balanced with the inculcation of character as part of the attitude domain. Good character involves understanding, caring and acting according to ethics [2].

Student character relies heavily on the elements of cognition and conation. If the element of cognition is related to thought, then the aspect of conation lies between the elements of cognition and affection. The aspect of cognition can be measured by giving the instrument in the form of a test. While the aspect of conation can be measured using an instrument conation. [3] mentions four aspects of conation that are vital for students and educators to understand, including fact finder (finding facts), follow-through (following the pattern), quick start (initiating) and implementer (implementing). Each type of conation shows the types of actions to be taken by students based on their abilities. Conation is defined as a mental process that activates and directs behaviour and actions. Developing conation, especially aspects of the fact finder, follow thru, quick start and implementer, will be very important for the success of learning in the 21st century [4].

The results of observations at schools in Banda Aceh revealed that teachers encountered difficulty in assessing conations due to the scarcity of instruments developed that can measure the type of
students’ conation comprehensively. Besides, the teachers also had difficulty in character assessment. The character values to be achieved in learning biology are, honest, responsible, caring for the environment, curious, meticulous, careful, diligent, objective, critical and creative, which are usually measured using direct observation techniques through students’ activity and their behaviour. Teachers rarely create assessment instruments to measure the value of characters in students as a whole. Even some teachers assume that the students who have good cognitive values must also have good character values. This is premature judgment since some students with good cognitive scores may not necessarily have good character scores, and vice versa students who have poor cognitive scores, not necessarily have good characters. This is consistent with the research conducted by [5], who argues that the students who have good cognitive skills may not necessarily apply their knowledge in the real world.

Some research related to the development of instruments has been carried out, such as Suhalli research [6] on the development of student and character assessment instruments in mathematics and the results of the development had a reliability value of 0.989 with a very high category. Another study that was conducted by [7] who developed a science process skill-based test for secondary schools also resulted in the reliability of 0.88. However, the research on the type of students’ conation has never been done, so there is no instrument to measure the type of conation in students. Therefore, it is necessary to develop an instrument based on conation that can be used to measure the level of students’ conation and be a reference for teachers in assessing student character.

The next procedure to be done after developing the instrument is to validate that the developed instrument thoroughly evaluates the quality of the assessment. The most important consideration in evaluating the quality of tests as a measurement is validity. [8] defines content validity as validity that is estimated through testing the suitability or relevance of test content through rational analysis by a competent panel or expert judgment. The more scale items reflect the measured area, the higher the content validity. Or in other words, content validity is a function of how well the dimensions and elements of a concept have been described. In line with [8] in [9], the Standards for Education and Psychological Testing defines validity as the extent to which evidence and theory support the interpretation of the results of test scores obtained from soft tests.

This article discusses evidence of validity based on the content of conation-based assessment tests in biology subject for high school students. This article aims to discuss the evidence of content validity. This type of validity can increase reader confidence in the assessment that has been developed. By reading this article, readers can also understand how to collect and analyse evidence of validity based on test content to evaluate the use of tests for specific purposes.

2. Method
In the early stages of developing an instrument, the purpose of content validation is to reduce the variation in the potential for instrument error making and increase the likelihood of obtaining a construct validity index in further studies. 5 Experts were asked to rate the representation of content based on the construction and clarity of the instrument. The assessment instrument evaluated in this study was "conation-based assessment of high school biology learning. The conation instrument consists of 39 question items. Data collection was carried out using a 4 - point scale validation sheet. Each expert was given a conation instrument containing 39 questions that were assessed based on aspects of construct and clarity (Table 1). [10] formulated the formula Aiken V to calculate the content-validity coefficient based on the results of the assessment of an expert panel of n people on an item in terms of the extent to which the item represented the measured construct. The formula proposed by Aiken is \[ V = \frac{\sum s}{[n(c-1)]}. \] The “s” value is obtained from the rating given by expert subtract the integer assigned to the lowest validity category (lo). While “c” is the integer assigned to the highest validity category.
Table 1. Criteria for Measuring Content Validity

| Criteria | Description |
|----------|-------------|
| a. Construction | 1 = poor, 2 = fair, 3 = good, 4 = Very good |
| b. Clarity | 1 = not clear, 2 = item need revision, 3 = clear but need minor revision, 4 = clear |

Criteria for evaluating the content validity of assessments are analysed separately through a rating scale. Content validity index is calculated using the Aiken V formula with the criterion ≥ 0.89. When the Aiken index value is ≥ 0.89 and is statistically significant, there is agreement among experts that the item is relevant to a particular content area. When the Aiken index value is < 0.89 and is statistically significant, there is agreement that the item is not relevant to specific content.

Table 2. Example of Expert rating task assessing item construction

| Item | Goal | How well does the item measure its construction? |
|------|------|-----------------------------------------------|
| 1    | I would ask the teacher or friend if there is material that is not understood | 1 (poor) 2 (fair) 3 (average) 4 (good) |
| 2    | I want to find information through the internet as completely as possible to explore a material | |
| 3    | I want to use concept maps to help me learn | |
| 4    | I want to formulate a problem to prove a concept or theory | |
| 5    | I want to gather various learning resources to solve problems in a learning concept | |

Directions: Please read each item and its associated benchmark. Rate how well the item constructed using the rating scale provided. Check the list rating for each item

3. Results and discussion

By using a rating scale approach, we can get a picture or evidence of the extent to which the contents of the test can measure certain objectives. The evidence is used to establish that the test includes a representative or critical example of the relevant content domain, and the contents do not include content outside that domain. Examples of how the data are summarized according to the criteria are illustrated in Table 3. The Aiken index ranges from zero to one and shows the proportion of experts who judge. Analysis of the relevance of item content with five judges must produce a coefficient of V equal to or above 0.87 to be statistically significant [10]. This value is taken from the right-tell probability table selected values of the validity coefficient (V) provided by Aiken.
Table 3. Example summary of Item Clarity Results

| Item | Goal                                                                 | V  |
|------|----------------------------------------------------------------------|----|
| 7    | I would ask the teacher or friend if there is material that is not    | 0.91* |
|      | understood                                                           |    |
| 12   | I want to find information through the internet as completely as      | 0.78 |
|      | possible to explore a material                                       |    |
| 13   | I want to use concept maps to help me learn                          | 0.94* |
| 19   | I want to formulate a problem to prove a concept or theory           | 0.61 |
| 39   | I want to gather various learning resources to solve problems in a   | 0.94* |
|      | learning concept                                                    |    |
|      | **Average for Item clarity**                                         | 0.86 |

Notes: Statistics based on 5 Expert and rating scale where 1= not clear, 4 = clear.
* p<0.05

Analysis results show that the convention-based instrument assessment in biology subject has a good representation related to the extent to which the item has clarity with the measured domain. The index values on items 12 and 19 have a lower index than other items (<0.87) so that the clarity of the item is invalid, while items no. 7 and 39, which have an index of content validity equal to or greater than (>0.87) then it is considered to have adequate validity. The overall results of items that have been assessed by experts using the Aiken V index will be displayed in table 4.

Table 4. Aiken’s V values of conation instrument

| Item | (V)  | Item  | (V)  |
|------|------|-------|------|
| 1    | 0.82 | 21    | 0.89*|
| 2    | 0.72 | 22    | 1.00*|
| 3    | 0.89*| 23    | 0.78 |
| 4    | 0.89*| 24    | 0.61 |
| 5    | 0.89*| 25    | 0.78 |
| 6    | 0.89*| 26    | 0.78 |
| 7    | 0.94*| 27    | 0.89*|
| 8    | 0.83 | 28    | 0.85 |
| 9    | 0.76 | 29    | 0.96*|
| 10   | 0.94*| 30    | 0.96*|
| 11   | 0.79 | 31    | 0.93*|
| 12   | 0.78 | 32    | 0.89*|
| 13   | 0.94*| 33    | 0.89*|
| 14   | 1.00*| 34    | 0.94*|
| 15   | 0.94*| 35    | 0.61 |
| 16   | 1.00*| 36    | 0.83 |
| 17   | 1.00*| 37    | 0.89*|
| 18   | 1.00*| 38    | 0.94*|
| 19   | 0.61 | 39    | 0.94*|
| 20   | 0.94*|      |      |

Note: the critical value of V according to Aiken’s (1985) table of critical values is 0.87.

The conation instrument that has been developed is a content validity testing process which was discussed with three experts in education and two experts in psychology. Validation is intended so that the instruments developed can genuinely measure the conation of students. The results of the
validation of the conation instrument assessed by experts using a 4-point scale show that of the 39 instrument items, 26 were stated to meet good validity criteria and 12 items did not meet the adequate content validity criteria. The index value of each item is analysed using the Aiken’s V. The formula Aiken’s V value ranges from 0 - 1. When the Aiken index value is ≥ 0.87, there is agreement among experts that the item is relevant to a particular content area. If the Aiken index value is ≤ 0.87, experts agree that the item is irrelevant to the particular content, so the item is not valid because it does not meet content validity criteria. The agreement value of 0.87 is said to meet the valid criteria taken from the right-tell probability table selected values of the validity coefficient (V) provided by Aiken by looking at the number of raters and rating scale. The greater the number of raters used, the smaller the specified criteria index and vice versa, the fewer number of rater used, the greater the validity criteria index specified. But according to Azwar [11], the coefficient between 0.64 to 1 can be considered to have adequate content validity.

4. Conclusion
Content validity needs to be done to determine the feasibility or relevance of the test content through rational analysis by a competent panel or through expert judgment that can help ensure the validity of the concept to give readers and researchers confidence about the appropriateness of the instrument both in terms of concepts and operations. The validity described in this paper focuses on issues related to evaluating content validity. Evaluating content validity requires a reputable expert to check whether the test item is assessing the specified content [12]. Content domain representation is very important to show the validity of conclusions derived from test scores, and the more detailed the assessment is done, the more it is believed that the measuring instrument is more precise [13].

Overall, of the 39 item tests developed, 22 instruments have adequate valid criteria. Thus, those instruments can be used to measure types of conation of high school students. The result of this study can benefit the students and teachers to understand their learning competence and learning style that is important in the teaching-learning process. Further research is necessary to develop the instrument than can optimally measure students’ conation types.

5. Reference
[1] Stemen K 2010 Pengembangan Evaluasi Pembelajaran Online (Jurnal Pendidikan dan pengajaran 43 20).
[2] Mardapi D 2008 Teknik Penyusunan Instrumen Tes dan Non-tes (Mitra Cendikia : Jakarta).
[3] Kolbe K 1990 The conative connection (MA: Addison-Wesley Publishing Company, Inc)
[4] Huit W & Cain S 2005 Conation As An Important Factor of Mind. Journal Educational Physiology Valdosa, G.A: Valdosa State University 1
[5] Darmawan E 2013 Pengaruh PBL Terhadap Sikap dan Hasil Belajar Jurnal Lentera Sains 3 4.
[6] Shahali E H and Lilia H 2010 Development and Validations of a Test of Integrated Science Process Skills (Procedia Sosial and Behavior Science 9 146).
[7] Oeng Ek Tek 2011 The Development and Validations of An All Encompassing Malaysian-Based Science Process Skill Test for Secondary Schools (Journals of Science and Mathematics Education 34 236
[8] Sekaran U 2006 Metode Penelitian Bisnis (Jakarta: Salemba empat)
[9] Pada A U T 2015 Content Validity of Creative Thinking Skills Assessment Proceeding of International Conference On Research, Implementation And Education Of Mathematics And Sciences 10
[10] Aiken L R 1985 Three Coefficients for Analyzing The Reliability And Validity of Rating (Educational and Psychological Measurement 45 134
[11] Azwar S 2012 Realibilitas dan Validitas Edisi 4 (Yogyakarta: Pustaka Pelajar)
[12] Lawshe C H 1975 A Quantitative Approach to Content Validity (Personnel Psychology 28, 575).

[13] Murphy K R and Davidshofer C O 1991 Psychological Testing: Principles and Applications (New Jersey: Prentice-Hall)