Characteristics of CPS-Based Assessment Instrument for Critical Thinking Ability in Stoichiometry

Agustini, Usman Rery, Lenny Anwar
Chemistry Education, Graduate Study Program – FKIP, Universitas Riau
Jl. HR. Soebrantas, Km. 12.5, Pekanbaru, 28293, Indonesia
agustinihusniwan@gmail.com

Abstract. The purpose of this study was to determine the characteristics (validity, reliability, difficulty indices and discrimination indices) of the items in the developed assessment instruments. Subjects in the study were 30 students of class X Science 1 SMA Negeri 2 Bantan. The samples in this study were chosen by purposive sampling technique, in which the samples were selected based on the researchers’ consideration. Then, the approach used to develop the assessment instrument was Research and Development (R&D) adapting Borg & Gall Model. In this study, the feasibility test was carried out to see the construct validity, reliability, difficulty indices, and discrimination indices of the instrument. The CPS-based assessment instrument for critical thinking ability on Stoichiometry was developed in the form of essay questions. This instrument produced 10 items that were declared valid and reliable with a reliability value of 0.749. Thus, this study concluded that CPS-based assessment instrument for critical thinking ability provided more comprehensive information about the characteristics of the instruments that have been developed.

1. Introduction
Assessing the learning outcomes conducted by educators on an on-going basis aims to monitor the process and progress of students’ learning as well as to improve the effectiveness of learning activities [1]. Assessing the learning process and outcomes is an inseparable part of the planning and implementation of learning undertaken by the teachers. Assessment is the last part of learning that aims to determine the achievement of the goals set in the curriculum and to make decisions on all students for the next learning phase [2].

Assessment instruments are an integral part of an assessment process in learning. Assessment instruments include tests and grading systems. These instruments are designed to determine the level of students’ understanding after learning a competency [3]. Achieving the actual goals of chemistry learning requires the use of assessment instruments that not only cover memorization and understanding, but also require assessments that train students’ thinking skills [4]. A well-designed assessment instruments which are developed based on the level of thinking ability can improve students' thinking abilities, especially critical thinking [5].

Thinking critically is skill activity that can be done well or otherwise, and a good critical thinking will fulfill various intellectual standards, like clarity, relevance, adequacy, coherence etc [6]. Critical thinking skills is one of the thinking skills that defined by [7] as self-guided, self-disciplined thinking which attempts to reason at the highest level of quality in a fair-minded way. Ennis in [8] gave a definition-critical thinking is thinking by argument and reflection by pressing the decision making about what to believe and to do. The purpose of critical thinking is to make people away from wrong and rushed decision so it can be guaranteed. According to Costa in [9] in thinking learning there are 3
ways, they are teaching for thinking, teaching of thinking and teaching about thinking. The same thing was stated by [10] in Foundation of Critical Thinking, critical thinking skills is an intellectual process about conceptualization, implementation, analysis, synthesis and evaluation actively and expert towards information that gained from observation, experience, reflection, thinking or communication as a guidance to believe and to do. This skills is marked by intellectual values that is universal, which are clarity, accuracy, consistency, carefulness, suitability, right evidence, a good thinking, deepness, width, and justice. Critical thinking can not be separated from education and is an important cognitive ability so schools try to improve it [11].

The results of interviews with several chemistry teachers at SMA Negeri in Bengkalis revealed that teachers never made an assessment instrument to measure students’ critical thinking skills. The assessment instrument was used to measure the achievement of learning indicators without considering the ability to think critically; unfortunately, the assessment instrument used by teachers was still measuring aspects of memorization and understanding. The teachers adopted the questions from various sources. Stoichiometry is the basic material that must be understood by students in learning chemistry. For this reason, students must have good analytical and mathematical skills in order to solve the calculation problems correctly. In stoichiometry, there is a material about the concept of mole which is a basic material or concept in chemical calculations so that it becomes one of the essential chemical materials in general [12]. The development of learning achievement indicators for Stoichiometry should be adjusted to the indicators of critical thinking skills and CPS in order that a valid assessment instrument can be produced and it can measure students’ critical thinking skills.

A characteristic refers to a distinctive feature of someone or something. An instrument will be reliable if it gives a relatively stable and consistent score. The higher the accuracy and the precision of the assessment results, the lower the level of error in making the assessment. Then, the lower the error, the more consistent of the results of an assessment will be. The best way to assess good characteristics is by checking the validity and the most important validity of teacher-made tests is curricular validity (content validity). After that, the reliability of the instrument should also be checked. One of the indicators for tests that have high reliability is that most of the items also have high discrimination indices [13]. However, each study has its own characteristics. For example, research conducted by Amalia revealed that the validity of the research instrument which was in the form of essay analysis was good and valid [14]. Thus, it can be stated that the reliability of the essay test was in the category of good. Moreover, there are typical processes from the development of instruments to measure critical thinking skills in the material about chemical bonding using Rasch Model, which were in defining constructs, types of statistical analysis, and Wright's Map analysis. The developed instrument had moderate reliability and it consisted of 12 valid items. The analysis using Rasch Model provided more comprehensive information on the characteristics of the instruments that had been developed [15].

This study focused on the characteristics of the assessment instrument based on the items developed by the researchers. By developing CPS-based assessment instrument, which had good characteristics, for critical thinking ability, it was expected that the instrument would give a positive influence on students’ critical thinking skills on Stoichiometry.

2. Methodology
Based on the purpose, this study was classified as a qualitative descriptive study. Subjects of the study were 30 students of class X Science 1 of SMA Negeri 2 Bantan. The samples were chosen through a purposive sampling technique, in which the samples were selected based on the researchers’ consideration. Then, the approach used to develop the assessment instrument was Research and Development (R&D) adapting Borg & Gall Model. In this study, the feasibility test was carried out to see the construct validity, reliability, difficulty indices, and discrimination indices of the instrument.

Construct validity in this study refers to the ability of the instrument to measure the material to be measured; it presented the degree of accuracy between the actual data and the data collected by the researchers, which were carried out by using Pearson Correlation. Meanwhile, reliability meant measuring the extent to which the measuring instrument used for respondents’ answers in the study was able to reach the answer to the research problem objectively by using the method of calculating
the reliability index by the repeat method that was through Cronbach Alpha correlation. Then, difficulty indices were measured based on the students’ ability to answer a question. Meanwhile, discrimination indices were conducted to find out the quality of the items between the upper group who answered correctly and the lower group who answered correctly.

3. Results and Discussion
After the questions validated by the material experts, to see the feasibility of the instrument, it was tested to 30 students of class XI IPA 1 SMA Negeri 2 Bantan, Kab. Bengkalis on the academic year 2019/2020. The feasibility test was performed by the students in 90 minutes. They were asked to work on items about CPS-based critical thinking skills and the results of students' response were used to determine the characteristics of the questions namely construct validity, reliability, difficulty indices, and discrimination indices. All of these characteristics were analysed by using SPSS 24.0 program.

3.1 Construct validity
The construct validity testing was performed by using SPSS software for Windows version 24.0. The following are the results of the validity testing in this study, which can be seen in table 1 below:

| No. | R Calculated | Conclusion |
|-----|--------------|------------|
| 1   | 0.521        | Valid      |
| 2   | 0.683        | Valid      |
| 3   | 0.507        | Valid      |
| 4   | 0.428        | Valid      |
| 5   | 0.402        | Valid      |
| 6   | 0.251        | Invalid    |
| 7   | 0.314        | Invalid    |
| 8   | 0.447        | Valid      |
| 9   | 0.613        | Valid      |
| 10  | 0.405        | Valid      |
| 11  | 0.241        | Invalid    |
| 12  | 0.404        | Valid      |
| 13  | 0.188        | Invalid    |
| 14  | 0.442        | Valid      |
| 15  | 0.193        | Invalid    |

Table 1 shows that 10 of 15 items contained in the instrument were valid while the other 5 items were invalid. The invalid items were omitted since they could not be used in this study. It can be seen that the value of Pearson correlation in the table of each item > 0.361 if n = 30; therefore, there were 10 questions used as the data in this study. The invalid items were questions number 6, 7, 11, 13, and 15. Those 5 items were invalid due to insufficient time to do the questions, the level of difficulty of the questions which was quite difficult compared to other questions, less detailed answers given by students. For example, the aspect of the critical thinking ability measured in question no. 6 was analyzing the arguments and the aspect of CPS characteristics was finding facts meaning that students were expected to be able to analyze the arguments contained in the problem so that they could prove the facts in accordance with the answers given in solving problems presented on the questions. Data of this study found that the answers given by students reflected that the students were not able to think critically in solving the problems. This was observed from the facts presented by the students.

3.2 Reliability Testing
Reliability testing was performed by Cronbach Alpha correlation using SPSS for Windows version 24.0. The results of the reliability testing showed that the tested variable obtained a Cronbach alpha coefficient of 0.749, which was greater than 0.6. Thus, it could be concluded that the instrument was
reliable and feasible to use. The value of 0.6 was derived from the theory of reliability decision making according to Rochaety that the minimum requirement of a correlation coefficient was 0.6 because it was considered to have a safe point in determining instrument reliability and was also generally widely used in research while a value of 0.5 was considered reliable but in a poor level [16].

3.3 Difficulty Indices and Discrimination Indices

In this study, the items analysis of the instruments for assessing analytical thinking abilities and chemical literacy included difficulty indices and discrimination indices [17]. The analysis of difficulty indices were conducted to 10 valid items as displayed in Table 2.

### Table 2. Results of Difficulty Indices and Discrimination Indices of the Valid Items

| No | Items Number | Difficulty Indices | Discrimination Indices | Categories          |
|----|--------------|--------------------|------------------------|--------------------|
| 1  | 1            | 0.300              | 0.400                  | Difficult, Satisfactory |
| 2  | 2            | 0.292              | 0.583                  | Difficult, Good     |
| 3  | 3            | 0.742              | 0.283                  | Easy, Satisfactory  |
| 4  | 4            | 0.633              | 0.500                  | Moderate, Good      |
| 5  | 5            | 0.633              | 0.333                  | Moderate, Satisfactory |
| 6  | 8            | 0.650              | 0.300                  | Moderate, Satisfactory |
| 7  | 9            | 0.650              | 0.433                  | Moderate, Good      |
| 8  | 10           | 0.742              | 0.217                  | Easy, Satisfactory  |
| 9  | 12           | 0.758              | 0.217                  | Easy, Satisfactory  |
| 10 | 14           | 0.650              | 0.300                  | Moderate, Satisfactory |

Table 2 shows that the difficulty index of the items varies from the category of easy, moderate and difficult. There were 3 items categorized as easy items; 5 items were categorized as moderate items; and 2 items were categorized as difficult items. Syamsudin [18] concludes that a good test is one containing questions that are not too easy and not too difficult. The item which has too high or too low difficulty index has academically inappropriate function. This is because questions that are too easy will decrease the students’ interest in learning. Similarly, questions that are too difficult will discourage students in learning and they even can be frustrated. Table 3 shows the difficulty indices of the 10 questions so the results of the difficulty indices are displayed in table below:

### Table 3. Results of Difficulty Indices of the Items

| Criteria of Difficulty Indices | Difficulty Indices Index | Items Number | Number of Questions | %   |
|-------------------------------|--------------------------|--------------|---------------------|-----|
| Easy                          | 0.71 - 1.00              | 3, 8, 9      | 3                   | 30% |
| Moderate                      | 0.31 - 0.70              | 4, 5, 6, 7, 10 | 5                  | 50% |
| Difficult                     | 0 - 0.30                 | 1, 2         | 2                   | 20% |

Table 3 shows the results of difficulty indices that 30% items were categorized as easy items, 50% items were categorized as moderate items, and 20% items were categorized as difficult items. The data shows that 3 items were difficult, 5 items were medium, and 2 items were easy. Sudjana [19] asserts that difficulty indices means analyzing the difficulty of test items so that the questions can be classified as easy, moderate, and difficult. The assumption to obtain good quality questions, in addition to fulfilling validity and reliability, is a balance in the difficulty indices of the items. This means the proportional amount of the questions that are easy, moderate, and difficult. A comparison for easy – medium – difficult items can be 2 – 6 – 2 which means that 20% are easy items, 60% are moderate items, and 20% are difficult items. Another comparison that has similar proportions as the above is 3 – 4 – 3, meaning that 30% are easy, 40% are moderate, and 30% are difficult [20].
Based on the results of difficulty indices, it was obtained that there were 20% easy items, 60% moderate items, and 20% difficult items. This finding is approaching the theory that the proportional ratio for easy-medium-difficult items can be 2 – 6 – 2 which means that 20% of the items are easy, 60% are moderate, and 20% are difficult. Meanwhile, the proportion of 3 – 5 – 2 (30% of easy items, 50% of moderate items, and 20% of difficult items) is almost approaching, which means that the questions contained in the developed instrument have quite good difficulty indices.

Table 2 shows the difference in the discrimination index of 10 items that were tested for construct validity. Table 4 presents the discrimination indices of the questions.

**Table 4. Results of Difficulty Indices of the Items**

| Criteria of Discrimination Indices | Discrimination Index | Items Number | Number of Questions | % |
|------------------------------------|----------------------|--------------|---------------------|---|
| Poor                               | 0.00 ≤ Dp ≤ 0.20     | -            | -                   | - |
| Satisfactory                       | 0.20 ≤ Dp ≤ 0.40     | 1, 3, 5, 6, 8, 9, 10 | 7 | 70 |
| Good                               | 0.40 ≤ Dp ≤ 0.70     | 2, 4, 7      | 3                   | 30 |
| Excellent                          | 0.70 ≤ Dp ≤ 1.00     | -            | -                   | - |

Table 4 demonstrates that for the discrimination indices, 70% of the items were classified as satisfactory and 30% of the items were classified as good. Wulandari, et al., [21] concluded that the smaller the discrimination index, the lower the quality of the questions, and conversely, the higher the discrimination index, the better the quality of the questions. The requirements for discrimination indices are a) the question is accepted if it has high quality or the discrimination index > 0.20; b) the question is accepted for moderate quality if it has discrimination index 0 ≤ Dp ≤ 0.20, c) the question is rejected if it has low quality or the discrimination indices < 0 or (-). Moreover, the questions are accepted if the number students who can answer the questions correctly from the upper group are higher than the number of students who can answer correctly from the weak group. The results of data analysis revealed that the items had discrimination indices in the satisfactory category with the index 0 ≤ Dp ≤ 0.20 assuming that the questions were in the difficult category and were being answered by intelligent students.

The characteristics results of the instrument for assessing critical thinking skills on stoichiometry can be seen in table 5.

**Table 5. Characteristics Results of CPS-Based Assessment Instrument for Critical Thinking Skills on Stoichiometry**

| Item Number | Validity | Reliability | Difficulty Indices | Discrimination Indices | Categories |
|-------------|----------|-------------|--------------------|------------------------|------------|
| 1           | 0.521    | 0.701       | 0.300              | 0.400                  | Qualified  |
| 2           | 0.683    | 0.697       | 0.292              | 0.583                  | Qualified  |
| 3           | 0.507    | 0.738       | 0.742              | 0.283                  | Qualified  |
| 4           | 0.428    | 0.729       | 0.633              | 0.500                  | Qualified  |
| 5           | 0.402    | 0.736       | 0.633              | 0.333                  | Qualified  |
| 6           | 0.251    | -           | 0.508              | 0.217                  | Not Qualified |
| 7           | 0.314    | -           | 0.500              | 0.267                  | Not Qualified |
| 8           | 0.447    | 0.744       | 0.650              | 0.300                  | Qualified  |
| 9           | 0.613    | 0.715       | 0.650              | 0.433                  | Qualified  |
| 10          | 0.405    | 0.731       | 0.742              | 0.217                  | Qualified  |
| 11          | 0.241    | -           | 0.417              | 0.133                  | Not Qualified |
| 12          | 0.404    | 0.738       | 0.758              | 0.217                  | Qualified  |
Table 5 demonstrates that questions number 6, 7, 11, 13, and 15 were rejected because they did not fulfill the requirement of validity. It has been stated that the questions is rejected if one of the characteristics is not fulfilled. The characteristics of the accepted questions include:

- The results of the analysis obtained valid and reliable questions, difficulty indices categorized as easy but the discrimination indices categorized as good, satisfactory, and bad.
- The results of the analysis obtained valid and reliable questions, difficulty indices categorized as moderate but the discrimination indices categorized as good or satisfactory.
- The results of the characteristics obtained valid and reliable questions, difficulty indices categorized as difficult but the discrimination indices varied from satisfactory or bad.

4. Conclusion

Based on the results explained above, it can be concluded that this study produced a CPS-based assessment instrument for critical thinking ability on Stoichiometry in the form of essay questions. The developed assessment instrument created 10 valid and reliable items with a reliability value of 0.749 with the very high category. Then, the difficulty indices that 30% items were categorized as easy items, 50% items were categorized as moderate items, and 20% items were categorized as difficult items. The discrimination indices, 70% of the items were classified as satisfactory and 30% of the items were classified as good. The CPS-based assessment instrument for critical thinking ability provided more comprehensive information about the characteristics of the instruments that had been developed and were feasible to be implemented.

References

[1] Saptorini 2012 Strategi Pembelajaran Kimia *Kimia FMIPA UNNES*, Semarang
[2] Susetyo B 2015 Prosedur Penyusunan dan Analisis Tes untuk Penilaian Hasil Belajar Bidang Kognitif *Refika Aditama*, Bandung
[3] Prasasti S and Basuki 2012 Pengembangan Instrumen Asesmen Berpikir Kritis melalui Membaca untuk Siswa SD/MI UM, Malang *Jurnal Universitas Negeri Malang* 48(2) 1-12
[4] Lissa 2012 Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi Materi Sistem Respirasi Dan Ekskresi Universitas Negeri Semarang, Semarang *Jurnal Lembaran Ilmu Kependidikan* 41(1) 27-32
[5] Amalia N F 2014 Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Siswa SMA pada Materi Asam Basa, Universitas Negeri Semarang, Semarang *Jurnal Inovasi Pendidikan Kimia* 8(2) 1380-1389
[6] Fisher A 2009 Berpikir Kritis-Sebuah Pengantar Erlangga, Jakarta
[7] Elder L and Paul R 2008 Critical Thinking Development: A Stage Theory with Implications for Instruction
[8] Hassoubah I Z 2004 Developing Creativ and Critical Thinking Skill (Cara Berpikir Kreatif dan Kritis) Nuansa, Bandung
[9] Liliasari 2005 Membangun keterampilan berpikir manusia Indonesia melalui pendidikan sains, Pidato pengukuhan Jabatan Guru Besar Tetap dalam Ilmu Pendidikan IPA pada FPMIPA UPI, Bandung
[10] Scriven M and Paul R 1987 Critical Thinking-in The 8th Annual International Conference on Critical Thinking and Education Reform, C.A 7(9)
[11] Zhou Q, Huang Q and Tian H 2013 Developing Students’ Critical Thinking Skill by Task-Based Learning in Chemistry Experiment Teaching Creative Education 4(12) 40–45
[12] Rohmah S 2011 Pengaruh Problem Solving dalam Meningkatkan Hasil Belajar Kimia Siswa terhadap Konsep Mol dalam Stoikiometri *UIN Syarif Hidayatullah* Jakarta
[13] Arikunto S 2013 Prosedur Penelitian: Suatu Pendekatan Praktik *PT. Rineka Cipta* Jakarta
[14] Amalia N F 2014 Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Siswa SMA pada Materi Asam Basa, Universitas Negeri Semarang, Semarang Jurnal Inovasi Pendidikan Kimia 8(2) 1380-1389

[15] Sabekti, Widhia A and Khoirunnisa F 2018 Using Rasch Measurement to Develop an Instrument to Assess Students' Critical Thinking Ability in Chemical Bonding Topic Universitas Maritim Raja Ali Haji, Riau Jurnal Zarah 6(2) 68–75

[16] Roichaety E, Tresnati R and Latief A M 2012 Metodologi Penelitian Bisnis dengan Aplikasi SPSS Edisi I Mitra Wacana Media Jakarta

[17] Ani R, Lina K, Sunyoto E N and Arif W 2016 Developing an Instrument of Scientific Literacy Assessment on the Cycle Theme International Journal Of Environmental & Science Education 11(12) 5718-5727

[18] Syamsudin 2012 Pengukuran Daya Pembeda, Taraf Kesukaran dan Pola Jawaban Tes (Analisis Butir Soal) STIT Muhammadiyah Pacitan, Pacitan Jurnal At-Tajdid 1(2) 187-198

[19] Sudjana N 2009 Penilaian Hasil Proses Belajar Mengajar Remaja Rosdakarya Bandung

[20] Monika N R 2019 Pengembangan Instrumen Asesmen Kemampuan Berpikir Kreatif Peserta Didik pada Materi Minyak Bumi dengan Model pembelajaran CPS, Thesis yang tidak diterbitkan, University of Riau, Pekanbaru

[21] Wulandari, Annisaul K, Jimmi C and Rasmiwetti 2019 The Development Of Assessment Instrument Using Teslet Models In Stereochemistry Materials International Journal Of Educational Best Practices (Ijebp) 3(1) April 2019 Issn: 2581-0847 Doi: 10.31258/Ijebp.V3n1.P29-42, Universitas Riau, Riau