Evaluation of chemical literacy assessment instruments in solution materials

H K Muchtar*, Nahadi, and Hernani

Departemen Pendidikan Kimia, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No. 229, Bandung 40154, Indonesia

*Corresponding author email: anifahkurniam@gmail.com

Abstract. The purpose of this study was to produce information about assessment instruments quality to assess chemical literacy of high school students in solution material developed by researcher. Research method used was descriptive with quality parameters tested were empirical validity, reliability, readability, distinguishing power, difficulty level and distractors function. Research participants were 26 high school students class XI in Bandung. Chemical literacy assessment instrument that were tested consisted of 49 multiple choice items, 12 essay items and 26 attitude scale statements. Study results showed that empirical validity values of multiple choice items, essays and attitude scale respectively in range 0.11–0.87, 0.4–0.8, 0.1–0.74. Reliability values for multiple choice items, essays and attitude scales were 0.93, 0.87, and 0.94, respectively. Readability test scores for multiple choice items, essays and attitude scales were respectively 97%, 99%, 100%. The results of distinguishing power test for multiple choice and essays items were in range of 0.07–0.85 and 0.3–0.8. Difficulty level test results for multiple choice items were 16% difficult category, 61% moderate category, and 23% easy category while for essay items were 9% difficult category, 75% moderate category, and 16% easy category. Distractor function test results for multiple choice items were 37 items which the distractor was still required to be improved.

1. Introduction

Today the world is in the 21st century, which results an expansion of knowledge that impacts in life. This has resulted humans must have the understanding of science and technology in living their lives so that they are not outdated [1]. The impact of science for society can be seen in various aspect in life, such as social, political, educational, technological and economic [2]. But the development of science and technology raises issues that threaten the survival of organism in the world such as global warming, the use of harmful additives into food, water pollution and air pollution. To improve this, it is necessary to prepare a community that has an understanding of science and technology that is more environmentally responsible. This can be done through science learning that can deliver students have 21st century skills. This is appropriate with the 2013 curriculum currently applied in Indonesia. In the 2013 curriculum, students must have the skills needed in the 21st century, so that they can form human resources that have competitiveness, competence and skill. If the 2013 curriculum is applied in science learning, it can train students to use scientific knowledge and abilities to solve problems in the world [3]. This requires the need to have scientific literacy [4].

Scientific literacy is currently the main goal in science learning even in science education [4,5,6]. Scientific literacy related with the ability to use conceptual knowledge of science and the ability to
distinguish between scientific data and data from other scientific disciplines [7]. Scientific literacy is the gateway to achieving scientific and technological progress and economic survival that can be achieved through science education [2].

Chemistry is part of science and one of the important branches of science. Chemistry topics generally study matter and understand material properties that are important in many scientific disciplines such as health sciences, geography, physics, environmental sciences and economics [4,8]. Understanding chemistry is very important, because nature is greatly influenced by chemistry and is filled with chemical products [9]. Therefore, the goal of learning chemistry must consider the problems that exist in life, so students can use conceptual understanding of chemistry to solve problems that exist in life [10]. This ability is called chemical literacy. Thus, current chemistry learning must aim to encourage the development of student chemical literacy effectively [11,12].

Several studies about chemical literacy have been done, that are the efforts to improve students' chemical literacy through learning [5,13], identification chemical literacy of teacher candidate [6, 14, 15], the effect of argumentation on student chemical literacy [16], identification chemical literacy profiles of student [17], identification of student chemical literacy [18,19], identification the meaning of the term chemical literacy [11,20,21] and the development chemical literacy assessment of student [4,22]. In addition, there are also studies that not only develop chemical literacy assessment instruments, but also assessment instruments for chemical literacy and generic science skills [23].

To identify chemical literacy, a chemical literacy assessment instrument is needed. The importance of chemical literacy assessment instruments is based on the fact that the achievement of chemical learning requires assessment instruments that not only assess understanding and memorization, but also assess students' ability to apply concepts that have been learned when they have problems [23]. Today it is difficult to find suitable instruments to assess students' chemical literacy [4]. Research on chemical literacy is very important so that chemical learning can effectively improve students' chemical literacy. Based on this, researchers are interested to develop assessment instruments for chemical literacy because these instruments are especially needed for high school students. The instruments which developed need to be analyzed the quality with certain parameters.

2. Method

The research method used in this study was descriptive with quality parameters tested were empirical validity, reliability, readability, distinguishing power, difficulty level and distractors function. Participants in this study were 26 high school students of class XI in one of the high schools in the city of Bandung who had studied solution material which included the concept of acid base, buffer solution, hydrolysis of water by salt, solubility and solubility product constants.

The chemical literacy assessment instrument which were analyzed was developed by the researchers themselves. The instruments developed consisted of 55 multiple choice items to assess knowledge and understanding of chemical content, 12 essay items to assess knowledge and understanding of chemical relations with technology and social, analytical thinking applications, application of reasoning and 28 statements for assess aspects of attitude. The items and attitude scale which had developed then their contents validation tested with the CVR (Content Validity Ratio) method. Items and attitude scale statements that were declared valid then corrected (if needed) based on the advice given by the validator.

Based on the results of the content validity test there were 6 multiple choice items with CVR value 0.6, while the other were 1. For a number of validators of five, the minimum CVR value for each item was 0.99 [24]. Thus, the item was declared valid or was appropriate with criteria for content validity if the CVR value > 0.99 and the item was declared invalid or was not appropriate with criteria for content validity if the CVR value < 0.99. For essay items, there were 12 valid items which have CVR value 1 Thus, there were 49 valid multiple choice items and 12 valid essays items. The results of the content validity test for the attitude scale, there were 28 statements which have CVR value 1. In addition, the CVI value for multiple choice item was 0.96, for essay item was 1 and for attitude scale was 1.
3. Result and Discussion
The researcher analyzed the quality of the chemical literacy assessment instruments of high school students in solution material.

3.1 Empirical Validity, Reliability and Readability Test Results
Chemical literacy assessment instruments were tested for empirical validity by correlation statistics (Table 1 and 2). The validity of multiple choice items were calculated by the biserial point correlation formula, the validity of the essay items and attitude scale were calculated by the product moment correlation formula. An instrument could be said to have empirical validity if it had been tested into the school then calculated using the correlation formula. In general, if the correlation value was > 0.3, then the item was valid [25]. Based on the results of the empirical validity test there were 6 invalid multiple choice items, there were items number 11, 19, 20, 22, 29 and 43, while for essays all items were valid (Table 1). The results of the empirical validity test for the attitude scale were 2 invalid statements with a correlation value were 0.27 (statement 2) and 0.1 (statement 25) (Table 2). A valid instrument, it could be said that the instrument can assess students' chemical literacy. Thus, there were 43 items multiple choice items, 12 essays items and 26 statements attitude scale which were valid.

| No Item | Correlation Value | No Item | Correlation Value | No Item | Correlation Value | No Item | Correlation Value |
|---------|-------------------|---------|-------------------|---------|-------------------|---------|-------------------|
| 1       | 0.87              | 13      | 0.42              | 25      | 0.31              | 37      | 0.33              | 49      | 0.61              |
| 2       | 0.67              | 14      | 0.45              | 26      | 0.59              | 38      | 0.47              | 1       | 0.7               |
| 3       | 0.81              | 15      | 0.45              | 27      | 0.34              | 39      | 0.45              | 2       | 0.7               |
| 4       | 0.79              | 16      | 0.49              | 28      | 0.57              | 40      | 0.64              | 3       | 0.7               |
| 5       | 0.82              | 17      | 0.46              | 29      | 0.18              | 41      | 0.30              | 4       | 0.5               |
| 6       | 0.62              | 18      | 0.46              | 30      | 0.33              | 42      | 0.54              | 5       | 0.6               |
| 7       | 0.74              | 19      | 0.19              | 31      | 0.31              | 43      | 0.11              | 6       | 0.4               |
| 8       | 0.41              | 20      | 0.13              | 32      | 0.52              | 44      | 0.52              | 7       | 0.8               |
| 9       | 0.74              | 21      | 0.41              | 33      | 0.68              | 45      | 0.43              | 8       | 0.7               |
| 10      | 0.75              | 22      | 0.20              | 34      | 0.45              | 46      | 0.31              | 9       | 0.6               |
| 11      | 0.17              | 23      | 0.67              | 35      | 0.48              | 47      | 0.37              | 10      | 0.7               |
| 12      | 0.56              | 24      | 0.64              | 36      | 0.41              | 48      | 0.43              | 11      | 0.7               |
|         |                   |         |                   |         |                   |         |                   | 12      | 0.7               |

Table 2. Empirical Validity Test Results of Attitude Scale

| No statement | Correlation Value | No statement | Correlation Value | No statement | Correlation Value | No statement | Correlation Value |
|--------------|-------------------|--------------|-------------------|--------------|-------------------|--------------|-------------------|
| 1            | 0.58              | 8            | 0.55              | 15           | 0.70              | 22           | 0.72              |
| 2            | 0.27              | 9            | 0.68              | 16           | 0.64              | 23           | 0.69              |
| 3            | 0.59              | 10           | 0.54              | 17           | 0.69              | 24           | 0.44              |
| 4            | 0.53              | 11           | 0.67              | 18           | 0.65              | 25           | 0.10              |
| 5            | 0.50              | 12           | 0.60              | 19           | 0.67              | 26           | 0.70              |
| 6            | 0.59              | 13           | 0.72              | 20           | 0.74              | 27           | 0.66              |
| 7            | 0.65              | 14           | 0.68              | 21           | 0.68              | 28           | 0.71              |

Based on the results of the multiple choice reliability test using the KR-20 formula, the correlation value was 0.93 and that was included in the high category [26], while the essay uses the Cronbach's Alpha formula, the Alpha value was 0.87 and belongs to the high category. Furthermore, for the attitude scale using the Cronbach's Alpha formula, the Alpha value was 0.94 and was included a very
high category. A high reliability value means that the level of consistency of this test was high when repeatedly tested at different times with the same sample, so that this test could be trusted to use.

Based on the results of multiple choice, essays and attitude scale readability test respectively were 97%, 99%, 100%. This shows that the chemistry literacy assessment instrument developed almost all students could understand the assessment instruments which is given to them.

3.2 Distinguishing Power, Difficulty Level and Distractor Function Test Results

Difficulty level test was conducted to classify items that were developed including difficult, moderate or easy categories, while the distinguishing power test was conducted to find out the ability of item to distinguish between high-ability students and low-ability students (Table 3). Based on the results of the distinguishing power test there were 20 multiple choice items with very good categories, 22 items with good categories and 1 item with enough categories. One item with a enough category could be corrected to be able to distinguish students who were in the upper and lower groups or could also be omitted. In this study, the item was omitted because there were other items whose indicators were the same which had better D values and the items were enough to measure the indicators that had been set. Twelve essay items had good categories. Items with high distinguishing power were positively correlated with the overall test results. In other words, the items were answered correctly by most groups of students who scored high on the test and were answered incorrectly by most students who scored low on the test, so they could distinguish students in the upper and lower groups.

Table 3. Distinguishing Power and Difficulty Level Test Results

| No Item | D Value | P Value | No Item | D Value | P Value | No Item | D Value | P Value | No Item | D Value | P Value | No Item | D Value | P Value |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1       | 0.85    | 0.42    | 12      | 0.38    | 0.73    | 23      | 0.23    | 0.12    | 34      | 0.30    | 0.46    | 1      | 0.33    | 0.51    |
| 2       | 0.69    | 0.42    | 13      | 0.23    | 0.88    | 24      | 0.38    | 0.81    | 35      | 0.62    | 0.46    | 2      | 0.31    | 0.60    |
| 3       | 0.77    | 0.38    | 14      | 0.23    | 0.73    | 25      | 0.46    | 0.46    | 36      | 0.30    | 0.46    | 3      | 0.35    | 0.75    |
| 4       | 0.69    | 0.5     | 15      | 0.38    | 0.34    | 26      | 0.23    | 0.81    | 37      | 0.46    | 0.23    | 4      | 0.30    | 0.40    |
| 5       | 0.69    | 0.42    | 16      | 0.30    | 0.76    | 27      | 0.46    | 0.46    | 38      | 0.23    | 0.92    | 5      | 0.30    | 0.71    |
| 6       | 0.54    | 0.5     | 17      | 0.30    | 0.31    | 28      | 0.46    | 0.23    | 39      | 0.46    | 0.46    | 6      | 0.30    | 0.48    |
| 7       | 0.62    | 0.3     | 18      | 0.30    | 0.69    | 29      | 0.23    | 0.81    | 40      | 0.30    | 0.34    | 7      | 0.38    | 0.23    |
| 8       | 0.77    | 0.88    | 19      | 0.77    | 0.54    | 30      | 0.38    | 0.81    | 41      | 0.23    | 0.65    | 8      | 0.35    | 0.83    |
| 9       | 0.42    | 0.23    | 20      | 0.62    | 0.38    | 31      | 0.23    | 0.81    | 42      | 0.23    | 0.57    | 9      | 0.30    | 0.29    |
| 10      | 0.54    | 0.27    | 21      | 0.07    | 0.19    | 32      | 0.23    | 0.12    | 43      | 0.46    | 0.53    | 10     | 0.30    | 0.63    |
| 11      | 0.38    | 0.81    | 22      | 0.42    | 0.38    | 33      | 0.23    | 0.2     | 11      | 0.30    | 0.85    | 1      | 0.30    | 0.48    |

Furthermore the results of the difficulty level test of multiple choice items were 26 items including the medium category, 10 items including easy categories and 7 items including difficult categories, while for essays, 9 items included medium category, 2 items including easy categories, and 1 item including difficult categories. The proportion for this item was quite good because the proportion was close to 3: 5: 2 (30% easy category items, 50% moderate category items and 20% difficult category items). For multiple choice items 23% including easy category, 61% including moderate category and 16% including difficult categories. For essays 16% were easy categories, 75% were moderate categories and 9% were difficult categories.

A distractor could function well if chosen by more than or equal to 5% of all students, there were chosen especially from the lower group, chosen more by the lower group than the upper group, the number of upper group who chose a distractor was less than the number of upper group who chose the answer key. Based on the results of the distractor function test, there were 12 items which had distractors that were chosen less than 5% of all students (Table 4). Then there were 24 items with the number of upper group and lower group who chose at that distractor was same, namely item number 4 (d), 6 (b, d), 7 (c), 9 (a), 10 (e), 12 (e), 15 (e), 17 (c), 18 (b, c), 20 (a), 24 (c, d), 26 (a), 27 (b, c, d), 30 (d), 31 (b, d), 32 (e), 34 (c), 35 (d), 36 (e), 38 (a, d), 39 (e), 40 (e), 42 (d, e). In addition, there were 10 items with a number of lower group who chose at that distractor was less than upper group at that
The chemical literacy assessment instrument developed consists of multiple choice items to assess aspects knowledge and understanding chemical content, there were 42 items, essay items to assess aspects knowledge and understanding of the relationship between chemistry, technology and society, the application of analytical thinking and the application of reasoning there were 12 items and to assess the attitude aspect there were 26 statements.

Based on the development test results, a chemical literacy assessment instruments which were developed were valid and reliable, so that this instrument was suitable to be used to assess students' chemical literacy specifically in solution materials. This was appropriate with the results of other research [4], that the developed chemical literacy assessment instrument consisting of multiple choice test to assess aspects of knowledge and understanding of chemical content. To examine aspects of knowledge and understanding of chemical content it was suitable to use multiple choice tests that were flexible in measuring various cognitive levels, the chance to be answered correctly through smaller guessing than true-false forms and open stem and could include distractors in the form of misconceptions about certain concepts [27]. Essay test to assess aspects of knowledge and understanding of the relationship between chemistry, technology and society, the application of analytical thinking and the application of reasoning. Essay were suitable for assessing this aspect because the items involved high-level thinking skills and essay tests could provide an opportunity to freely show the breadth of knowledge and depth of understanding on certain concepts. Aspects of attitude on chemical literacy researchers adopted aspects of attitude from PISA, so that there were 3 aspects of attitudes that were evaluated, namely interest in chemistry, support for scientific inquiry, and responsibility for resources and environment using attitude scales [28]. The construction of the chemical literacy assessment instrument which was developed similar with assessment instrument which was developed by Thummathong and Thathong, Schwartz and Celik, the instrument have to assess 5 aspect chemical literacy.

4. Conclusion
The results of the analysis quality of chemical literacy assessment instruments show: (1) Empirical validity values of multiple choice items, essays and attitude scale respectively in range 0.11-0.87, 0.4-0.8, 0.1-0.74; (2)Reliability values for multiple choice items, essays and attitude scales respectively 0.93, 0.87, 0.94; (3)Readability test scores for multiple choice items, essays and attitude scales respectively 97%, 99%, 100%; (4)Results of distinguishing power test for multiple choice and essays items were in range 0.07-0.85 and 0.3-0.8; (5)Difficulty level test results for multiple choice items
were 16% difficult category, 61% moderate category, and 23% easy category while for essay items were 9% difficult category, 75% moderate category, and 16% easy category; (6) Distractor function test results for multiple choice items were distractors of 37 items that need to be fixed.

5. References

[1] Serhat Y R E Z and Çakir M 2006 Turkish J. Sci. Edu. 3 8-23
[2] Oludipe D and Awokoya J O 2010 Effect of cooperative learning teaching strategy on the reduction of students’ anxiety for learning chemistry Journal of Turkish science education 7 1-30
[3] Kementrian Pendidikan dan Kebudayaan 2013 Kurikulum 2013 kompetens idasar Sekolah Menengah Atas/Madrasah Aliyah Jakarta: Kementrian Pendidikan dan Kebudayaan
[4] Thummathong R and Thathong K 2016 Construction of a chemical literacy test for engineering students Journal of Turkish Science Education 13 3185-98
[5] Cigdemoglu C and Geban O 2015 Improving students’ chemical literacy levels on thermochemical and thermodynamics concepts through a context-based approach Chemistry Education Research and Practice 16 2 302-17
[6] Celik S 2014 Chemical literacy levels of science and mathematics teacher candidates Australian Journal of Teacher Education 39 1-1
[7] Barlia L 2016 Patterns of conceptual change process in elementary school students’ learning of science Journal of Turkish Science Education 13 2 49-60
[8] Nahadi, Siswaningsih W and Muchtar H K 2017 American Sci. Pub, 23 10555-10558
[9] Gilbert JK and Treagust DF 2009 Towards a coherent model for macro, submicro and symbolic representations in chemical education In Multiple representations in chemical education 333-350
[10] Grüber W, Nentwig P, Becker HJ, Sumfleth E, Pitton A, Wollweber K, and Jorde D 2001 Scientific literacy: From theory to practice In Research in science education-past, present, and future 61-70
[11] Shwartz Y, Ben-Zvi R, and Hofstein A 2006 The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students Chemistry Education Research and Practice 7 4 203-25
[12] TsaparlisG 2000 The states-of-matter approach (SOMA) to introductory chemistry Chemistry Education Research and Practice 1 1 161-8
[13] Rahayu S 2017 Pengaruh Laju Penginjeksiyan Doping Fluor Terhadap Aktivitas Fotokatalis Nano titania Pros. Sem. Kim.
[14] Hernani, Saefulloh, and Mudzakir A 2017 The first year pre-service teachers’ chemical literacy in individual learning case using the fuel cell technology based-chemical enrichment book In AIP Conference Proceedings 1868 1 030007
[15] Sujana A, Permanasari A, Sopandi W, and Mudzakir A 2014 Literasi kimia mahasiswa PGSD dan guru IPA sekolah dasar Jurnal Pendidikan IPA Indonesia 3 1
[16] Cigdemoglu C, Arslan HO, and Cam A 2017 Argumentation to foster pre-service science teachers’ knowledge, competency, and attitude on the domains of chemical literacy of acids and bases Chemistry Education Research and Practice 18 2 288-303
[17] Hernani, Ulum LL, and Mudzakir A 2017 Artificial muscles’ enrichment text: Chemical Literacy Profile of pre-service teachers in AIP Conference Proceedings 1868 1 030005
[18] Shwartz Y, Ben-Zvi R, and Hofstein A 2006 The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students Chemistry Education Research and Practice 7 4 203-25
[19] Witte D and Beers K 2003 Testing of chemical literacy (Chemistry in Context in the Dutch National Examinations) Chemical Education International 4 1 1-5
[20] Shwartz Y, Ben-Zvi R and Hofstein A 2005 International. J. Sci. Edu. 27 323-344
[21] Missen R W and Smith WR 1989 A question of basic chemical literacy? *Journal of Chemical Education* **66** 3 217
[22] Yusmaita E and Nasra E 2017 Perancangan Assesmen Literasi Kimia Dengan Menggunakan Model of Educational Rekonstruktion (MER) PadaTema:“Air SebagaiPelarut Universal” *JurnalEksaktaPendidikan(JEP)* **12** 49-55
[23] Sumarni W 2016 Preliminary Analysis of Assessment Instrument Design to Reveal Science Generic Skill and Chemistry Literacy *International Journal of Evaluation and Research in Education* **5** 4 331-40
[24] Lawshe CH 1975 A quantitative approach to content validity *Personnel* **28** 4 563-75
[25] Wiersma W and Jurs S G 2009 Research Methods in Education an Introduction *Pearson International Edition* **9**
[26] Doran RL 1980 Basic Measurement and Evaluation of Science Instruction*National Science Teachers Association* **1742** 20009 471-14764
[27] Domyancich J M 2014 *J. Chem. Edu.* **91** 1347–1351
[28] Bybee R and McCrae B 2011 Scientific literacy and student attitudes: Perspectives from PISA 2006 science *International Journal of Science Education* **33** 1 7-26