Validation of instrument to measure chemical literacy ability in Islamic senior high school students

Mahbub Alwathoni*, Sulistyo Saputro, Ashadi, Mohammad Masykuri

Department of Chemistry Education, Universitas Sebelas Maret, Surakarta, Indonesia.

*Email: mahboeb@student.uns.ac.id

Abstract. Advances in science and technology, mainly Information and Communication Technology (ICT), have a significant impact on the management of education today. It enables the development of an assessment system to facilitate the teaching and learning process. For this reason, there is a great need to develop assessment instruments. The purpose of this study is to validate the instrument of chemical literacy ability of Islamic high school students (Madrasah Aliyah) in Central Java in Indonesia. This instrument is in the form of chemical literacy questions in computer media (computer-based tests). Data needed to analyze the model quantitatively was obtained selectively from a sample of 67 respondents. Data validity and reliability were evaluated using the Rasch model. The findings show that Cronbach-α is 0.82 and reaches an acceptable PT-Measure range. Outfit Mean Square or Outfit Mean Square (MNSQ) and standard outfit z (ZSTD) values. It shows that the reliability of items was 0.95; the reliability of people was 0.82. Cronbach’s Alpha (α) measures the interaction between respondents and items, and this value indicates that the instrument used is in perfect and practical condition with a high level of reliability so that it can use in actual research. Chemical literacy assessment results show that the ability of chemical literacy in Islamic high school students is still low, of the 67 respondents, the average ability of chemical literacy is 48.50.

1. Introduction
The 21st century, marked as the era of literate society, requires graduates who can compete in the current era of information disclosure. In order to compete in the current era, education must be in line with current demands [1,2]. Thus, reforming the 21st-century education system is very important and immediate. The 21st-century education curriculum, for example, must be designed by 21st-century education standards, while the demand for results that must produce is graduates who are ready to work outside the disciplines they are studying [3,4]. The 21st-century education curriculum designs according to 21st-century education standards. 21st-century education standards can see from several reports of economic development cooperation organizations (OECD), which organize international student assessment programs (PISA). The OECD (2016a) reports that there are four main components in the form of skills that students must have in the 21st century, 1) Digital-Age literacy; 2) Inventive thinking; 3) Effective communication, and 4) high productivity [5].

Literacy abilities are not only on the ability to read, hear, write, and oral communication, but more than that, literacy skills in the digital era emphasize literacy abilities that connect to another. Digital era literacy includes scientific literacy, technological literacy, economic literacy, visual literacy, information literacy, and global awareness [2,6]. PISA-OECD scientific literacy in the Analytical Framework 2018 is the ability and skills to use knowledge contextually, identify problems, and draw
conclusions based on empirical evidence to make decisions related to changes in everyday life. This definition of scientific literacy is multi-dimensional in many aspects of measurement; content science, context science, and process science. Science content discusses certain natural phenomena and changes that occur due to human activities [7]. The science process is the ability of students to use scientific knowledge and understanding in answering questions or solving problems, and context of science discusses essential issues in daily life. Science literacy encourages students to be able to use science knowledge and apply it to solve problems related to the subject matter (content) which studied. Science literacy can develop through several scientific disciplines, such as chemistry, physics, and biology [8]. Educating people to think scientifically is one of the goals in science education reform. Need a clear definition of the framework for this education. According to DeBoer, without a precise definition, reforming science education will not achieve the goals expected by DeBoer (2000) [9]. Bybee (1997) [10] proposed a framework for defining and identifying scientific literacy. According to this framework, an individual can have several levels of scientific literacy based on the subject and context. Nominal, functional, conceptual, and multi-dimensional scientific literacy is a component of this framework.

Scientific literacy is a broad learning concept. Therefore, teaching for all subjects including chemistry must contribute to the aim of training people who are scientifically literate [11]. Chemical literacy is the ability of students to connect all information in chemical problems, and the ability to solve everyday problems [2,12]. Chemistry is an important basis for many dimensions of life and has many potential benefits for the future of the world. A good understanding of chemistry allows one to understand and explain the phenomena and problems of the world. It develops basic knowledge of how to live in the world, to deal with the problems of everyday life and how to make decisions regarding our actions as individuals. Unfortunately, many chemistry education programs fail to achieve many of these rather demanding goals. This failure is also due to the lack of development of instruments to measure the ability of adequate chemical literacy [13]. Some research results show that the ability of chemical literacy in Indonesia is still low. the ability of chemical literacy at the nominal level and on the macroscopic scale is very good, but for the microscopic scale students are still experiencing difficulties [13-14]. Measurement of chemical literacy abilities that still use this conventional model (paper-based assessment) has illustrated the ability of chemical literacy in high school students in Indonesia. Likewise, the ability and knowledge of chemistry teachers about the components of chemical literacy are still lacking, causing the learning objectives to improve the ability of chemical literacy to be less fulfilled [15].

Today, information and communication technology (ICT) plays an important role in the educational, social and economic changes that characterize the current knowledge society [16]. Skills called the 21st century, including communication, creativity, literacy, and ICT, are increasingly important in the education curriculum [17]. To cope with these changes, students must master a variety of clear skills and attitude changes; teachers and schools use ICT in their practice to improve student skills and as an important source of information [18]. The integration of ICT for teaching and learning has become an important task for schools throughout the world [19]. Research on the integration of ICT in education has become an important and broad research topic in educational technology. The use of ICTs in education is a diverse domain characterized by various objectives, assumptions, research designs, and data collection and analysis methods. Therefore, teachers and schools in the 21st century must develop ICT-based assessment instruments. The development of ICT-based research aims to make it more efficient, effective, accountable, and pay attention to environmental aspects by reducing paper usage.

The purpose of this study is to develop and validate instruments to measure the ability of chemical literacy in Islamic high school students in Central Java, Indonesia. The results of this study are expected to produce a valid computer-based chemical literacy assessment instrument to be able to measure the actual ability of students' chemical literacy. Then, the results of the study are used to develop teaching materials, assessment instruments, and learning models that are appropriate to empower the ability of scientific literacy (chemical literacy) for chemistry teachers in Islamic
secondary schools and overall for the improvement of education in Indonesia. The subject tested on
the chemical literacy ability test in this study is the research method, the nature of chemistry, the role
of chemistry in daily life. The focus of the assessment of the ability of chemical literacy in this study is
10th-grade Islamic high school. Chemical literacy ability should begin in the early grades (10th-grade
high school), so there is plenty of time to improve learning models while studying in high school to
improve chemical literacy abilities.

2. Research Methods
An instrument has been designing to measure the understanding of chemical literacy in Madrasah
aliyah students in Central Java, Indonesia. The instrument included 40 items which divided into four
constructions. This research conducted in Central Java, Indonesia. The participants are students of
Islamic High Schools (Madrasah Aliyah) in Central Java. The number of participants was 67 students
from 3 schools (Madrasah Aliyah 1 Semarang, Madrasah Aliyah 1 Demak, and Madrasah Aliyah 2
Grobogan). Test equipment to measure the ability of chemical literacy in the form of multiple-choice
through computer media (computer-based tests). 80% of chemical literacy questions in the form of
text and images and 20% in the form of videos with a duration of 10 minutes. Software to support
computer-based tests using the XAMPP local network. Figure 1. show examples of question items in
the form of videos and data collection activities. The video illustrates the role of chemistry in daily
life, which is observed by students as the video running; students choose one of the answers via the
selection button. 40 Students completed chemistry literacy questions for 120 minutes.

![Example of chemical literacy items in video format](image1)

**Figure 1.** Examples of chemical literacy items in video format and data collection activities.

After all, the participants did the test, and the researcher began to analyze the results. In this study,
the Rasch Modeling provides the basis for establishing instrument measurements (a test of chemical
literacy ability). Through calibration of item difficulty and people's ability, Winsteps software
mathematically converts raw ordinal data into logit size while simultaneously assessing the overall suitability of the instrument [20]. The results of this iterative process produce relevant output tables that present diagnostic evaluations of relevant instruments for reporting statistical fit and psychometric properties of this instrument. The Rasch Model Analysis is used to investigate the validity and reliability of the instruments developed. The Rasch model analysis is used to test the construct validity using the analysis of item polarity, dimensions, and item compatibility. Instruments to measure the ability of chemical literacy, based on PISA (2015) and Swartz (2006) [21], are presented in Table 1.

Table 1. Aspects of Chemical Literacy Test and types of assessment

| Aspects     | Descriptions                                                                 | Types of assessment | Item Number   |
|-------------|------------------------------------------------------------------------------|---------------------|---------------|
| Knowledge   | An understanding of the central facts in everyday life, basic concepts, and explanatory theories that build the foundation of scientific knowledge. | Multiple-choice     | P1,P2,P3,P4,P5,P6, P7,P8,P9,P10 |
| Context     | Local, national and global issues. Current issues or issues that already occur that require an understanding of science and technology. | Multiple-choice     | K11,K12,K13,K14, K15,K16,K17,K18, K19, K20 |
| Competency  | The ability to explain natural phenomena scientifically, study, understand and design scientific research. | Multiple-choice     | C21,C22,C23,C24, C25,C26,C27,C28, C29,C30 |
| Attitude    | Assessment of attitudes towards scientific and technological interests, assessing scientific approaches to research, perceptions and moral awareness, sense of responsibility, and scientific attitude. | Multiple-choice     | S31,S32,S33,S34, S35, S36,S37, S38,S39,S40 |

3. Results and Discussion
The reliability of instruments to measure the ability of chemical literacy in this study shown in Table 2. Acceptable reliability based on the approach of the RASCH model, Cronbach's Alpha (α) is between 0.71-0.99, this figure shows the best level. Research findings on measuring instruments found that the reliability obtained based on Cronbach Alpha (α) was 0.82 (reliability of people) and 0.95 (reliability of items). According to the Summary Statistics of Table 2, the real Person Reliability index (0.82) indicates that the consistency of person responses is right [17]. Cronbach's Alpha (α) is measuring the interaction between respondents and items, and this value shows instruments used are in perfect condition and effectively with a high level of reliability thus can use in the actual research. Research findings indicate that respondents responding to the instrument in the study are likely to be high. High item reliability estimates can indicate items have excellent latent variables. As such instrument design on this study may regard as a reliable instrument for use with a different group of respondents. The value of the Cronbach Alpha coefficient is 0.82, according to the Rasch Model computation describes the interaction between 67 persons and the 40 items. The reliability score of 0.82 classified as useful as described under the instrument quality criteria in Sumintono and Widhiarso (2014) [22]. This score indicates that there is a high level of interaction between respondents and items. An instrument that has the right psychometric internal consistency is considered a very reliable instrument. the reliability of the instrument eventually becomes important and the results of the research can be scientifically justified. The statistical summary provides overall info on the quality of whole student (respondents) response patterns, the variety of the instruments used, as well as the interactions between respondents and items.
Table 2. Summary of Measured Person and Item

| SUMMARY OF 67 MEASURED Person |  |  |  | INFIT |  |  | OUTFIT |  |  |
|-------------------------------|---|---|---|------|---|---|-------|---|---|
| TOTAL SCORE                  | COUNT | MEASURE | ERROR | MNSQ | ZSTD | MNSQ | ZSTD |
| S.D.                         | 6.0 | 40.0 | 1.00 | .04 | .22 | 1.2 | .45 | 1.1 |
| MAX.                         | 36.0 | 40.0 | 3.07 | .59 | 1.59 | 2.6 | 2.48 | 3.4 |
| MIN.                         | 8.0 | 40.0 | -1.95 | .38 | .37 | 2.6 | .36 | 2.1 |

| REAL RMSE | .42 TRUE SD | .91 SEPARATION | 2.16 Person RELIABILITY | .82 |
| MODEL RMSE | .40 TRUE SD | .92 SEPARATION | 2.27 Person RELIABILITY | .84 |

| S.E. OF Person MEAN = .12 |

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82

| SUMMARY OF 40 MEASURED Item |
|-------------------------------|---|---|---|------|---|---|-------|---|---|
| TOTAL SCORE                  | COUNT | MEASURE | ERROR | MNSQ | ZSTD | MNSQ | ZSTD |
| S.D.                         | 17.2 | 67.0 | 1.36 | .08 | .13 | 1.0 | .32 | 1.0 |
| MAX.                         | 63.0 | 67.0 | 3.15 | .56 | 1.30 | 1.7 | 2.25 | 2.9 |
| MIN.                         | 4.0 | 67.0 | -3.21 | .26 | .68 | 2.5 | .43 | 1.7 |

| REAL RMSE | .35 TRUE SD | 1.52 SEPARATION | 4.36 Item RELIABILITY | .95 |
| MODEL RMSE | 34 TRUE SD | 1.52 SEPARATION | 4.51 Item RELIABILITY | .95 |

| S.E. OF Item MEAN = .25 |

UMEAN= .0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = .09
2680 DATA POINTS, LOG-LIKELIHOOD CHI-SQUARE: 2563.66 with 2574 d.f., p=.5536
Global Root-Mean-Square Residual (excluding extreme scores): .3963
Capped Binomial Deviance = .2077 for 2680.0 dichotomous observations

The Rasch model also shows the separation value, which indicates the differences in the level of respondents and items. The separation of the respondents showed a good separation of the item difficulty level appropriate to the Linacre (2005) [18], which describes the separation of more than 2.0 is a good value. The three psychometrics attribute are an Outfit mean square value (0.5 <Outfit MNSQ <1.5), a Z-standard Outfit value (-2.0 <ZSTD <+2.0) and a Point Measure Correlation Value (0.4 <Pt Measure Corr <0.85). Item that has the value beyond of these three psychometrics attribute can categorize as misfit items that need to be revised and re-test again. The findings show that the initial assessment is an effective and appropriate method to assess the understanding of teachers, where the Cronbach-α is 0.82 and achieve the acceptable ranges of PT-Measure, Mean Square Outfit or Outfit Mean Square (MNSQ) and z-standardized values (ZSTD) Outfit. The reliability of the instruments was found to be 0.95.

An important consideration in Rasch's analysis is "fit". The quality control mechanism "fit", evaluates how well the data show in accordance with the Rasch model. If the data deviates significantly, the cause needs to be considered, and inappropriate respondents or items can be deleted or revised. It would be useful to consider compatibility analysis as a step to investigate whether the items of an instrument involve one trait and if individual responses are suitable for a particular calculation and communication from someone who measures along a single trait. Two statistics that can be used to assess it are Infit and outfit [23]. For the introductory analysis that we present, the outfit will be used. Outfit means-square value, z-standard outfit, and point measure correlation are the criteria used to see the level of suitability of items (item fit). If the items in the three criteria are not fulfilled, it can be ensured that the items are not useful, so it needs to be deleted or corrected. This will ensure that the students' level of understanding of the chemistry literacy items is appropriate.
Table 3. Item misfit statistics provided in a Winsteps output table

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | MEASURE | MODEL S.E. | INFIT MNSQ | ZSTD | OUTFIT MNSQ | ZSTD | PT-MEASURE CORR | EXP. | EXACT MATCH OBS | EXACT MATCH EXP | Item |
|--------------|-------------|-------------|---------|-----------|-----------|------|------------|------|-----------------|------|-----------------|----------------|------|
| 34           | 4           | 67          | 3.15    | .56       | .79       | .4   | .43        | .4   | .50             | .50  | .94             | .94            | S34  |
| 37           | 4           | 67          | 3.15    | .56       | .79       | .4   | .43        | .4   | .50             | .50  | .94             | .94            | S34  |
| 34           | 4           | 67          | 3.15    | .56       | .79       | .4   | .43        | .4   | .50             | .50  | .94             | .94            | S34  |
| 37           | 4           | 67          | 3.15    | .56       | .79       | .4   | .43        | .4   | .50             | .50  | .94             | .94            | S34  |
| 32           | 7           | 67          | 2.47    | .44       | .95       | .0   | .67        | .6   | .45             | .45  | .90             | .90            | S32  |
| 33           | 9           | 67          | 2.08    | .49       | .20       | .3   | .76        | .7   | .49             | .49  | .86             | .86            | S36  |

Table 3. shows the order of the mismatch of items that provide important information about the instrument being analyzed. Measurement items are stated in logit units. With the coding used for the analysis of this report, the top item code (S34) indicates items that are more difficult to answer, while the bottom item code (S36) indicates items that are easier to answer. Table 3., which provides item measures, reveals that item S34 and S37 is the most challenging item to endorse on the instrument and item S36 is the most comfortable item to endorse. More specifically, item S34 is 3.15 logits, and item S36 is −3.21 logits. The average logit value of the item is 0.0. This instrument measures what had determined. The average logit item value of 0.00 random assessment was determined to be announced 50:50. Items that meet the MNSQ criteria are in the range of 0.5 - 1.5. For the ZSTD outfit criteria, the ZSTD criteria range is -2.0 +2.0. Ten statement items met two criteria; 10 items only met MNSQ outfit criteria. While items that did not meet the MNSQ outfit criteria and ZSTD outfit of 2 items (K15 and K12), the two statements need further revision by improving their editors.

Figure 2. This Map can provide strong visuals of person-to-person relationships on the same logit interval scale. The Wright Map helps researchers to (a) assess the strengths and weaknesses of instruments, (b) document item hierarchies, (c) compare theories with observed data, and (d) provide clinical guidance to practitioners [20]. Parts of the Wright Map designed with Winsteps for the
instrument data set are provided in Figure 2. In this report, we only display the respondents and items section from the Wright Map in one plot. We have found that for an analysis of Wright Maps for ranking scale, it would be helpful to get started with Wright Maps, which provides item sizes. Through Rasch's analysis, through WinStep software, responses to the Person-Item Distribution Map were obtained, as shown in Figure 2. References for student abilities that refer to question items are plotted on the left, and for the level of difficult questions are plotted on the right-hand side. On the same logit scale in line with Latent Character Theory. Variable Map is identical with names such as Person-Item or Item-Person Map and Wright Map, as mentioned in Bond and Fox (2015) [24]. Variable Map is a Winsteps technique that can map respondents' conceptual understanding of items with varying degrees of difficulty. The Wright Map shows that some respondents had a good understanding by answering their agreement with the items in the questionnaire. Respondents 20P, 67P, and 66P are outside the limits of two standard deviations (T) which indicate a good understanding of all items. Respondents who are above the two standard deviations (T) indicate their ability of chemical literacy and its components. The results of this study indicate that the distribution of levels of understanding of chemical literacy in students is uneven. Only a few students understand and have chemical literacy skills.

Furthermore, research needs to be conducted on the learning model conducted by chemistry teachers, because chemical literacy abilities must refer to a framework for measuring actual chemical literacy abilities according to the PISA (2015) or Swartz (2006) assessment. Rasch's analysis also provides a method for validating the instrument to match the level of difficulty of the item with the level of individual respondents who took the test [25,26]. Table 4, shows, very few S34 and S37 items (6%) were answered correctly by students. The questions number S34 and S37 shown in Table 4.

Table 4. Questions of Chemical Literacy Test

| No. | Questions List | Aspects |
|-----|----------------|---------|
| S34 | Additives are chemicals from nature or synthesized products that added to food for specific purposes, examples of additives shown in the following information; (a.) formalin, (b) benzoic acid, (c) metal yellow, (d) curcumin, (e) chlorophyll, (f) rhodamin B. Food additives which are not banned for use in accordance with BPOM regulations of the Republic of Indonesia are ... | Moral awareness, a sense of responsibility, and scientific attitude. |
| S37 | Garbage is a global problem that is quickly sought by the world's population. Minister of Maritime Affairs and Fisheries Susi Pudjiastuti said Indonesia is the most significant contributor of plastic waste in the world which thrown into the sea. Based on data obtained from the Indonesian Plastic Industry Association (NAPLAS) and the Central Statistics Agency (BPS), plastic waste in Indonesia reaches 64 million tons/year in which 3.2 million tons is plastic waste that thrown into the sea. Plastic waste that enters the sea can split into small particles called microplastics with a size of 0.3 - 5 millimetres. This microplastic is very easy to be consumed by marine animals and causes poisoning so that the marine ecosystem threatened. (Source: kompas.com/published 23 Nop 2018). The accurate statement of how simple can be done to reduce inorganic waste that is difficult to decompose is... | Moral awareness, a sense of responsibility, and scientific attitude. |

Overall, the results of measuring the ability of chemical literacy with this instrument are still low. The average measurement results show a value of 48.50. Measurement results of every aspect of chemical literacy; knowledge of chemical literacy is 45.3, context of chemical literacy is 46.8, chemical literacy competence is 58.4, and scientific attitude is 43.2. The results of measuring scientific attitudes show the lowest score. Scientific attitude is the assessment of attitudes towards science
shown by an interest in science and technology, assessing scientific approaches to appropriate investigations, and perceptions and awareness of environmental problems. Moral awareness, sense of responsibility.

Figure 2. The Wright Map (Person–Item Map)

Table 5 shows Unidimensionality, and this picture is relevant to know whether an instrument can measure what must be measured, in this case, it is to measure the chemical literacy ability of Islamic high school students (Madarasah Aliyah). Unidimensionality in Rasch Modeling refers to invariant measurements [27]. Therefore, Unidimensionality is a fundamental requirement in building validity. The Unidimensionality condition is critical to determine the estimated Rasch Modeling parameters [28,29]. Rasch modelling uses principal component analysis of residues, which measures the extent to which the diversity of instruments measures what must be measured. Table 5 shows the results of diversity data measurement is 37.1% which is not much different from the expected value of 36.8%.
Differences in measurement values with expectations of less than 20% indicate that Unidimensionality is acceptable.

### Table 5. Standardized Residual Variance

|                          | -- Empirical -- | Modeled |
|--------------------------|-----------------|---------|
| Total raw variance       | 106.5 100.0%    | 100.0%  |
| Raw variance explained by measures | 39.5 37.1% | 38.8%  |
| Raw variance explained by persons | 10.6 10.0% | 9.9%   |
| Raw variance explained by items | 28.9 27.1% | 26.9%  |
| Raw unexplained variance (total) | 67.0 62.9% | 100.0% 63.2% |
| Unexplained variance in 1st contrast | 7.5 7.0% | 11.1% |
| Unexplained variance in 2nd contrast | 5.3 5.0% | 7.9% |
| Unexplained variance in 3rd contrast | 4.1 3.9% | 6.1% |

### 4. Conclusion

The purpose of this report is to contribute conversations around crucial knowledge and skills for schools in the 21st century. The report also examines essential areas of assessment and evaluation of crucial knowledge and skills for the 21st century, especially the ability of scientific literacy (chemical literacy). Science literacy is the ability and skill recommended by PISA in 21st-century education. The component of scientific literacy can be developed based on subject matter with scientific discipline, so scientific literacy is the same as chemical literacy, physical literacy, or biological literacy. Science literacy as a capability expected in 21st-century education needs to be supported by the development of accurate scientific literacy ability assessment instruments. Assessments can be conventional (paper-based) or technology and information-based (computer). Advances in information technology and computers in the 21st century should have involved technology. The research findings found that the reliability obtained based on Cronbach Alpha (a) was 0.82 (reliability of people) and 0.95 (reliability of items). Cronbach's Alpha (α) measures the interaction between respondent and item, and this value indicates that the instrument used is in perfect and practical condition with a high level of reliability so that it can use in actual research. The results of this study indicate that there are still many students who lack chemical literacy abilities. The results of the assessment of the ability of chemical literacy of 67 respondents, the average is 48.50. The results of chemical literacy ability assessments are still low, and it expected that teachers could apply approaches or learning models that support learning objectives in the form of scientific literacy abilities and skills.

### 5. Acknowledgement

We thank our colleagues from the chemistry teachers meeting in central Java Indonesia (MGMP) for helping with this research, even though they may not agree with all interpretations of this research report. We also thank the 10th-grade students involved in this study.

### References

[1] Halah Ahmed Alismail and Patrick McGuire, (2015). 21st Century Standards and Curriculum: Current Research and Practice. Journal of Education and Practice. ISSN 2222-1735 (Paper) ISSN 2222-288X. Vol.6, No.6.

[2] Zehavit Kohen, Orit Herscovitz, and Yehudit Judy Dori, (2019). How to promote chemical literacy? On-line question posing and communicating with scientists. Chemistry Education Research and Practice. Royal Society of Chemistry. DOI: 10.1039/c9erp00134d.

[3] Rungrat Thummathong, Kongsak Thathong, (2018). Chemical literacy levels of engineering students in Northeastern Thailand. Kasetsart Journal of Social Sciences 39 (2018) 478e487. https://doi.org/10.1016/j.kjss.2018.06.009.
[4] World Economic Forum (2015). *New Vision for Education Unlocking the Potential of Technology*. 91-93 route de la Capite CH-1223 Cologny/Geneva Switzerland. http://www.weforum.org

[5] OECD (2016a) PISA (2015). *Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*. Doi: 10.1787/9789264255245-en.

[6] Sothayapetch, Lavonen and Juuti, (2013). *A Comparative Analysis of PISA Scientific Literacy Framework in Finnish and Thai Science Curricula*, Science Education International, 24(1), pp. 78–97. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1015827 &login.asp&lang=es&site=ehost-live.

[7] OECD (2016b). *PISA 2018 Draft Analytical Frameworks*, (May 2016). Available at: http://www.oecd.org/pisa/data/PISA-2018-draft-frameworks.pdf.

[8] Celik, S. (2014). *Chemical Literacy Levels of Science and Mathematics Teacher Candidates*, Australian Journal of Teacher Education, 39(1), http://doi:10.14221/ajte.2014v39n1.5.

[9] DeBoer, G. E., (2000). *Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform*. Journal of Research in Science Teaching, 37(6), 582-601. http://dx.doi.org/10.1002/0022-3778(200008)37:6<582::AID-TEA5>3.0.CO;2-L.

[10] Bybee, R W, (1997). *Achieving scientific literacy: From purposes to practices*; ERIC.

[11] Shwartz, Y., Ben-Zvi, R., & Hofstein, A. (2006a). *Chemical literacy: What does this mean to scientists and school teachers?*. Journal of Chemical Education, 83(10), 1557. http://dx.doi.org/10.1021/ed083p1557.

[12] Cigdemoglu, C and Geban, (2015). *Improving students’ chemical literacy levels on thermochemical and thermodynamics concepts through a context-based approach*. (Chemistry Education Research and Practice, 2015), pp. 302-317.

[13] Muntholib, Suhadi Ibu, Sri Rahayu, Fauziatul Fajaroh, Sentot Kusairi, and Bambang Kuswandi, (2019). *Chemical Literacy: Performance of First Year Chemistry Students on Chemical Kinetics*. Indones. J. Chem., xxxx, xx (x), xx – xx. Accepted: April 15, 2019. DOI: 10.22146/ijc.43651.

[14] Shifi Syarifa Fahmina, N Y Indriyanti W, A E Setyowati, M Masykuri, and S Yamtinah, (2019). *Dimension of Chemical Literacy and its Influence in Chemistry Learning*. IOP Conf. Series: Journal of Physics: Conf. Series 1233 (2019) 012026. doi:10.1088/1742-6596/1233/1/012026.

[15] Fetchiyatun Ni’mah, (2019). Research trends of scientific literacy in Indonesia: Where are we?. Jurnal Inovasi Pendidikan IPA, 5 (1), 2019, 23-30. https://doi.org/10.21831/jipi.v5i1.20862.

[16] Kozma, R, (2008). *Comparative analysis of policies for ICT in education*. In J. Voogt & G, Knezek (Eds), International handbook of information technology in primary and secondary education (pp. 1083–1096). New York: Springer.

[17] Voogt, J., & Pareja Roblin, N. (2010). *21st-century skills: Discussion paper*. Retrieved from http://www.internationalsymposiumoneducationalreform.com/storage/21st%20Century%20Skills.pdf

[18] Anderson, R E, (2008). *Implications of the information and knowledge society for education*. In J. Voogt & G. Knezek (Eds), International handbook of information technology in primary and secondary education (pp. 5–22). New York: Springer.

[19] Vanderlinde, R., & van Braak, J. (2015). Measuring ICT use and contributing conditions in primary schools. British Educational Research Journal, Vol. 46, 1056–1063.

[20] Linarce, J M (2012). *A user’s guide to Winsteps Ministepl Rasch-model computer programs [version 3.74.0]*, Chicago IL: Winstep.com.

[21] Shwartz, Y, Ben-Zvi, R, & Hofstein, A (2006a). *Chemical literacy: What does this mean to scientists and school teachers?*. Journal of Chemical Education, 83(10), 1557. http://dx.doi.org/10.1021/ed083p1557
[22] Sumintono, and Widhiarso, (2014). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi). Cimahi, Indonesia: Trim Komunikata Publishing House. (in Indonesian).

[23] Linacre, J M, (2011). *A User’s guide to WINSTEPS Ministeps; Rasch-model Computer Program*. Program Manual 3.73.

[24] Boone, W J, Staver, J R, & Yale, M S (2014). *Rasch Analysis in the Human Sciences*. Netherlands: Springer.

[25] Bond, T G, & Fox, C M (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd Edition). New York: Routledge.

[26] Hendriks, J., Fyfe, S., Styles, I., Skinner, S., Merriman, G. (2012). Scale construction utilizing the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. The Australasian Medical Journal, *5*(5), 251-261. Doi:10.4066/AMJ.2012.952.

[27] Tractenberg, R E, Gushta, M. M., Mulroney, S. E. Weissinger, P. A. (2013). *Multiple-choice questions can be designed or revised to challenge learners' critical thinking*. Advances in Health Sciences Education *18*(5), 945-961. DOI: 10.1007/s10459-012-9434-4.

[28] Corinne J Perera, Bambang Sumintono, Jiang Na (2018). *The Psychometric Validation of The Principal Practices Questionnaire Based on Item Response Theory*. International Online Journal of Educational Leadership, 2018. Vol. 2, No. 1, 21-38

[29] John R. Thompson, Joseph C Cappelleri, Christine Getter, Andreas Pleil (2007). *Enhanced Interpretation of Instrument Scales Using the Rasch Model*. Drug Information Journal, Vol. *41*, pp. 541–550, 2007 • 0092-8615.