Construct Validity of Science Motivation and Beliefs Instrument (SLA-MB): A Case study in Sumedang, Indonesia

A Rachmatullah¹, R P Octavianda², M Ha¹, N Y Rustaman³ and S Diana³

¹Division of Science Education, College of Education, Kangwon National University, Chuncheon-si, South Korea
²Department of Bio-Management, School of Life Sciences and Technology, Institut Teknologi Bandung, Jl. Ganesha No.10, Bandung, Indonesia
³Department of Biology Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No. 229, Bandung 40154, Indonesia
Email: arifraach@gmail.com

Abstract. Along with numerous instruments developed and used in science education researches, some of those instruments have been translated to local language in the country where the instruments were used. Most of researchers that used those translated instruments did not report the quality of those translated instruments. One of the instruments is the Scientific Literacy Assessment (SLA) including the Science Motivation and Beliefs (SLA-MB) as part of the SLA. In this study, the SLA-MB has been translated into Indonesian Language (Bahasa). The purpose of this study is to investigate the SLA-MB instrument that has been translated to Indonesian language from the view of dimensionality, reliability, item quality and differential item functioning (DIF) based on IRT-Rasch analysis. We used Conquest and Winstep as the program for IRT-Rasch analysis. We employed quantitative research method with school-survey on this study. Research subjects are 223 Indonesian Middle school students (age 13-16), with 64 boys and 159 girls. IRT-Rasch analysis of the SLA-MB Indonesian version indicated that a three-dimensional model fit significantly better than one-dimensional model, and the reliability of each dimensions are about 0.60 to 0.82. As well as those findings, fit values of all items are acceptable, moreover we found no DIF for all of the SLA-MB items. Overall, our study suggests that Indonesian version of SLA-MB is acceptable to be implemented as research instrument conducted in Indonesia.

1. Introduction
The rapid development of science and technology in the 21st century demands humans not only have the basic literacy such as reading, writing and counting but requires another ability that makes human can survive and follow this century which full of science and technological products. Basic capability that enables humans to survive in the rapid development of science and technology is called as scientific literacy. As described by the Organisation for Economic Cooperation and Development or OECD [1], scientific literacy is the ability to understand scientific knowledge, identify scientific issues, draw conclusions based on the evidences and make decisions on how the human impact the environment. Thus, scientifically literate society can make decisions and participate actively in the discussions about scientific issues which are currently being ardently discussed [2]. Unfortunately,
based on the results from international survey, such as the Program for International Student Assessment (PISA) which is a popular triennial international survey, from 2000 until 2012 Indonesian students achievement on scientific literacy are always categorized in the low category, even in the 2012 survey from 65 countries that participated on PISA Indonesian students achievement on scientific literacy was placed on rank 64 [3]. The Indonesian Government is not standing still to follow up the results of this “long-standing” low achievement. It is proved in line with the publication of a new Curriculum 2013, which one of the goals of Indonesian education is to increase the performance of Indonesian students in the International surveys and studies especially in PISA [4]. However, scientific literacy is not only deal with the cognitive aspects, but also affective aspects which is from several studies and some experts prove and mentioned that affective could significantly affect cognitive aspects [5],[6]. One of the affective component that has been scrutinized and has a relationship with scientific literacy is the motivation to learn science [7].

The motivation to learn science is one’s internal state which arises, directs and sustains in learning-science behaviours [7]. Motivation is essentially an elaboration of Bandura’s social cognitive theory which assumes that students will learn effectively when it is self-regulated, which takes place when students can control their motivation against what they learn and it will eventually lead towards their learning outcomes [7],[8]. Making students become motivated self-learner is the goal of every teacher [9], because students, who have become a motivated self-learner in science classes, will have much curiosity to solve and discuss vigorously on science-related problems. This kind of capability will eventually lead students to implement it on their daily lives which are multi-discipline. It is aligned with one of the definitions and purposes of scientific literacy which stated that individuals can implement the methods, concepts and theories of science in their daily lives. Hence, the motivation to learn science has an important role in improving the student’s scientific literacy, especially Indonesian students, which besides the results from PISA, the results from Rachmatullah et al. showed that Indonesian student’s motivation towards science is still low which is in the value of 60% [10].

In order to evaluate Indonesian students’ attitude and motivation in learning science, we need a valid, reliable and qualify instrument which meets the psychometric standard of instruments, so that the results which are obtained from the instrument can be trusted. Moreover, recently Fives et al. developed a set of instruments to measure scientific literacy of middle school students, and the instrument is not only measure students cognitive aspect, but also it measures student motivation towards science which is called Scientific Literacy Assessment-Science Motivation and Beliefs (SLA-MB) [6]. In previous studies, we have been using SLA-MB which has been translated into Indonesian (Bahasa), and we only reported the validation of its content, but did not report the validation of the construct [10]. In fact the construct validity is the important thing that should be reported because as Messick stated about construct validity, it provides evidence to support the trustworthiness of the score interpretation in the explanatory concept that describes the relationship between scores in the performance and test with other variables. In short, construct validity is evidence-based to interpret the score that obtained from the students or participants [11].

Based on Messick [11] in addition to the content validity that has mentioned above, there are five other aspects of the Construct Validity namely substantive, structural, generalizability, external and consequential aspects. A substantive aspect of the construct validity is the empirical evidence that related to the consistency of the responses on the instrument [12]. Structural aspect of construct validity is an aspect that related to the consistency of responses that it could describe a particular domain or specific dimensions that lie in the instrument. Aspects generalizability assumes that an instrument must be fair from a variety of respondents, no group of respondents who get disadvantage when doing the instrument. External aspects of construct validity sees the correlation between the developed instrument with other similar instruments (e.g. same content), while consequences aspects of construct validity predicts how the consequences or effect on respondents after doing the instruments. For more information about the construct validity see Messick "Validity of Psychological Assessment"[11]. In order to determine the construct validity of an instrument, especially in the realm
of social sciences namely psychology and education including science education, the most well-known methods are through Classical Test Theory (CTT) and Item Response Theory (IRT).

CTT assumes that the individual’s attitude scores on the rating scale form a linear combination with total score, which in CTT each item is presumed to have the same difficulty level and standard error [13], so that when analyzing rating scale data (e.g. Likert scale) is often considered as interval data, and sum it directly from the raw data. In fact the numbers from the Likert-scale are ordinal data, which the differences between one points to another is exactly not known. As an example of the unclear distinction "strongly agree", "agree", "disagree", whether it was true that "agree" to "strongly agree" to have one point different attitudes? Point or score is a number belonging to the categorical data or namely ordinal data, which can not be used when testing statistical parametric tests [13], because only the data in the form of interval and ratio one which can be used in the statistic tests [14]. Thus, with the CTT method most researches are not same, and the results of the instrument, which were analyzed using CTT, can only be used in the same research subjects and can not be used for wider subjects [15].

IRT is one of psychometric methods to examine an instrument to be more reliable, valid and widely used [16]. IRT such as Rasch analysis is conducted in this study will convert ordinal data into ratio scale data and produce item parameters and person’s parameters which are ratio data too [16]. The data which have been analyzed through Rasch analysis will become equal with other types of measurements such as height measurement, and so forth that is continuous data and has specific unit. The data that have been analyzed using Rasch method will have a unit called the logit Rasch (logarithm of odds) which is in interval/ratio level [13],[17]. Therefore, recently many science education researchers mainly those who relate to educational assessment have started using IRT as a method of testing the reliability, validity and legality of the instrument and began to leave the CTT method [18]. For more details about Rasch models see Boone et al [19]. Generally speaking, based on the above explanation the purpose of this study is to uncover the construct validity (dimensionality, generality and reliability) of SLA-MB which has been translated into Indonesian by using IRT-Rasch model where the method was not used by the instruments’ developers.

2. Research methods
Previously this research had a purpose to uncover Indonesian middle school students’ scientific literacy in one district called Sumedang, so the data collection procedure referred to how PISA collects the data, in which the subject is the student with average age of 13-16 years. In Indonesia, students with that age-range are in the ninth grade of middle school. Therefore, the research subjects are 223 middle school students of grade ninth consisting of 159 girls and 64 boys. Students who used directly as the subject of our research are derived from eight middle schools in four sub-regions in Sumedang district. Four of eight those middle schools are A accredited schools and the other fourschools are the school with B accreditation, so that each sub-region have representation in both A and B accreditation. Because it is only used for the pilot study and the purpose of generalization, we only collected samples one class from each school.

As has already been mentioned and described in the previous section about research instrument that we used, we employed one part of scientific literacy instruments that have been developed by Fives et al [6] which is the Scientific Literacy Assessments-Science Motivation and Beliefs. At the time of developing the SLA-MB, Fives et al divided it into three indicators which are the value of science, self-efficacy and personal epistemology. SLA-MB is formatted rating scale with Likert- scale 1 to 5 types. Before employing SLA-MB as the research instrument, it was translated first from English into Indonesian by experts in the field of English-Indonesian and then re-checked by an expert in the field of science education for its contextualization fit. Table 1 is listed each item on each indicator of SLA-MB both Original English version and Indonesian version.

The construct validity of an instrument or assessment consists of six aspects: content, substantive, structural, generalizability, external and consequential. In this intensive research we did not examine all of these aspects, but we just examined three aspects that we believe are the crucial aspects and highly

---

The above text is a natural representation of the image provided.
correlated with the definition of construct validity. The third aspects that we have done are substantive, structural and generalizability. Before analyzing these aspects, at the time of coding data, we reversed the score or points items that have a negative attitude. Then after that, firstly we examined the structural aspects. Exploring structural aspects, we employed the dimensionality test based on IRT-Rasch analysis. We tested two types of Rasch models for SLA-MB instruments, whether such instruments unidimension or multidimension (three-dimension) that is accordance with the validity of the content. The higher chi-square (X²) value and lower value on Final Deviance and the Akaike Information Criteria (AIC) than other models indicated that the model fit well to the instrument. Examining substantive aspects, we conducted an analysis of reliability by using IRT-Rasch (through PV-reliability) and CTT (Cronbach's Alpha), in addition we also reported item fit indices of each dimension by reporting mean-square (MNSQ), which the benchmark for MNSQ rating scale is between 0.7-1.4 [19]. And lastly, in order to uncover the generalizability aspects, we did Differential Item Functioning (DIF) gender analysis and examines whether there is gender bias for each item or not. According to Boone et al. items contained bias could be seen from the DIF contrast > 0.64 [19]. We used software Conquest 4.5.0, SPSS 22 and Winstep 3.68.2 to analyze our data.

| Indicator                          | Original Version [6]                                                                 | Indonesian Version [10]                       | Code |
|-----------------------------------|--------------------------------------------------------------------------------------|----------------------------------------------|------|
| **Value of Science**              | In general, I find working on science assignments useful.                             | Secara umum, saya melakukan suatu kerja dalam tugas sains. | VS1  |
|                                   | Compared to most of your other activities, how useful is what you learn in science?  | Dibandingkan dengan banyaknya aktivitas kamu, seberapa berguna kamu belajar?               | VS2  |
|                                   | For me, being good in science is                                                    | Bagi saya, menjadi pintar dalam sains adalah                                               | VS3  |
|                                   | Compared to most of your other activities, how important is it for you to be good at science? | Dibandingkan dengan aktivitas lain, seberapa pentingkah untuk pintar dalam sains?           | VS4  |
|                                   | How much do you like doing science?                                                 | Seberapa suka-kah kamu dalam mengerjakan sains?                                             | VS5  |
|                                   | Some things that you learn in school help you do things better outside of class, that is, they are useful. | Beberapa hal yang kamu pelajari di sekolah membantu kamu untuk melakukan hal yang lebih baik di luar kelas, berarti ini sangat berguna. Sebagai contoh, mempelajari tumbuhan yang kamu tanam di kebun. Secara umum, seberapa berguna-kah apa yang kamu pelajari di dalam sains? | VS6  |
|                                 | **Self-Efficacy**                                                                  |                                               |      |
|                                   | I know when to use science to answer questions.                                     | Saya tahu kapan saatnya menggunakan sains untuk menjawab suatu masalah                     | SE1  |
|                                   | I can use science to make decisions about my daily life.                            | Saya dapat menggunakan sains untuk membuat suatu keputusan mengenai kehidupan sehari-hari saya. | SE2  |
|                                   | I know how to use the scientific method to solve problems.                          | Saya tahu bagaimana menggunakan metode ilmiah untuk memecahkan suatu masalah.               | SE3  |
|                                   | It is easy for me to tell the difference between scientific findings and advertisements. | Sangat mudah bagi saya untuk membedakan antara temuan ilmiah dan yang bukan.                | SE4  |
|                                   | When I do my work in science class, I am able to find the important ideas.          | Ketika saya melakukan pekerjaan saya di kelas sains, saya dapat menemukan ide-ide penting.   | SE5  |
|                                   | I can use math to answer scientific questions.                                      | Saya dapat menggunakan matematika untuk menjawab pertanyaan ilmiah.                         | SE6  |
|                                   | I can tell the difference between observations and conclusions in a story.          | Saya dapat membedakan antara observasi dan kesimpulan pada sebuah cerita.                  | SE7  |
|                                   | It is easy for me to make a graph of my data.                                       | Sangat mudah untuk membuat grafik dari data yang saya dapat.                                | SE8  |
|                                   | Everybody has to believe what scientists say.                                      | Semua orang harus percaya apa yang dikatakan ilmuwan.                                       | EB1  |
|                                   | All questions in science have one right answer.                                     | Semua pertanyaan dalam sains hanya mempunyai satu jawaban benar.                            | EB2  |
|                                   | Scientific knowledge is always true.                                                | Pengetahuan ilmiah selalu benar.                                                            | EB3  |
|                                   | In science, you have to believe what the science books say about stuff.             | Dalam sains, kamu harus mempercayai sesuatu yang dikatakan dalam buku sains.                | EB4  |
|                                   | The most important part of doing science is coming up with the right answer.        | Bagian terpenting dalam mengerjakan sains yaitu dengan mendapatkan jawaban yang benar.       | EB5  |
|                                   | Whatever the teacher says in science class is true.                                | Apapun yang guru katakan dalam kelas sains adalah benar.                                    | EB6  |
|                                   | Scientists pretty much know everything about science; there is not much more to know.| Ilmuwan sangat mengetahui mengenai segala sesuatu tentang sains, sehingga tidak ada lagi keraguan. | EB7  |
|                                   | If you read something in a science book, you can be sure it’s true                  | Jika kamu membaca sesuatu di buku sains, kamu dapat meyakini itu benar.                      | EB8  |
|                                   | Once scientists have a result from an experiment that is the only answer            | Begitu para ilmuwan mendapatkan hasil dari sebuah penelitian, itu menjadi satu-satunya jawaban. | EB9  |
|                                   | Scientists always agree about what is true in science.                             | Para ilmuwan selalu setuju tentang apa yang benar dalam sains.                               | EB10 |
|                                   | Only scientists know for sure what is true in science.                              | Hanya para ilmuwan yang tahu pasti apa yang benar dalam sains.                              | EB11 |

Table 1. Original version and Indonesian version of SLA-MB.
3. Findings

As can be seen on Table 2, a dimensionality comparison test between unidimension and three-dimension of the SLA-MB is showed that SLA-MB fit the Rasch model with three dimensions. This is supported by the acquisition of three-dimensional has higher $X^2$ value than unidimension model, which are 976.76 and 904.62, respectively. In addition three-dimensional model also has lower FinalDeviance and lower AIC (13432.49 and 13500.49 respectively) compared to SLA-MB unidimension model which has 13712.81 for Final Deviance and 13770.81 for AIC.

| Model           | $X^2$  | df   | Sig.   | Final Deviance | AIC            |
|-----------------|--------|------|--------|----------------|----------------|
| One Dimension   | 904.62 | 24   | 0.00   | 13712.81       | 13770.81       |
| Three Dimensions| 976.76 | 22   | 0.00   | 13432.49       | 13500.49       |

Substantive aspect of SLA-MB was examined through reliability test, which the results are shown on Table 3, based on Table 3 the reliability values of each dimensions through CTT methods (Cronbach’s alpha) and IRT-Rasch (PV-reliability) are above 0.6, which means that the results obtained from instrument can be accepted.

Indicator Value of Science has a reliability value on CTT and IRT-Rasch 0.691 and 0.735 (respectively), for self-efficacy has value of reliability 0.603 and 0.662, and for the Personal Epistemology dimension has reliability value 0.785 and 0.817. In addition to value the reliability, we examined the substantive aspects to explore the fit of each item with the Rasch model, the results are shown on the value weighted and unweighted MNSQ on Table 4. For dimensions Value of Science, the MNSQ value of each item ranged from 0.87 – 1.24 logit, for the dimensions of self-efficacy MNSQ values ranged from 0.79 – 1.21 logit and for personal epistemology dimension ranged from 0.65 – 1.31 logit. According to the MNSQ value ranged between 0.7-1.4 logit for almost all of the items in each dimension of SLA-MB, it means that the items fit Rasch model.

| Model               | Indicators       | Reliability | Cronbach's Alpha | PV Reliability |
|---------------------|------------------|-------------|------------------|----------------|
| Three-Dimension     | Value of Science | 0.691       | 0.735            |                |
|                     | Self-Efficacy    | 0.603       | 0.662            |                |
|                     | Personal Epistemology | 0.785   | 0.817            |                |

As shown on Table 4, we did not find any DIF gender or gender bias from all of SLA-MB items. The DIF contrast values of each item are still in the value of lower than 0.64. DIF contrast that is ranged the values of science dimension is from 0.03 to 0.27, on the dimensions of self-efficacy is ranged from 0.03 to 0.54 and Personal Epistemology from 0.09 to 0.35.
### Table 4. Item fits for three dimensions and differential item functioning (DIF) gender of SLA-MB Items.

| Dimensions          | Item Code | Estimates | Unweighted MNSQ | Weighted MNSQ | DIF Gender Contrast |
|---------------------|-----------|-----------|-----------------|---------------|---------------------|
| Value of Science    | VS1       | 0.75      | 0.87            | 0.87          | 0.12                |
|                     | VS2       | -0.21     | 1.17            | 1.16          | 0.27                |
|                     | VS3       | -1.08     | 0.94            | 0.97          | 0.03                |
|                     | VS4       | 0.24      | 0.99            | 1.01          | 0.24                |
|                     | VS5       | 0.94      | 1.06            | 1.07          | 0.07                |
|                     | VS6       | -0.64     | 1.16            | 1.24          | 0.12                |
|                     | SE1       | -0.18     | 0.79            | 0.80          | 0.54                |
|                     | SE2       | -0.16     | 0.96            | 0.96          | 0.09                |
|                     | SE3       | 0.22      | 0.80            | 0.80          | 0.21                |
|                     | SE4       | 0.43      | 0.85            | 0.85          | 0.22                |
|                     | SE5       | -0.47     | 1.02            | 1.01          | 0.03                |
|                     | SE6       | 0.19      | 1.21            | 1.20          | 0.35                |
|                     | SE7       | 0.23      | 1.01            | 1.02          | 0.08                |
|                     | SE8       | -0.26     | 1.08            | 1.09          | 0.09                |
|                     | EB1       | -0.04     | 1.05            | 1.03          | 0.15                |
|                     | EB2       | 0.06      | 1.23            | 1.23          | 0.35                |
|                     | EB3       | 0.10      | 0.84            | 0.84          | 0.21                |
|                     | EB4       | 0.16      | 0.66            | 0.65          | 0.09                |
|                     | EB5       | -0.84     | 1.19            | 1.21          | 0.33                |
|                     | EB6       | 0.37      | 0.85            | 0.85          | 0.12                |
|                     | EB7       | -0.03     | 0.93            | 0.92          | 0.12                |
|                     | EB8       | 0.07      | 0.79            | 0.78          | 0.31                |
|                     | EB9       | -0.11     | 1.28            | 1.28          | 0.17                |
|                     | EB10      | -0.24     | 0.93            | 0.93          | 0.17                |
|                     | EB11      | 0.50      | 1.31            | 1.29          | 0.23                |

### 4. Discussions

The validity of an instrument is an overall evaluative assessment of the empirical and theoretical evidence supporting the feasibility and suitability of interpretation and feedback based on the test scores or other parts of the assessment [11]. Therefore, the validity of an instrument is not only as a supporting instrument or assessment but also as a reflection of what we explore through the instruments. A valid instrument will produce a finding, data and other things that we searched through it by reflecting what it should be reflected in and obtained from the instrument. Based on our findings on the construct validity of SLA-MB that have been described in the previous section, in this section we will discuss the extent to which aspects of construct validity are traced from this study might be explained.

First, the structural aspect of construct validity is related to the consistency of response so that from participants that can describe a particular domain or specific dimensions. Based on Table 2, it is shown that the numbers of domains that exist on the SLA-MB based IRT Rasch analysis are three domains, which indeed it is in accordance with the content validity that has been validated by the original developers of SLA-MB. Those three domains or dimensions are the value of science, self-efficacy and personal epistemology. The importance of disclosure the domains or dimension besides indeed reveal the structural form of an assessment or research instruments, with the disclosure of an instrument dimensionsionality will provide clarity regarding what are the variables that traced from research instruments that we used. Basically an instrument or assessment can uncover one or more variables whether it will be knowledge, performance or attitude of the respondents [18]. Why we should use a statistical analysis, in this case IRT-Rasch analysis, and why do not it just divided based on its content? As explained previously that the construct validity not only deal with the theoretical aspect (in this case the content), but it should be dealt with empirical aspect which is reflected from data sample. Thus, the validity of the content and structural in this case are complemented to each other. The implications of the acquisition of an instrument dimensionality is to uncover the boundaries.
of an item in instruments that measure the properties that are supposed to be measured by the instrument which is in the domain or the same dimensions with other items that measure the same variable.

The results of the dimensionality investigation will impact on the quality of every item that exists in every dimension. Based on the results that shown in Table 4 especially on the fit indices or MNSQ of each item, we found that almost all items has MNSQ values are met the benchmark. It shows that every item is fit with Rasch models, in which case each item fit with every dimension they are occupied. But, we found one of the items on the Personal Epistemology dimension that overfit Rasch model, the item is EB4 with MNSQ 0.66. We assume that the item is still fit the Rasch model because the relative value of 0.66 is still being near the cutoff. In addition to the items fit, substantive aspects of an instrument can also be explained by the reliability of the instrument or the internal consistency of the instrument. With sufficient reliability, then a predictable instrument can provide consistent measurement results [18]. Based on the results of reliability test through CTT with Cronbachs’ alpha and reliability based on the Rasch model with the PV-Value, every dimension of SLA-MB possessed reliability values which belong to the instrument which has good internal consistency [20].

The last finding of our study is Differential item Functioning (DIF) gender of SLA-MB. The purpose of analyzing DIF gender is to examine whether one gender group benefited at the time doing on each item in the SLA-MB and other gender groups disadvantaged. Based on the results that shown in TABLE 4, we found that there is no big gender bias from all of the items (DIF contrast <0.64). It reveals that the SLA-MB instrument is not an instrument that is biased or in favor of one gender group, it also supports the generality of SLA-MB. The absence of DIF or bias of the instrument SLA-MB implies that the obtained survey results can perform a comparative study of gender because no significant gender bias in each item. In contrast, if we found gender bias, for example, some items more easily answered by girls than boys, then the findings could not do a comparative analysis of gender, because it is quite predictable that girls might have higher results.

5. Conclusions and limitations

Based on our findings, the exploration of construct validity on SLA-MB instrument using IRT-Rasch, it was found that the SLA-MB fit with the three dimensions Rasch model which consist of the Value of Science dimension, Self-efficacy dimension and the Personal Epistemology dimension, which corresponds to the validity of the content that have been analyzed by SLA-MB developers. Based on substantive aspects, SLA-MB has acceptable value of internal consistency and has MNSQ values that indicate each item of SLA-MB fit with Rasch models. In addition, in the aspect of generalizability, there is no gender bias of each item of SLA-MB, so that each item has the same difficulty level if done well by female students and male students. With such results, we conclude that the SLA-MB Indonesian version can be used as an instrument to explore motivation, values and beliefs of Indonesian students towards science.

Although this study has done to produce and reveal the construct validity of the SLA-MB using IRT-Rasch, but not all aspects of construct validity are revealed. Therefore, for further study more aspects of the disclosure of the construct validity are strongly advised to obtain the better validity and legitimacy of instruments that fulfill the pre-requisites of psychometric and psychology.

6. References

[1] OECD 2006 Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA (OECD Publishing) pp 1-187
[2] Glynn S M, Taasoobshirazi G and Brickman P 2007 Journal of Research in Science Teaching 44 1088-107
[3] OECD 2014 PISA 2012 Results in Focus: What 15-Year-Olds Know and What They can do with What They Know (OECD Publishing) pp 1-39
[4] Indonesian Ministry of Education 2013 Dokumen Kurikulum 2013
[5] Holbrook J and Rannikmae M 2007 International Journal of Science Education 29 1347-62
[6] Fives H, Hubnuer W, Bimbaum A S and Nicolich M 2014 *Science Education* 98 549–80
[7] Glynn S M, Brickman P, Norris A and Taasoobshirazi G 2011 *Journal of Research in Science Teaching* 48 1159-76
[8] Bandura A 2006 *Applied psychology: New Frontiers and Rewarding Careers* (London: Psychology Press) pp 53-79
[9] Druger 2006 Experiential learning in a large introductory biology course *Handbook of College Science Teaching* (Arlington: National Science Teacher Association Press) pp 37-43
[10] Rachmatullah A, Diana S and Rustaman N Y 2016 *International Seminar on Mathematics, Science and Computer Science Education 2015 (Bandung)* (New York: AIP Conference Proceedings) ed Hidayat et al pp 080008-1–080008-5
[11] Messick S 1995 *American Psychologist* 50 741
[12] Loevinger J 1957 *Psychological Report*
[13] Sondergeld T A and Johnson C C 2014 *Science Education* 98 581-613
[14] Bode R K and Wright B D 1999 Rasch measurement in higher education *Higher Education: Handbook of Theory and Research* (Netherlands: Springer Netherlands) pp 287-316
[15] Hambleton R K 2000 *Medical Care* 38 60-5
[16] Bond T and Fox C M 2015 *Applying the Rasch model: Fundamental Measurement in The Human Sciences* (New York: Routledge) pp 1-382
[17] Boone W J, Townsend J S and Staver J 2011 *Science Education* 95 258-80
[18] Neumann I, Neumann K and Nehm R 2010 *Int. J. Science Education* 33 1373-405
[19] Boone W J, Staver J R and Yale M S 2005 *Rasch Analysis in The Human Sciences* (Dordrecht: Springer) pp 69-91
[20] Field A 2013 *Discovering Statistics using IBM SPSS Statistics* (London: Sage Publication)