Testing validity inferences of science motivation questionnaire (SMQ-II) instrument: Rasch-based analysis with Indonesian secondary students

R Q Aini1*, A Rachmatullah2, F Roshayanti3, S Shin4, JK Lee5, MS Ha1

1Department of Science Education, Kangwon National University, Chuncheon, Republic of Korea
2Department of STEM Education, NC State University, North Carolina, USA
3Department of Biology Education, PGRI Semarang University, Semarang, Indonesia
4Division of Biology Education, Chungbuk National University, Cheongju, Republic of Korea
5Division of Science Education, Chonbuk National University, Jeonju, Republic of Korea

*corresponding author: qurota.rahmi@gmail.com

Abstract. Examining science motivation is essential for science learning in the classroom, even more, science motivation is often related to students’ achievement and performance. The test of science motivation has broadly employed and some of the instrument has been translated into a local language where the instrument used. The current study explores the validity evidence of Science Motivation Questionnaire II (SMQ-II) by IRT-Rasch based analysis using Test Analysis Models ‘TAM’ Package v. 3.1–45 in R software from Indonesian context. This study reveals construct validity of SMQ-II instrument based on The Standards for Educational and Psychological Testing using Messick’s (1989) validity theories and construct validity. The data collected from 984 Indonesian middle and high school students. As result, the current study had SMQ-II validity evidence which includes (1) Dimensionality of the instrument using Exploratory Factor Analysis and Rasch-likelihood ratio approach, (2) Expected A Posterior (EAP) reliability and Weighted Maximum Likelihood (WLE) estimation, (3) Item fit and item properties, (4) Wright Map using Rasch modeling. The result of construct analysis shows that the instrument was appropriate with a theory where the construct was categorized into five dimensions, all the items of SMQ-II fit the rasch modelling, and the internal reliability of each dimension is categorized as good. In general, the current study suggests that SMQ-II instrument of Indonesian version is adequate to be implemented in Indonesia.

1. Introduction
Motivation is one of the predictors of students’ attitude toward science in the classroom, even more, it can activate students perceived career in the science field. For example, students who are motivated in science had related to their achievement [1,2] had better performance in science classroom and had willingness to continue in the STEM field [3]. Research on students’ motivation has been widely used, particularly in science classroom context [1,4,5]. For the appropriate measurement, Glynn et al. [6,7] developed Science Motivation Questionnaire to measure students’ motivation toward science. The
instrument provides construct to measure non-observable variables of student’s motivation to learn science. According to the instrument, the construct of students’ motivation to learn science is divided into five dimensions, which are intrinsic motivation, self-efficacy motivation, and self-determination motivation, grade motivation, and career motivation.

In the case of developing and evaluating an assessment, validity is crucial and fundamental considerations. Standard and guideline of measurement instrument were developed from various organization and discipline, including education [8]. In Standards for Educational and Psychological Testing [9], the standards were had guide influenced by Messick’s [10] validity theory. With several validity theories, Messick [10] introduced internal structure validity aspects of validity evidence and point out the account of process modeling in items responds and its theories. Furthermore, the measurement validity study also have construct validity to indicate if the instrument is appropriate to measure some knowledge domain [11]. Widely used methods such as Classical Test Theory (CTT) and Item Response Theory (IRT) are methods to determine the construct validity of the instrument.

Classical Test Theory or CTT emphasize the dependent value for the distribution of samples using the overall raw test score. Even though using CTT is easy to use, the CTT still lack of theoretical assumptions with the testing condition [12]. Rasch modeling is one of the Items Respond Theory as the method for the psychometrics. In general, IRT-Rasch considers the person ‘ability’ and item ‘difficulty’ [13]. Therefore, in psychometrics, IRT-Rasch modeling works by converting the raw score from ordinal scale level into interval level, and use logits (logarithm of odds) as the unit of the measure. By converting the raw data and using the interval level, the analysis is more advance using parametric one. Another strength of IRT-Rasch modeling is predicting the missing data from the probabilistic of the Rasch measurement. The correlation of converted the raw data into logit is close to 1 [12]. The use of IRT-Rasch model for testing the validity, reliability, and the analysis has been widely used for educational assessment related.

The previous study of validity of SMQ instrument is also limited by the evidence based on responses process, particularly, for Indonesian response. IRT-Rasch analysis could explore the construct validity and the internal validity evidence of SMQ-II instrument in Indonesian middle and high school students according to Messick validity theory. Likewise, the current study used R software with Test Analysis Models ‘TAM’ package as a cutting-edge accessible way of exploring validity evidence of psychometric aspects in science education. Using the Indonesian participants, the purpose of the paper is to add and introduce science educator researcher and educator to use IRT-Rasch method and measurement using R TAM package, as well as providing the evidence of SMQ-II Indonesian version validity.

2. Methods
Total of 984 data was collected from high school level in 10th grade (n=588) and middle school level in 8th grade (n=365). 44% of the participants were male, and 46% of the population were female. 25 items of SMQ-II [7] were used using 5 Likert-scale. The instrument were translated into Indonesian. Original version and the translated version are shown in Table 1.

| Table 1. Original version and Indonesian version of SMQ-II |
|----------------------------------|----------------------------------|----------------------------------|
| Dimensions                      | Item name                        | Indonesian Version               |
|---------------------------------|----------------------------------|----------------------------------|
| Intrinsic motivation            | Intrinsic motivation 1           | For me learning science is interesting |
|                                 | Intrinsic motivation 2           | I want to know about scientific discoveries |
|                                 | Intrinsic motivation 3           | The science that I learned is very relevant / supports my life |
|                                 | Intrinsic motivation 4           | Learning science makes my life meaningful |
|                                 | Intrinsic motivation 5           | I enjoy learning science         |
| Career motivation               | Career motivation 1              | Studying science will help me get a high job |
|                                 | Career motivation 2              | Knowing science will give me career benefits |
|                                 | Career motivation 3              | Understanding science will be beneficial and beneficial in |
Career motivation

Career motivation 4
I will use my natural abilities to solve problems in my career

Career motivation 5
My career will involve science

Self-determination

Self-determination 1
I studied science seriously

Self-determination 2
I prepared myself well for the Sciences test and laboratory performance tests

Self-determination 3
I worked hard to study science

Self-determination 4
I spend a lot of time studying science

Self-determination 5
I used many strategies while studying science

Self-efficacy

Self-efficacy 1
I believe I can get an "A" / "100" in science lessons

Self-efficacy 2
I'm sure I can do the science test well

Self-efficacy 3
I believe that I can master science and scientific skills in it

Self-efficacy 4
I'm sure I can understand science

Self-efficacy 5
I'm sure I can work on projects and science practicums well

Grade motivation

Grade motivation 1
The thing that means in my opinion is getting a high score during the test and the Science performance test

Grade motivation 2
It is important for me to get an "A" or high in science

Grade motivation 3
I thought about the value that I would get in science lessons

Grade motivation 4
Getting science lessons in class is important to me

Grade motivation 5
I want to get higher science exam results than my other friends

The current study examined the construct of validity and validity evidence of the instrument using IRT-Rasch modelling. The construct validity was examined by Rasch model fit, Rasch dimensionality analysis, item fit, and person reliability [11]. Internal structure validity evidence to see whether the instrument was measure several traits based on the construct items was examined by Rasch dimensionality test, person and item reliability [11]. Item properties to measure students’ item performance characteristics were examined by item measures, item fit statistics, and wright map [11]. The current study used Principal Component Analysis (PCA) by and Factor Analysis (FA) by data reduction to measure the latent variable. In IRT-Rasch analysis will estimate and produce of the parameter of person ‘ability’ and item ‘difficulty’. The current analysis also use the “item” as a constraint when producing the ‘item’ difficulty and ‘person’ ability. The constraint used as the anchor as with average 0. When the model parameters are produced, the validation of the model also needs to check if the data fit with the IRT-rasch model [13]. Furthermore, to check whether the items of instrument appropriate and efficient, the plot of Wright Map is also examined to see if the difficulties of the items align with the person abilities. R package TAM v 3.1-45 used and provide all the information in this current study.

3. Result and Discussion

In this study, two kind dimensionality test approach were used to determine the construct of the instrument. Table 2 shows the exploratory factor analysis using promax rotation to examine the relations between observed and latent variable. This approach also used to examine students answer of Indonesian version questionnaire. The estimates of factor analysis using ‘TAM’ R package with marginal maximum likelihood estimation. Five value was used to indicate factors, as a result, five dimension of the instrument were categorized each item of the same construct in one dimension. Table 1 showed that in each item in the construct are assembled in the same dimensions, which in each item in the intrinsic construct in dimension 1, items grade in dimension 2, items career in dimension 3, self-
determine in dimension 4, and self-efficacy in dimension 5. According to Glynn et al. [7] both construct career and grade are fall under external motivation, which according to Table 1, dimension 2, not only grade construct that assembled but also 3 items in career construct also applies to dimension 3. This shows that there are still similar ones between construct career and grade.

**Table 2. Exploratory Factor Analysis result (promax rotation)**

|                     | Dim01 | Dim02 | Dim03 | Dim04 | Dim05 |
|---------------------|-------|-------|-------|-------|-------|
| Intrinsic motivation 1 | 0.802 |       |       |       |       |
| Intrinsic motivation 2 | 0.43  |       |       |       |       |
| Intrinsic motivation 3 | 0.458 | -0.146|       |       |       |
| Intrinsic motivation 4 | 0.634 |       |       |       |       |
| Intrinsic motivation 5 | 0.794 |       |       |       |       |
| Career motivation 1   |       | 0.154 | 0.625 |       |       |
| Career motivation 2   |       | 0.92  |       |       |       |
| Career motivation 3   |       | 0.958 |       |       |       |
| Career motivation 4   | -0.173| 0.472 | 0.137 |       |       |
| Career motivation 5   | -0.146| 0.653 |       |       |       |
| Self-determination 1  | 0.122 |       |       | 0.345 | 0.13 |
| Self-determination 2  |       |       |       | 0.621 |       |
| Self-determination 3  |       |       |       | 0.745 |       |
| Self-determination 4  |       | -0.112|       | 0.709 | -0.103|
| Self-determination 5  | 0.125 |       |       | 0.438 | 0.126|
| Self-efficacy 1       |       |       |       |       | 0.839 |
| Self-efficacy 2       |       |       |       |       | 0.742 |
| Self-efficacy 3       |       |       |       |       | 0.829 |
| Self-efficacy 4       | 0.106 |       |       |       | 0.586 |
| Self-efficacy 5       |       |       |       |       | 0.664 |
| Grade motivation 1    |       | 0.61  |       | 0.12  |       |
| Grade motivation 2    |       | 0.919 |       |       |       |
| Grade motivation 3    |       | 0.458 | 0.104 | 0.151 |       |
| Grade motivation 4    |       | 0.627 |       |       |       |
| Grade motivation 5    |       | 0.412 |       |       | 0.13  |

The second approach is dimensionality test using Rasch likelihood approach. As shown in table 3, the result also supports that multi-dimension model is fit according to the theory than unidimensional model. According to the result, 5-dimension has lower AIC and Deviance (53361.65 and 53061.65, respectively) compared to 1-dimension model (56713.61 and 56461.61, respectively).

**Table 3. Item dimensionality test results using a Rasch analysis (TAM)**

| Model     | loglike  | Deviance | Npars | AIC      | BIC     |
|-----------|----------|----------|-------|----------|---------|
| 1         | 1-dimension | -28230.8 | 56461.61 | 126 | 56713.61 | 57326.06 |
| 2         | 5-dimension | -26530.8 | 53061.65 | 150 | 53361.65 | 54090.75 |

After the investigation of the construct dimensionality, the quality of every item was processed in every dimension to measure the aspect of substantive from the validity. The amount of model fit of the response was analyzed the Item Fit by the range of Infit and Outfit result. In TAM ‘R’, the item fit statistics are generated by the weighted mean square (MNSQ). According to Wright and Linacre (1994), the range of MNSQ were between 0.50 until 1.50. As a result, all the items of SMQ-II show the fit indices with the benchmark. Furthermore, the internal consistency of the instrument was examined based on reliability. Table also 4 show the result of item reliability using the result of IRT-Rasch Expected A Posteriori (EAP) reliability as item reliability, WLE as person reliability, compared
with CTT Cronbach's Alpha of each dimension. All the coefficient of the IRT-rasch model (Table reliability are above 0.70, which show that the instrument are accepted. According to table 5, lowest item difficulty were grade_5, career_1, intrinsic_2, determine_1, efficacy_4 (-0.417, -0.365, -0.238, -0.226, -0.154, respectively) and the highest item difficulty were determine_4, grade_1, determine_2, career_5, career_4 (0.179, 0.179, 0.195, 0.287, 0.312, respectively).

Table 4. Psychometrics properties of Items

| Items                      | N   | M      | Item Difficulty | Infit | Outfit | Cronbach's Alpha | EAP Reliability | WLE Reliability |
|----------------------------|-----|--------|-----------------|-------|--------|------------------|-----------------|-----------------|
| Intrinsic motivation 1     | 954 | 3.726  | -0.046          | 0.889 | 0.877  |                  |                 |                 |
| Intrinsic motivation 2     | 954 | 3.958  | -0.238          | 1.273 | 1.239  |                  |                 |                 |
| Intrinsic motivation 3     | 954 | 3.517  | 0.119           | 1.041 | 1.042  | 0.784            | 0.818           |                 |
| Intrinsic motivation 4     | 954 | 3.547  | 0.133           | 0.921 | 0.91   |                  |                 |                 |
| Intrinsic motivation 5     | 954 | 3.582  | 0.031           | 0.825 | 0.819  |                  |                 |                 |
| Career motivation 1        | 954 | 3.913  | -0.365          | 1.068 | 1.035  |                  |                 |                 |
| Career motivation 2        | 954 | 3.779  | -0.106          | 0.861 | 0.835  |                  |                 |                 |
| Career motivation 3        | 954 | 3.768  | -0.128          | 0.727 | 0.706  | 0.871            | 0.869           |                 |
| Career motivation 4        | 954 | 3.464  | 0.312           | 1.071 | 1.088  |                  |                 |                 |
| Career motivation 5        | 954 | 3.431  | 0.287           | 1.023 | 1.03   |                  |                 |                 |
| Self-determination 1       | 954 | 3.749  | -0.226          | 0.993 | 0.994  |                  |                 |                 |
| Self-determination 2       | 954 | 3.436  | 0.195           | 0.932 | 0.929  |                  |                 |                 |
| Self-determination 3       | 954 | 3.607  | -0.068          | 1.03  | 1.013  | 0.786            | 0.839           |                 |
| Self-determination 4       | 954 | 2.956  | 0.179           | 1.11  | 1.12   |                  |                 |                 |
| Self-determination 5       | 954 | 3.465  | -0.08           | 0.813 | 0.818  |                  |                 |                 |
| Self-efficacy 1            | 954 | 3.701  | 0.039           | 0.905 | 0.888  |                  |                 |                 |
| Self-efficacy 2            | 954 | 3.694  | 0.085           | 0.915 | 0.923  |                  |                 |                 |
| Self-efficacy 3            | 954 | 3.684  | 0.015           | 0.719 | 0.72   | 0.861            | 0.867           | 0.816           |
| Self-efficacy 4            | 954 | 3.877  | -0.154          | 0.764 | 0.771  |                  |                 |                 |
| Self-efficacy 5            | 954 | 3.618  | 0.015           | 0.831 | 0.836  |                  |                 |                 |
| Grade                      | 953 | 3.843  | 0.179           | 1.032 | 1.026  | 0.816            | 0.830           |                 |
motivation 1
Grade motivation 2 954 4.037 0.002 0.963 0.923
Grade motivation 3 954 3.879 0.163 1.14 1.149
Grade motivation 4 954 3.969 0.074 1.069 1.031
Grade motivation 5 954 4.294 -0.417 1.277 1.204

The further relation between the item difficulty of each item and person ‘ability’ also showed in the Figure 1. Figure 1a evince the overall plot histograms of person ability and item parameters in the same graph and scale. Figure 2b showed the distribution of item difficulty and person ability in each dimension. The result in figure 1b (left) showed that each dimension had a different distribution of difficulty level. Figure 1a (right) showed that in each item had different scale value in each category. According to the result, both figures show that students tend to assume that this instrument is easy to fill with the tendency to respond with a high scale. However, in such of instrument to measuring latent traits in students, it should be noted that there will also be a response bias in which students tend to agree to the statement on the questionnaire. The advantages of the resulting wright map can measure the real value of each item. For example, the value on scale (or category5) on grade_5 items (has the lowest item difficulty value, -0.417) has a category value lower than the item value career_4 (0.312) (Figure 2)

![Figure 1](image1.png)

**Figure 1.** a) Histogram of SMQ II IRT-Rasch parameter. The item difficulty run as five dimensional model and the person ability runned as unidimensional using Weighted Maximum Likelihood Estimation (WLE). Both use the same “items difficulty” as the constraint of the modelling b) WrightMap of SMQ run as five dimensional model using Weighted Maximum Likelihood Estimation (WLE) parameter.

4. Conclusion
Based on the result of this study, it can be concluded that SMQ-II Indonesian version is appropriate to use based on the validity test using IRT-Rasch analysis with the large sample of participants (n>700). Using two approach of dimensionality test, the structural validity of this instrument showed that SMQ-II consist of five dimension in accordance with the theory which are Intrinsic motivation, Career motivation, Self-determination motivation, Self-efficacy motivation, and Grade motivation. The fit items in the instrument were also fit with the rash modelling. Furthermore, item and person reliability were also categorized as good internal consistency. Provided with the Wright Map, the characteristics
of each items in the current instrument can be measured. With the result, the further study of science motivation can be measured using the current translated SMQ-II instrument to measure Indonesian students.

5. References

[1] Bryan RR Glynn SM and Kittleson JM 2011 Motivation, achievement, and advanced placement intent of high school students learning science Science education 95 6 pp.1049-1065.

[2] Singh K Granville M and Dika S 2002 Mathematics and science achievement: Effects of motivation, interest, and academic engagement The journal of educational research 95 6 pp.323-332.

[3] Robnett RD and Leaper C 2013 Friendship groups, personal motivation, and gender in relation to high school students' STEM career interest Journal of Research on Adolescence 23 4 pp.652-664.

[4] Vedd-Weiss D and Fortus D 2012 Adolescents' declining motivation to learn science: A follow-up study Journal of Research in Science Teaching 49 9 pp.1057-1095.

[5] Sanfeliz M and Stalzer M 2003 Science motivation in the multicultural classroom The Science Teacher 70 3 p.64.

[6] Glynn, S.M., Taasoobshirazi, G. and Brickman, P., 2007. Nonscience majors learning science: A theoretical model of motivation. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 44(8), pp.1088-1107.

[7] Glynn SM Brickman P Armstrong N and Taasoobshirazi G 2011 Science motivation questionnaire II: Validation with science majors and nonscience majors Journal of research in science teaching 48 10 pp.1159-1176.

[8] Zambo BD and Chan EK eds 2014 Validity and validation in social, behavioral, and health sciences New York (US): Springer International Publishing.

[9] American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (US) 1999. Standards for educational and psychological testing. Amer Educational Research Assn.

[10] Messick S 1989 Meaning and values in test validation: The science and ethics of assessment Educational researcher 18 2 pp.5-11.

[11] Tornabene RE Lavington E and Nehm RH 2018 Testing validity inferences for genetic drift inventory scores using Rasch modeling and item order analyses Evolution: Education and Outreach 11 1 p.6.

[12] Adedoyin OO Nenty HJ and Chilisa B 2008 Investigating the invariance of item difficulty parameter estimates based on CTT and IRT Educational Research and Reviews 3 3 pp.083-093.

[13] Wu M and Adams R 2007 Applying the Rasch model to psycho-social measurement: A practical approach (Melbourne: Educational Measurement Solutions)