World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Assessment of students' science literacy skills using diagnostic items on the concept of energy with Rasch modeling

Syarifa A Alatas *, Mohamad Jahja and Muhammad Yusuf

Gorontalo State, University Postgraduate Program, Indonesia.

World Journal of Advanced Research and Reviews, 2024, 21(01), 1356–1362

Publication history: Received on 28 November 2023; revised on 06 January 2024; accepted on 08 January 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.21.1.2697

Abstract

This study aims to describe the quality of diagnostic items of the concept of Energy developed based on Rasch modeling and describe students' science literacy skills on the concept of Energy. The research method used in this research is descriptive. Research data were collected through online and offline tests and analyzed using the Rasch model and diagnostic items. The results showed that (1) The science literacy test instrument made was valid with an excellent category and a reliability value of 0.97, meaning that the quality of the items in the instrument was very good. This instrument can be used to assess students' abilities at all levels of ability, from those with low to high abilities; (2) For Person reliability or the quality of students' answers, the reliability value is below 0.30, meaning that the consistency of answers from students is low; (3) The quality of item questions based on the Rasch model from the item measure statistics results that all items meet the MNSA outfit criteria and no negative PTMEA Corr occurs. That is, all items are non-deviant, appropriate, and valid. This is also confirmed based on the bubble chat, where there are no items that are overfit or underfit; (4) The Rasch model makes very thorough and good measurements to measure the difficulty level of items and person's ability; (5) Based on the measurement results, it was found that the science literacy ability of students at SMPN 1 Paleleh was divided into low ability as much as 1.6%, moderate 53.6%, high 43.8%, very high 0.82%.

Keywords: Diagnostic Items; Science Literacy; Rasch Modeling; Descriptive Research

1. Introduction

Indonesia ranked lowest among 41 countries in the 2018 *Program for International Student Assessment* (PISA) report in terms of student achievement in reading, mathematics, and science. The OECD (2019) PISA test scores for Indonesian students in these three areas are shown in Figure 1. It appears that these scores have not changed significantly over the past 20 years. These scores are averages of scores from different regions across Indonesia, which have disparities in several factors, such as gender, regional/school origin, and socioeconomics. PISA data does not mention the origin of the region or school, so efforts to improve science literacy must be preceded by regional evaluation efforts covering a large enough area (several districts/cities).

Some research on evaluating science literacy skills was conducted by Laliyo, *et al*. (2020) on a small scale (within one school) and by Soeharto, *et al*. (2021) in several secondary schools using various kinds of questions developed from PISA questions and other questions. Generally, the analysis method used is still unable to reveal any differences in scores in terms of various factors. Sumintono (2014) said that the Rasch Model could map these differences to the person and item level.

Students' low level of science literacy is one of the educational problems in Indonesia. Based on observations in the field and experiences about science learning that have been carried out at SMP Negeri 1 Paleleh, Buol Regency, it was found

* Corresponding author: Syarifa A Alatas

that science (Physics) teachers have a lack of understanding in science knowledge and problem recognition. This certainly makes it difficult for students to draw conclusions in learning activities. In addition, the learning model applied at school is still teacher-centered when learning takes place.

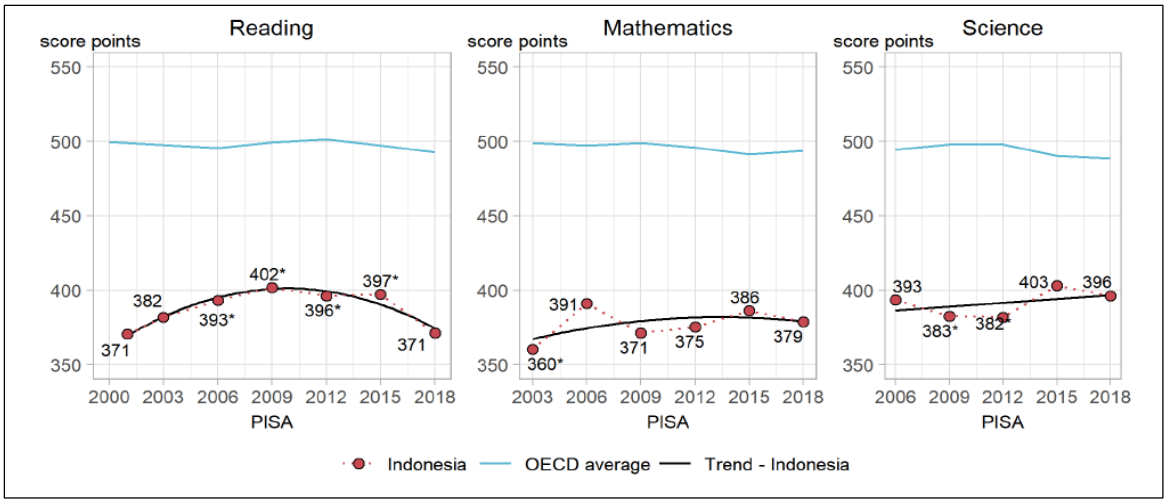

Figure 1 Indonesia's PISA Reading, Mathematics, and Sciences Scores for the Period 2000-2018

One of the science materials that is difficult for students to understand is the concept of Energy. The difficulty of understanding the concept of energy occurs because, in the learning process, the relationship between the concepts in it is less mastered by students so it is still difficult for them to understand it. By understanding the concept of energy, learners can make better decisions about the use of energy resources and reduce their negative impact on the environment. Therefore, it is important to evaluate learners' science literacy skills in understanding the concept of energy. Measurement of learners' science literacy skills can be done using the Rasch model and diagnostic items. These two models can provide more precise and objective measurements compared to other measurement methods.

The Rasch model is a measurement model used to measure individual ability based on item difficulty. The Rasch model assumes that individual ability and item difficulty can be measured on the same dimension and individual ability can be measured continuously (Andrich, 2010). According to Aydin, *et al.* (2013), the Rasch model can be used to measure students' science literacy skills. This model allows a more accurate assessment because it takes the difficulty value of each question and the ability of students to answer questions. This is emphasized by Soeharto (2021), that Rasch measurement is based on the interaction between items and individuals described by mathematical equations. People who have high abilities should answer questions with easier difficulty levels.

According to Laliyo, et al. (2022), analysis using the Rasch model is a very useful tool for evaluating the validity and reliability of measurement instruments used in educational research. This is also supported by Soeharto, *et al.* (2019) who stated that in fact, the Rasch model proved to have significant advantages in understanding and measuring test characteristics, both for the purpose of classroom assessment and other research. This means that the Rasch model has proven to be very useful in understanding and measuring test characteristics, both for classroom assessment and research purposes. Soeharto and Csapo (2022) said that the use of the Rasch model in a study results in measurements that can be analyzed in detail, so as to provide specific information about the characteristics of individuals and items in the measurement. This is very useful in the development of assessment instruments. So, by using this model, the science literacy ability of students can be known more accurately and can be used as a basis for developing appropriate learning programs.

Diagnostic items are a type of test item designed to identify students' misunderstanding of certain concepts or skills. Diagnostic items can provide more detailed information about students' level of understanding and assist teachers in designing effective teaching strategies (Misbahuddin, 2014). According to Herman and Roychoudhury (2015), diagnostic items can be used to evaluate learners' understanding of difficult science concepts.

Based on this description, the researcher conducted a study with the title "Evaluation of the Science Literacy Ability of Learners Using the Rasch Model and Diagnostic Items on the Concept of Energy." The purpose of this study is to describe the quality of diagnostic items on the concept of Energy developed based on Rasch modeling and describe the science literacy skills of students on the concept of Energy.

2. Methods

The research method used in this research is descriptive. The stages of this research are divided into six stages, namely the instrument design stage, the diagnostic item design stage, the assessment rubric preparation stage, the measurement model design stage, the assessment and validation stage, and the data analysis stage. Data was collected through online and offline tests. The data from the test results of students' science literacy skills were then analyzed using the Rasch model and diagnostic items.

3. Results

3.1. Results of Data Analysis Using the Rasch Model

3.1.1. Effectiveness of measuring instruments

The first step to elaborate on the effectiveness of measuring instruments is to measure person and item reliability. This is done to collect information on the extent to which measurements produce consistent information in displaying latent properties or one-dimensionality of the measured variables (Sumintono & Widhiarso, 2015). The results of the analysis are presented in the form of summary statistics in Table 1.

Table 1 Summary Statistic (N=244)

Person	244 INPUT	244 MEASURED			INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	5.2	12.0	-.37	.71	1.00	.0	1.01	.0
P.SD	1.9	.0	.86	.09	.27	1.0	.49	1.0
REAL RMSE	.72	TRUE SD	.47	SEPARATION	.66	Person RELIABILITY		.30
Item	12 INPUT	12 MEASURED			INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	105.3	244.0	.00	.15	1.00	.0	1.01	-.1
P.SD	44.3	.0	.93	.01	.05	1.1	.12	1.2
REAL RMSE	.15	TRUE SD	.92	SEPARATION	6.05	Item RELIABILITY		.97

The results in Table 1 show that the person's reliability value is 0.30 with a separation value of 0.66. This means that the *person's* response to items is still very weak or the student's ability to answer questions is inconsistent. In addition, the resulting Cronbach Alpha Coefficient (KR-20) value is 0.32 which indicates poor interaction between students and the test. However, the test items or instruments used produced a relatively high item separation index value of 6.05 which is equivalent to an item reliability value of 0.97. This indicates that the item's consistency is very good, or the item is considered capable of meeting the criteria of one-dimensionality. In other words, the item has an excellent performance in defining the measured variable. This is confirmed by the *outfit* value results, where the items are within the acceptable range for multiple-choice tests (Bond & Fox, 2015; Herrmann-Abell & DeBoer, 2011).

3.1.2. Instrument validity based on Rasch measurement

Figure 2 displays the measurement information graph to show the measurement reliability. The higher the end of the information function graph, the measurement reliability value tends to increase. At the intermediate level (-0.81 logits to +0.81 logits), the measurement information is very high.

Figure 2 shows that students with science literacy ability occupy the largest population. This indicates that the science literacy instrument can optimally generate information for students with moderate ability levels. These results mean that the instrument has high measurement reliability (Bond & Fox, 2015; Kim & Wilson, 2019).

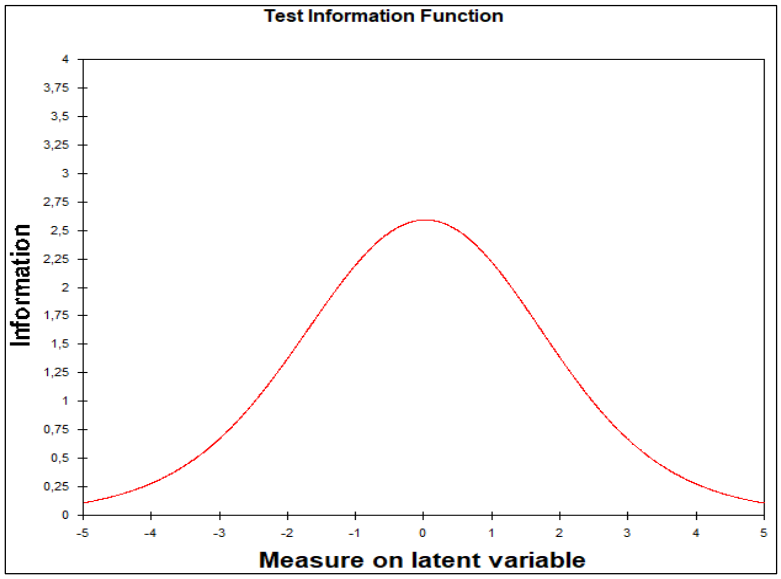

Figure 2 Test Information and Student Ability

3.1.3. The quality of the developed diagnostic items on the concept of Energy based on Rasch modeling

To determine the quality of items based on Rasch modeling, it is seen in the item statistics table, that what needs to be considered is the MNSQ and ZSTD infit values, MNSQ and ZSTD outfit, and PTMEA Corr. The results of the item suitability analysis are shown in Table 2 as follows:

Table 2 Item Statistics

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	143	244	-.76	.14	1.02	.44	.98	-.22	.36	.37	63.5	67.5	S1
2	158	244	-1.06	.14	1.04	.73	1.03	.35	.32	.37	68.0	70.2	S2
3	138	244	-.66	.14	1.02	.45	1.05	.74	.34	.37	65.6	66.9	S3
4	80	244	.47	.15	1.04	.74	1.01	.17	.30	.34	66.8	70.2	S4
5	58	244	.97	.16	.98	-.21	.91	-.66	.34	.31	77.5	76.8	S5
6	93	244	.20	.14	1.04	.75	1.06	.85	.31	.35	60.7	67.7	S6
7	188	244	-1.75	.16	1.00	.00	.97	-.21	.34	.34	77.9	78.2	S7
8	50	244	1.19	.17	.98	-.21	.88	-.77	.33	.29	80.7	79.8	S8
9	117	244	-.26	.14	.88	-2.64	.85	-2.60	.50	.37	69.7	65.5	S9
10	41	244	1.45	.18	1.02	.24	1.32	1.74	.21	.27	84.0	83.3	S10
11	122	244	-.35	.14	.91	-2.07	.90	-1.62	.47	.37	69.3	65.3	S11
12	76	244	.56	.15	1.08	1.26	1.14	1.38	.24	.34	67.6	71.3	S12
MEAN	105.3	244.0	.00	.15	1.00	-.04	1.01	-.07			70.9	71.9	
P.SD	44.3	.0	.93	.01	.05	1.12	.12	1.18			7.0	5.8	

From the item statistics displayed in Table 2, it resulted that all items met the MNSA *Outfit* criteria and no negative PTMEA Corr occurred. This means that all items are non-deviant, appropriate, and valid, which is excellent item quality. Although some items did not meet any of the criteria, this did not necessarily lower the quality of the items. For example, items with infit and outfit Zstd values do not meet the standard, namely question number 9.

3.1.4. ICC Model

Good item quality can also be seen through the ICC graph. The graph in Figure 3 displays students' abilities based on the easiest to the most difficult questions.

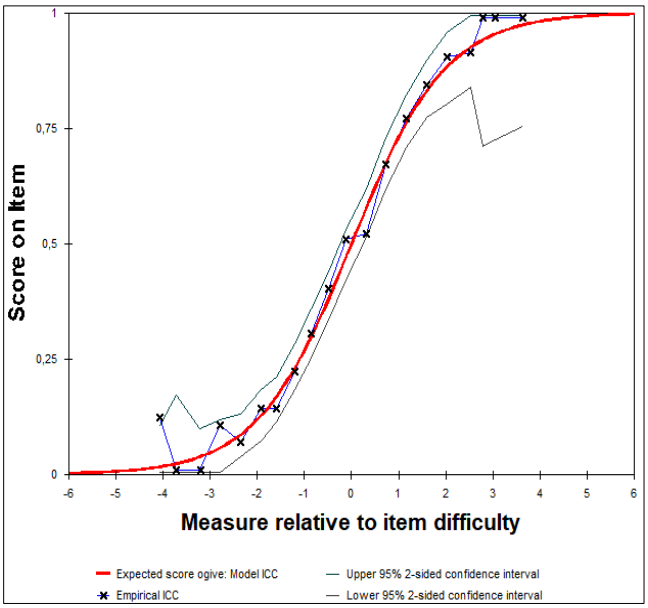

Figure 3 Respondents' Answer Patterns

Based on Figure 3, it is explained that the pattern of respondents' answers is close to the Rasch model pattern, which means that the instrument developed is in accordance with the Rasch model. This means that students with low ability will not answer questions with high difficulty and students with high ability are able to answer questions with high difficulty.

3.2. Science Literacy Skills of Students on the Concept of Energy

The results from Table 3 show that 2 (0.82%) females and 0 (0%) males are classified as having very high ability, 79 (32.3%) females are classified as having high ability and 28 (77.8%) males have high ability. In addition, 77 (31.5%) females and 54 (22.13%) males were classified as having medium ability, and 1 (0.4%) female and 3 (1.2%) males were classified as having low ability.

Table 3 Students' Science Literacy Ability Based on *Logit Value*

Demographics	Very high, LVP > Mean Logit + 2SD	High, Mean Logit + 2SD ≤ LVP < Mean Logit	Moderate, Mean Logit ≥ LVP > Mean Logit - 2SD	Low, LVP < Mean Logit - 2SD	Total
Gender					
Female	2	79	77	1	159
Male	0	28	54	3	84
Total	2	107	131	4	244

Based on Table 3, it is known that students' science literacy skills are divided into 4 categories, which are low, medium, high, and very high. Students with low ability were 1.6%, moderate 53.6%, high 43.8%, and very high 0.82%.

4. Discussion

The developed multiple-choice diagnostic test shows preliminary results that the test is valid and reliable to identify students' science literacy skills on the concept of Energy for grades VII, VIII, and IX in Junior High School (SMP). Item analysis showed that all items can cover students' abilities from low to high ability. However, the stable value for item *misfit* depends on the sample size (Boone et al., 2013; Khine, 2020).

Testing the effectiveness of the instrument used begins with testing the reliability and validity of the instrument. The result is that the items used are reliable with **excellent** categories based on Rasch calculations, but on the other hand, the *person* or respondent under study is not reliable or the consistency in answering questions is low, so it can be concluded that the person's ability still does not meet the reliability requirements. To measure the validity of the instrument, the information test graph is used (Figure 2). Based on the graph displayed the graphical pattern of the test item is in accordance with Rasch modeling, which is also reinforced based on the results of the *Wright Map* person and item. In the *Wright Map*, items and persons are divided into several categories, ranging from persons who answered all the answers correctly, persons with moderate ability, and persons who answered all the answers incorrectly. As for items, there are items that are difficult for students to answer, items with moderate difficulty categories, and items with easy categories.

The development of science literacy test results in 21 schools with a total of 1538 respondents who became the target of the researchers found that science literacy skills in the knowledge and competency domain of students were divided into 4 categories of low, medium, high, and very high. Students with low ability were 1.6%, medium ability was 53.6%, high ability was 43.8%, and very high ability was 0.82%. The findings in Table 3 in detail in this study show that, based on gender, there are more high-ability students in female gender than male. However, in general, students' abilities are still at a low level, as seen from the weak consistency of students' answers. It is necessary to improve students' science literacy skills, especially in the material of the Law of Conservation of Energy and Potential Energy.

This result is in line with Sutrisna's research (2021) which states that students' science literacy level is still low (below 37%). In addition, it is also in line with research by Wati, et al. (2019) which states that the results of the analysis are based on the Rasch model with a *Mean Square Statistics* (MNSQ) value of 0.95, Z *Standard (*ZSTD) is 0.0, *Pearson reliability is 0.*77, and individual raw score test *reliability CRONBACH ALPHA* (KR-20) is 0.86. A lower logit indicates students have lower abilities. This proves that students' problem-solving skills are very poor. Students' problem-solving ability needs to be improved through classroom learning.

Diagnostic tests are instruments used to find learners' learning difficulties. Each test is made to determine one or more learners' disabilities. Teachers' strategies in diagnosing learners also vary depending on the learning difficulty. Therefore, teachers should understand how to start a lesson and the knowledge that should be emphasized. If this is not done, learners' weaknesses will not be discovered and learning programs cannot be created. In this study, the test used is a multiple-choice test of 12 items, where the diagnostic test has an excellent category based on Rasch measurement. The diagnostic items were divided into 4 concepts of energy, namely Kinetic Energy, Mechanical Energy, Potential Energy, and the Law of Conservation of Energy, and distributed based on literacy domains, namely science knowledge and science competence. The results showed that the items can measure students' abilities from low to high.

Improving science literacy in the concept of Energy in targeted areas needs to be done through efforts to improve literacy skills in classrooms. One simple way that can be done is by reading before learning is done. In addition, efforts to improve teacher competence are also an evaluation of the low science literacy skills of students in some of these target schools.

5. Conclusion

Based on the description of the research results obtained, the following conclusions can be drawn:

- The science literacy test instrument is valid with an excellent category and a reliability value of 0.97, meaning that the quality of the items in the instrument is very good. This instrument can be used to assess students' abilities at all levels of ability, from low to high ability.
- For Person reliability or the quality of students' answers, the reliability value is below 0.30, meaning that the consistency of answers from students is low.
- The quality of question items based on the Rasch model of item measure statistics resulted that all items met the MNSA outfit criteria and no negative PTMEA Corr occurred. This means that all items are non-deviant, appropriate, and valid. This was also confirmed based on the bubble chat, where no items were overfit or underfit.
- The Rasch model does a very thorough and good job of measuring both item difficulty and Person ability.
- Based on the measurement results, it was found that the science literacy ability of students at SMPN 1 Paleleh was divided into low ability as much as 1.6%, moderate 53.6%, high 43.8%, and very high 0.82%.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Andrich, D. 2010. Rasch models for measurement: Invariance, sufficiency, and optimality. Routledge.

[2] Aydin, A., Kukul, V., Baykul, Y. 2013. An Investigation Of The Psychometric Properties Of The Rasch Rating Scale Model Under The Polytomous Scoring. Educational Sciences: Theory & Practice, 13(3), 1533-1540. DOI: 10.12738/estp.2013.3.1599

[3] Bond, T. G., & Fox, C. M. 2015. Applying the Rasch model: Fundamental measurement in the human sciences. 3rd ed. New York and London: Routledge Taylor & Francis Group.

[4] Herrmann-Abell, C. F., & Deboer, G. E. 2016. Using Rasch modeling and option probability curves to diagnose students' misconceptions. In Paper presented at the 2016 American Educational Research Association Annual Meeting, Washington, DC (pp. 1-12). Retrieved from https://files.eric.ed.gov/fulltext/ED572821.pdf

[5] Herman, B. C., & Roychoudhury, A. 2015. Exploring The Utility Of The Rasch Model For Analyzing Multiple-Choice Questions In Higher Education. Journal of Applied Measurement, 16(3), 334-345. DOI: 10.1016/j.jam.2015.06.003

[6] Laliyo, L. A. R., Sumintono, B., Panigoro, C. 2022. Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model. Heliyon, 8(3). https://doi.org/10.1016/j.heliyon.2022.e09126

[7] Misbahuddin, M. 2014. Pengembangan Tes Diagnostik dalam Pembelajaran Biologi. Jurnal Pendidikan Biologi Indonesia, 1(1), 1-12.

[8] OECD. 2019. Country Note Programme for International Student Assessment (PISA) Result Combined Executive Summaries Volume I, II & III. Paris: OECD Publishing.

[9] OECD. 2019. What Students Know and Can Do. Vol. I, PISA 2009 at a Glance.

[10] Soeharto, S. & Csapó, B. 2021. Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. Heliyon, 7(11).

[11] Soeharto, S. 2021. Development of A Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach. Journal of Turkish Science Education, 18(3), 351–370.

[12] Sumintono, B. & Widhiarso, W. 2014. Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi). Bandung: Trim Komunikata Publishing House

[13] Sumintono, B., & Widhiarso, W. 2015. Aplikasi pemodelan rasch pada assessment pendidikan. Trim Komunikata. Bandung: Trim Komunikata Publishing House

[14] Sutrisna, N. 2021. Analisis Kemampuan Literasi Sains Peserta Didik SMA DI Kota Sungai Penuh. Inovasi Penelitian, 1(12), 2683.

[15] Wati, M., Mahtari, S., Hartini, S., Amalia, H., 2019. A Rasch Model Analysis on Junior High School Students' Scientific Reasoning Ability. iJIM, 13(7). doi: https://doi.org/10.3991/ijim.v13i07.10760