Assessing competency levels of students on learning of energy and weather concepts using partial credit model

IOP Conf. Series: Journal of Physics: Conf. Series **1318** (2019) 012086 doi:10.1088/1742-6596/1318/1/012086

Assessing competency levels of students on learning of energy and weather concepts using partial credit model

E Murniarti[1], S Solihatun[2] and I B Rangka[2, *]

[1] Universitas Kristen Indonesia, Jl. Mayjen Sutoyo No.2, RT.5/RW.11, Cawang, Kramatjati, Jakarta Timur, Indonesia
[2] Universitas Indraprasta PGRI, Jl. Nangka No. 58 C (TB. Simatupang), Kel. Tanjung Barat, Kec. Jagakarsa, Jakarta, Indonesia

*itsarbolo.rangka@unindra.ac.id

Abstract. This research described a competence levels of students on learning natural sciences in elementary school. Specifically, we measure student's capacity in understanding of energy and weather concepts. This research involved 64 students on elementary school. The data takes from class promotion test with consisting multiple choice questions (MCQ) and Essay Test. Data analysis was produced using the Partial Credit Model (PCM) Rasch measurement. The Rasch PCM specifies that each item has its own scale rating structure, so the two test types can be analysed simultaneously. The results showed that in general the students had difficulty to concluding the observation that the motion of the object was influenced by its shape and size; and pointed out of the relation between cloud and weather conditions. On the separate part, students clearly realize how to conserve the energy in everyday life; and interpret the effects of weather for human life. Furthermore, the MCQ test is problematic than the Essay Test for students. The average male student has better ability than the female students to comprehend the learning materials.

1. Introduction

The learning issue of natural sciences in elementary school means to frame up the students' awareness, understanding, and positive reaction toward the theories of sciences and technology in its function on daily life [1, 2]. One of the crucial natural science learning materials which demonstrated to students is energy and weather concepts and their applications. In general, learning the primary concepts of energy aims to enable students to express the origin of energy, proposition of objects, and their practicality. On the other hand, learning the concept of weather designs to set up students to define the relation between climate change, and weather to human life.

Student competencies are defined as the capabilities (cognitive, affective, and psycho-motoric) that students have after getting school exercises over a time. Student competence in learning requires to be a varied thing [3] dealing with competence is the primary reason to evaluate the success of students in learning. Commonly, learning of natural science becomes more problematic if not justified by teacher understanding [4]. On the other hand, teachers' ability to assess student learning outcomes is also fundamental. During this moment, the classical scoring approach (CTT) is so largely applied by teachers to decide whether students succeed or fail in learning [5]. The CTT approach also remains the primary benchmark for assessing students capabilities [6]. It is realized that regardless of the form of tests given

to the students, it should be noticed that the structure of natural sciences learning materials in elementary schools is complex, and potentially made stressful, frustrating, and causes low self-esteem for students [7]. Practically this demands that teachers are needed to provide a complete judgment to determine the proper student's ability and package it measurably to be communicated to affected parties. The clarity and objectivity of the assessment will support students to determine their learning activities and encourage teachers to determine proper methods to enhance the quality of learning in the classroom [4]. In the last 30 years the CTT approach has come out to be pulled out by education practitioners because it consists of many vulnerabilities [8, 9], and less raised measurements that are seen at "objective" [5, 10]. On the other hand, Rasch analysis as an Item Response Theory (IRT) family expanded so rapidly as a robust approach to examining individual capabilities as latent attributes objectively [11-14]. Rasch provides a linear measure of measurement in units of Logit [15, 16]. The developing of PCM inspired from Andrich Rating Scale [17] to determine data with compound scoring patterns [6, 18, 19]. Unlike the Rasch modeling to the dichotomy and polytomy, the Rasch PCM implementation can draw students' actual capabilities by combining MCQ and Essay Test simultaneously to provide linear measurements [16]. The practice of PCM is viewed as an appropriate method of determining students ability [20] and other educational context. Study conducted by Neumann, and Nehm [21], also Liu, Lee and Linn [22], [23], shows if Rasch modeling can be used for examined how students, classes, and teacher characteristics affect student inquiry science learning dan assessment. Furthermore, findings from Dorfner, Förtsch and Neuhaus [24] confirm if the Rasch PCM can identify of Three basic dimensions of instructional quality.

This research aims to evaluate (1) competence levels of students on learning of energy and weather concepts, and (2) the property test of natural sciences (MCQ and Essays) which provided to students in elementary schools through PCM. It is realized that the use of PCM as a tool for analyzing student achievement has not been widely used. This paper presents how to implement Rasch PCM in assessing student learning abilities in Natural Sciences. Expressing competency levels of elementary students can produce a rich insight for teachers to understand students' learning ability in learning the concept of energy and weather and its application in everyday life. On the other hand, teachers can also figure out how the influence of the form and quality of the questions given to the quality of student answers when the exam is realized.

2. Method

The research covered 64 students of third grade elementary school in Jakarta. Students who are exposed in this research are students who fully study and pick up the natural sciences test in the academic year 2018/2019. Natural sciences exam material consists of 40 items and sets into 2 types, i.e. MCQ has 35 items and Essay Test has 5 items. Students are allowed 90 minutes to settle the exam questions provided.

Fourthly questions on exam are constructed to evaluate students' competence in working (1) observation that the motion of the object is influenced by shape and size; (2) a description of the results of observations about the effects of heat energy, motion, and vibration in everyday life; (3) identification of energy sources and their practices; (4) understanding the windmill to produce the form of wind energy can be transformed into motion energy; (5) application of sustaining energy in everyday life; (6) description of the appearance of the earth's surface in the surrounding environment; (7) an explanation of the relationship between cloud and weather conditions; (8) description of the effects of weather on human life, and; (9) identification of human ways of preserving and protecting nature in the environment.

We explore the student exam results based on the key answers which has been prepared by the teacher on the MCQ test type, and the estimate rubric on the Essay Test. Data analysis using PCM was performed via WINSTEPS 4.1.0 Computer Program of Rasch Model [25]. The results of data analysis as a whole can be accessed through https://osf.io/2tkvd/ Open Sciences Framework [26]. Furthermore, all information regarding the students as respondents in this research is credential and is operated responsibly.

3. Result and discussion

3.1. Psychometrics properties of natural sciences test: a partial investigation

The first focus on this analysis is to investigating the quality of the tests administered to students. Estimation of PCM is performed simultaneously with the quality of the global test that consists of aspects of unidimensional, reliability, and rating structure.

Table 1. Summary of quality of natural sciences test (I=40; N=64).

Estimation (1)	Value	Estimation (2)	Value
Items reliability	.93	Raw var. explained by persons	25.1%
Persons-Items Reliability	.98	Raw var. explained by Items	30.0%
Separation Index of Items	3.77	Coherence Measure to Category (0)	70%
Mean Person	+1.00	Coherence Category to Measure (0)	57%
Raw variance explained by measures	55.1%	Coherence Measure to Category (1)	77%
Raw var**.** unexplained by measures	44.9%	Coherence Category to Measure (1)	85%

Based on table 1 indicated the overall quality of Natural Sciences Test is impressive (α = .93). It is also accompanied by the quality of the interaction between the person and the items as test was applied reaches the results not much different (α =. 98). Students are prepared to perform the 5 Essay and 35 MCQ test well (+1.00 logit). Relate to the raw variance explained by measures (55.1%), then the test used was unidimensional (> 40%). Deeply, the accuracy of measurements on a one-time observation or data collection at both the correct answer (1) and the incorrect answers (0) each show a significant percentage. However, the ability of items of Natural Sciences Test is only 25% of measuring student's ability, and 30% of energy concept and measure weather and its application in everyday life. This reflects that called for improvement against the content material in a test that tested subjects to students.

3.2. Measuring of outlier-sensitive fit on students work and test materials

To measure the competence level of the product, formerly we conclude the investigation of the consistency of the students when the test held place. It is highly important to separate out which students actually serious or consistent and which are not consistent in dealing with the test.

Table 2. Students and items outfit (I=40; N=64).

Students #	Outfit MNSQ	Items #	Outfit MNSQ
64P	1.54	E5	2.11
54P	2.44	E1	1.46
37L	1.54	S5	1.29
29P	1.71		
01P	2.40		
27L	1.65		

Refer to table 2 in notice there are 6 students who have expressed a pattern of response did not match against the test is given. The six students were not prepared to because of its ability to give an answer being estimated test that did not fit in with their actual abilities (+0.5 < Outfit MNSQ > +1.5). This happens due to the misconception of students in understanding learning materials or test is given. This diagnosis be essential so that the 6 students can be handling by the subjects-teachers at school. Furthermore, the examination was also performed against items used in the test; where this inspection serves to identify items that are not functioning in the measurement of the test material concept of energy and weather and its application in everyday life. In table 2 is also learned there are 3 items that outfit (Outfit MNSQ > 1.26). This means that the 3 items that are less well used to measure students ' ability in the material test of the concept of energy and weather and its application in everyday life. Students and items that they would outfit we didn't cover in the next analysis in this study.

3.3. Competence levels of students on learning of energy and weather concepts

After we reviewed Outfit value on students work and test materials, we return to analysis of 58 student ability. To show which students have strong and poor abilities, we rank based on the logit person value. Of the 58 students who picked up the test, we paid attention to students who occupied the top 3 and lowest positions.

Table 3. Students rank: top and bottom (N=58).

Rank #	Students Code	Measure	Rank #	Students Code	Measure
1	41P	+1.64	56	18P	-.43
2	21P	+1.49	57	16L	-.83
3	36L	+1.49	58	19P	-1.29

Based table 3 we identified students 41P who have the competence of most high (+1.64 Logits) among all students who pick up the test. Instead, 19P is a student who receives the lowest competence (-1.29 Logits) among all students who pick up the test. Furthermore, we also investigate of the test material has the highest and lowest levels of complexity for all student of the test.

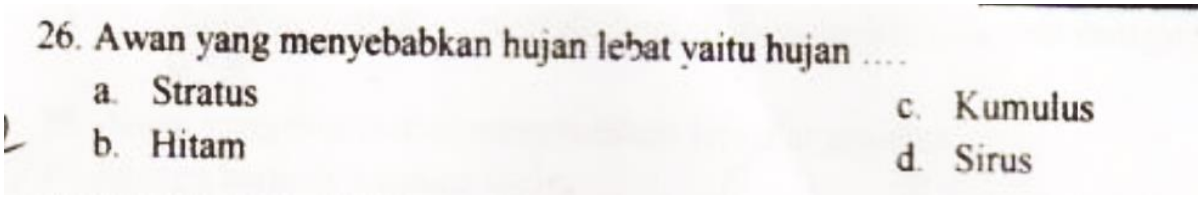


Figure 1. Item with the highest difficulty levels for all students (Measure = +2.70 Logits).

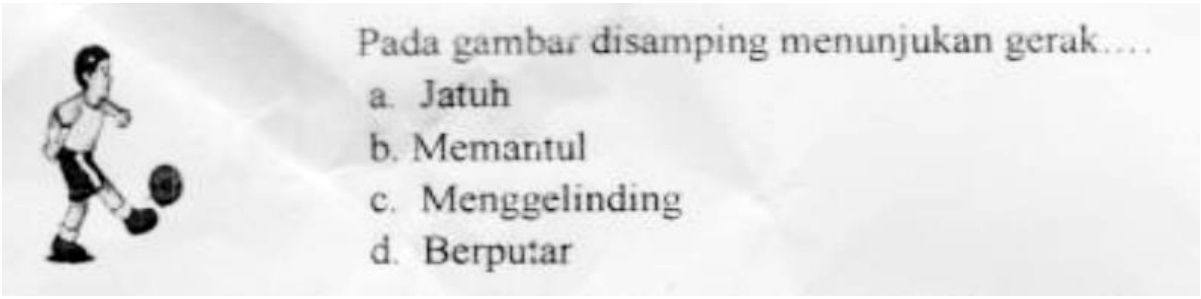


Figure 2. Item with the highest difficulty levels for all students (Measure = +2.70 Logits).

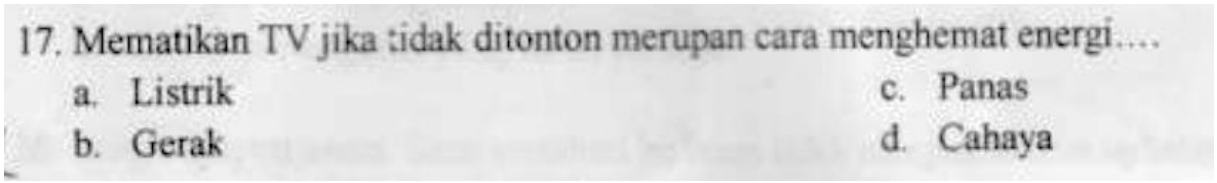
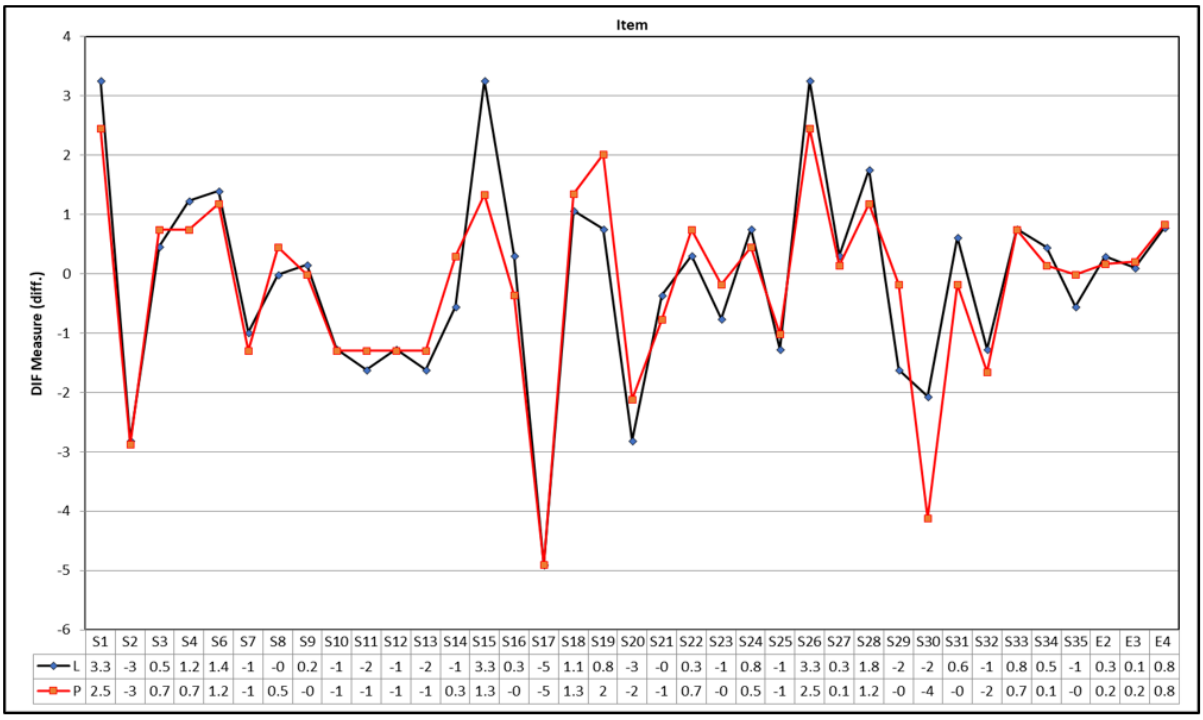


Figure 3. Items with the lowest difficulty level for all students (Measure = -4.89 Logits).

Figure 1 (item No. 26) and figure 2 (Item No. 1) is a difficult question to be solved by all students. In comparison, figure 3 (Item No. 17) is the easiest question solved by all students. If we paid attention to the level of difficulty to question No. 26 and No. 01 is 3 times higher complex than question No. 17. This also confirms that the questions in the MCQ formats are more complicated than Essay tests format for all students. Relates to the Standard Competence of material test, item No. 26 aims to measure students' ability in interpreting the relationship between cloud and weather. Meanwhile, item No. 01

aims to measure students' ability in the determining observations that the motion of the objects influenced by the shape and size. Further, item No. 17 aims to measure students ' ability in how to applied conserve energy in daily life. Furthermore, the 37 items test given admitted the male students (-0.12 Logits) has higher ability rather than female students (0.20-Logits).



	S1	S2	S3	S4	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32	S33	S34	S35	E2	E3	E4
L	3.3	-3	0.5	1.2	1.4	-1	-0	0.2	-1	-2	-1	-2	-1	3.3	0.3	-5	1.1	0.8	-3	-0	0.3	-1	0.8	-1	3.3	0.3	1.8	-2	-2	0.6	-1	0.8	0.5	-1	0.3	0.1	0.8
P	2.5	-3	0.7	0.7	1.2	-1	0.5	-0	-1	-1	-1	-1	0.3	1.3	-0	-5	1.3	2	-2	-1	0.7	-0	0.5	-1	2.5	0.1	1.2	-0	-4	-0	-2	0.7	0.1	-0	0.2	0.2	0.8

Figure 4. Comparison of the male and female competency in each of the items.

More specifically, in the Essay Test (item E2) male students has greater ability than female students. Instead, on the item E3 female students is greater than male students. The other interesting feature, on the item E4 students' ability of men and women is the same. Comparison of detailed information on competence of students both (male and female) in each of the test items are displayed in figure 4.

This research reveals some crucial points on assessing student learning abilities. The first finding referred to the test construction of learning outcome is called for to be recovered. Study organized by Downing [27] reported that poor items that were indicative of a systematic error between the measurement constructs; possible to affected the instrument's validity and disadvantaged students who picked up the exam. The second finding, there were 6 students who expressed different responses to the test. This means that students' misconceptions related to the material being questioned. In addition, a good test to measuring students' abilities can also be applied as a tool to identify misconceptions of students learning. When misconceptions can be found, the teacher can provide assistance to students [28]. In the third finding, it is unfortunate that students have not been able to recognize the relationship between cloud and weather, and also the concept of moving objects. Study conducted by Henriques that normally children in elementary school have different ways to thinking and tend to be naive in recognizing the nature, weather, climate and atmosphere and also interpreting moving objects [29]. In this section it is also known that PCM can estimate MCQ and Essay tests properly, and maintain information on differences in abilities between male and female students.

4. Conclusion

This research demonstrated that in general the students had difficulty in concluding the observation that the motion of the object was influenced by its shape and size; and explained the relationship between cloud and weather conditions. Instead, an effort to conserve energy in everyday life; and define the

effects of weather for human life is an aspect that is most easily understood by students. Referring to psychometrics properties of Natural Sciences Test found that the quality of test items is already acceptable. However, it still required some improvement to enhance the capability test in order to reach a stronger measurement attribute. The other interesting thing in this study it was found that male students had a better ability than female students in completing the exam is provided.

Acknowledgement
We thank to Mrs. Wahyu Eka Prasetyaningtyas who had provided facilities during this research was conducted. The authors declared no conflict of interest and no grants were accepted in this research.

References

[1] Samatowa U 2006 *Bagaimana membelajarkan IPA di sekolah dasar*: PT Pustaka Indonesia Press)
[2] Rustaman N, Sekarwinahyu M, Sutarno N, Widiasih W, Rahayu U, Budiastra A, Ratnaningsih A, Sholihin H H, Wahyuningsih T and Mujadi M 2014 *PDGK4503/MODUL* (Jakarta: Universitas Terbuka)
[3] Gresalfi M, Martin T, Hand V and Greeno J 2009 Constructing competence: An analysis of student participation in the activity systems of mathematics classrooms *Educational studies in mathematics* **70** 49-70
[4] Georgiou S N, Christou C, Stavrinides P and Panaoura G 2002 Teacher attributions of student failure and teacher behavior toward the failing student *Psychology in the Schools* **39** 583-95
[5] Kline T 2005 *Psychological testing: A practical approach to design and evaluation*: Sage)
[6] Sumintono B and Widhiarso W 2015 *Aplikasi pemodelan rasch pada assessment pendidikan*: Trim Komunikata)
[7] Alizamar A, Ifdil I, Fadli R P, Erwinda L, Zola N, Churnia E, Bariyyah K, Refnadi R and Rangka I B 2018 The Effectiveness of Hypnotherapy in Reducing Stress Levels *Addictive Disorders & Their Treatment* **Inpress**
[8] De Champlain A F 2010 A primer on classical test theory and item response theory for assessments in medical education *Medical education* **44** 109-17
[9] Crocker L and Algina J 2008 *Introduction to classical and modern test theory* (USA: Cengage Learning)
[10] Sumintono B and Widhiarso W 2014 *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*: Trim Komunikata Publishing House)
[11] Wright B D 1977 Solving measurement problems with the Rasch model *Journal of Educational Measurement* **14** 97-116
[12] Engelhard Jr G 2013 *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*: Routledge)
[13] Fan X 1998 Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics *Educational and Psychological Measurement* **58** 357-81
[14] Folastri S, Rangka I B and Ifdil I 2017 Student's Self-concept Profile Based on Gender: a Rasch Analysis. In: *9th International Conference for Science Educators and Teachers (ICSET 2017),* (Semarang: Atlantis Press) pp 264-9
[15] Wilson M, Allen D D and Li J C 2006 Improving measurement in health education and health behavior research using item response modeling: introducing item response modeling *Health Education Research* **21** i4-i18
[16] Bond T G and Fox C M 2013 *Applying the Rasch model: Fundamental measurement in the human sciences*: Psychology Press)
[17] Masters G N 1982 A Rasch model for partial credit scoring *Psychometrika* **47** 149-74
[18] Masters G N 2016 *Handbook of Item Response Theory, Volume One*: Chapman and Hall/CRC) pp 137-54
[19] van der Linden W J and Hambleton R K 2013 *Handbook of modern item response theory*: Springer Science & Business Media)

[20] Sabina E and Susanne B 2010 Students' use of decision-making strategies with regard to socioscientific issues: An application of the Rasch partial credit model *Science Education* **94** 230-58
[21] Neumann I, Neumann K and Nehm R 2011 Evaluating Instrument Quality in Science Education: Rasch- based analyses of a Nature of Science test *International Journal of Science Education* **33** 1373-405
[22] Liu O L, Lee H-S and Linn M C 2010 Multifaceted assessment of inquiry-based science learning *Educational Assessment* **15** 69-86
[23] Liu O L, Lee H-S and Linn M C 2011 An Investigation of Explanation Multiple-Choice Items in Science Assessment *Educational Assessment* **16** 164-84
[24] Dorfner T, Förtsch C and Neuhaus B J 2018 Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction *Learning and Instruction* **56** 42-53
[25] Linacre J M 2018 Winsteps® Rasch measurement computer program (Version 4.1.0). (Beaverton, Oregon: Winsteps.com)
[26] Rangka I B 2018 Data Sets: Expressing competency levels of elementary students on learning of energy and weather concepts and their applications using partial credit model. Open Sciences Framework)
[27] Downing S M J A i H S E 2005 The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education **10** 133-43
[28] Treagust D F 1988 Development and use of diagnostic tests to evaluate students' misconceptions in science *International Journal of Science Education* **10** 159-69
[29] Henriques L 2002 Children's Ideas About Weather: A Review of the Literature *School Science and Mathematics* **102** 202-15