Item_Analysis_of_Heat_Transfer _Concept_Using_Rasch_Model.p df

_by_

PAPER · OPEN ACCESS

Item Analysis of Heat Transfer Concept Using Rasch Model in Elementary School

To cite this article: N Hermita *et al* 2021 *J. Phys.: Conf. Ser.* **2049** 012058

View the article online for updates and enhancements.

Item Analysis of Heat Transfer Concept Using Rasch Model in Elementary School

N Hermita [1,*], S Sakinah[1], T T Wijaya[2], R Vebrianto[3], J A Alim[1], Z H Putra[1], N Fauza[1], D A Dipuja[1], J Pereira[4], C Jihe[4]

[1]Faculty of Teacher Training and Education, Universitas Riau, Pekanbaru, Indonesia.
[2]School of Mathematical Sciences, Beijing Normal University, Beijing, China.
[3]Fakultas Tarbiyah, UIN SUSKA Riau, Pekanbaru, Indonesia.
[4]Department Mathematics and Statistics, Guangxi Normal University, Guilin, China.

* neni.hermita@lecturer.unri.ac.id

Abstract. This study aims to analyze test items of heat transfer instructional materials using the Rasch model to obtain information on the Wright map, item measure, item fit, and item reliability. This study applied a quantitative-descriptive research design. The subjects in this study were 75 fifth graders of State Elementary School 192 of Pekanbaru. The students were asked to do 11 essay items. Students' responses on the essay items were collected through Google Form. The collected data were then analyzed using the Rasch model on the Winstep® application version 3.73. Based on the data analysis, item 2 and 3 were the most difficult items, and item 10 was the easiest item. All items are at a moderate difficulty level. There was an invalid item of 11 items. The item reliability value was 0.75 which meant sufficient. In conclusion, the 10 test items (1 is invalid) are selected for inclusion in the achievement test to measure the students' understanding of heat transfer instructional materials.

1. Introduction

Natural Science is the sciences that study surrounding natural phenomena using scientific procedures to reach a conclusion [1]. Natural science is one of the main subjects in Indonesia's education curriculum [2]. In elementary school, natural science on the topic of heat transfer is learned in grade 5. The students learn temperature and heat, heat transfer around us, and the impacts of heat on life [3]. Heat transfer is transferring energy (heat) from one area to another due to the different temperatures in those areas. There are three heat transfer mechanisms: conduction, convection, and radiation [4]. Conduction is the process by which heat or electricity passes through a solid object. Convection is the flow of heat through a gas or a liquid, and radiation is the sending out of heat in waves.

Natural science is an exact science that is those sciences that admit of absolute precision in their results. By learning natural science, students develop their intellectual skills, critical thinking skills, and accuracy skills. Furthermore, students often ask to do experiments to improve students' thinking skills and problem-solving skills. This is in line with the objectives of the 2013 curriculum, one of which is to produce students whose soft skills and hard skills are balanced [5]. In the 2013 curriculum, teachers are expected to develop practice questions that train students' thinking skills [6]. In developing practice questions, it is necessary to do item analysis to assess the quality of those items and the test as a whole [7].

Item analysis is a test question assessment to obtain a set of questions with good quality. Item analysis is a systematic procedure that provides precise information on the test items [8]. Item analysis is a systematic procedure that provides specific information about the analyzed items [9].

The purpose of item analysis is to improve the quality of test items and identify their shortcomings for learning improvement [10]. Test items can bring out information that can be used to produce higher-quality test items. The process of item analysis aims to examine and review each item by collecting information from students' responses to obtain high-quality test items before they are used [11]. This quantitative analysis includes the analysis of validity, reliability, difficulty, discrimination, and distraction of test items [12].

We saw that in practice item analysis are rarely conducted. That is why the materials, item construction, language, validity, reliability, difficulty, discrimination, and distraction are often considered low. In other words, there is no accountability in its quality. A validity test is a proofing process to the extent to which evidence and theory support the instrument being developed. Valid test items mean that they can be used to measure what they are supposed to measure. A reliable test item is a test item that produces the same data when it is used on the same objects at different times [13]. In other words, it does not make much difference in information. The item difficulty level is simply the percentage of students who answer an item correctly. This study provides information on the Wright Map, item measure, item fit, and item reliability of test items using the Rasch model. The Rasch model can immediately measure validity and reliability based on the probability principle [14].

Other advantages of using the Rasch model are predicting missing data based on a systematic response pattern, producing standard error measurement values for the instruments which can increase the accuracy of calculation, and performing three-way calibration: the measurement scale, respondents, and items [15]. The development of functional assessments based on Rasch measurement models is becoming a preferred method among rehabilitation professionals for constructing tests that traditional methods [16]. The test items investigated in this study had never been validated by any validator and had never been tested on students so that the quality of the test items was unknown. The researcher investigated the test items used to measure the students' understanding of heat transfer using the Rasch model to obtain information on the Wright map, item measure, item fit, and item reliability.

2. Method

This study applied a quantitative-descriptive research design that aims to describe events statistically. This study analyzed 11 essay items in thematic lessons. The subjects in this study were 75 fifth graders of State Elementary School 192 of Pekanbaru. The researcher collected the students' responses to the essay items using Google Forms. The collected data were then analyzed using the Rasch model on the Winstep® application version 3.73. The process of this study was divided into three phases: early phase, implementation phase, and final phase.

3. Results and Discussion

3.1. Style and spacing

AA Wright Map provides a picture of a test by placing the difficulty of the test items on the same measurement scale as the students' ability. Figure 2 is the Wright Map that compares the students and the test items. The Wright Map is organized as two vertical histograms. The left side of the map shows the distribution of the measured ability of the students from most able at the top (i.e., student 56, 57, and 71) to least able at the bottom (i.e., student 20). The items on the right side of the map are distributed from the most difficult at the top (i.e., item 2 and 3) to the least difficult at the bottom (i.e., item 10). The most difficult test items could be answered correctly by 60 students (80%). The student 2, 15, 65, and 68 are opposite to the item 2 and 3 on the map. It means the difficulty of the items and the students' ability were comparable, so the students had approximately a 50% probability of answering the items correctly. The items 10 was probably answered incorrectly by the student 9, 10, and 20 and was probably answered correctly or incorrectly by the student 12 and 18. Based on the

map, it can be concluded that the test items are generally not too difficult for the students. This means that the students understand the instructional materials of heat transfer.

According to Herppich & Wittwer [17], the accuracy in assessing a learning process will have a strong impact on what students have learned in class. Therefore, in developing test items, it is necessary to do an item analysis to avoid students' misconceptions about the items [18].
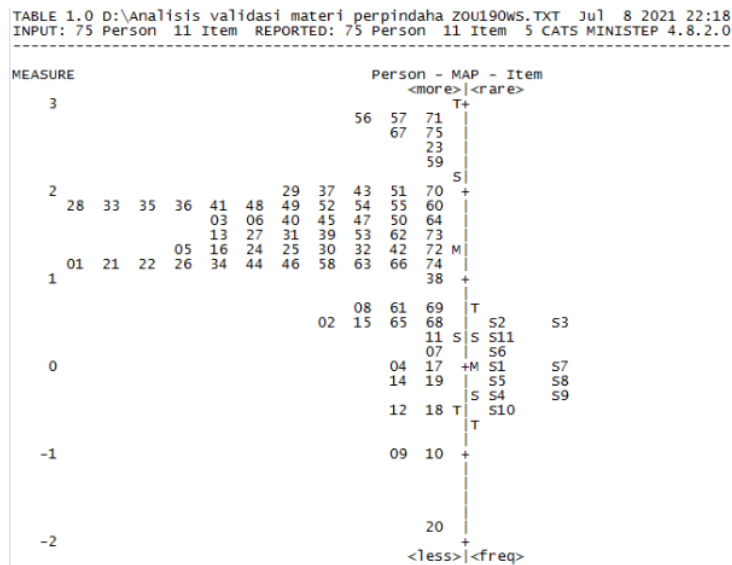

Figure 1. The Wright Map.

3.2. Item Measure

Figure 2. Item Measure.

An item measure provides information on the difficulty level of each item. The column of entry number contains information on the items' numbers, and the numbers are sorted based on the difficulty level. The most difficult item was item 3, and the least difficult item was item 10. Weller [19] divide the difficulty level of test items into three categories: easy, moderate, and difficult. An item will be said to be easy if the measure score is $-2.00 < b < -1.00$. An item will be said to be moderate if the measure score is $-1.00 < b < +1.00$. An item will be said to be difficult if the measure score is $+1.00 < b < +2.00$. Table 1 presents information on the difficulty level of each item based on the score interpretation.

Table 1. Difficulty Levels (Logit)

Test Items	Level of Difficulty	Category
3	0:55	Moderate
2	0:45	Moderate
11	0:35	Moderate
6	0:12	Moderate
1	0:01	Moderate
7	-0.04	Moderate
5	-0.17	Moderate
8	-0.21	Moderate
4	-0.26	Moderate
9	-0.30	Moderate
10	-0.50	Moderate

All items were at a moderate difficulty level. According to Widyaningsih et al. [20], a good question should have a moderate difficulty level. Similarly, Kurniasi et al. [21] state that good questions should not be too easy or difficult. Easy questions make it difficult for students to stimulate their critical thinking while difficult questions decrease their motivation because they are not able to do so easily. Furthermore, Susanto et al. [22] state that questions with a moderate difficulty level can be stored up and used to measure students' abilities in upcoming years. It can be concluded that the quality of the test items is good because they are at a moderate difficulty level.

3.3. Item Fit

```
TABLE 10.1 D:\Analisis validasi materi perpindah ZOU190WS.TXT  Jul  8 2021 22:18
INPUT: 75 Person  11 Item  REPORTED: 75 Person  11 Item  5 CATS MINISTEP 4.8.2.0
--------------------------------------------------------------------------------
Person: REAL SEP.: 1.69  REL.: .74 ... Item: REAL SEP.: 1.75  REL.: .75

        Item STATISTICS:  MISFIT ORDER

-------------------------------------------------------------------------------
|ENTRY   TOTAL   TOTAL           MODEL|  INFIT  |  OUTFIT |PTMEASUR-AL|EXACT MATCH|      |
|NUMBER  SCORE   COUNT  MEASURE   S.E.|MNSQ ZSTD|MNSQ ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|-----------------------------------------------------------------------------------|
|    9     190     75    -.30     .15|1.73 4.20|1.68 3.92|A .51   .54| 30.7  44.7| S9   |
|    3     153     75     .55     .16|1.31 1.86|1.24 1.45|B .50   .49| 40.0  51.4| S3   |
|   11     161     75     .35     .15|1.13  .85|1.11  .72|C .45   .50| 41.3  49.9| S11  |
|    2     157     75     .45     .15| .97 -.12|1.05  .37|D .51   .50| 56.0  50.4| S2   |
|    6     171     75     .12     .15| .91 -.58| .88 -.76|E .59   .52| 46.7  47.5| S6   |
|   10     199     75    -.50     .15| .90 -.67| .88 -.82|F .53   .55| 52.0  45.4| S10  |
|    4     188     75    -.26     .15| .89 -.73| .89 -.76|e .60   .54| 49.3  44.6| S4   |
|    1     176     75     .01     .15| .74 -1.87| .86 -.92|d .41   .52| 62.7  46.3| S1   |
|    7     178     75    -.04     .15| .82 -1.26| .86 -.90|c .53   .53| 53.3  46.2| S7   |
|    5     184     75    -.17     .15| .82 -1.24| .82 -1.28|b .62   .53| 54.7  45.2| S5   |
|    8     186     75    -.21     .15| .80 -1.45| .81 -1.30|a .55   .53| 53.3  45.1| S8   |
|-----------------------------------------------------------------------------------|
|MEAN    176.6    75.0     .00     .15|1.00  -.1|1.01   .0|           | 49.1  47.0|      |
|P.SD     14.1     .0      .32     .00| .28  1.7| .25  1.5|           |  8.5   2.4|      |
-------------------------------------------------------------------------------
```

Figure 3. Item Fit.

Item fit explains whether an item functions normally to take measurements or not. The outfit means-square (MNSQ) and the infit z-standard (ZSTD) are the criteria used to see the item fit level. An item will be said to be fit if the MNSQ score is 0,5 < MNSQ < 1,5, and the ZSTD score is -2,0 < ZSTD < +2,0. Based on the score interpretation, item 9 is not fit. According to Palimbong et al. [23], an item that does not meet the MNSQ and ZSTD criteria cannot be used to measure students' achievement. Furthermore, Marfu'i [24] state if there are items that do not fit, this indicates a student's misconception in answering the items. If there is an item that does not fit, the item needs to be revised or replaced. According to Weller et al [19], providing test items that meet the item fit criteria ensure that students' achievement is measured using good quality test items. Therefore, teachers are expected to provide good-quality test items so that the resulting scores show the real students' achievement.

3.4. Item Reliability

```
      SUMMARY OF 11 MEASURED Item
-----------------------------------------------------------------------
|          TOTAL                    MODEL |   INFIT   |   OUTFIT   |
|          SCORE   COUNT   MEASURE   S.E. | MNSQ  ZSTD| MNSQ  ZSTD |
|---------------------------------------------------------------------|
| MEAN    176.6    75.0     .00      .15  | 1.00  -.09| 1.01  -.02 |
| SEM       4.4     .0      .10      .00  |  .09   .54|  .08   .48 |
| P.SD     14.1     .0      .32      .00  |  .28  1.70|  .25  1.51 |
| S.SD     14.7     .0      .34      .00  |  .29  1.78|  .26  1.58 |
| MAX.    199.0    75.0     .55      .16  | 1.73  4.20| 1.68  3.92 |
| MIN.    153.0    75.0    -.50      .15  |  .74 -1.87|  .81 -1.30 |
|---------------------------------------------------------------------|
| REAL RMSE   .16 TRUE SD   .28 SEPARATION  1.75  Item  RELIABILITY  .75 |
| MODEL RMSE  .15 TRUE SD   .28 SEPARATION  1.87  Item  RELIABILITY  .78 |
```

Figure 4. Item Reliability.

Reliability means the extent to which the results of a measurement can be trusted [25]. A measure must meet the criteria of reliability and validity. A measure is said to have high reliability if it produces similar results under consistent conditions. The summary statistics out table in figure 5 provides information on the reliability of the test items. Ishar [26] divide the reliability of a measure into five categories: weak, sufficient, good, very good, and exceptional.
<0.67   : Weak

0.67-0.80   : Sufficient
0.8-0.90   : Good
0.91-0.94   : Very Good
>0 ,94   : Exceptional

If the reliability value is in the sufficient category, it means a measure can yield the same results on repeated trials. Furthermore, Farradillah [27] states that test items used to measure students' achievement must have a minimum reliability coefficient of 0.70. Indrayani, Djuniadi, and Ridlo [28] also state that a measure is reliable if the reliability coefficient is above 0.70. Based on the summary statistics out table in figure 5, the reliability score was 0.75 which means a sufficient category. In conclusion, the test items are reliable enough and thus can be used to measure students' achievement

4. Conclusion and Recommendation
This study aims to analyze 11 essay items of heat transfer instructional materials using the Rasch model. Based on the data analysis, item 2 and 3 were the most difficult items, and item 10 was the easiest item. All items are at a moderate difficulty level. There was an invalid item of 11 items. The item reliability value was 0.75 which meant sufficient. In conclusion, the 10 test items (1 is invalid) are selected for inclusion in the achievement test to measure the students' understanding of heat transfer instructional materials.

References
[1]   Suhandi A Muslim Samsudin A Hermita N and Supriyatman 2018 Effectiveness of the use of question-driven levels of inquiry based instruction (QD-LOIBI) assisted visual multimedia supported teaching material on enhancing scientific explanation ability senior high school students *J. Phys. Conf. Ser.* **1013** 1
[2]   Hermita N *et al* 2018 Level conceptual change pre-service elementary teachers on electric current conceptions through visual multimedia supported conceptual change *J. Phys. Conf. Ser.* **1013** 1
[3]   Hermita N Alpusari M Noviana E Kurniaman O Widyanthi A and Suhandi A, 2019 A Study of Prospective Primary School Teachers' Alternatif Conception in Heat and Temperature *J. Phys. Conf. Ser.* **1351** 1
[4]   Anam R S Widodo A and Sopandi W 2017 Representation of Elementary School Teachers on Concept of Heat Transfer *J. Phys. Conf. Ser.* **895** 1
[5]   Wijaya T T Purnama A and Tanuwijaya H 2020 Pengembangan Media Pembelajaran Berdasarkan Konsep Tpack pada Materi Garis dan Sudut Menggunakan Hawgent Dynamic Mathematics Software *JPMI – J. Pembelajaran Mat. Inov.* **3** 3 p. 205–214
[6]   Rosaini Budiyono and Pratiwi H 2019 How does HOTS based problem posing model improve students' attitudes toward mathematics? *IOP Conf. Ser. Earth Environ. Sci.* **243** 1
[7]   Rahayu W Putra M D K Iriyadi D Rahmawati Y and Koul R B 2020 A Rasch and factor analysis of an Indonesian version of the Student Perception of Opportunity Competence Development (SPOCD) questionnaire *Cogent Educ.* **7** 1
[8]   Zhang D Orrill C and Campbell T 2015 Using the Mixture Rasch Model to Explore Knowledge Resources Students Invoke in Mathematic and Science Assessments *Sch. Sci. Math.* **115** 7 p. 356–365
[9]   Yang S C Tsou M Y Chen E T Chan K H and Chang K Y 2011 Statistical item analysis of the examination in anesthesiology for medical students using the Rasch model *J. Chinese Med. Assoc.* **74** 3 p. 125–129

[10] Arikunto S, 2012 *Dasar-dasar Evaluasi Pendidikan* Jakarta: Bumi Aksara.

[11] Van Zile-Tamsen C 2017 Using Rasch Analysis to Inform Rating Scale Development *Res. High. Educ.* **58**, 8 p. 922–933

[12] Rosiqoh R and Suhendi E 2021 Using Rasch model analysis to analyse students' mastery of concept on newton law *J. Phys. Conf. Ser.* **1731**, 1

[13] Hidri S 2018 Assessing spoken language ability: A many-facet Rasch analysis *Second Lang. Learn. Teach.* 9783319628837 p. 23–48

[14] Cheong S M Ling M Wong J and Ling K 2018 *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings* Springer Singapore

[15] Othman H Asshaari I Bahaludin H Nopiah Z M and Ismail N A 2012 Application of Rasch Measurement Model in Reliability and Quality Evaluation of Examination Paper for Engineering Mathematics Courses *Procedia - Soc. Behav. Sci.* **60**, 2009 p. 163–171

[16] Haspel R L Lin Y Fisher P Ali A and Parks E 2014 Development of a validated exam to assess physician transfusion medicine knowledge *Transfusion* **54**, 5 p. 1225–1230

[17] Herppich S and Wittwer J 2018 Preservice teachers' beliefs about students' mathematical knowledge structure as a foundation for formative assessments *Teach. Teach. Educ* **76** 242–254

[18] Ramadhani R & Fitri Y 2020 Validitas E-Modul Matematika Berbasis EPUB3 Menggunakan Analisis Rasch Model *J. Gantang* **5**, 2 p. 95–111

[19] Weller J A Dieckmann N F Tusler M Mertz C K Burns W J and Peters E 2013 Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach *J. Behav. Decis. Mak.* **26**, 2 p. 198–212

[20] Widyaningsih S W Yusuf I Prasetyo Z K and Istiyono E 2021 The Development of the HOTS Test of Physics Based on Modern Test Theory *Int. J. Instr.* **14**, 4

[21] Kurniasi E R Y Y and Karennisa F 2020 Analisis Soal Ulangan Harian Matematika Kelas IX SMP Negeri 1 Toboali *J. Ilmu Pendidik. STKIP Kusuma Negara* **12**, 1 p. 43–52

[22] Susanto H Rinaldi A and Novalia 2015 Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika *J. Japan Soc. Respir. Endosc.* **6**, 2

[23] Palimbong J Mujasam M and Allo A Y T 2019 Item Analysis Using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK Negeri 2 Manokwari *Kasuari Phys. Educ. J.* **1**, 1 p. 43–51

[24] Marfu'i L N R Ilfiandra and Nurhudaya 2019 The analysis of critical thinking skills test in social-problems for physics education students with Rasch Model *J. Phys. Conf. Ser.* **1280**, 5

[25] Ridwan I I Ali R Mohamed I I Adam M Z and Elfadil N 2017 Rasch measurement analysis for validation instrument to evaluate students technical readiness for embedded systems *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON* p. 2115–2119

[26] Ishar N I M and Masodi M S 2012 Students' perception towards quality library service using Rasch measurement model *ICIMTR 2012 - 2012 Int. Conf. Innov. Manag. Technol. Res.* p. 668–672

[27] Faradillah A and Febriani L 2021 Mathematical Trauma Students' Junior High School Based on Grade and Gender *Infin. J.* **10**, 1 p. 53

[28] Indrayani L Djuniadi and Ridlo S 2017 Pengembangan Instrumen Penilaian Afektif Peminatan Peserta Didik SMA Negeri 1 Semarang *J. Res. Educ. Res. Eval.* **6(1)** 39–45

12%
SIMILARITY INDEX

11%
INTERNET SOURCES

8%
PUBLICATIONS

5%
STUDENT PAPERS

| Exclude quotes | On | Exclude matches | < 2% |
| Exclude bibliography | On | | |