The Reliability Analysis for Force Concept Inventory

Mazlena Binti Murshed, Fatin Aliah Phang,
Muhammad Abd Hadi Bin Bunyamin and Ibnatul Jalilah Binti

Abstract--- *Force Concept Inventory (FCI) is a diagnostic test that is commonly used to measure the concept of knowledge based on force and motion. This study examines the validity and reliability of FCI using a new item analysis method. Rasch Measurement Model(RMM) enables the construction of linear scales for individuals (person) and items from raw test scores. This model can provide valuable insights into the structure and function of the test regarding the difficulty distribution of the test item. It can also predict to what extent it fits the model, and the items also can determine its basic structure. The data for Rasch's analysis were obtained from a pilot study involving 78 high school students (16 to 17 years old) who took physics subject. The reliability of FCI items is high, and it is proven that it can measure conceptual understanding. Thus, this test can determine the level of conceptual understanding of force and motion.*

Keywords--- *Force Concept Inventory (FCI), Pilot Study, Testing, Rasch Stochastic Model, Force, and Motion.*

I. INTRODUCTION

Force Concept Inventory (FCI) is an instrument used to assess students' understanding of force and Newton's law (Hestenes, Wells, & Swackhamer, 1992). FCI is one of the most reliable physics tests to assess the effectiveness of teaching in introductory physics courses (Persson, 2015; Savinainen, 2005). Educators and educational researchers have widely used the FCI test for a variety of purposes. For example, to identify student difficulties (Fadaei & Mora, 2015; Rosenblatt, 2012) and to compare the effectiveness of the curriculum and pedagogy (Heller & Heller, 2010; Lin & Singh, 2011). This test is one of a variety of test options designed to assess conceptual understanding in physics (Hestenes, Wells, & Swackhamer, 1992; Thornton & Sokoloff, 1998). The original version of FCI was published in 1992. Meanwhile, the revised version was developed and uploaded on the web in 1995.

Additionally, the new version has 30 questions, while the original FCI has only 29 questions. All FCI questions have five optional choices of answer. In each of these questions, the other four answers are distractors that represent student misconceptions. This test has been accessed from the website https://www.physport.org/assessments/ with the permission of the admin of the site to be adapted according to the syllabus of force and motion in Malaysia.

This type of question is easy to administer and check. However, some criticisms suggest that multiple-choice tests do not test students' cognitive levels. Furthermore, the multiple-choice test only tests the low cognitive level of memorization (Ghafar, 2000). This opinion is in line with the view from Arhin (2015), which states that multiple-

Mazlena Binti Murshed, Ministry of Education, Pusat Pentadbiran Kerajaan Persekutuan, Putrajaya, Malaysia. E-mail: enn7997@yahoo.co.uk
Fatin Aliah Phang, Centre for Engineering Education, Universiti Teknologi Malaysia, Malaysia. E-mail: p-fatin@utm.my
Muhammad Abd Hadi Bin Bunyamin, School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Malaysia. E-mail: mabhadi@utm.my
Ibnatul Jalilah Binti Yusof, School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Malaysia. E-mail: ijalilahyusof@hotmail.com

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

choice tests cannot test students with higher-level teaching processes such as analysis, synthesis, and assessment. One of the reasons is that teachers are challenging to build and teachers have limited ability to build high-quality multiple-choice tests. Therefore, the existing test was chosen because it proved to be capable of measuring the level of mastery of the concept of force and motion, namely the Force Concept Inventory (FCI) test.

FCI has been translated into the Malay language by Jaafar Jantan (2002) to make it easier for students to understand the items presented. This version of FCI also been adapted to the structure of the Malaysia physics syllabus by Ahmad Tarmimi dan Shahrul Kadri (2016) and is known as FCIspm. In this study, the FCIspm and the curriculum specification of the form four physics lessons in Malaysia are used to review the scope of the FCI topic that is appropriate for students. The scope of the item is as in Table 1.

Table 1: Coverage of Force and Motion Topics for FCI

Topics	Items No.
Linear Motion	19,21,22,23,24
Impulse and Impulsive Force	4,28
Effect of Force	8,9,10,20,25,26
Gravitational Force	1,2,3,13
Forces in Equilibrium	11,15,16,17, 18,27,29,30
Projectile Motion	12,14
Circular Motion	5,6,7

Based on Table 1, all FCI questions were grouped according to the distribution of topics for force and motion titles. However, two topics are lacking in the scope of the subject and the motion, which is the projectile motion and the circular motion. Therefore, it would be unfair to test these two concepts against fourth-grade students. Therefore, the questions (5,6,7,12 and 14) that covered the topics were eliminated. Therefore, the total FCI score is 25 after removal. Students who answered 18 questions correctly (60%) or above were said to have mastered the Newtonian concept of force (Hestenes & Wells, 1992; Mohammad Mubarrak, Siti Fairuz, & Norezan, 2013). Therefore, in this study, a percentage score of 60 and above refers to students who master the concept of force and motion well or high (Mohammad Mubarrak et al., 2013).

As FCI is a recognized assessment tool in physics educational research, it is crucial to evaluate and monitor its function using in-depth analysis. The analysis used is the Rasch Measurement Model (RMM), which employed One-Parameter Logistics that provide a deeper understanding of students' ability from this diagnostic test (Abdul Latif et al. , 2015). This model, only measures for the ability of students and item difficulty (Sumintono, 2018), also assumed that the level of item discrimination is constant (Magno, 2009). RMM generates independently of the test items as well as the test respondent (Sumintono, 2018).

RMM used the logarithmic method to generate the same interval (equal distance) measurement scale named logit (log odds unit). Through this data type interval, the Rasch model develops a logit ruler that determines the relationship between the ability of the student and the level of difficulty of the item (Emanuela Ene & Ackerson, 2018; Magno, 2009).

Application of the Rasch model through Winsteps (Linacre, 2012) is the consequence of fundamental principles deemed essential and indispensable. It will provide estimates of person and threshold locations on the latent variable

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

scale. Rasch solves on how to have the right measurement with a valid instrument.

II. Data Collection And Analysis

The concept of reliability of a research instrument is that it must be stable and accurately predict (Kumar, 2011). In other words, reliability is the degree of accuracy of an instrument for measuring. There are two types of instrument reliability, internal and external reliability (Kumar, 2011). In this study, the reliability of the FCI test was measured using internal reliability. Cronbach's Alpha scores were used to assess FCI internal reliability. Cronbach's alpha was noted as the most common measure of the internal reliability of several items that assessed the same construct (Flanders, 2014; Mcdermott, 2009). Moreover, the Rasch model will provide the item reliability and also item separation which provide the discrimation of itemThe data obtained from this pilot was to determine the reliability of this type. Nevertheless, the instrument must be established in advance of the face and content validity from three physics experts teachers before conducting the pilot study.

Two schools were selected for this pilot test. The school principals were contacted, informed about the study, and request permission to test students taking physics subjects. Both principals allowed the pilot tests to be conducted on their students. This test was conducted in October 2018 with the presence of a school-going researcher to administer, monitor, and compile tests and student response sheets. This pilot study was conducted on 78 students who had the same criteria as the actual study participants (Cohen, Manion, & Morrison, 2007). They took physics lessons and learned the subject of force and motion.

According to the study of Mohammad Mubarrak et al. (2013) who used FCI test allotted 30 minutes to answer all questions. This study also provides for the same period for this test. This is due to students are unfamiliar with the types of questions they ask, such as long-term questions and language barriers. Students also do not make any preparations for this test.

This pilot test was conducted without recording the student's name. However, each student was provided with a code to identify their test scores. Students are required to mark their answers on a particular answer sheet. After the test has been marked, and the score has been obtained, the participating schools are notified of the student test scores and are asked to send the information to their students. No incentives, such as grades, are offered to students to take the test. However, the purpose of the study and the importance of the test were explained to the students before the researcher conducted the test. Students generally show interest in tests and want to know their scores.

All students completed the study of force and motion in April, and the test was conducted in October the same year. Therefore, the average time between the teaching and learning sessions and the FCI test is six months. Although the learning and test intervals were large, during this time, students learned other physics topics and took mid-year exams involving concepts of force and motion. Therefore, these concepts should have been used by the students for several months. They learn the mechanics, and students' understanding of these concepts must be in-depth through applications in different contexts. Data collected were then analyzed using WINSTEPS 3.72 software to carry out a Rasch model using a logit measurement unit.

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

III. Findings And Discussions

Rasch model analysis can measure FCI diagnostic test findings with more in-depth explanations. An essential feature of the Rasch model is that it allows users to create an in-between scale for scores for items and persons. The first requirement of this analysis is that the variables will be measured by the test (Planinic, Ivanjek, & Susac, 2010). As an example of Newton's concept of force and motion and law. These variables are expressed and described in a selected set of test items. A person's position throughout the measured variable is described by a measure called person ability, which provides information about the tendency of the measured variable possessed by the individual.

Person reliability values mean that the assumptions of individual capabilities in the sample are consistent despite being given different sets of items but still measuring the same constructs (Linacre, 2012). An item's reliability value means that the item is sufficient to measure what it wants to measure (Azrilah, Mohd Saifudin, & Azami, 2013). A separation index is an index that classifies persons or items into groups. Person separation index values and items higher than value 2 are considered good (Linacre, 2012).

The summary statistics in Figure 1 show the reliability and indices of the person. This statistic also shows the item's compatibility with Rasch's model. Based on Figure 1, the individual distribution is 5.50 logit (the spread between the maximum measurement of 0.74 and the minimum of -4.76) with the person separation index, the G for the FCI test is 0.69 (G < 2.0). This index value indicates that the ability of individual items to discriminate is low. The items of FCI were not sensitive enough to decsriminate between low and high performance (Gothwal, Wright, Lamoureux, & Pesudovs, 2010; Linacre, 2012)While the person reliability was 0.32 (KR20> 0.5), it was less than the Cronbach's alpha KR20 = 0.37. This indicates low reliability. The main findings of this study are the person mean, µperson = -1.47logit, which is lower than the mean item, Min Item = 0.17. These values clearly indicate that the level of understanding of the student's concept is lower than expected.

SUMMARY OF 78 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	5.2	25.0	-1.67	.57				
S.D.	2.4	.0	.74	.17				
MAX.	15.0	25.0	.60	1.84				
MIN.	.0	25.0	-4.76	.45	.70	-1.9	.54	-1.4

REAL RMSE	.61	TRUE SD	.42	SEPARATION	.69	PERSON RELIABILITY	.32
MODEL RMSE	.60	TRUE SD	.44	SEPARATION	.74	PERSON RELIABILITY	.35

S.E. OF PERSON MEAN = .08

PERSON RAW SCORE-TO-MEASURE CORRELATION = .95
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .37

Figure 1: Analysis of Person Separation and Reliability Index for FCI Test

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

Figure 2 shows the reliability and indices of the items. Based on summary in Figure 2 the item separation index is 2.29 (> 2.0) and the item reliability is 0.84 (> 0.8). It shows high reliability (Bond & Fox, 2007). Item distribution is good with 5.40logit with a standard deviation of item 1.18. These FCI items are tests that measure the level of understanding of student concepts.

```
     SUMMARY OF 25 MEASURED (EXTREME AND NON-EXTREME) ITEM
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      16.2      78.0         .17     .40                                 |
| S.D.      10.3        .0        1.29     .31                                 |
| MAX.      32.0      78.0        4.14    1.82                                 |
| MIN.        .0      78.0       -1.26     .24        .90   -1.2    .75   -1.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .52 TRUE SD    1.18  SEPARATION  2.29  ITEM   RELIABILITY  .84 |
|MODEL RMSE    .51 TRUE SD    1.18  SEPARATION  2.31  ITEM   RELIABILITY  .84 |
| S.E. OF ITEM MEAN = .26                                                     |
-------------------------------------------------------------------------------
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.91
1848 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1634.48 with 1748 d.f. p=.9745
Global Root-Mean-Square Residual (excluding extreme scores): .3769
```

Figure 2:Analysis of Item Separation and Reliability Index for FCI Test

Figure 3 shows a map of individual items for the FCI pilot test for a secondary school taking physics subjects. The left side of the map shows the distribution of student abilities, and the right shows the distribution of item difficulty. Items are labeled as Q1-Q25. M, S, and T are labels for mean values, one standard deviation, and two standard deviations for each distribution.

When ratings of items and persons are obtained, they arrange on a vertical line (Figure 3) that measures an individual's ability to test. This map can show an individual's capabilities and the difficulty of the items on the same logit scale. To the right of the vertical line is the FCI item arranged by difficulty level. The order of the items in the highest order is the complexity of the item, and at the bottom of the map is the easier item (Azrilah et al. 2013).

Based on Figure 3, Q3 and Q24 are the simplest items with a measured value of -1.26 (Figure 4), respectively. Meanwhile, Q21 is the hardest item with a measurement value of 4.14 (Figure 4). There is one student of the highest ability and another of the lowest ability. Based on the individual items map, most students are not able to answer these items correctly. 0This indicates that students' level of understanding of the concepts of force and motion is low.

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

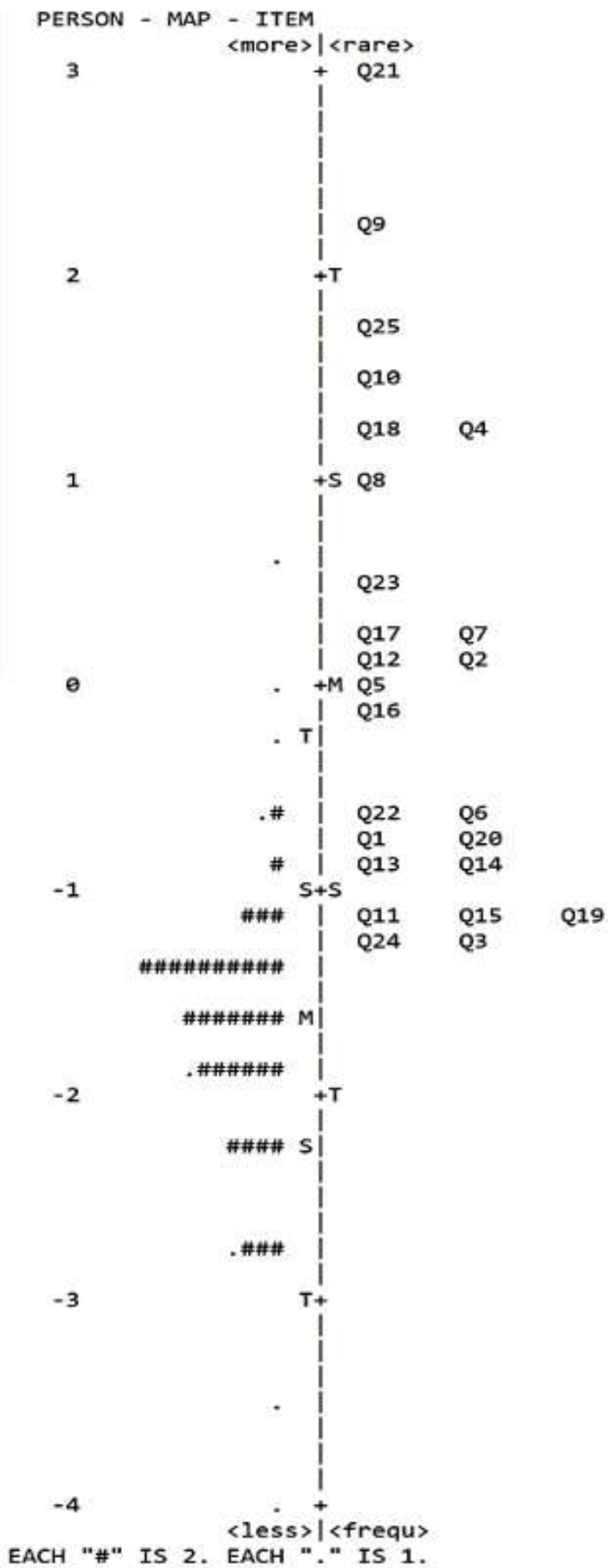

Figure 3: Individual Item Map for Secondary School Students. Each “#” Represents Two Students and “.” Represents One Student Each

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
21	0	78	4.14	1.82	MAXIMUM MEASURE				.00	.00	100.0	100.0	Q21
9	2	78	2.21	.72	1.03	.3	.94	.2	.07	.11	97.4	97.4	Q9
25	3	78	1.78	.60	1.05	.3	1.02	.3	.06	.13	96.1	96.1	Q25
10	4	78	1.47	.52	1.08	.3	1.38	.8	.00	.15	94.8	94.8	Q10
4	5	78	1.23	.47	1.04	.2	1.16	.5	.09	.17	93.5	93.5	Q4
18	5	78	1.23	.47	.94	.0	.78	-.3	.25	.17	93.5	93.5	Q18
8	6	78	1.02	.43	.94	-.1	.83	-.2	.26	.18	92.2	92.2	Q8
23	9	78	.55	.36	.90	-.3	.75	-.7	.35	.21	89.6	88.3	Q23
7	11	78	.31	.34	1.13	.6	1.22	.8	.05	.22	84.4	85.9	Q7
17	11	78	.31	.34	.98	.0	1.21	.8	.19	.22	87.0	85.9	Q17
2	13	78	.10	.31	1.09	.5	1.02	.2	.14	.23	81.8	83.4	Q2
12	13	78	.10	.31	1.03	.2	.96	-.1	.21	.23	81.8	83.4	Q12
5	14	78	.00	.31	1.10	.6	1.42	1.6	.04	.24	80.5	82.2	Q5
16	15	78	-.09	.30	.97	-.1	1.01	.1	.27	.24	83.1	81.0	Q16
6	21	78	-.56	.27	1.06	.5	1.30	1.7	.14	.27	76.6	73.9	Q6
22	22	78	-.63	.26	.98	-.1	.98	-.1	.29	.27	72.7	72.7	Q22
1	23	78	-.70	.26	.91	-.8	.88	-.8	.39	.28	75.3	71.6	Q1
20	23	78	-.70	.26	1.02	.3	.99	.0	.25	.28	70.1	71.6	Q20
14	25	78	-.83	.25	.97	-.3	.94	-.4	.33	.28	70.1	69.5	Q14
13	26	78	-.90	.25	1.03	.3	1.00	.1	.26	.28	63.6	68.4	Q13
11	29	78	-1.08	.25	.95	-.6	.93	-.6	.36	.29	72.7	65.4	Q11
15	30	78	-1.14	.24	.92	-1.0	.88	-1.2	.41	.29	61.0	64.5	Q15
19	30	78	-1.14	.24	.95	-.6	.94	-.5	.35	.29	66.2	64.5	Q19
3	32	78	-1.26	.24	1.03	.4	1.04	.4	.26	.30	61.0	62.9	Q3
24	32	78	-1.26	.24	.91	-1.2	.87	-1.4	.42	.30	66.2	62.9	Q24
MEAN	16.2	78.0	.17	.40	1.00	.0	1.02	.0			79.7	79.4	
S.D.	10.3	.0	1.29	.31	.07	.5	.17	.8			11.5	11.6	

Figure 4: Displayed are the Total Raw Score, Item Measure in Logit, Rasch Standard Error, Infit and Outfit MNSQ Statistics, and Point- Measure Correlation for Each Item

In general, Rasch model provides two reliability - person reliability and item reliability. Acceptable reliability values are between 0.6 and 0.8 (Bond & Fox, 2007; Linacre, 2012). It is known that person reliability is independent of sample size. Person reliability was found to be due to the distribution of students' understanding of the concepts of force and motion that are about similar. This finding is consistent with the of person-items map from the study of Planinic et al. (2010) which indicates that the distribution of students who can accurately answer the test item is low. Similar problems have been encountered for high school students such as in the United States (Hake, 1997, 1998) and Iran (Fadaei & Mora, 2015). Most students find the FCI conceptual questions difficult and this, in line with their findings Fadaei & Mora (2015), where most students only get 21% of the FCI test. Other Malaysian studies using this test also reported difficulty in understanding the concept of force and motion as the study of Ahmad Tarmimi & Shahrul Kadri (2016) and Mohammad Mubarrak et al.(2013).However, the physicist's view of FCI testing is simple and straightforward (Rosenblatt, 2012).

IV. CONCLUSION

The analysis shows that this FCI has a high reliability of the item. Moreover, it can determine the level of conceptual understanding. Item split values indicate that two levels of good and bad distinguish students' understanding. As such, FCI is an important instrument that has contributed significantly to the development of physics education research and the changing practice of physics teaching worldwide. It is a measuring instrument in

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020

the study of physics education, and therefore it needs to be carefully measured, examined, and monitored, just like other measuring instruments in physics. The reliability of FCI items nevertheless, is high and is able to separate an individual's capabilities.In conclusion, FCI is the best test to measure the level of understanding of concepts of force and motion.

V. ACKNOWLEDGEMENT

The project is supported by the Ministry of Education under the Fundamental Research Grant Scheme (FRGS) vot number R. J130000.7853.5F140 and Universiti Teknologi Malaysia under the International / Industry Incentive Grant (IIIG) vot number Q.J130000.3053.01M59.

REFERENCES

[1] Adibah, A. L., Ibnatul Jalilah, Y., Nor Fadila, M. A., Libunao, W. H., & Siti Sarah, Y. (2016). Multiple-choice items analysis using classical test theory and rasch measurement model. *Man in India*, *96*(1–2), 173–181.

[2] Ahmad Tarmimi, I., & Shahrul Kadri, A. (2016). Tahap Kefahaman Dan Salah Konsep Terhadap Konsep Daya Dan Gerakan Dalam Kalangan Pelajar Tingkatan Empat. *Jurnal Fizik Malaysia*, *37*(1), 01090–01101.

[3] Arhin, A. K. (2015). The Effect of Performance Assessment-Driven Instruction on the Attitude and Achievement of Senior High School Students in Mathematics in Cape Coast Metropolis , Ghana. *Journal of Education and Practice*, *6*(2), 112–114.

[4] Azrilah, A. A., Mohd Saifudin, M., & Azami, Z. (2013). *Asas Model Pengukuran Rasch Pembentukan Skala & Strktur Pengukuran*. *Bangi: Penerbit Universiti Kebangsaan Malaysia.*

[5] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model : Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Earlbaum Associates. https://doi.org/10.1111/j.1745-3984.2003.tb01103.x

[6] Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (sixth edit). New York: Routledge.

[7] Emanuela Ene, & Ackerson, B. J. (2018). Assessing Learning in Small Sized Physics Courses. *Phys. Rev. Phys. Educ. Res.*, *14*(010102).

[8] Fadaei, A. S., & Mora, C. (2015). An Investigation About Misconceptions in Force and Motion in High School. *US-China Education Review*, *5*(1), 38–45.

[9] Flanders, S. T. (2014). *INVESTIGATING FLEXIBILITY, REVERSIBILITY, AND MULTIPLE REPRESENTATIONS IN A CALCULUS ENVIRONMENT*. University of Pittsburgh.

[10] Ghafar, M. N. A. (2000). *Penyelidikan Pendidikan*. Skudai, Johor.: Penerbit Universiti Teknologi Malaysia.

[11] Gothwal, V. K., Wright, T., Lamoureux, E. L., & Pesudovs, K. (2010). Psychometric properties of visual functioning index using Rasch analysis. *Acta Ophthalmologica*, *88*(7), 797–803.

[12] Hake, R. R. (1997). Evaluating conceptual gains in mechanics: A six thousand student survey of test data. In *AIP Conference Proceedings* (Vol. 399, pp. 595–604). *AIP*.

[13] Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.

[14] Heller, K., & Heller, P. (2010). *Cooperative Problem Solving in Physics A User ' s Manual Can this be true ?*

[15] Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, *30*(3), 159–166.

[16] Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, *30*(3), 141–158.

[17] Khairani, A. Z., & Abd. Razak, N. (2015). Modeling a Multiple Choice Mathematics Test with the Rasch Model. *Indian Journal of Science and Technology*, *8*(12). https://doi.org/10.17485/ijst/2015/v8i12/70650

[18] Kumar, R. (2011). *Research Methodology:a step-by-step guide for beginners*. *SAGE Publication*. London. https://doi.org/http://196.29.172.66:8080/jspui/bitstream/123456789/2574/1/Research%20Methodology.pdf

[19] Lin, S. Y., & Singh, C. (2011). Using isomorphic problems to learn introductory physics. *Physical Review Special Topics - Physics Education Research*, *7*(2), 1–16.

[20] Linacre, J. M. (2012). *A user's guide to WINSTEPS: Rasch Model Computer Programs*.

https://doi.org/ISBN 0-941938-03-4

[21] Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *Online Submission*, (May).

[22] Mcdermott, M. A. (2009). *The impact of embedding multiple modes of representation on student construction of chemistry knowledge by*. University of Iowa.

[23] Mohammad Mubarrak, M. Y., Siti Fairuz, D., & Norezan, I. (2013). the Level of Understanding of Students and Teachers in the Concept of Force and Motion. *International Conference on Social Science Research, ICSSR 2013, (e-ISBN 978-967-11768-1-8), 2013*(June), 1083–1090.

[24] Persson, J. R. (2015). Evaluating the Force Concept Inventory for different student groups at the Norwegian University of Science and Technology, 1–13.

[25] Planinic, M., Ivanjek, L., & Susac, A. (2010). Rasch model based analysis of the Force Concept Inventory. *Physical Review Special Topics - Physics Education Research*, *6*(1), 1–11.

[26] Rosenblatt, R. (2012). Identifying and addressing student difficulties and misconceptions : examples from physics and from materials science and engineering. *Proquest Llc*, (March).

[27] Savinainen, A. (2005). *High School Students ' Conceptual Coherence of Qualitative Knowledge in the Case of the Force Concept*. *Nordic Studies in Science Education* (Vol. 1).

[28] Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. In *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*.

[29] Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, *66*(4), 338–352.

DOI: 10.37200/IJPR/V24I5/PR201677
Received: 05 Feb 2020 | Revised: 01 Mar 2020 | Accepted: 18 Mar 2020