JURNAL PENDIDIKAN DAN PENGAJARAN
Volume 57 Nomor 1 2024, 127-138
E-ISSN: 2549-2608; P-ISSN: 2301-7821
DOI: https://doi.org/10.23887/jpp.v57i1.57883

Is it True That My Students don't Understand the Static Fluid Concepts? Rasch Modeling Perspective

Moh. Irma Sukarelawan[1], Sriyanto[2], Miftakhatun Sa'adah[3], Jimmy Jupri[4], Fitri Nur Hikmah[5], Eko Nursulistiyo[6]*

[1,3,4,6] Department of Physics Education, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Indonesia
[2] State of Senior High School 2 Bantul, Yogyakarta, 55714, Indonesia
[5] Universitas Islam Negeri Antasari, Banjarmasin, Indonesia
*Corresponding author: fitri.nurhikmah@uin-antasari.ac.id

Abstrak

Analisis pemahaman konsep siswa dapat dilakukan dengan menggunakan pendekatan Classical Test Theory (CTT) dan Item Response Theory (IRT). CTT hanya dapat menjelaskan pemahaman konsep siswa pada tingkat kelompok, sedangkan IRT dapat menjelaskan sampai pada tingkat individu. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi pemahaman konsep siswa dengan menggunakan model Rasch. Penelitian ini menggunakan jenis penelitian survei dan dilaksanakan di kelas XI. Pemahaman konsep siswa dikumpulkan menggunakan 10 soal pilihan ganda pada materi fluida statis. Tingkat pemahaman konsep dianalisis secara bertahap menggunakan pemodelan Rasch. Analisis pemahaman dimulai dengan menganalisis tingkat kesulitan soal, kemudian dilanjutkan dengan analisis deskriptif. Peta Wright digunakan untuk melihat hirarki antara pemahaman konsep siswa dengan tingkat kesulitan soal yang digunakan. Terakhir, menggunakan peta person diagnostic untuk melihat pemahaman konsep siswa secara detail. Hasil analisis menunjukkan bahwa pemahaman konsep siswa secara umum berada pada kategori tinggi dan sedang. Peta person diagnostic telah mengidentifikasi pola jawaban benar dan salah dari siswa yang tidak menunjukkan kemampuan sebenarnya. Maka, perlu dianalisis letak yang tepat dari kelemahan dan kelebihan pemahaman konsep siswa. Hal ini berimplikasi pada kesesuaian rencana pembelajaran yang akan digunakan guru.

Kata Kunci: *Teori Tes Klasik, Pemahaman Konseptual, Peta Person Diagnostic, Pemodelan Rasch*

Abstract

Students' conceptual understanding can be analysed using the Classical Test Theory (CTT) and Item Response Theory (IRT) approaches. CTT can only explain students' conceptual understanding at the group level, while IRT can at the individual level. Therefore, this research evaluates students' conceptual understanding using the Rasch model. This research used a survey research type and was carried out in class XI. Students' conceptual understanding was collected using 10 multiple-choice questions on static fluid material. The level of concept understanding was analyzed in stages using Rasch modelling. Understanding analysis begins by analyzing the difficulty level of the questions and then continues with descriptive analysis. The Wright map is used to see the hierarchy between students' understanding of concepts and the difficulty level of the questions used. Finally, a person diagnostic map will be used to see students' understanding of concepts in detail. The analysis results show that students' conceptual understanding is generally in the high and medium categories. Person diagnostic maps have identified patterns of correct and incorrect answers from students that do not reflect their true abilities. So, it is necessary to analyze the exact location of the weaknesses and strengths of students' conceptual understanding. This has implications for the suitability of the learning plan that the teacher will use.

Keywords: Classical Test Theory, Conceptual understanding, Person diagnostic map, Rasch modeling

History:
Received : July 07, 2023
Accepted : December 06, 2023
Published : April 25, 2024

Publisher: Undiksha Press

1. INTRODUCTION

Evaluation of students' understanding of physics concepts can be done before, during, or after the implementation of learning. Evaluation of conceptual understanding before learning is used to identify students learning needs and assist teachers in designing appropriate learning models or strategies (Atasoy & Kaya, 2022; English et al., 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022). The teacher can predict which part of the concept needs to be emphasized. This is done as a preventive effort against students' low understanding of concepts. Evaluation of conceptual understanding during learning will make it easier for the teacher to evaluate the success of implementing the teaching carried out at that time (Cizek & Lim, 2023; Rasmitadila et al., 2020). The teacher can immediately provide feedback if it is found that there are students who need help understanding the concept being discussed. Meanwhile, the evaluation of conceptual understanding carried out after learning can be used as material for self-reflection and designing other models or strategies that are more appropriate (McInerney & Kerrigan, 2022; Mohamadi, 2018). Teachers can also design remedial learning for students who have not mastered the concept well or provide enrichment for students who have mastered the concept (Chen, 2011).

Understanding the concept of physics is one indicator of student learning success. Improper understanding of concepts will have a negative impact on an incomplete understanding of more complex concepts. It can even lead to misconceptions for students. Students who have a good understanding of concepts do not just know but can apply them (Pranata, 2016; Shidik, 2020). A good conceptual understanding will help students solve problems or find solutions to their problems. Students with a strong understanding of physics concepts can convey new ideas in various representations (Riwanto et al., 2019; Saputro et al., 2019). Historically, researchers and teachers have measured students' understanding of physics concepts using conventional multiple-choice assessments (Saputro et al., 2019). This test was chosen for practical reasons. In its journey, the use of conventional multiple-choice began to be refined into various models because it needed to provide more information. Conventional multiple choice has been modified into a reasoned multiple-choice, two-tier multiple choice three-tier multiple choice to four-tier multiple choice (Diani et al., 2019; Fenditasari et al., 2020; He et al., 2022; Irwansyah et al., 2018; Ivanjek et al., 2021; Sukarelawan et al., 2019; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Suma et al., 2018; Umam et al., 2020). All modifications that have been made are to get an understanding of the concept that is close to the actual state of the students. This is the impact of the limited use of classical test theory analysis.

Classical Test Theory (CTT) and item response theory (IRT) are two approaches that can be used to conduct individual assessments (Ayanwale et al., 2022; Jabrayilov et al., 2016). The analysis of conceptual understanding using the classical test theory approach has several limitations. The classical test theory uses a group-centered approach so that only a little information is obtained about students' individual understanding of concepts. Therefore, a new approach is needed that can provide an overview according to what is needed by the teacher in an effort to improve the quality of the learning that will be carried out. An alternative to the CTT is Rasch Modeling, based on item response theory (IRT). Rasch modeling is individual-centered. This means that the Rasch modeling can analyze down to the individual level so that it allows the teacher to obtain an overview of students' conceptual understanding according to their circumstances (Sumintono & Widhiarso, 2014). This will have implications for the suitability of the lesson plan that the teacher will prepare. In addition, IRT has a much smaller measurement error than the CTT approach (Gorter et al., 2020; Magno, 2009). Therefore, this study aims to evaluate students' understanding of concepts using Rasch modeling. We took the static fluid material as an example because the time this research was conducted corresponds to that material.

2. METHODS

This research is a type of survey research that was conducted at one of the State Senior High Schools in the city of Yogyakarta (Pandey & Pandey, 2015). The survey was carried out on class XI students who took Physics as a subject. The survey was carried out before students received Static Fluid material. The number of students surveyed was 35 students. The selection of students was carried out because of the ease of getting respondents. Initial mastery of the concept was evaluated using 10 multiple-choice questions on static fluid material. The questions used were developed from class XI physics books which will be used for the next semester. Before being used, the questions were evaluated by colleagues. Suggestions from colleagues are used to refine the questions used. Questions that have been corrected are formatted in the form of a Google form. This is done to facilitate the administration and recording of student answers.

Testing students' initial conceptual understanding is carried out about 20 minutes before learning is carried out. Before the google form was shared, we explained that this test was carried out to map the state of understanding the initial concept. It is intended that students answer without feeling pressured, and the answers given can describe the actual condition of students' conceptual understanding. The data that has been collected is administered using the help of MS Excel. Then, students' initial conceptual understanding was analyzed using the Winsteps 4.1.6 application (Linacre, 2021). Students' initial understanding was analyzed descriptively then the hierarchical level of students' conceptual understanding was visualized using a Wright map (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). In the final section, we use the person diagnostic map to look in detail at the weaknesses and strengths of individual student's mastery of concepts. In this case, we took a sample of two students with the highest and lowest knowledge levels.

3. RESULTS AND DISCUSSION

Result

The quality of the items used is shown in Table 1, and a summary of the suitability of the items used is shown in Table 2.

Table 1. Item Quality Summary

Index	Nilai
Mean Item	0.00
Std. Dev. Item	2.42
Reliabilitas Item	
Strata	4.61
Reliability	0.91
Dimensionality: Raw variance explained by measures	51.0%

Table 2. Item Fit Summary

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	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	23	35	.60	.42	.89	-.46	1.00	.10	.56	.51	85.7	76.7	Q1
2	29	35	-.65	.52	1.18	.65	.79	-.16	.42	.47	77.1	85.4	Q2
3	22	35	.77	.41	1.37	1.74	1.35	1.19	.27	.50	60.0	75.4	Q3
4	30	35	-.93	.55	.69	-.91	.32	-1.08	.67	.45	91.4	86.8	Q4
5	11	35	2.44	.40	1.23	1.43	1.44	1.09	.21	.40	71.4	70.5	Q5
6	34	35	-2.96	1.05	.73	-.05	.13	-.58	.41	.24	97.1	97.1	Q6
7	32	35	-1.65	.66	.61	-.90	.20	-.88	.62	.38	94.3	91.4	Q7
8	34	35	-2.96	1.05	.73	-.05	.13	-.58	.41	.24	97.1	97.1	Q8
9	1	35	5.50	1.03	1.06	.37	.72	.25	.13	.16	97.1	97.1	Q9
10	27	35	-.17	.47	1.05	.28	.75	-.43	.51	.50	77.1	82.4	Q10
MEAN	24.3	35.0	.00	.65	.96	.2	.68	-.1			84.9	86.0	
P.SD	10.2	.0	2.42	.26	.25	.8	.46	.7			12.3	9.2	

Based on Table 1, several important informations were found. The item average is 0.00, with a standard deviation of 2.42, describing the distribution of student responses to these items. High item reliability reached 0.91 and the strata of 4.61, indicating that the instrument used was very consistent and reliable. In addition, it was found that the Raw variance explained by measure was 51.0%, indicating the instrument's ability to explain variations in the data. Finally, in measuring the degree of concordance between student responses and items, the MnSq infit and outfit values were in the range of 0.61 to 1.37 and 0.13 to 1.44, reflecting the varying difficulty levels of the items in the instrument.

Item Difficulty Level

The difficulty level of the questions used to evaluate students' conceptual understanding of static fluid material is shown in Table 3.

Table 3. Item Difficulty Level

Item	Description	Logit	Level of difficulty
Q1	Definition of Fluid	0.60	Difficult
Q2	Applying a Static Fluid	-0.65	Easy
Q3	Type of Magnitude of Pressure	0.77	Difficult
Q4	Pressure Unit Conversion	-0.93	Easy
Q5	Hydrostatic Pressure	2.44	Very Difficult
Q6	Archimedes' Law	-2.96	Very Easy
Q7	Surface Tension	-1.65	Easy
Q8	Density	-2.96	Very Easy
Q9	Capillarity Events	5.50	Very Difficult
Q10	Viscosity	-0.17	Easy
Mean		**0.00**	
P.SD		**2.42**	

The analysis results show various information regarding the item's difficulty level. The average difficulty level of the questions is 0.00 logit with a standard deviation of 2.42. In this collection of questions, 2 out of 10 questions, namely Q6 and Q8, are classified as a very easy difficulty level with a logit value of -2.96. A total of 4 questions, or 40%, can be categorized as easy questions. Meanwhile, the questions that fall into the difficult and very difficult categories are 20% each. Students also identified Q9 as the most difficult question, with a logit score 5.50.

Description of Students' Conceptual Understanding

Descriptively, the percentage of students' initial understanding of static fluid material is summarized in Table 4.

Table 4. Descriptive Students' Conceptual Understanding

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item
1	0	0	12	34	.41	1.13	.34	.8	.8	-.56	Q1
	1	1	23	66	1.90	1.02	.22	1.0	1.3	.56	
2	0	0	6	17	.23	.82	.36	1.0	.7	-.42	Q2
	1	1	29	83	1.63	1.21	.23	1.4	1.3	.42	
3	0	0	13	37	.94	1.31	.38	1.3	1.3	-.27	Q3
	1	1	22	63	1.65	1.17	.25	1.5	1.5	.27	

4	0	0	5	14	-.69	.33	.17	.5	.2	-.67	Q4
	1	1	30	86	1.74	1.01	.19	.9	.9	.67	
5	0	0	24	69	1.21	1.28	.27	1.1	1.1	-.21	Q5
	1	1	11	31	1.79	1.15	.36	1.5	1.6	.21	
6	0	0	1	3	-1.68	.00		.4	.1	-.41	Q6
	1	1	34	**97**	1.48	1.17	.20	.7	.8	.41	
7	0	0	3	9	-1.20	.34	.24	.4	.2	-.62	Q7
	1	1	32	91	1.63	1.03	.19	.6	.7	.62	
8	0	0	1	3	-1.68	.00		.4	.1	-.41	Q8
	1	1	34	**97**	1.48	1.17	.20	.7	.8	.41	
9	0	0	34	**97**	1.36	1.28	.22	1.1	1.0	-.13	Q9
	1	1	1	3	2.32	.00		1.8	.7	.13	
10	0	0	8	23	.20	1.03	.39	.9	.6	-.51	Q10
	1	1	27	77	1.74	1.11	.22	1.2	1.1	.51	

Based on the results from Table 4, it can be seen that questions Q6, which relates to "Archimedes' Law," and Q8, which relates to "Density," had the highest percentage of correct answers. The average distance between the number of correct and incorrect answers to the two questions is 3.16 logits. However, the most difficult question is Q9, which tests understanding capillarity events. Only 3% of students could answer this question correctly, and the average ability gap between correct and incorrect answers in Q9 was 0.96 logit.

Hierarchy of Students' Conceptual Understanding

The hierarchy of students' conceptual understanding is illustrated through the Wright map in Figure 1.

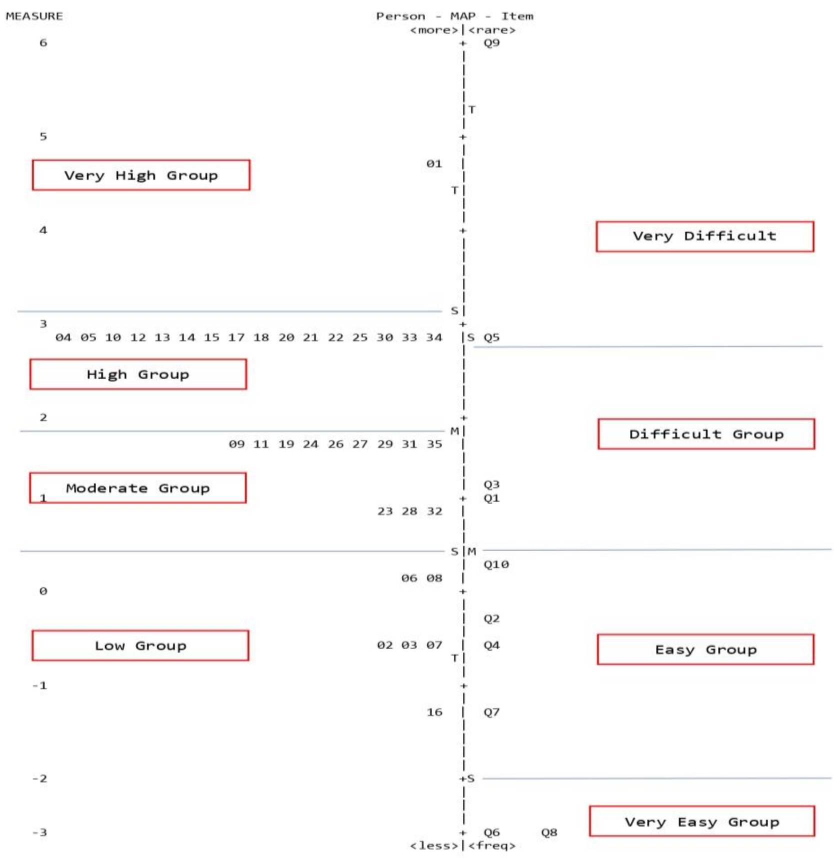
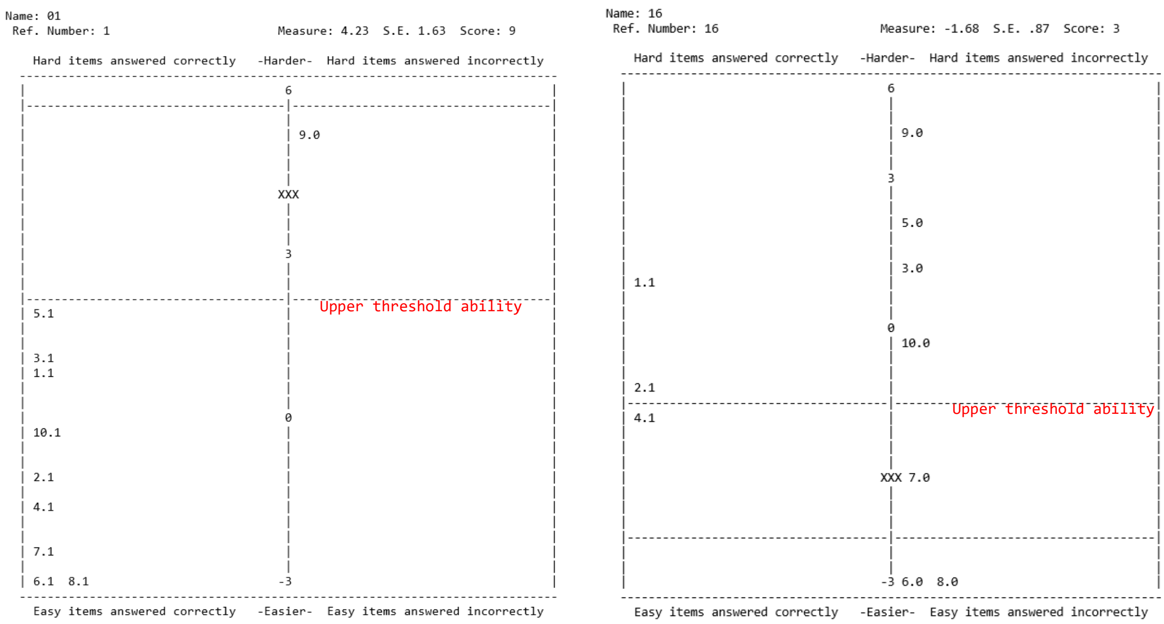


Figure 1. Hierarchy of Students' Initial Conceptual Understanding by Wright Map

Figure 1 is a visual representation that maps student abilities and the difficulty level of the questions simultaneously on one logit ruler. In the picture, most students, namely 16 people, greatly understand the concept. They have more than a 50% chance of mastering one question, namely Q5, which is included in the group of very difficult questions. Apart from that, the possibility of questions Q6, Q7, and Q8 being mastered by all students exceeds 50%. However, question Q9 shows a probability of less than 50% to be mastered by all students.

Analysis of Individual Student Abilities

Based on Figure 1, students with the highest and lowest conceptual understanding are students in codes 01 and 16. The results of the person diagnostic map analysis are shown in Figure 2.



(a) (b)

Figure 2. Diagnostics of Students' Individual Conceptual Understanding

Figure 2 is a diagnostic map display that describes the condition of students' abilities. Part Figure 2(a) visualizes the ability of student 01, who has the highest conceptual understanding with a logit of 4.23 and achieved a score of 9 in the evaluation. Meanwhile, in Figure 2(b), the condition of students' abilities with the lowest understanding of concepts can be seen. Student 16 had a conceptual understanding of -1.68 logits and managed to get a score of 3 on the assessment.

Discussions

Item Quality

Table 1 shows that the questions used are of good quality. The item separation index is 3.21, meaning that the items' difficulty levels can be grouped into four levels. At the same time, the reliability value is 0.91, which indicates that the reliability of the item is in the excellent category (Adams et al., 2020; Linacre & Wright, 2012). While the item match rate in Table 2 refers to the MnSq infit and outfit values. The MnSq infit value of all items is in the range of 0.61 - 1.37, and the MnSq outfit value is in the range of 0.13 - 1.44. Some items with outfit MnSq values outside the acceptance range (0.50 – 1.50) are retained because all Pt.Mea.Corr values to support them. Pt.Mea.Corr positive value indicates that all items function in the same direction to predict latent traits (Boone et al., 2014; Liu et al., 2022).

The items' dimensionality is indicated by the Raw variance explained by measures, which is 51.0%. Unidimensionality describes items in a measuring tool to measure a single ability (Baghaei, 2013; Tabatabaee-Yazdi et al., 2018). Recommended value > 20%.

Item difficulty level

Table 3 shows that the distribution of the difficulty level of the questions is almost even, from very easy to very difficult. The difficulty level of these questions is grouped based on the Logit Value of Item (Adams et al., 2021; Puspitasari et al., 2022; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The average difficulty level of the questions was 0.00 logit with a standard deviation of 2.42. As many as 2 out of 10 questions (Q6 and Q8) are at a very easy difficulty level with a logit value of -2.96. A total of 4 questions (40%) are in the easy category. While questions that are in the category of difficult and very difficult each as much as 20%. According to students, the most difficult question was Q9, with a logit score of 5.50.

Description of students' conceptual understanding

Based on Table 4, questions Q6, "Archimedes' Law," and Q8, "Density," have the highest percentage of correct answers. These two concepts have been mastered by 97% of students. The mean capability distance for the number of correct and incorrect answers to these two questions is 3.16 logit each. The most difficult question is Q9 about capillarity events. The concept of capillarity can only be answered by 3% of students. The mean ability distance between right and wrong answers in Q9 is 0.96 logit. The concepts of Archimedes' law, density, and capillarity have been introduced in the science curriculum at the junior high school level. However, the concept of Archimedes' law and density is easier for students to understand than the concept of capillarity. In contrast to the findings reported by previous studies analyzed the understanding of 11 students' concepts on fluid material using three-tier multiple choice (Irwansyah et al., 2018; Koto & Ilhami, 2023). They reported that the understanding of class XI students on the concept of capillarity was in second place, understood by 35.04% of students. At the same time, the concepts of Archimedes and density are in the order of six and eight.

The initial screening step is the percentage of students who answered correctly or incorrectly in Table 4. Do the correct answers or wrong answers given by students describe their abilities? We must look again through the Wright map to understand the concepts described in Table 4. The distribution of students' conceptual understanding can be investigated accurately through the hierarchical location of student abilities and the questions' difficulty level in the Wright map (Hikmah et al., 2021; Nursulistyo et al., 2022).

Hierarchy of students' Conceptual understanding

Figure 1 maps the students' abilities and difficulty level of the questions simultaneously in one logit rule. The logit ruler values range from -3 to 6 logits. The left part of Figure 1 visualizes the level of students' understanding. Student understanding is distributed from the bottom (students with low knowledge) to the top (students with high knowledge). The right part of Figure 1 visualizes the items' difficulty levels according to students. Items that are easiest for most students to answer will be placed at the bottom, and items that are most difficult for most students to answer will be placed at the top (Karini et al., 2022; Sukarelawan, Gustina, et al., 2021). Most students (16 people) have a high understanding of the concept. They have a > 50% chance of mastering 1 question (Q5) in the very difficult item group. At the same time, most students in the moderate group have a >50% chance of mastering all the questions in the difficult group. Meanwhile, groups of students who have low abilities have the opportunity to master questions in easy and very easy groups.

If an in-depth analysis of the results shown in Table 2 is carried out, the chances of questions Q6, Q7, and Q8 being mastered by all students are > 50%. This is indicated by the locations Q6, Q7, and Q8, which are below the ability locations of all students. Question Q9 has a probability of < 50% being mastered by all students. From another perspective, code 01 students have the highest conceptual understanding because they are in the highest logit location than the others. Student 01 has a > 50% chance of mastering the nine questions tested. At the same time, code 16 students have the lowest conceptual understanding because their logit location is the lowest of all students. However, Wright showed that student 16 had the opportunity to master three questions (Q6, Q7, and Q8).

Analysis of individual student abilities

Figure 2 is a person diagnostic map display. The person diagnostic map can estimate the pattern of responses or answers for each student. The student's "correct" response pattern can be detected, whether because of the student's ability or because of guessing or cheating. Likewise, the pattern of "wrong" responses from students, whether purely because they did not understand or were not careful. The map is divided into two parts, namely the left and right. The left side of the map shows questions that were answered “correctly” (coded with number 1), and the right part shows questions answered “incorrectly” by individual students (coded with number 0). The symbol “xxx” is the location of the student’s ability, and the two horizontal dashed lines indicate the upper limit of the student’s ability.

More specifically, maps/graphs can be divided into four sections. The top-right section (quadrant 1) is a question with a higher difficulty level than the student’s ability and is answered “incorrectly.” The top-left section (quadrant 2) is a question with a higher difficulty level than the student’s ability and is answered “correctly.” The bottom-left (quadrant 3) are questions that have a lower level of difficulty than the student’s ability and are answered “correctly.” While the bottom-right are questions that have a low level of difficulty and are answered “incorrectly.”

Figure 2(a) visualizes the condition of student 01's ability as the student who has the highest conceptual understanding. Logit conceptual understanding of student 01 was 4.23 and received a score of 9. Nine items that were answered correctly were actually understood by student 01. This can be seen from the location of the nine items below their ability limits. Whereas question Q9 could not be answered by student 01 even though the difficulty level of question Q9 was within reach of his ability. Therefore, it is likely that student 01 needed to be more careful in answering question Q9. Teachers need to use a personal approach to train students' accuracy. So, based on Rasch's modeling, student 01 understands all the concepts being tested.

Figure 2(b) illustrates the condition of the student's ability with the lowest ability to understand the concept. Student 16 has a conceptual understanding of -1.68 logit and gets a score of 3. The investigation results using the person diagnostic map show that the locations of 2 of the three questions (Q1 and Q2) that can be answered are above their ability. There are indications of guesses made by student 16 on both questions. While questions Q6, Q7, and Q8 were below their ability limit, they could not be answered correctly. This indicates that student 16 needed to be more careful and thorough in answering Q6, Q7, and Q8. So, based on Rasch's modeling, 16 students understand four concepts (Q4, Q6, Q7, and Q8). Therefore, teachers need to use a personal approach to student 16 to emphasize six concepts that still need to be understood.

4. CONCLUSION

Rasch modeling is a technique of analyzing student abilities oriented toward individual needs. The level of understanding of individual students' concepts can be traced through the description stage of concept mastery, analysis of the Hierarchy of understanding concepts through a Wright map, then an analysis of students' abilities through a personal diagnostic map. The analysis results using Rasch modeling can identify conceptual understanding based on the thresholds of individual student abilities so that students learn according to their needs. The impact of this research is that the learning needs of individual students can be clearly identified so that the type of learning needed by students can be fulfilled. This study has presented how to evaluate conceptual understanding using Rasch modeling, which is comprehensively oriented to students' individual conditions. However, this research has yet to be continued to identify the types of answers given by students, whether from cheating or carelessness. Therefore, future researchers can use Rasch modeling to identify the types of student answers using the Guttman matrix or Scalogram. Combining the Scalogram and the diagnostic person map will make it easier for teachers to organize additional learning such as remedial, enrichment, or re-teaching.

5. REFERENCES

Adams, D., Chuah, K. M., Sumintono, B., & Mohamed, A. (2021). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: a Rasch analysis. *Asian Education and Development Studies*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/AEDS-05-2020-0100.

Adams, D., Joo, M. T. H., Sumintono, B., & Pei, O. S. (2020). Blended Learning Engagement in Public and Private Higher Education Institutions: A Differential Item Functioning Analysis of Students' Backgrounds. *Malaysian Journal of Learning and Instruction*, *17*(1), 133–158. https://doi.org/10.32890/mjli2020.17.1.6.

Atasoy, V., & Kaya, G. (2022). Formative assessment practices in science education: A meta-synthesis study. *Studies in Educational Evaluation*, *75*, 101186. https://doi.org/10.1016/j.stueduc.2022.101186.

Ayanwale, M. A., Chere-Masopha, J., & Morena, M. C. (2022). The Classical Test or Item Response Measurement Theory: The Status of the Framework at the Examination Council of Lesotho. *International Journal of Learning, Teaching and Educational Research*, *21*(8), 384–406. https://doi.org/10.26803/ijlter.21.8.22.

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, *28*(3), 1087–1103. https://doi.org/10.1007/s10212-012-0157-y.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands.

Chen, L.-H. (2011). Enhancement of student learning performance using personalized diagnosis and remedial learning system. *Computers & Education*, *56*(1), 289–299. https://doi.org/10.1016/j.compedu.2010.07.015.

Cizek, G. J., & Lim, S. N. (2023). Formative assessment: an overview of history, theory and application. In *International Encyclopedia of Education(Fourth Edition)* (pp. 1–9). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.09002-3.

Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-Tier Diagnostic Test With Certainty of Response Index on The Concepts of Fluid. *Journal of Physics: Conference Series*, *1155*, 012078. https://doi.org/10.1088/1742-6596/1155/1/012078.

English, N., Robertson, P., Gillis, S., & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, *113*, 101964. https://doi.org/10.1016/j.ijer.2022.101964.

Fenditasari, K., Jumadi, Istiyono, E., & Hendra. (2020). Identification of misconceptions on heat and temperature among physics education students using four-tier diagnostic test. *Journal of Physics: Conference Series*, *1470*(1), 012055. https://doi.org/10.1088/1742-6596/1470/1/012055.

Gorter, R., Fox, J.-P., Riet, G. Ter, Heymans, M. W., & Twisk, J. W. R. (2020). Latent growth modeling of IRT versus CTT measured longitudinal latent variables. *Statistical Methods in Medical Research*, *29*(4), 962–986. https://doi.org/10.1177/0962280219856375.

He, P., Zheng, C., & Li, T. (2022). Upper Secondary School Students' Conceptions of Chemical Equilibrium In Aqueous Solutions: Development and Validation of a Two-Tier Diagnostic Instrument. *Journal of Baltic Science Education*, *21*(3), 428–444. https://doi.org/10.33225/jbse/22.21.428.

Hikmah, F. N., Sukarelawan, M. I., Nurjannah, T., & Djumati, J. (2021). Elaboration of high school student's metacognition awareness on heat and temperature material: Wright map in Rasch model. *Indonesian Journal of Science and Mathematics Education*, *4*(2), 172–182. https://doi.org/10.24042/ijsme.v4i2.9488.

Irwansyah, Sukarmin, & Harjana. (2018). Analysis Profile of Student Misconceptions on The Concept of Fluid Based Instrument Three-Tier Test. *Journal of Physics: Conference Series*, *1097*, 012020. https://doi.org/10.1088/1742-6596/1097/1/012020.

Ivanjek, L., Morris, L., Schubatzky, T., Hopf, M., Burde, J.-P., Haagen-Schützenhöfer, C., Dopatka, L., Spatz, V., & Wilhelm, T. (2021). Development of a two-tier instrument on simple electric circuits. *Physical Review Physics Education Research*, *17*(2), 020123. https://doi.org/10.1103/PhysRevPhysEducRes.17.020123.

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, *40*(8), 559–572. https://doi.org/10.1177/0146621616664046.

Karini, R. A., Fikroh, R. A., & Cahyani, V. P. (2022). Identification of Students' Misconceptions on Hydrocarbon Material Using a Four-Tier Multiple Choice Diagnostic Test. *Jurnal Pendidikan Kimia Indonesia*, *6*(2), 79–87. https://doi.org/10.23887/jpki.v6i2.39022.

Koto, I., & Ilhami, D. (2023). *High-School Students' Conceptual Understanding of Fluid Dynamics Following Online Learning During the Coronavirus Pandemics*. Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-012-1_16.

Linacre, J. M. (2021). *Winsteps® (Version 4.6.1) [Computer Software]*.

Linacre, J. M., & Wright, B. D. (2012). *A User's Guide to WINSTEPS Ministeps Rasch Model Computer Programs*. Mesa Press.

Liu, F., Zhang, Z., Lin, B., Ping, Z., & Mei, Y. (2022). Assessing the psychometric properties of the Chinese return-to-work self-efficacy questionnaire using Rasch model analysis. *Health and Quality of Life Outcomes*, *20*(1), 27. https://doi.org/10.1186/s12955-022-01929-7.

Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, *1*(1), 1–11.

McInerney, I., & Kerrigan, E. C. (2022). Teaching Predictive Control Using Specification-based Summative Assessments. *IFAC-PapersOnLine*, *55*(17), 236–241. https://doi.org/10.1016/j.ifacol.2022.09.285.

Mohamadi, Z. (2018). Comparative effect of online summative and formative assessment on EFL student writing ability. *Studies in Educational Evaluation*, *59*, 29–40.

https://doi.org/10.1016/j.stueduc.2018.02.003.
Nursulistyo, E., Indratno, T. K., Dwiastuti, E., Arifiyanti, F., Puspitasari, A. D., Abdullah, N. S. Y. binti, & Sukarelawan, M. I. (2022). Perception Scale of Online Learning in the Indonesian Context During the Covid-19 Pandemic: Psychometric Properties Based on the Rasch Model. *Indonesian Review of Physics*, *5*(2), 49–56. https://doi.org/10.12928/irip.v5i2.6544.
Pandey, P., & Pandey, M. M. (2015). *Research Methodology: Tools and Techniques*. Bridge Center.
Pranata, E. (2016). Implementasi Model Pembelajaran Group Investigation (GI) Berbantuan Alat Peraga Untuk Meningkatkan Kemampuan Pemahaman Konsep Matematika. *JPMI (Jurnal Pendidikan Matematika Indonesia)*, *1*(1), 34–38. https://doi.org/10.26737/jpmi.v1i1.80.
Puspitasari, A. D., Sukarelawan, M. I., Damayanti, E. N., Syifa, A., & Fitri, F. (2022). Model pembelajaran predict observe explain dalam edmodo untuk meningkatkan pemahaman konsep fisika di SMP: Analisis stacking. *Berkala Fisika Indonesia : Jurnal Ilmiah Fisika, Pembelajaran Dan Aplikasinya*, *13*(1), 31–40. https://doi.org/10.12928/bfi-jifpa.v13i1.23204.
Rasmitadila, Aliyyah, R. R., Rachmadtullah, R., Samsudin, A., Syaodih, E., Nurtanto, M., & Tambunan, A. R. S. (2020). The perceptions of primary school teachers of online learning during the covid-19 pandemic period: A case study in Indonesia. *Journal of Ethnic and Cultural Studies*, *7*(2), 90–109. https://doi.org/10.29333/ejecs/388.
Riwanto, D., Azis, A., & Arafah, K. (2019). Analisis Pemahaman Konsep Peserta Didik dalam Menyelesaikan Soal-Soal Fisika Kelas X Mia SMA Negeri 3 Soppeng. *Jurnal Sains Dan Pendidikan Fisika (JSPF)*, *15*(2), 23–31. https://doi.org/10.35580/jspf.v15i2.11033.
Saputro, D. E., Sarwanto, S., Sukarmin, S., & Ratnasari, D. (2019). Pre-services science teachers' conceptual understanding level on several electricity concepts. *Journal of Physics: Conference Series*, *1157*, 032018. https://doi.org/10.1088/1742-6596/1157/3/032018.
Shidik, M. A. (2020). Hubungan antara Motivasi Belajar dengan Pemahaman Konsep Fisika Peserta Didik MAN Baraka. *Jurnal Kumparan Fisika*, *3*(2), 91–98. https://doi.org/10.33369/jkf.3.2.91-98.
Sukarelawan, M. I., Gustina, E., Ayu, S. M., Sofiana, L., & Sriyanto, S. (2021). Workshop penelitian alternatif di masa pandemik covid-19 bagi guru-guru SMA/SMK. *Peran Perguruan Tinggi (PT) Dalam Meningkatkan Kapastias Masyarakat Di Era Pandemi*, 1538–1544.
Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, *10*(4), 486–495. https://doi.org/10.15294/jpii.v10i4.27114.
Sukarelawan, M. I., Jumadi, J., & Rahman, N. A. (2019). An Analysis of Graduate Students' Conceptual Understanding in Heat and Temperature (H&T) Using Three-Tier Diagnostic Test. *Indonesian Review of Physics*, *2*(1), 9–14. https://doi.org/10.12928/irip.v2i1.910.
Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., Indratno, T. K., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, *10*(2), 216–224. https://doi.org/10.20527/bipf.v10i2.12209.
Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi, J. (2022). A

Shift in the Conceptual Understanding of Physics Students Through the Wright Map. *Jurnal Pendidikan Dan Pengajaran*, *55*(1), 127–141. https://doi.org/10.23887/jpp.v55i1.38342.

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, *5*(1), 1–8. https://doi.org/10.30599/jipfri.v5i1.856.

Suma, K., Sadia, I. W., & Pujani, N. M. (2018). The identification of the 11 th grade students' prior knowledge of electricity concepts. *Journal of Physics: Conference Series*, *1040*, 012038. https://doi.org/10.1088/1742-6596/1040/1/012038.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. *International Journal of Instruction*, *11*(2), 129–144. https://doi.org/10.12973/iji.2018.11210a.

Umam, A., Suparmi, & Sukarmin. (2020). Using two tier based concept test to analysis profile of student understanding on the concept of simple harmonic motion. *Journal of Physics: Conference Series*, *1567*(3), 032076. https://doi.org/10.1088/1742-6596/1567/3/032076.