JPF | Volume 13 | Number 1 | 2025 | 1 - 13

p - ISSN: 2302-8939
e - ISSN: 2527-4015

Jurnal Pendidikan Fisika

https://journal.unismuh.ac.id/index.php/jpf

DOI: 10.26618/jpf.v13i1.15687

Analysis of Concept Understanding Test Items on Static Fluid Material Using Rasch Model

Ivan Danar Aditya Irawan[1)], Ristanti Mulia Indraloka[2)], Nur Akhyar Basri[3)], Ummi Salmah[4)], Parno[5)*]

Physics Education Department, State University of Malang, Malang, 65145, Indonesia

**Corresponding author: parno.fmipa@um.ac.id*

Received: August 06, 2024; Accepted: November 11, 2024; Published: January 08, 2025

Abstract – *Conceptual understanding is a vital component in physics education, particularly for static fluid concepts, which are often sources of misconceptions among students. Common misunderstandings include incorrect interpretations of hydrostatic pressure and Pascal's law. This study aims to develop and analyze a conceptual understanding test instrument for static fluid materials using the Rasch Model. A descriptive quantitative research design was employed, involving 75 eleventh-grade students from three high schools in Lumajang and Malang, selected through cluster random sampling. The instrument comprised 16 multiple-choice questions based on eight conceptual indicators and underwent expert validation by two physics teachers. Data were analyzed using the Winstep application to assess validity, reliability, difficulty level, and item discrimination. The results revealed that 12 out of 16 test items met the validity criteria, with an expert validation score of 98.3% categorized as "very valid." Item reliability was rated at 0.96 (very good), while person reliability was 0.47 (very poor), indicating significant variations in student responses. The difficulty levels were balanced: 2 very easy items, two easy items, five moderate items, one difficult item, and two very difficult items. Discrimination analysis grouped respondents into two categories and items into seven distinct groups, showcasing the instrument's effectiveness in identifying variations in student understanding. In conclusion, the developed instrument is valid and reliable for assessing students' conceptual understanding of static fluid topics. The study highlights the need for further validation with larger and more diverse samples to enhance the instrument's applicability across broader educational contexts.*

Keywords: *conceptual understanding; item analysis; Rasch model; static fluid*

I. INTRODUCTION

Conceptual understanding is defined as students' ability to analyze and synthesize knowledge while relating concepts to various contexts and situations (Banda & Nzabahimana, 2021). Empirical data indicate that students often achieve low learning outcomes in physics due to difficulties in understanding the subject (Hakiki et al., 2015). Students struggle to explain fundamental static fluid concepts, including hydrostatic pressure, Pascal's law, and Archimedes' law, leading to misconceptions (Atika et al., 2023; Jamaludin & Batlolona, 2021; Mutmainnah et al., 2023). For example, students mistakenly believe that cross-sectional area influences hydrostatic pressure (Atika et al., 2023; Irma et al., 2022). Conceptual understanding plays a crucial role in physics learning, particularly in bridging the gap between theory and practice (Shishigu et al., 2018). It also serves as a benchmark for evaluating students' learning outcomes (Purwanto, 2010). Therefore, a reliable instrument is necessary to accurately measure students' learning outcomes during the learning process.

Conceptual understanding is a key component of metacognitive development, associated with higher-order thinking skills and active engagement in the learning process (Mills, 2016). It is reflected in physics evaluation outcomes, serving as a reliable measure of learning success (Gunawan et al., 2018; Rahmawati & Suryadi, 2019). There is a strong positive relationship between conceptual understanding and students' learning outcomes, where the latter is heavily influenced by the former (Muslichatun et al., 2021; Syaifullah et al., 2020). Thus, students' conceptual understanding must be analyzed to identify misconceptions and align them with scientific principles (Grob et al., 2017; Zvoch et al., 2021). Students with strong conceptual understanding typically achieve better learning outcomes. Therefore, appropriate instruments are required to accurately measure the outcomes of efforts to improve conceptual understanding (Scott & Schumayer, 2017). The use of valid and reliable instruments is essential for ensuring that conceptual understanding tests accurately reflect students' understanding (Matondang, 2009; Adha et al., 2023).

In reality, the evaluation and assessment process plays a central role in developing and measuring conceptual understanding (Ceran & Ates, 2020). However, many conceptual understanding tests lack valid and reliable instruments. This is often attributed to the complex and time-consuming nature of validation processes (Sari, 2020).This finding aligns with Rismaulhijjah & Kuswanti (2022), who note that limited resources and analytical skills hinder test item analysis. On the other hand, teachers often face difficulties in accessing valid and empirically tested test instruments. Brief interviews with high school physics teachers in Lumajang Regency reveal that many conceptual understanding tests rely on teacher-made instruments based on available indicators. These tests frequently rely on practice questions used in classroom lessons. This observation aligns with Nafsih et al. (2020), who highlight that teachers rarely validate and test their assessment instruments for reliability. Poorly analyzed and repeatedly used test instruments risk becoming ineffective and unreliable for assessing students' conceptual understanding (Setyaedhi et al., 2023). Thus, many existing instruments fail to deliver optimal results (Misbah et al., 2022; Nehru et al., 2022).

As a solution, test items must undergo thorough analysis to ensure their quality and validity. This analysis can be performed using the Rasch Model. The Rasch Model goes beyond statistical analysis, offering a comprehensive perspective on educational measurement (Planinic et al., 2019). Thus, the use of the Rasch Model in instrument validation will provide a more comprehensive understanding and be consistent with the concept of measurement (Bond & Fox, 2013). By applying the Rasch Model, ordinal data can be transformed into ratio data with improved accuracy (Aryanti et al., 2020). The Rasch Model carefully examines various aspects such as response types, suitability of each item and respondent, dimensions, difficulty levels, and item bias. Testing can also be done on different aspects such as ensuring item consistency across different age groups and genders, or even in the form of scales used (Shea et al., 2009). Several advantages ensure that the analyzed items can consistently apply in various usage contexts.

This research aims to develop a multiple-choice conceptual understanding test instrument using the Rasch Model. The study focuses specifically on static fluid materials as taught during the research period. The goal is to create a valid and reliable test instrument for assessing conceptual understanding of static fluid materials, validated both by experts and empirically.

II. METHODS

This study employed a descriptive quantitative research approach. The aim was to describe and evaluate the quality of conceptual understanding test items on static fluid material using the Rasch Model. The study participants consisted of 11th-grade science students who had previously studied static fluid material. The sample was selected using cluster random sampling, comprising 75 students from three senior high schools in Lumajang and Malang Regencies: SMAN Bululawang, SMAN Klakah, and SMAN Singosari. The sample size was considered sufficiently representative for statistical analysis, ensuring both accuracy and practicality. The instrument used in this study was a conceptual understanding test consisting of 8 indicators and 16 test items, as detailed in Table 1.

Table 1. Indicators of concept understanding test items

No	Level	Indicator	Number
1	C1	Identify the properties and types of fluids	1,9
2	C4	Analyze the relationship between depth and hydrostatic pressure in a liquid in a container	2,10
3	C4	Find the relationship between variables in hydrostatic pressure at a certain depth	3,11
4	C3	Apply hydrostatic pressure in everyday life	4, 12
5	C3	Apply Pascal's law in hydraulic jack applications in everyday life	5, 13
6	C3	Apply the concept of viscosity in daily phenomena	6, 14
7	C3	Analyze floating, sinking, and submerged objects	7, 11
8	C3	Analyze the relationship between density and conditions in a liquid	8, 16

The data collection took place in May 2024 across the selected sample. Data were analyzed using the Rasch Model, following criteria outlined by Austvoll-Dahlgren et al. (2017), Prasetya & Pratama (2023), and Sumintono & Widhiarso (2015). The procedural steps of the research are shown in Figure 1.

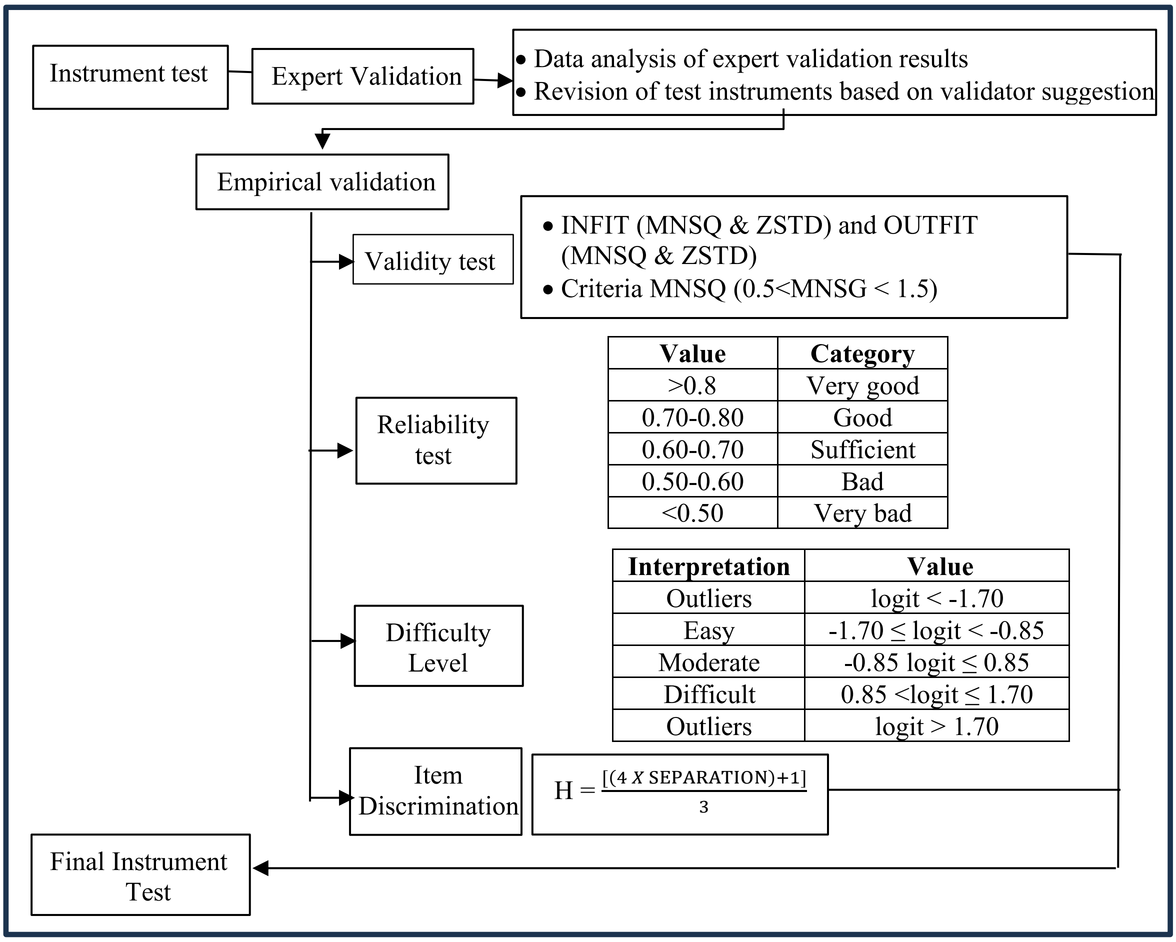

Figure 1. Research procedure

The test instrument underwent expert validation by two high school physics teachers. Revisions were made based on expert feedback and recommendations. The instrument was then pilot-tested with high school students. The test was conducted using a multiple-choice format. The responses were evaluated based on a predefined rubric for multiple-choice assessments. The Rasch Model analysis was performed using the Winstep application. The analysis focused on assessing the validity, reliability, difficulty levels, and discrimination index of each test item. The Item validity was assessed using OUTFIT MNSQ and OUTFIT ZSTD criteria. Reliability was evaluated using Person Reliability and Item Reliability scores. The Item difficulty was analyzed using measure (logit) values in relation to standard deviation thresholds. The Item discrimination was assessed using the equating formula (H). The finalized test instrument, validated and analyzed through the Rasch Model, was deemed suitable for effectively measuring students' conceptual understanding of static fluid material.

III. RESULTS AND DISCUSSION

The conceptual understanding test instrument was validated by two high school physics teachers. One validator was based at one of the sample schools, while the other was a high school physics teacher from East Java. Validation was carried out using a Likert scale questionnaire. Validators assessed aspects such as content, construct, language, and appropriateness. The validation results are summarized in Table 2.

Table 2. Expert validation results

Validators	Content	Construct	Language	Appropriateness	Average
Validators 1	99.0 %	97.0 %	94.0 %	98.0 %	97.0 %
Validators 2	100.0 %	100.0 %	98.0 %	100.0 %	99.5 %
Average	99.5 %	98.5 %	96.0 %	99.0 %	98.3 %

The validation results indicate that the conceptual understanding test instrument is highly valid, with an average score of 98.3%. The content aspect achieved the highest score at 99.5%, indicating strong alignment with the static fluid concepts taught in schools. The appropriateness aspect scored 99.0%, showing alignment with the indicators of conceptual understanding proposed by Anderson and Krathwohl (2001). The construct aspect achieved 98.5%, reflecting suitable presentation, clear wording, effective answer cues, and appropriate image usage. Furthermore, language aspect received the lowest score, at 96.0%. These validation results offered valuable insights for revising test items before empirical validation.

Table 3. Validator's improvements to test items

Validator's	No Item	Revised
1	2	Revised "Illustration image" to "Image"
	3	Combined repeated sentences in each answer option into the question
	5	Corrected the writing of units with exponents
	7	Added captions to the presented images in the question
	8	Simplified the wording of the question to facilitate understanding of the purpose or instruction given
	10	Adjusted the wording of the question to the presented illustration
	13	Combined repeated sentences in each answer option into the question
2	5	Changed the wording of the test item to more communicative language
	6	Changed the wording of the test item to more communicative language
	7	Added captions to the presented images in the question

The item revisions were made following the validators' suggestions and feedback. Following the revisions, the conceptual understanding test instrument underwent empirical validation with high school students who had studied static fluid material. Empirical validation was conducted using both paper-based tests and Google Forms. The dual format accommodated the geographical distribution of the sample schools. Test results were tabulated using Microsoft Excel based on an assessment rubric. Research Data analysis was performed using the Rasch Model with the WinSteps application. The Analysis focused on determining validity, reliability, difficulty levels, and item discrimination. The complete analysis results are presented as follows.

1. Item Question Validity Test

Validity was assessed using the Rasch Model via the WinSteps application, focusing on INFIT MNSQ and ZSTD values. The OUTFIT MNSQ and ZSTD values were also analyzed. Criteria for these values are detailed in Figure 1. Interpretation of the results of the item question validity analysis is presented in Table 4.

Table 4. Item question validity

No Item	INFIT		OUTFIT		Interpretation
	MNSQ	ZSTD	MNSQ	ZSTD	
1	0.81	-0.33	0.32	-1.04	Invalid
2	0.87	-0.47	0.64	-0.87	Valid
3	0.88	-0.52	0.68	-0.88	Valid
4	0.77	-1.29	0.55	-1.62	Valid
5	1.23	1.24	1.76	2.18	Invalid
6	1.04	0.38	1.03	0.24	Valid
7	1.00	0.09	1.60	1.79	Invalid
8	1.03	0.40	1.06	0.44	Valid
9	0.96	-0.39	0.90	-0.58	Valid
10	1.07	0.59	1.15	0.74	Valid
11	1.03	0.38	1.05	0.37	Valid
12	0.96	-0.42	1.29	1.71	Valid
13	1.03	0.22	1.85	2.25	Invalid
14	0.93	-0.29	1.17	0.61	Valid
15	0.91	-0.16	0.69	-0.40	Valid
16	0.97	-0.09	1.10	0.39	Valid

The results of the validity test indicate that all items meet the validity criteria based on INFIT MNSQ and INFIT ZSTD. All items are within the accepted range: (0.5 < MNSQ < 1.5) and (-2.00 < ZSTD < +2.00). However, a different outcome emerges when analyzed using OUTFIT MNSQ and OUTFIT ZSTD criteria. There four items 1, 5, 7, and 13 do not meet the OUTFIT MNSQ criteria. Additionally, items 5 and 13 fall outside the acceptable range for OUTFIT ZSTD. In total, 12 items were deemed valid, while four items (1, 5, 7, and 13) were classified as invalid.

The invalid items are items number 1, 5, 7, and 13. Item 1 recorded an MNSQ value below 0.5, while items 5, 7, and 13 exceeded 1.5. Furthermore, items 5 and 13 did not meet the ZSTD criteria. hese invalid items were either revised or removed from the instrument. Solutions to address these issues, invalid items can be revised, replaced, or removed to ensure the instrument achieves optimal validity.

2. Instrument Reliability Test

Reliability measures the consistency and accuracy of a test instrument in assessing the intended construct (Alwan & Alshurideh, 2022; Grgic et al., 2020). Reliability analysis in the Rasch Model uses two key criteria: Person Reliability and Item Reliability (Sumintono & Widhiarso, 2015). The reliability values are presented in Table 5.

Table 5. Instrument reliability for concept comprehension test

Type	Reliability value	Category
Person Reliability	0.47	Very Bad
Item Reliability	0.96	Very Good

Based on Table 5, the person reliability score of 0.47 falls into the 'Very Bad' category, indicating a low level of response consistency among students. This low score may stem from inconsistencies in student responses and assessment patterns (Djidu et al., 2023). Population size and diversity can also affect Person Reliability scores (Tinôco et al., 2019). Consequently, the internal validity of the results may not be optimal, and generalizations could be inaccurate. Low Person Reliability might also result from students' difficulties in understanding or completing the test. Factors such as students' understanding of the material, motivation, and psychological conditions (e.g., confidence) may contribute to these inconsistencies. However, in contrast, Item Reliability achieved a score of 0.96, placing it in the 'Very Good' category. This score indicates that the test items demonstrate a high level of consistency and reliability in measuring conceptual understanding. This suggests that the instrument can consistently measure students' understanding of static fluid concepts across repeated uses and different student samples.

The disparity between person reliability and item reliability does not inherently imply poor instrument quality. The focus should remain on Item Reliability, as it reflects the consistency of the test instrument rather than variations in student responses. In conclusion, the reliability analysis indicates that the test instrument falls into the 'Very Good' category and can reliably measure students' conceptual understanding of static fluid concepts.

3. Difficulty Level

The difficulty level of test items indicates how likely a respondent is to answer a question correctly (Beatty et al., 2021). This analysis helps determine whether the test items can effectively differentiate between test-takers with varying abilities and assess their likelihood of answering correctly (Andriani et al., 2023). In the Rasch Model's measurement theory, item difficulty indices are interpreted according to the criteria outlined by Lestari and Yudhanegara (2018).

Table 6. Results of item difficulty level through Rasch modeling with SD logit of 1.70

Measure value (logit)	Interpretation	Item number
Measure logit < -1.70	Very easy (Outliers)	2, 15
-1.70 ≤ Measure logit < -0.85	Easy	4, 16
-0.85 ≤ Measure logit ≤ 0.85	Moderate	8, 9, 10, 11, 12
0.85 < Measure logit ≤ 1.70	Difficult	6
Measure logit > 1.70	Very difficult (Outliers)	3, 14

Rasch analysis measured the difficulty level of each test item in logit units (Planinic et al., 2019). According to the criteria, items were grouped into five difficulty categories: very easy (outliers), easy, moderate, difficult, and very difficult (outliers). The distribution shows two very easy items (2 and 15). Two items (4 and 16) fall into the easy category. The majority of items (5 in total: 8, 9, 10, 11, and 12) are categorized as moderate. There One item (6) is classified as difficult. Finally, two items (3 and 14) fall into the very difficult category. The analysis indicates a diverse distribution of item difficulty, ranging from very easy to very difficult. In summary, the distribution includes 2 very easy items, 2 easy items, 5 moderate items, 1 difficult item, and 2 very difficult items. The majority of the items fall into the moderate category. Which aligns with Purnasari et al. (2021), who argue that an ideal instrument should have a balanced difficulty level, avoiding extremes.

4. Item Discrimination

Item discrimination refers to a test item's capacity to distinguish between high-performing and low-performing students. In Rasch modeling, discrimination analysis examines individual ability levels to identify differences between students who answered items correctly and those who did not. Additionally, respondent separation indices are also used to identify distinct respondent groups. Higher separation values indicate better quality in distinguishing between different respondent groups and item difficulties. This is because it can help identify these groups more accurately. The strata equation (H) is also used to obtain more accurate separation:

$$H = \frac{[(4\,x\,SEPARATION)+1\,)}{3} \qquad (1)$$

The analysis yielded a respondent separation value of 0.94, resulting in H = 1.59. Rounding H to 2 suggests that respondents can be classified into two distinct groups. The item separation value was 5.00, resulting in H = 7. This indicates that the test instrument can be divided into seven groups of items. A high discrimination index enhances the test instrument's effectiveness in measuring student abilities accurately. Furthermore, item discrimination analysis results provide insights for improving and refining less effective test items.

5. Implications

The analysis of the conceptual understanding test instrument on static fluid material resulted in 12 valid test items. Teachers can utilize this instrument as an effective assessment tool for evaluating student learning activities. It can also serve as a valuable tool for future researchers seeking to analyze students' conceptual understanding of static fluid material. Data derived from this instrument offer reliable and consistent interpretations. The multiple-choice format simplifies the assessment process for teachers by using clear and straightforward rubric requirements. This valid instrument addresses the scarcity of reliable, high-quality tools for measuring students' conceptual understanding in schools. With this instrument, teachers can efficiently identify students' strengths and weaknesses in grasping static fluid concepts. Improving the quality of assessment instruments will ultimately enhance the overall quality of education. The balanced distribution of item difficulty levels carries significant implications for the design of effective physics instruction in schools. By using this instrument, teachers can better recognize areas where students excel and where they face challenges in understanding the material. Enhancing the quality of assessment tools supports more targeted, evidence-based instructional designs and fosters improvements in overall educational outcomes.

Analyzing item difficulty distribution allows educators to ensure that test items address a wide spectrum of cognitive skills, from lower-order to higher-order thinking. If test items are overly easy or excessively difficult, assessments may fail to accurately capture students' understanding and abilities. A balanced difficulty distribution enables teachers to design instructional strategies that accommodate students' varied learning needs. Low-difficulty items can serve as introductory exercises to build student confidence, while moderate- and high-difficulty items encourage deeper application and analysis of concepts. The results of the analysis can also guide curriculum adjustments, ensuring alignment with students' abilities and academic progress. This alignment creates a supportive and engaging learning environment, ultimately enhancing the quality of physics education

IV. CONCLUSION AND SUGGESTION

The analysis results indicate that, out of 16 test items analyzed, 12 items met the validity criteria based on INFIT and OUTFIT MNSQ and ZSTD values. The reliability analysis revealed a 'Very Bad' score for Person Reliability and a 'Very Good' score for Item Reliability. The difficulty level analysis demonstrated a balanced distribution of easy, moderate, and difficult items. Specifically, the distribution consists of 2 very easy items, 2 easy items, 5 moderate items, 1 difficult item, and 2 very difficult items. The Item discrimination analysis classified respondents into two distinct groups and the test instrument into seven item groups. The study successfully developed a valid and reliable conceptual understanding test instrument suitable for evaluating students' learning outcomes in schools.

The analysis confirmed that the majority of test items are valid and suitable for assessing students' conceptual understanding. Nevertheless, the study faced limitations, primarily the relatively small sample size. Future studies should consider larger sample sizes across diverse regions to enhance result accuracy and improve instrument quality. Despite these limitations, this study effectively analyzed the conceptual understanding test instrument and addressed invalid items through refinement or removal.

ACKNOWLEDGMENTS

The author expresses heartfelt gratitude to all parties who have provided support and contributions in the completion of this research.

REFERENCES

Adha, A., Rera, A., & Farisi, S. A. (2023). Analysis of the TGT cooperative learning model in physics learning: in terms of the implementation of procedures and principles. *International Journal of Education and Teaching Zone*, *2*(1), 51–61. https://doi.org/10.57092/ijetz.v2i1.70

Alwan, M., & Alshurideh, M. T. (2022). The effect of digital marketing on purchase intention: Moderating effect of brand equity. *International Journal of Data and Network Science*, *6*(3), 837–848. https://doi.org/10.5267/j.ijdns.2022.2.012

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assesing: a revision of bloom's taxonomy of educational objectives*. Addison Wesley Longman, Inc

Andriani, R., Widya., Fadieny, N., & Permana, N. D. (2023). Development of conceptual understanding student tests to the basic physics subject; a rasch model. *Co-Catalyst: Journal of Science Education Research and Theories*, *1*(1), 43–54. https://doi.org/10.33830/cocatalyst.vol1iss1.4891

Aryanti, R. D., Sari, E. Y. D., & Widiana, H. S. (2020). Aplikasi model rasch pada skala work engagement (utrecht work engagement scale). *Jurnal Psikologi Integratif*, *8*(2), 61–77.

Atika, S. D., Santhalia, P. W., & Sudjito, D. N. (2023). STEAM integrated project-based learning exploration against understanding the concept of static fluids. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5357–5364. https://doi.org/10.29303/jppipa.v9i7.2905

Austvoll-Dahlgren, A., Guttersrud, Ø., Nsangi, A., Semakula, D., & Oxman, A. D. (2017). Measuring ability to assess claims about treatment effects: A latent trait analysis of items from the "claim evaluation tools" database using Rasch modelling. *BMJ Open*, *7*(5), 1-10. https://doi.org/10.1136/bmjopen-2016-013185

Banda, H. J., & Nzabahimana, J. (2021). Effect of integrating physics education technology simulations on students' conceptual understanding in physics: A review of literature. *Physical Review Physics Education Research*, *17*(2), 1-18. https://doi.org/10.1103/PhysRevPhysEducRes.17.023108

Beatty, A., Berkhout, E., Bima, L., Pradhan, M., & Suryadarma, D. (2021). Schooling progress, learning reversal: Indonesia's learning profiles between 2000 and 2014. *International Journal of Educational Development, 85,* 1-16. https://doi.org/10.1016/j.ijedudev.2021.102436

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press. https://doi.org/10.4324/9781410614575

Ceran, S. A., & Ates, S. (2020). Conceptual understanding levels of students with different cognitive styles: An evaluation in terms of different measurement techniques. *Eurasian Journal of Educational Research*, *88*, 149–178. https://doi.org/10.14689/ejer.2020.88.7

Djidu, H., Istiyono, E., & Widihastuti, W. (2023). Quality of performance assessment instruments for educators in higher education: Implementation of factor analysis and generalizability theory. *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: E-Saintika*, *7*(2), 144–159. https://doi.org/10.36312/esaintika.v7i2.716

Grgic, J., Lazinica, B., Schoenfeld, B. J., & Pedisic, Z. (2020). Test–retest reliability of the one-repetition maximum (1RM) strength assessment: A systematic review. *Sports Medicine – Open*, *6*(1), 1-6. https://doi.org/10.1186/s40798-020-00260-z

Grob, R., Holmeier, M., & Labudde, P. (2017). Formative assessment to support students' competences in inquiry-based science education. *Interdisciplinary Journal of Problem-Based Learning*, *11*(2), 1-7. https://doi.org/10.7771/1541-5015.1673

Gunawan, G., Nisrina, N., Suranti, N. M. Y., Herayanti, L., & Rahmatiah, R. (2018). Virtual laboratory to improve students' conceptual understanding in physics learning. *Journal of Physics: Conference Series*, *1108*, 1-6. https://doi.org/10.1088/1742-6596/1108/1/012049

Hakiki, R. S., Arsyad, M., & Khaeruddin, K. (2015). Upaya meningkatkan hasil belajar fisika siswa kelas VII SMP Negeri 30 Makassar melalui model pembelajaran kooperatif tipe snowball throwing. *Jurnal Pendidikan Fisika*, *3*(3), 285–294. https://doi.org/10.26618/jpf.v3i3.281

Irma, Z. U., Kusairi, S., & Yuliati, L. (2022). Level of students' conceptual understanding of static fluid. *Journal of Physics: Conference Series*, *2392*, 1-11. https://doi.org/10.1088/1742-6596/2392/1/012020

Jamaludin, J., & Batlolona, J. R. (2021). Analysis of students' conceptual understanding of physics on the topic of static fluids. *Jurnal Penelitian Pendidikan IPA*, *7*(SpecialIssue), 6–13. https://doi.org/10.29303/jppipa.v7ispecialissue.845

Lestari, K. E., & Yudhanegara, M. R. (2018). *Penelitian pendidikan matematika: Panduan praktis menyusun skripsi, tesis, dan laporan penelitian dengan pendekatan kuantitatif, kualitatif, dan kombinasi disertasi dengan model pembelajaran dan kemampuan matematis*. Refika Aditama.

Matondang, Z. (2009). Validitas dan reliabilitas suatu instrumen penelitian. *Jurnal Tabularasa*, *6*(1), 87–97.

Mills, S. (2016). Conceptual understanding: A concept analysis. *The Qualitative Report*, *21*(3), 546–557. https://doi.org/10.46743/2160-3715/2016.2308

Misbah, M., Trisnowati, E., Rahim, A., & Zb, A. (2022). Investigating problem solving and mathematical connections in solving the fermi-dirac equation. *International Journal of Education and Teaching Zone*, *1*(2), 106–115. https://doi.org/10.57092/ijetz.v1i2.36

Muslichatun, M., Ellianawati, E., & Wardani, S. (2021). Analisis pemahaman konsep dan hasil belajar siswa dalam pembelajaran konsep rangka manusia berbantuan media interaktif berbasis android. *Jurnal Profesi Keguruan*, *7*(1), 142–150.

Mutmainnah, B., Amni, Z., & Wahyudi, M. (2023). development of android-based mobile learning media using ADDIE model on static fluid topics for high school 11 th grade students. *Konstan: Jurnal Fisika dan Pendidikan Fisika*, *8*(2), 125–133. https://doi.org/10.20414/konstan.v8i02.471

Nafsih, N. Z., Festiyed., & Mufit, F., (2020). Meta analisis pengembangan instrumen tes keterampilan proses sains dalam pembelajaran fisika dan IPA. *Jurnal Penelitian Pembelajaran Fisika*. 6(1). 89-96. https://doi.org/10.24036/jppf.v6i1.108965

Nehru, N., Ananda, R., Zb, A., & Navalyyan, D. (2022). The analysis of mathematical critical thinking ability and mathematical creativity: Judging from the process of deriving the fermi-dirac formula. *International Journal of Education and Teaching Zone*, *1*(2), 87–96. https://doi.org/10.57092/ijetz.v1i2.33

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 1-14. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Prasetya, W. A., & Pratama, A. T. (2023). Item quality analysis using the Rasch model to measure critical thinking ability in the material of the human digestive system of Biology subject in high school. *Jurnal Penelitian dan Evaluasi Pendidikan*, *27*(1), 76–91. https://doi.org/10.21831/pep.v27i1.58873

Purnasari, P. D., Silvester, S., & Lumbantobing, W. L. (2021). Pengembangan instrumen asesmen higher order thingking skills (hots) ditinjau dari gaya belajar siswa. *Sebatik*, *25*(2), 571–580. https://doi.org/10.46984/sebatik.v25i2.1607

Purwanto, M. (2010). *Evaluasi hasil belajar.* Pustaka Pelajar

Rahmawati, M., & Suryadi, E. (2019). Guru sebagai fasilitator dan efektivitas belajar siswa. *Jurnal Pendidikan Manajemen Perkantoran*, *4*(1), 49-54. https://doi.org/10.17509/jpm.v4i1.14954

Rismaulhijjah, W. A., & Kuswanti, N. (2022). Anslisis butir soal ulangan harian hasil pengembangan guru materi sistem gerak manusia kelas XI IPA. *Bioedu: Berkala Ilmiah Pendidikan Biologi*, *11*(3), 643–661. https://doi.org/10.26740/bioedu.v11n3.p643-661

Sari, D. K. (2020). Analisis instrumen penilaian kemampuan pemodelan matematis pada kelas fisika menggunakan rasch model. *Mega: Jurnal Pendidikan Matematika*, *1*(1), 47–52. https://doi.org/10.59098/mega.v1i1.182

Scott, T. F., & Schumayer, D. (2017). Conceptual coherence of non-Newtonian worldviews in Force Concept Inventory data. *Physical Review Physics Education Research*, *13*(1), 1-12. https://doi.org/10.1103/PhysRevPhysEducRes.13.010126

Setyaedhi, H. S., Mustaji., & Fitri, C. (2023). Empirical quality of final exam questions in a learning management system-based course. *JPI (Jurnal Pendidikan Indonesia)*, *12*(1), 78–88. https://doi.org/10.23887/jpiundiksha.v12i1.52262

Shea, T. L., Tennant, A., & Pallant, J. F. (2009). Rasch model analysis of the depression, anxiety and stress scales (DASS). *BMC Psychiatry*, *9*, 1–10. https://doi.org/10.1186/1471-244X-9-21

Shishigu, A., Hailu, A., & Anibo, Z. (2018). Problem-based learning and conceptual understanding of college female students in physics. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 145–154. https://doi.org/10.12973/ejmste/78035

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. (1st ed., Vol. 1). Trim Komunikata.

Syaifullah, S., Suharto, S., & Suryanti, Y. (2020). Hubungan pemahaman konsep trigonometri dengan hasil belajar fisika pada materi vektor. *Seminar Nasional Pendidikan STKIP Kusuma Negara II*, 381–387.

Tinôco, M. A., Gouveia, É. R., Ihle, A., Marques, A., Gouveia, B. R., & Kliegel, M. (2019). The cognitive telephone screening instrument (cogtel): A reliable and valid tool for the assessment of cognitive functioning in the Brazilian elderly. *Revista Brasileira de Geriatria e Gerontologia*, *22*(1), 1-8. https://doi.org/10.1590/1981-22562019022.180130

Zvoch, K., Holveck, S., & Porter, L. (2021). Teaching for conceptual change in a density unit provided to seventh graders: A comparison of teacher- and student-centered approaches. *Research in Science Education*, *51*(5), 1395–1421. https://doi.org/10.1007/s11165-019-09907-8