

Development of An Assessment Instrument To Analyze Conceptual Abilities And Daily Life In PJOK Subject Questions Based On The Rasch Model

Muhammad Izzul Afriza¹, Heny Setyawati²

¹ Universitas Negeri Semarang, Indonesia; imronnugrohosaputro@students.unnes.ac.id

² Universitas Negeri Semarang, Indonesia; henysetyawati@mail.unnes.ac.id

Article Info

History Articles

Received:

July 2025

Accepted:

July 2025

Published:

Maret 2026

Keywords:

Minimum Competency
Assessment, Conceptual
Ability, Daily Life, Rasch
Model.

Abstract

This study aims to develop an assessment instrument to analyze conceptual understanding and application in everyday life (daily life) in Physical Education, Sports, and Health (PJOK) based on the Rasch Model. The research method used was Research and Development (R&D), adapting the development model from Gall, Gall, and Borg. The research procedure included a preliminary study, instrument design, expert validation, small-scale trials, large-scale trials, and dissemination of the final product. The developed instrument consisted of 16 essay questions that integrated conceptual mastery and application in everyday life contexts. Data were analyzed using content validity (Aiken's V), reliability (Cronbach's Alpha), and item characteristic analysis using the Rasch Model with the assistance of Winsteps software. The results showed that the instrument met the criteria for validity and reliability, and the items demonstrated a good level of fit. This instrument can be used by PJOK teachers as an evaluation tool capable of comprehensively describing students' minimum competencies.

© 2026 Universitas Negeri Semarang

✉ Address correspondence:

Universitas Negeri Semarang, Kampus Sekaran, Gunungpati,
Semarang
E-mail: (imronnugrohosaputro@students.unnes.ac.id)

p-ISSN 2252-648X
e-ISSN 2502-4477

INTRODUCTION

Explaining Education in Indonesia continues to transform in line with changing times and societal needs. One strategic policy implemented is the elimination of the National Examination (UN) and its replacement with the National Assessment (AN), which consists of three main components: the Minimum Competency Assessment (AKM), the Character Survey, and the Learning Environment Survey. The AKM specifically measures students' literacy and numeracy skills, reflecting their critical thinking capacity and problem-solving abilities in everyday life.

Although the AKM has been widely implemented, the implementation of assessments measuring minimum competencies in Physical Education, Sports, and Health (PJOK) remains very limited. PJOK not only focuses on physical skills but also teaches important concepts related to health, physical activity, and the application of positive attitudes in everyday life. The Independent Curriculum requires students to develop critical, creative, collaborative, and communicative thinking skills (the 4Cs), making it crucial to develop assessment instruments that measure not only factual knowledge but also conceptual skills and their application in real-life contexts.

A problem identified in the current evaluation process is the limited availability of instruments that can holistically assess students' ability to understand physical education (PJOK) concepts and apply them in everyday life. Furthermore, existing evaluations tend to be normative in nature, lacking a scientific approach that can guarantee the validity and reliability of the instruments. One solution to this challenge is to use the Rasch Model in developing assessment instruments. The Rasch Model offers advantages in analyzing question difficulty, item suitability, respondent reliability, and detecting bias in items.

This research aims to develop a Rasch Model-based assessment instrument that can measure students' conceptual and practical application skills in the subject of PJOK. It is hoped that the results of this development will serve as a reference for PJOK teachers in conducting more meaningful evaluations and encouraging improvements in the quality of learning in accordance with the objectives of the Independent Curriculum.

METHODS

Make This research employed a research and development (R&D) method, adapting the ten-stage development model of Gall, Gall, and Borg (2003). This model was chosen because it can produce a product in the form of an evaluation instrument that has been validated theoretically and empirically. The research focused on developing a minimum competency assessment instrument that measures conceptual and daily life skills in Physical Education (PJOK) using the Rasch Model approach.

The research design followed the following steps:

1. Identification of potential and problems through observation and interviews with elementary school PJOK teachers regarding the need for a valid and reliable assessment instrument.
2. Initial data collection, in the form of a literature review and data collection related to learning outcomes (CP) and learning objectives (TP) in PJOK Phase C.
3. Designing the initial product, in the form of a grid, essay questions, and scoring guidelines that measure conceptual and daily life skills.
4. Expert validation, involving one educational evaluation expert, one physical education lecturer, and one certified physical education teacher, using a content validation instrument.
5. Product revision based on expert input to refine the instrument.
6. Small-scale pilot testing, conducted with 52 students from two elementary schools.

7. Product revision after the small-scale pilot testing based on data analysis.
8. Large-scale pilot testing, conducted with 104 students from four elementary schools.
9. Final revision of the instrument with selection of valid and reliable items.
10. Dissemination of the final product, in the form of an assessment instrument and guidelines for its use.

The research subjects consisted of two trial groups: (1) a small-scale trial involving 52 fifth-grade students from two elementary schools, and (2) a large-scale trial involving 104 students from four different elementary schools.

Data collection instruments: interviews, expert validation questionnaires, and observation sheets. The main assessment instrument: 16 essay questions measuring conceptual and daily life skills. The scoring guidelines were developed based on a rubric.

Content validity was analyzed using Aiken's V formula to assess the appropriateness of the content by experts. Instrument reliability was calculated using the Cronbach's Alpha coefficient. Item characteristic analysis was conducted using the Rasch Model with Winsteps software to determine:

Person Reliability and Item Reliability, Item Fit (MNSQ and ZSTD), Item Difficulty (logit), Person Fit, Differential Item Functioning (DIF) to detect bias based on gender and school location.

An instrument is declared suitable if it meets the criteria for validity, reliability, and fit of the Rasch model.

RESULTS AND DISCUSSION

Well Content validation was conducted by three validators: an educational evaluation expert, a Physical Education lecturer, and a certified Physical Education teacher. Validity was assessed using a 1–5 Likert scale and analyzed using Aiken's V. The Aiken's V values for all items were within the range of 0.85–0.95,

indicating that all items met high content validity criteria. This indicates that the instrument is content-appropriate in measuring conceptual and daily life skills in Physical Education.

The instrument's reliability was analyzed using Cronbach's Alpha. In a small-scale pilot test with 52 students, a Cronbach's Alpha value of 0.864 indicated excellent reliability. A large-scale pilot test with 104 students yielded a Cronbach's Alpha of 0.889, also considered high. Thus, the instrument has excellent internal consistency in measuring student competency.

Item Characteristics Analysis Using the Rasch Model Analysis using Winsteps software showed the following results:

- Item Fit: Most items had Outfit Mean Square (MNSQ) values in the range of 0.7–1.3 and ZSTD between -2 and +2. No items had MNSQ values outside the 0.5–1.5 range, thus all items were deemed fit to the Rasch model.
- Item Difficulty: The distribution of item difficulty levels was relatively balanced, with logit values ranging from -1.23 to +1.40. This indicates that the instrument is capable of measuring low to high levels of ability.
- Person Reliability: The person reliability value reached 0.87, indicating consistency in measuring ability across students.
- Item Reliability: The item reliability value was 0.93, indicating high stability in measuring item characteristics.

Differential Item Functioning (DIF) DIF analysis showed no significant differences between student groups based on gender or school location. All items had probability values above 0.05, so the instrument can be categorized as free from group bias.

The results of the study indicate that the developed instrument meets the criteria for validity and reliability, with good item characteristics. This finding aligns with research by Sumintono and Widhiarso (2015), which stated that the Rasch Model is

capable of producing high-quality instruments for educational evaluation.

The use of contextual essay questions based on daily life encourages students to think critically, relate physical education (PJOK) concepts to real life, and enhance problem-solving skills. This instrument aligns with the policy direction of the Independent Curriculum, which demands more meaningful assessments with measurements oriented toward conceptual understanding and higher-order thinking skills.

Therefore, the developed instrument can be a solution for PJOK teachers to implement more objective and accurate assessments that support the comprehensive development of students' minimum competencies.

Table 1. Cognitive Domain Physical Education Test Items in Conceptual and Daily Life Abilities

No	Cognitive Domain		item
1	Analyzing daily physical activities	37,50%	6
2	Explain and implement a healthy lifestyle	50%	8
3	Explaining the importance of physical fitness for the body	12,50%	2
Total			16

Based on Table 1, the minimum competency assessment instrument, designed to measure conceptual and daily life skills, was tested to ensure its suitability for the desired skills. Content validity and reliability testing were conducted by a group of experts.

The minimum competency assessment instrument, specifically designed for fifth-grade students to measure conceptual and daily life skills, was tested for content validity by several experts. The content validity evaluation involved three experts who assessed the appropriateness of the material, structure, and language used in the instrument. The experts who evaluated the conceptual and daily life competency assessment instrument in the Physical Education (PJOK) questions included academics and practitioners. The results of the expert evaluation of the instrument were

then processed using the Aiken V formula prepared in Excel format. The data resulting from the content validity calculation using the Aiken V formula are presented in Table 2.

Table 2. Expert Agreement Coefficient

Index Item	Indeks Aiken's V	Description	Index Item	Indeks Aiken's V	Description
1	0,8	Valid	9	0,95	Valid
2	0,85	Valid	10	0,8	Valid
3	0,95	Valid	11	0,8	Valid
4	0,95	Valid	12	0,8	Valid
5	0,95	Valid	13	0,95	Valid
6	0,95	Valid	14	0,95	Valid
7	0,95	Valid	15	0,95	Valid
8	0,8	Valid	16	0,8	Valid

Berdasarkan data pada Tabel 2, koefisien kesepakatan antar pakar yang diperoleh kemudian dibandingkan dengan tabel koefisien validitas. Item versi Aiken dianggap valid jika koefisien kesepakatan antar melebihi nilai V. Tabel V Aiken menunjukkan bahwa nilai V untuk 3 asesor dan 3 pilihan skala validitas $> 0,80$. Sebaliknya, koefisien kesepakatan antar pakar dianggap tidak valid jika tidak mencapai nilai 0,80 dengan probabilitas kesalahan 0,04. Uji validitas konstruk membuktikan bahwa indikator yang dikembangkan secara akurat mengukur kemampuan variabel. Kemampuan konsep dan kehidupan sehari-hari melibatkan pemahaman. Konsep melalui kehidupan sehari-hari. Data yang digunakan untuk menguji validitas konstruk adalah hasil evaluasi kompetensi minimum pada skala kecil.

Analisis dilakukan dengan menggunakan perangkat lunak Winsteps. Validitas konstruk model Rasch dapat diamati melalui Polaritas Item. Nilai Korelasi Ukur Titik (Pt. Mea Cort) yang positif menunjukkan tidak ada konflik antara item yang diukur dan pengujian Unidimensionalitas konstruk juga dijelaskan pada Tabel 3.

Table 3. Unidimensionality Test

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)			
	-- Empirical --	Modelled	
Total raw variance in observations =	58.8 100.0%	100.0%	
Raw variance explained by measures =	34.8 68.0%	64.1%	
Raw variance explained by persons =	5.6 11.1%	10.5%	
Raw variance explained by items =	18.4 56.8%	53.6%	
Raw unexplained variance (total) =	16.6 32.0%	100.0%	35.9%
Unexplained variance in 1st contrast =	2.9 5.7%	17.5%	
Unexplained variance in 2nd contrast =	2.8 5.6%	17.4%	
Unexplained variance in 3rd contrast =	2.4 4.8%	15.1%	
Unexplained variance in 4th contrast =	1.4 2.9%	9.8%	
Unexplained variance in 5th contrast =	1.3 2.6%	8.3%	

Based on Table 3 above, the raw data variance measurement results are 68%. This value is not significantly different from the expected value of 64.91%. This indicates that the 20% unidimensionality requirement is met. A unidimensionality value of 40% is better, and a unidimensionality value of 60% is exceptional (Sumintono & Widhiarso, 2014). Unexplained variances were 5.7%, 5.6%, 4.8%, 2.9%, and 2.6%, respectively. This indicates that the variance unexplained by the instrument should ideally not exceed 15%, thus categorizing the instrument as good.

The second assumption is the assumption of local independence. This assumption is automatically proven by the evidence of unidimensionality in participant response data to a test (Retnawati, 2014). The assumption of local independence states that there is no correlation between test participants' responses to different test items (Kurniawan, 2019). In the explanation of unidimensionality above, it is proven that the test items tested have fulfilled unidimensionality so that they also fulfill local independence.

The Rasch model analysis was applied to the small-scale data, which consisted of 16 questions. The item analysis included evaluating item difficulty (item measure), item fit (item fit), individual abilities (person measure), and individual fit (person fit). Item Fit assesses whether the items function properly for measurement. The criteria used to evaluate item fit are outfit mean-square (MNSO), z-standard (ZSTD), and point measure correlation (Pt. Mean Corr). Items that do not meet these three criteria are considered suboptimal and need to be

revised or replaced. The results of the Item Fit analysis are detailed in Table 4.

Table 4. Output Item Fit

Item	Outfit		Pt. Measure Corr	No butir	Outfit		Pt. Measure Corr
	MNSQ	ZSTD			MNSQ	MNSQ	
9	3.30	1.7	-.06	4	.81	-.7	.62
14	3.22	1.6	-.03	3	.74	-.9	.80
16	3.01	2.9	-.02	7	.75	-.6	.70
1	1.97	3.0	.41	10	.72	-1.1	.81
2	1.92	1.7	.11	12	.60	-1.3	.61
8	1.91	1.5	.00	5	.50	-2.0	.60
13	1.15	.7	.73	15	.48	-1.9	.68
11	.95	-.2	.74	6	.48	-1.8	.68

The acceptable MNSQ score is $0.5 \leq \text{MNSQ} \leq 1.5$. The acceptable ZSTD score is $-2.0 < \text{ZSTD} < 2.0$. The acceptable Pt. Measure Corr score is $0.4 < \text{Pt. Measure Corr} < 0.85$ (Sumintono & Widhiarso, 2015). An item is considered fit or acceptable if it meets any of these three criteria.

The results of the Item Fit analysis showed that of the 16 items tested on 28 students, 10 were accepted, and 6 were discarded. The researcher discarded items 9, 14, 16, 1, 2, and 8 because they were deemed misfits.

The researchers then selected 10 questions from the 16 that met the criteria to be used as instruments in the large-scale trial phase. A total of 10 questions were selected for testing in the large-scale trial phase.

In the Item Measure analysis, the standard deviation reaches 1.80. If the value exceeds 0.0 logits plus the standard deviation, then the question is considered difficult. If the value falls between 0.0 logits plus the standard deviation and 0.0 logits minus the standard deviation, then the question is categorized as moderate. Meanwhile, if the value is less than 0.0 logits minus the standard deviation, then the question is classified as easy. Details regarding the Item Measure results and, the classification of difficulty levels can be seen in Table 5

Table 5. Output Item Measure

Item	Measure	Conclusion	Item	Measure	Conclusion
9	3.08	difficult	15	.24	moderate
14	3.08	difficult	5	-.18	moderate
8	1.56	moderate	1	-.79	moderate
2	1.37	moderate	4	-1.37	moderate
16	1.37	moderate	13	-2.16	easy
7	.61	moderate	10	-2.30	easy
6	.36	moderate	11	-2.36	easy
12	.36	moderate	3	-2.85	easy

Based on Table 4.8, of the 16 questions tested on 28 students in the small-scale test, there were 2 questions in the difficult category, namely questions number 9 and 14. A total of 10 questions were in the medium difficulty category. The remaining 4 questions were in the easy category, namely questions number 10, 11, and 3.

The results of the Output Person Measure show that the average person logit is -0.48 with a standard deviation of 2.44 on item point 1. A total of 4 students can Output Person Measure shows an average person logit of -1.74 and a standard deviation of 1.10. A total of 12 students can be classified as having a medium level of ability. While the remaining 16 students are classified as having low ability.

CONCLUSION

This research produced a valid and reliable assessment instrument to measure conceptual and practical skills in daily life in Physical Education (PJOK) based on the Rasch Model. The instrument development process, which included expert validation, small-scale trials, and large-scale trials, demonstrated that all items met the criteria for content validity and high reliability, and item characteristics consistent with the Rasch Model.

The developed instrument provides an accurate picture of student abilities and can be used by PJOK teachers as an effective diagnostic assessment tool. This instrument can help teachers design more targeted learning and improve the quality of learning outcome evaluations in accordance with the requirements of the Independent Curriculum.

It is recommended that teachers and education practitioners implement this instrument in their learning processes, and that future researchers develop similar instruments for other subjects and at different educational levels to ensure broader and more optimal application of minimum competency assessment.

ACKNOWLEDGEMENT

The author would like to thank all parties who contributed to the implementation of this research. Special thanks are extended to the expert validators who provided input and suggestions for improving the assessment instrument, as well as to the elementary school physical education (PJOK) teachers who provided the opportunity and support during the data collection process.

The author also expresses his appreciation to Semarang State University, specifically the Sports Education Study Program, Faculty of Sport Science, for providing facilities and guidance during the research.

Finally, the author hopes that this research will benefit the world of education and serve as a reference for the development of better learning evaluation instruments.

parties who contributed data in the history of sports in Indonesia.

REFERENCES

- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational Research: An Introduction* (7th ed.). Allyn and Bacon.
- Pusmenjar. (2020). *Pedoman Asesmen Kompetensi Minimum*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Retnawati, H. (2014). *Teori Respon Butir dan Aplikasinya*. Yogyakarta: Nuha Medika.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu*

- Sosial. Bandung: Trim Komunikata Publishing House.
- Permendikbudristek. (2022). Peraturan Menteri Pendidikan dan Kebudayaan Riset dan Teknologi Nomor 21 Tahun 2022 tentang Kurikulum Merdeka. Jakarta: Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Supriyadi, A., Susilaningsih, E., & Wibowo, R. (2022). Diagnostic Assessment in Kurikulum Merdeka: Concept and Implementation. *Journal of Educational Research and Evaluation*, 11(2), 213–223.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Springer.
- Purnamasari, R., Safitri, N., & Kurnia, D. (2023). Pengembangan Soal Asesmen Kompetensi Minimum (AKM) Literasi Numerasi Kelas 5 Sekolah Dasar. *Jurnal Basicedu*, 7(1), 787-797. <https://jbasic.org/index.php/basicedu>
- Puspasari, H., & Puspita, W. (2022). Uji Validitas dan Reliabilitas Instrumen Penelitian Tingkat Pengetahuan dan Sikap Mahasiswa terhadap Pemilihan Suplemen Kesehatan dalam Menghadapi Covid-19. *Jurnal Kesehatan*, 13(1), 65-71.
- Permendikbudristek. (2022). Standar Penilaian Pendidikan Permendikbudristek No. 21 tahun 2022. [Gurusumedang.Com. https://www.gurusumedang.com/2022/06/standar-penilaian-pendidikan.html](https://www.gurusumedang.com/2022/06/standar-penilaian-pendidikan.html)
- Purnama, D. N., & Alfaria, F. (2020). Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi Smk Berdasarkan Teori Tes Klasik Dan Teori Respons Butir. *Jurnal Pendidikan Akuntansi Indonesia*, 18(1), 36–46. <https://doi.org/10.21831/jpai.v18i1.31457>
- Purnama, D. N., & Alfaria, F. (2020). Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi Smk Berdasarkan Teori Tes Klasik Dan Teori Respons Butir. *Jurnal Pendidikan Akuntansi Indonesia*, 18(1), 36–46. <https://doi.org/10.21831/jpai.v18i1.31457>
- Rahayu, T. K., & Krisnawati, M. (2020). PENGARUH PENGGUNAAN BULU MATA PALSU TERHADAP KOREKSI MATA SIPIT UNTUK TATA RIAS PENGANTIN BRIDAL PADA PRODI TATA KECANTIKAN UNNES. *Beauty and Beauty Health Education*, 9(2), 219-222.
- Rahayu, I.S.D., & Purnawarman, P. (2019). The Use of Quizizz in Improving Students' Grammar Understanding through Self-Assessment. 254 (Conaplin 2018):102–106.
- Rahayu, R., & Ismawati, R. (2019). Pembelajaran Berbasis Masalah Pada Materi Pencemaran Lingkungan Sebagai Upaya Melatih Kemampuan Pemecahan Masalah Pada Siswa Smk. *IJNSE*, 2(2), 221-226. <https://doi.org/338475334/IJNSE.v2i2.2621>.
- Retnawati, h. (2014). Teori respons butir dan penerapannya: untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana.
- Resti, Y., & Kresnawati, E. (2020). “Peningkatan Kemampuan Numerasi Melalui Pelatihan Dalam Bentuk Tes Untuk Asesmen Kompetensi Minimum Bagi Guru Sdit Auladi Sebrang Ulu Ii Palembang.” *Jurnal Pendidikan* (November 2020): 18-19.
- Rahmi, Z., Ikhsan, M., & Susanti. (2022). Pengembangan Soal AKM (Assesmen Kompetensi Minimum) Numerasi untuk Siswa SMP. Skripsi tidak diterbitkan. Aceh: UIN Ar-Raniry.
- Ridlo, s. (2013). Pengembangan tes pengetahuan praktikum biologi berdasarkan graded response dan generalized partial credit. *Jurnal penelitian dan evaluasi Pendidikan*. 16, 166–182. <https://doi.org/10.21831/pep.v16i0.1111>
- Supriyadi, S., Lia, R. M., Rusilowati, A., Isaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan instrumen asesmen diagnostik untuk persiapan kurikulum merdeka. *Journal of Community Empowerment*, 2(2), 67-73.
- Susianita, A., & Koto, I. (2019). Pengaruh Penerapan Model Discovery Learning Melalui Media Nyata Terhadap Kemampuan Konseptual Dan Berpikir Kritis Siswa Pada Pembelajaran

- Ipa Kelas Iv Sd Negeri 45 Kota Bengkulu. Ejournal.unib.ac.id.
- Sumintono, b., & widhiarso, w. (2015). Aplikasi pemodelan rasch pada assessment pendidikan (issue october).
- Sumintono, b., & widhiarso, w. (2014). Aplikasi model rasch untuk penelitian ilmu-ilmu sosial.
- Sari, Ayang, Daulay, Syahnan, Putri, Yuliani, Y., Epriani, & Pita. (2021). Penghapusan Ujian Nasional Tahun 2021 Dalam Perspektif Guru SMA Di Kota Tebing Tinggi. Prosiding Seminar Nasional Pembelajaran Bahasa dan Sastra Indonesia (SemNas PBSI) - 3. pp. 213-220. ISSN 978-623-6984-07-9, <http://digilib.unimed.ac.id/id/eprint/41244>
- Sinambela, Pretty, Suhada, S., & Susilo, G. 2020. "Analisis Mengenai Dampak Penghapusan Ujian Nasional Terhadap Kelulusan Peserta Didik Jenjang Smp Di Era Pandemi Covid-19." In Seminar Nasional Dan Diskusi Panel Matematika, 281–90 . <http://www.proceeding.unindra.ac.id/index.php/DPNPMunindra/article/view/4730>
- Sugiono, S., Noerdjanah, N., & Wahyu, A. (2020). Uji validitas dan reliabilitas alat ukur SG posture evaluation. Jurnal Keterampilan Fisik, 5(1), 55-61.
- Sumintono, b., & widhiarso, w. (2015). Aplikasi pemodelan rasch pada assessment pendidikan (issue october).
- Schmid M., Brianza E. & Petko D. (2020). Efficient self-report measures for technological pedagogical content knowledge (TPACK): Constructing a reliable and valid short scale among pre-service teachers, Computers & Education, doi: <https://doi.org/10.1016/j.compedu.2020.103967>
- Susilaningsih, E., Aprilia, N., & Hayati, S (2022). The Effectiveness of Teaching Materials Loaded with Blended Learning with a Scientific Approach on Buffer Solution Learning. JIPK. 16 (1). <https://doi.org/10.15294/jipk.v16i1.33774>
- Susilaningsih, E., Supartono, S., Kristanto, T., Sariana, E., Azizah, P., & Natasukma, M. (2019). The Effectiveness of Multiple Representation Oriented Learning Material with Project Based Learning to Improve Students' Chemistry Learning Outcomes. ASSEHR: ICERI. Volume 30. 10.2991/iceri-18.2019.18
- Susilaningsih, E. (2021). Pengembangan Instrumen penilaian collaborative problem solving pada materi larutan elektrolit nonelektrolit dan analisisnya menggunakan Rasch model. Chemistry in Education, 10(1), 1-7.
- Susilaningsih, E., Fatimah, S., & Nuswawati, M. (2019). Analysis of students' conceptual understanding assisted by multirepresentation teaching materials in the enrichment program. KnE Social Sciences, 85-98.
- Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Rahayu, W., Suraji, S., & Sholihin, N. (2021, November). Analisis keefektifan pengabdian masyarakat pada kegiatan workshop AKM Guru SMP Mapel Bahasa Indonesia Kabupaten Semarang. In Unri Conference Series: Community Engagement (Vol. 3, pp. 84-91).
- Supriyadi, S., Suraji, S., Isnaeni, W., Rusilowati, A., Khumaedi, M., & Lia, R. M. (2020, December). Bimtek Penyusunan Tes Online Bagi Guru SD (Program Pengabdian Pada Masyarakat Bidang Pendidikan). In Seminar Nasional Pengabdian Kepada Masyarakat UNDIP 2020 (Vol. 1, No. 1).
- Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan Instrumen Asesmen Diagnostik untuk Persiapan Kurikulum Merdeka. Journal of Community Empowerment, 2(2), 67-73.
- Setiani, A. (2021, August). Evaluasi Instrumen Model dan Rancangan Model Entrepreneurship Berbasis Praktik Pembelajaran dalam Meningkatkan Kompetensi Profesional. In Posiding Seminar Nasional UNNES. UNNES.
- Safitri, R. K., & Dewi, N. R. (2020). The Development of Diagnostic Test with Testlet Model to Detect Difficult Learning Students in Temperature and Its Changes Materials.

- Unnes Science Education Journal, 9(2), 98-110.
- Sunarti, S., & Rusilowati, A. (2020). Pengembangan bahan ajar digital gerak melingkar berbantuan scratch berbasis science, technology, engineering, and mathematics. *UPEJ Unnes Physics Education Journal*, 9(3), 284-290.
- Tafryda, S. (2022). PENGEMBANGAN INSTRUMEN DARING UNTUK MENGUKUR KEMAMPUAN KOMUNIKASI SAINS BERBASIS MULTIREPRESENTASI PESERTA DIDIK PADA MATERI GERAK LURUS. *UPEJ Unnes Physics Education Journal*, 12(2), 92-102.
- Utami, C., & Anitra, R. (2019). MATA KULIAH MATEMATIKA SD Students Mathematical Understanding Ability Based on Initial Ability in Mathemcs for Elementary Sochool Subject. 11(02), 103–110.
- Wahyuningsih, t., raharjo, t., & masithoh, d. (2013). Pembuatan instrumen tes diagnostik fisika kelas xi. *Jurnal pendidikan fisika*, 1(1), 111–117.
- Widiawati, N. P., Pudjawan, K., & Margunayasa, I. G. (2015). Analisis pemahaman konsep dalam pembelajaran IPA pada siswa kelas IV SD di gugus II kecamatan banjar. *EJournal PGSD Universitas Pendidikan Ganesha*, 3(1), 12–14. <https://ejournal.undiksha.ac.id/index.php/JJPGSD/article/view/5847>
- Wiguna, C. S., Sumaatmadja, N., & Ningrum, E. (2018). Pengaruh Model Pembelajaran POE Terhadap Pemahaman Konsep dan Kemampuan Berpikir Kreatif Peserta Didik. *Geografi Gea*, 13 (1).
- Wahyuni, A., & Yusmaita, E. (2020). Perancangan instrumen tes literasi kimia pada materi asam dan basa kelas XI SMA/MA. *Edukimia e-ISSN: 2502-6399*. <http://edukimia.ppj.unp.ac.id/ojs/index.php/edukimia/>
- Yunita, M. (2016). Miskonsepsi Siswa Kelas X SMA Negeri 1 teluk Batang pada materi kalor dan perpindahannya. *Jurnal pendidikan dan pembelajaran*. 5(2).
- Yolanda, S., & Meilana, S. F. (2021). Pengaruh Aplikasi Quizizz Terhadap Minat Belajar IPA Siswa Kelas V di Sekolah Dasar. *Jurnal Educatio FKIP UNMA*, 7(3), 915-921.
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49-54.
- Sumintono, B., & Widhiarso, W. (2015). Pengantar Model Rasch: Aplikasi untuk Penelitian Ilmu-Ilmu Sosial. Bandung: Trim Komunikata Publishing House.
- Abdul kadir. (2015). Menyusun dan menganalisis tes hasil belajar abdul kadir. *Al- ta'dib*, 8(2), 70–81.
- Aiken, l. R. (1985). Evaluating ratings on bidirectional scales. *Educational and psychological measurement*, 45(2), 195–202. <https://doi.org/10.1177/001316448504500201>
- Amam, a. (2017). Penilaian kemampuan pemecahan masalah matematis siswa smp.
- Teorema: teori dan riset matematika, 2(1), 39. <https://doi.org/10.25157/.v2i1.765>
- Ambiel, r. A. M., noronha, a. P. P., & de francisco carvalho, l. (2015). Analysis of the professional choice self-efficacy scale using the rasch-andrich rating scale model. *International journal for educational and vocational guidance*, 15(3), 205–219. <https://doi.org/10.1007/s10775-015-9293-7>
- Andrich, d. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/bf02293814>
- Andrich, d. (2011). Rating scales and rasch measurement. *Expert review of pharmacoeconomics and outcomes research*, 11(5), 571–585. <https://doi.org/10.1586/erp.11.59>
- Angriani, a. D., nursalam, n., fuadah, n., & baharuddin, b. (2018). Pengembangan instrumen tes untuk mengukur kemampuan pemecahan masalah matematika siswa. *Auladuna: jurnal pendidikan dasar islam*,

5(2), 211.
<https://doi.org/10.24252/auladuna.v5i2a9.2018>

Anisa, w. N. (2015). Peningkatan kemampuan pemecahan masalah dan komunikasi matematik melalui pembelajaran pendidikan matematika realistik untuk siswa smp negeri di kabupaten garut. Jurnal penelitian pendidikan dan pengajaran matematika, 1(1).
<https://jurnal.unsil.ac.id/index.php/jp3m/article/view/wit11>

Arends, f., winnaar, l., & mosimege, m. (2017). Teacher classroom practices and mathematics performance in south african schools: a reflection on timss 2011. South african journal of education, 37(3).