

## THE DEVELOPMENT OF SUMMATIVE ASSESSMENTS FOR PPG STUDENTS IN THE SOCIOLOGY PROGRAM

**Seli Septiana Pratiwi, Herlina Ike Oktaviani**

*Pendidikan Profesi Guru (PPG), Pascasarjana, Universitas Negeri Malang*  
[seli.pratiwi.fis@um.ac.id](mailto:seli.pratiwi.fis@um.ac.id)

### Article History

*Received: 25 September 2025, Accepted: 13 November 2025, Published: 15 November 2025*

### Abstrak

Penelitian ini bertujuan untuk menganalisis tingkat kesulitan butir soal dalam mengukur potensi akademik mahasiswa Program Pendidikan Profesi Guru (PPG) Sosiologi, Universitas Negeri Malang. Pendekatan yang digunakan adalah kualitatif dengan analisis statistik melalui perangkat SPSS untuk memperoleh gambaran komprehensif terhadap karakteristik butir soal. Proses penelitian meliputi penentuan bentuk tes, pemetaan materi sosiologi, penyusunan kisi-kisi, penulisan dan penelaahan soal, pelaksanaan uji coba, serta analisis hasil tes. Subjek penelitian adalah mahasiswa PPG Sosiologi dengan analisis yang dilakukan meliputi daya pembeda, efektivitas pengecoh, tingkat kesukaran, validitas, dan reliabilitas butir soal. Daya pembeda dengan kategori baik lebih banyak daripada daya pembeda rendah. Berdasarkan hasil analisis diperoleh bahwa nilai reliabilitas menunjukkan instrument penilaian layak untuk digunakan kepada peserta tes. Ini berarti bahwa instrument soal tersebut secara internal relatif konsisten. Namun, ada beberapa item yang menunjukkan korelasi item-total koreksi rendah atau bahkan negatif. Hal ini merupakan bukti empiris bahwa instrument telah memenuhi standar konsistensi internal yang baik. Namun, untuk mencapai instrument yang berkualitas, perlu dilakukan penyempurnaan berkelanjutan melalui revisi item butir soal. Hal ini menunjukkan pentingnya standar minimum kualitas butir soal pada setiap lembaga pendidikan. Standar dapat berupa panduan teknis yang mengukur karakteristik soal yang baik seperti validitas empiris minimal, batas reliabilitas yang disarankan, dan prosedur revisi soal.

**Kata Kunci:** analisis butir; validitas; reliabilitas; tes; potensi akademik, pendidikan guru

### Abstract

*This study aims to analyze the level of item difficulty in measuring the academic potential of students in the sociology teacher professional education (PPG) Program at Universitas Negeri Malang. A qualitative approach was employed, utilizing statistical analysis through SPSS, to gain a comprehensive understanding of the item characteristics. The research process included determining the test format, mapping the sociology content, developing test blueprints, writing and reviewing test items, conducting tryouts, and analyzing the test results. The study involved Sociology PPG students as respondents. The analysis covered item discrimination, distractor effectiveness, item difficulty, validity, and reliability. The discriminatory power of the good category is greater than the discriminatory power of the low category. Based on the analysis results, the reliability value indicates that the assessment instrument is suitable for use with test participants. This means that the instrument is relatively internally consistent. However, there are several items that show low or even negative item-total correlation. This is empirical evidence that the instrument has met good internal consistency standards. However, to achieve a quality instrument, continuous improvement is necessary through item revisions. This demonstrates the importance of minimum standards for item quality in every educational institution. Standards can take the form of technical guidelines that measure good item characteristics such as minimal empirical validity, recommended reliability limits, and item revision procedures.*

**Keywords:** item analysis; validity; reliability; test; academic potential, teacher education

### To cite this article:

Pratiwi, S. S., & Oktaviani, H.I. (2025). The development of summative assessments for PPG students in the sociology program. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 8(4), 379–388. doi: [10.17977/um038v8i42025p379-388](https://doi.org/10.17977/um038v8i42025p379-388)

## INTRODUCTION

One way to improve teacher capacity is through the Teacher Professional Program (PPG), which refers to educational activities designed to prepare individuals to possess professional competence in teaching. The PPG program in Indonesia guides participants in mastering soft skills, forming an initial understanding of the curriculum, to good teaching practices. Soft skills taught in PPG programs foster self-confidence, collaboration, responsibility, respect for culture, mental resilience, and critical, creative, and positive thinking (Huda et al., 2021). The teaching profession is a field of work that requires expertise, skill, diligence, and knowledge; therefore, individuals involved in it need to possess the necessary knowledge, skill, and academic readiness (Supendi et al., 2023). Efforts to improve the quality of individuals working as teachers require to government to play a role as a decision-maker to improve teacher quality, and implementation of PPG programs can address these issues (Hoesny & Darmayanti, 2021). To date, challenges experienced by teachers related to professionalism have been addressed through PPG activities, which include both classroom and practical exercises, as well as those facilitated through a learning management system (LMS).

The PPG program is unique in that it divides participants into related fields. This ensures a more focused and directed learning process, aligned with the specifics of each field. Furthermore, the PPG program measures participants' abilities to determine whether they have achieved the program's objectives. One of the most common forms of assessment is through tests. These tests are used in the PPG program to demonstrate participants' competency attainment and mastery of cognitive, affective, and psychomotor skills. In this test-based assessment, participants not only work on conceptual questions but also on contextual questions relevant to everyday life and even problem-solving (Chien & Wu, 2020).

Test-based assessments can also measure academic potential and achievement, which can enhance cognitive, psychomotor, and affective skills (Lopez-Fernandez & Rodriguez-Illera, 2009). Tests also serve as a tool for measuring memory and are less risky than other forms of testing (Murphy et al., 2023). However, the use of test methods in its implementation has its drawbacks, particularly in terms of material mapping and item development. One test frequently used in the PPG program is the multiple-choice test. Multiple-choice tests are easy to score because they can be configured within the system and have definitive answers. However, assessments in the form of tests, particularly multiple-choice tests, present several challenges during their development. Problems encountered in item development include material that is dominated by one topic, interpretation of material that results in irrelevant questions, and the development of distractor options that do not function effectively. The development of item development results from material development, which also relates to the level of difficulty of each item. Each item has a different cognitive level to measure the test taker's ability.

The difficulty level of each item is divided into three categories: difficult, medium (ideal), and easy (Iñarrairaegui et al., 2022). The difficulty encountered by item compilers relates to item difficulty, including balancing items at difficult, medium (ideal), and easy levels. Question compilers tend to create more difficult or easier items, without considering the balance of item difficulty within a single item. The difficulty level of the item will also influence the question format, including the stimulus presented. The item difficulty will relate to the options presented in the multiple-choice test. Tests in the PPG program serve to select and test participants' knowledge and to more effectively refine their competencies. Therefore, item development needs to consider participant characteristics and curriculum standards to ensure effective and fair test administration (Hulaipah et al., 2023).

The PPG program applies to all fields, including the sociology PPG program. One of the challenges faced by prospective and prospective PPG participants in sociology is the test administered to determine graduation (Hutasoit & Purwasih, 2022). Implementing the PPG program in sociology presents significant challenges, including not all students possessing adequate knowledge of the program, varying perceptions and expectations of the program, and commitment and motivation to participate (Purwasih et al., 2023). This relates to participants' ability to follow the PPG program's learning activities and the evaluation process in the form of tests. The abstract nature of sociological studies poses a significant challenge for question writers. Therefore, question writers need to consider sociological concepts and theories, the characteristics of the test participants, the novelty of the phenomenon or case, and the distractors presented in multiple-choice questions.

Question writers need to pay attention to distractors presented in sociology questions to avoid ineffective distractors. Sociology questions tend to focus on the application of scientific concepts to a situation or phenomenon, requiring an understanding of the keywords in each question, including the question options. Distractors tend to be similar to the original answers. Furthermore, the abstract nature of sociological studies makes distractors invalid because they differ too much from the context of the question and the answer options. Invalid distractors will prevent participants from accessing information from previous memory, including when sociological material is processed into questions. If the distractors are not functional, they will hinder participants from interpreting and answering the questions (Schimmelfing & Persky, 2020).

Considering the explanations of the PPG program, the assessment in the form of tests, the construction of questions, and the test assessment for PPG sociology program participants, it is clear that the ability to interpret sociological knowledge into questions, mapping the level of question difficulty, and the distractors within each question are required. This article will discuss the level of question difficulty for the test assessment for PPG sociology program participants. This aims to demonstrate the diverse distribution of cognitive levels and how question compilers interpret sociological material into questions. Therefore, the purpose of this article is to demonstrate the level of discriminating power of questions, distractors within questions, item validity, and test reliability..

## METHOD

This research employed quantitative methods with the aid of SPSS analysis. The respondents were 76 students in the Sociology Education Program (PPG) studying at the State University of Malang. The research process involved determining the format of the test, mapping the material according to the distribution of sociology material, developing a test outline, writing questions, reviewing the questions, testing the questions, and analyzing the test results. Reliability testing used Cronbach's Alpha. This coefficient indicates that if a variable shows a Cronbach's Alpha value  $> 0.60$ , it can be concluded that the variable is reliable or consistent in its measurement (Anggraini et al., 2022). The test questions were based on the PPG summative test outline for the sociology field of study, as described in the following table 1.

Table 1 presents an outline of the summative test questions for teacher professional education in the Sociology field of study. Graduate competencies (*kompetensi lulusan/KL*) refer to the abilities that graduates of teacher professional education (PPG) must possess as preparation for their professional careers. The indicator column presents what emerges within students as a representation of their competency mastery. Meanwhile, the learning materials column presents the material that students must study and master to achieve these competencies. From Table 1, it was developed into 25 multiple-choice questions and 10 complex multiple-choice questions.

**Table 1. PPG Summative Test Outline for the Sociology Field of Study**

Graduate Competency	Indicator	Study Material
KL 1	1.1 Apply strategies for a safe and comfortable learning environment for students	Strategies for a safe and comfortable learning environment for students
	1.2 Apply effective, student-centered learning	Effective, Student-centered learning
	1.3 Conduct assessment, feedback, and reporting focused on students	Assessment, feedback, and reporting centered on students
KL 2	2.1 Demonstrate behaviors reflecting moral, emotional, and spiritual maturity in compliance with teachers' code of ethics	Behaviors reflecting moral, emotional, and spiritual maturity in line with the teachers' code of ethics
	2.2 Practice self-development through reflective habits	Self-development through reflective habits
	2.3 Cultivate the importance of placing students at the center of learning	Students as the center of learning
KL 3.	3.1 Collaborate to improve learning	Collaboration for learning improvement
	3.2 Involve parents/guardians and the community in learning	Involving guardians and the community in learning
	3.3 Participate in professional organizations and broader networks for improvement	Involvement in professional organizations and broader networks for improvement
KL 4.	4.1 Use learning content and teaching methods	Implementation of content and teaching methods
	4.2 Use knowledge to determine characteristics influencing students' learning	Learning strategies suited to students' characteristics
	4.3 Use knowledge of curriculum components and their application for lesson plan design	Curriculum implementation in learning

## RESULT

The competency achievement of PPG sociology participants was measured through a test-based assessment. Item development and analysis were conducted prior to administering the test to PPG Sociology participants. In this study, 76 participants answered the test items, with an average completion time of 1–2 hours. Based on the results of the item-total statistical test using SPSS on 35 items, an overview of each item's contribution to the overall consistency of the test was obtained. The analysis was conducted using two main indicators: the correlated item-total correlation and the Cronbach's Alpha value if an item is deleted. The correlated item-total correlation value indicates the extent to which each item has a positive relationship with the total test score. According to general criteria, items with a correlation greater than 0.30 are categorized as valid, while items below 0.30 require revision or deletion because they cannot adequately differentiate respondents with high and low ability. Based on the analysis, 25 items with correlation values above 0.30 were declared valid, and 10 items with correlations below 0.30 required improvement or replacement.

The Cronbach's Alpha value for item deletion was used to assess the effect of item deletion on the instrument's overall reliability. The overall alpha value of 0.798, which is considered high, indicates that the test has good internal consistency. No item deletion significantly increased the alpha, thus the overall instrument can be considered reliable and suitable for measuring the academic potential of PPG Sociology students. Test takers' ability to answer questions is partly determined by the item's discriminatory power. A variance analysis was conducted to determine the ability of each item to differentiate between high- and low-achieving test takers.

Discriminatory power is an important indicator in selecting test items because it indicates the extent to which an item can effectively measure differences in participant ability.

The results revealed that the discriminatory power for 35 items was classified as "adequate" and "poor" based on their discriminatory power correlation values. Twenty-seven items fell into the "adequate" category, with discriminatory power values ranging from 0.017 to 0.628, indicating the item's ability to adequately differentiate between high- and low-achieving participants. Eight items had poor discrimination power with low scores, some even negative at -0.130 and 0.000, indicating they were ineffective in differentiating test takers' abilities.

Items with poor discrimination power should be revised or replaced to improve test quality and the accuracy of measuring academic potential. Thus, the discrimination power results indicate that most items in the test instrument exhibited good quality in terms of discrimination power, although several items still need to be revised or replaced to improve measurement accuracy in the summative test for PPG students in sociology.

**Table 2. Item Categories Based on Distinction Power**

Highest		Ideal		Lowest	
Target	Realisation	Target	Realisation	Target	Realisation
	0		25		10

Table 2, which categorizes test items based on their discrimination power, shows that out of 35 test items, 25 were categorized as "ideal" and 10 as "low." Based on this, 71.43% of the test items had ideal discrimination power. This means that test takers' tendency to answer the questions was categorized as high. Meanwhile, 28.57% of the test items had low discrimination power.

The test also includes a difficulty level for each item. This is used to assess the level of thinking of test takers, including their ability to interpret questions. Difficulty level analysis is conducted to determine how easy or difficult each item is for test takers to answer. The difficulty index indicates the proportion of participants who answer an item correctly and is used to assess the balance of difficulty levels within a test item.

Based on the calculation results, the difficulty index ranges from 0.039 to 1.000. The grouping criteria refer to the guidelines, namely: 1) 0.00–0.30 is considered difficult, 2) 0.31–0.70 is considered moderate, and 3) 0.71–1.00 is considered easy. The calculation results show that 17 questions are classified as easy, with scores ranging from 0.711 to 1.000. Therefore, these questions have a low difficulty level, meaning they are easy for test takers to answer.

The analysis results for 16 questions are classified as moderate, with difficulty scores ranging from 0.303 to 0.671. Questions in this classification have a moderate level of difficulty, meaning they are quite challenging for test takers, although they can still be answered by the majority. On the other hand, the analysis results show that two questions fall into the difficult category, with scores of 0.039 and 0.118, respectively, indicating that PPG test takers are likely unable to answer the questions.

**Table 3. Item Categories Based on Item Difficulty**

Difficult		Moderate		Easy	
Target	Realisation	Target	Realisation	Target	Realisation
	2		17		16

Table 3, which categorizes test items based on difficulty, shows that 2 items fall into the difficult category, 17 items fall into the medium category, and 16 items fall into the easy category.

This indicates that 45.71% of test takers tended to find the questions easier to answer. Conversely, 48.57% of test takers tended to find the answers easier, even though they had to think carefully about the correct answer. 5.71% of test takers fall into the difficult category for item difficulty. Based on this analysis, test takers who answered the questions correctly fall into the high category.

Distractors are important items in multiple-choice test items. Distractors can be a trap for test takers when choosing an answer because they are made to be as similar as possible. However, distractors can also become the answer if the test designer is not careful and thorough in creating distractor options

**Table 4. Question Item Categories Based on Distractors**

All option functions		Some option function		Not all options function	
Target	Realisation	Target	Realisation	Target	Realisation
	7		21		7

Table 4 shows that of the 35 questions, 7 had all answer choices functioning properly, 21 had some answer choices functioning properly, and 7 had all answer choices not functioning properly. The calculation shows that 60% of the distractors were partially functioning. Conversely, 20% of the distractors were fully functioning, and 20% were not functioning properly. This indicates that the majority of the distractors in the questions were functioning properly. Test takers will repeatedly review the questions and answer choices before determining the correct answer, including comparing them with the distractors they created.

The questions given to PPG sociology test takers were analyzed to determine the appropriateness of each item. This included categorizing the questions as valid or invalid. Based on the 35 questions, 17, or 48.6%, fell into the valid category, with validity scores ranging from 0.298 to 0.502. Questions in this category are considered to accurately measure the abilities being tested. On the other hand, 18 questions fell into the invalid category, with low to negative validity values, such as -0.35 and -0.03, indicating that they were inaccurate or irrelevant in measuring the desired abilities. Invalid items need to be revised, replaced, or removed to improve the quality and accuracy of the measurement instrument.

## DISCUSSION

Based on the analysis, a reliability value of  $> 0.798$  indicates that the assessment instrument is suitable for use with test participants. This means that the instrument is relatively internally consistent. A reliable test indicates that participants with similar abilities will obtain similar scores when retested under similar conditions (Levine, 1955; Rahnev, 2025). However, several items showed low or even negative item-total correlations. Negative correlations indicate that the item is inconsistent with other items; perhaps the correct answer does not align with the ability construct being measured, or participants were confused by the options/phrasing of the item. These items need to be reanalyzed, revised, or even dropped to improve overall reliability. The sociology items are used to measure the pedagogical, professional, social, and personality competencies of test participants.

The items are not always balanced: some are too easy, others too difficult, so not all students are tested at the appropriate level (Wahab et al., 2023). A difficulty level of "moderate" or "difficult" is useful for differentiating student abilities, but items that are too easy or too difficult are less helpful in measuring overall ability (Sari & Rakhmawati, 2024). Some items do not meet the statistical criteria in the Rasch measurement model, such as INFIT/OUTFIT items that are too far from the ideal value. This means the item does not function as intended to measure the intended indicator. These items should be revised or removed from the scoring to avoid bias in the measurement results. Statistically, the instruments used generally demonstrate adequate reliability (e.g., reliability  $> 0.7$ ), indicating good internal consistency (Sari & Rakhmawati, 2024). However,

consistency alone is not enough; it must be ensured that each item is also valid and has good discriminatory power.

A reliable test minimizes random errors that can affect the measurement results. Therefore, the higher the reliability value, the smaller the measurement error rate, and the greater the confidence in the resulting score. A value of 0.798 indicates that the instrument has good measurement quality and can be used to consistently evaluate test takers' abilities. Accurate measurement will determine the quality of assessments for PPG sociology test takers, demonstrating the achievement of learning quality through validity, reliability, and fairness of key attributes. Validity is ensured by aligning assessments with learning outcomes and cognitive levels, while reliability is maintained through consistency and uniformity. Fairness considers individual characteristics and the testing context for valid interpretation (Rao & Banerjee, 2023). In this study, a high reliability value confirms that the instrument can fully and stably reflect the construct being measured.

The analysis process identified several items with low or even negative item-total correlations. This negative correlation indicates that the item behaves inconsistently or is not aligned with the ability construct measured by the entire instrument. A low correlation indicates that the item contributes little to the overall construct being measured, while a negative correlation indicates a mismatch between the item content and the measurement objectives. Items with a negative correlation can be interpreted as meaning that participants with low ability tend to answer correctly. This could also indicate possible errors in the formulation of the answer key, ambiguous wording of the question, or inappropriate context. Items with negative correlations should be reviewed by revising, modifying, or deleting them to avoid reducing the instrument's overall reliability.

Findings of inappropriate test items or imbalances in difficulty levels provide teachers with input to be more careful in creating questions: for example, evaluating questions through routine item analysis. Schools/institutions need to provide teacher training on test construction and item analysis techniques. At the policy level, minimum standards for test item quality can be established to ensure more objective and fair learning evaluations. Optimizing learning is necessary to improve learning outcomes and achieve learning targets. Relatedly, if learning practices are not optimal and innovative, learning outcomes are likely to be difficult to improve (Sari et al., 2024). The material in the summative sociology test is considered important and closely related to everyday life, so the material tested is aligned with material at the high school level (Parwati et al., 2025).

In general, questions designed for PPG students in sociology serve as an instrument to demonstrate the quality of professional teacher competency in the PPG program. Therefore, each test item must represent relevant learning indicators, such as a focus on professional skills, pedagogy, and application in good practice. Test takers need to focus on analysis and emphasize the ability to discover learning processes from various sources, formulate and solve problems, and collaborate in solving existing problems (Muzainuddin et al., 2024). This is based on the format of questions presented in summative tests, which are more practical than conceptual. Therefore, test takers need not only to master concepts and theories but also to understand how these concepts and theories are effectively implemented.

Research results demonstrate the importance of minimum standards for item quality in every educational institution. Standards can take the form of technical guidelines that measure good question characteristics, such as minimum empirical validity, recommended reliability limits, and question revision procedures. With these standards, the learning evaluation process will be more

objective, transparent, and accountable. Furthermore, a digitally managed question bank system can be a long-term solution for storing, selecting, and validating questions based on continuously updated empirical analysis results.

A reliability value of 0.798 is not just a statistic, but empirical evidence that the instrument has met good internal consistency standards. However, to achieve a quality instrument, continuous improvement is necessary through item revisions, increased item validity, balancing question difficulty levels, and training in instrument development based on empirical analysis. These steps will ensure that educational evaluation truly functions as a fair, objective, and useful diagnostic tool. The test is conducted to demonstrate the learner's adaptation process, in this case, PPG students in sociology, including overcoming challenges faced amidst professional duties, family, and study schedules that require adjustments (Halimah et al., 2024).

## CONCLUSION

The analysis of the PPG Sociology test items indicates that the instrument has good reliability, with a Cronbach's Alpha value of 0.798. This indicates that the questions developed are generally consistent in measuring participants' competencies. However, there are still a number of invalid items, unbalanced difficulty levels, and several distractors that do not function optimally. These findings indicate the need for improvements, especially for items with low or negative correlations, to improve the instrument's quality. Overall, this test instrument can be used to measure the competency achievements of PPG Sociology students, but revisions are still needed to make the items more representative and fair. Improvements can be made through re-evaluation of indicators, question wording, and adjustments to answer options. Thus, the test used is not only reliable but also valid and able to distinguish participants' abilities more accurately.

## ACKNOWLEDGEMENT

In compiling this article, we would like to thank the PPG study program at Malang State University for providing support in compiling articles that can be useful for education, especially in the field of learning evaluation.

## REFERENCES

- Anggraini, F. D. P., Aprianti, Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas. *Jurnal Basicedu*, 6(4), 6491–6504. <https://doi.org/10.31004/basicedu.v6i4.3206>
- Chien, S. P., & Wu, H. K. (2020). Examining influences of science teachers' practices and beliefs about technology-based assessment on students' performances: A hierarchical linear modeling approach. *Computers and Education*, 157(December 2019), 103986. <https://doi.org/10.1016/j.compedu.2020.103986>
- Halimah, N., Degeng, M. D. K., & Soepriyanto, Y. (2024). Peranan Tingkat Penerimaan Learning Management System Terhadap Kemampuan Self Regulated Learning. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 7(3), 118. <https://doi.org/10.17977/um038v7i32024p118>
- Hoesny, M. U., & Darmayanti, R. (2021). Permasalahan dan Solusi untuk Meningkatkan Kompetensi dan Kualitas Guru : Sebuah Kajian Pustaka. *Scholaria : Jurnal Pendidikan Dan Kebudayaan*, 11(2), 123–132. <https://ejournal.uksw.edu/scholaria/article/view/3595>
- Huda, N., Suyanto, Arifi, A., Putranta, H., & Azizah, A. N. M. (2021). Experiences of participants in teacher professional education on obtaining soft skills: A case study in Indonesia. *European Journal of Educational Research*, 10(1), 313–325. <https://doi.org/10.12973/EU-JER.10.1.313>

- Hulaipah, A., Syukri, M., & Indraswati, D. (2023). Analisis Kesulitan Guru Kelas IV dan V Dalam Menyusun Soal HOTS Pada Mata Pelajaran IPAS di SDN 2 Perampuan. *Jurnal Ilmiah Profesi Pendidikan*, 8(4), 2450–2460. <https://doi.org/10.29303/jipp.v8i4.1768>
- Hutasoit, E. F. J., & Purwasih, J. H. G. (2022). Tantangan Alumni Pendidikan Sosiologi Mengikuti Seleksi Pendidikan Profesi Guru Prajabatan. *Edukasi: Jurnal Pendidikan*, 20(2), 259–272. <https://doi.org/10.31571/edukasi.v20i2.4453>
- Iñarrairaegui, M., Fernández-Ros, N., Lucena, F., Landecho, M. F., García, N., Quiroga, J., & Herrero, J. I. (2022). Evaluation of the quality of multiple-choice questions according to the students' academic level. *BMC Medical Education*, 22(1), 1–7. <https://doi.org/10.1186/s12909-022-03844-3>
- Levine, R. (1955). Equating the score scales of alternate forms administered to samples of different ability. *ETS Research Bulletin Series*, 1955(2), i-118.
- Lopez-Fernandez, O., & Rodriguez-Illera, J. L. (2009). Investigating university students' adaptation to a digital learner course portfolio. *Computers and Education*, 52(3), 608–616. <https://doi.org/10.1016/j.compedu.2008.11.003>
- Muzainuddin, M., Setyosari, P., & Kuswandi, D. (2024). Pengembangan Web Untuk Model Resource Based Learning Dalam Pembelajaran Bahasa Inggris Kelas X. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 7(4), 209. <https://doi.org/10.17977/um038v7i42024p209>
- Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment. *Educational Psychology Review*, 35(3), 1–21. <https://doi.org/10.1007/s10648-023-09808-3>
- Parwati, N. N., Sari, N. K. A. I., & Suharta, I. G. P. (2025). Development Of Interactive E-Lkpd Based On Pbl To Improve Mathematical Problem Solving Skills In Trigonometry Material Tenth Grade High School Students. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 8(1), 001–013. <https://doi.org/10.17977/um038v8i12025p001>
- Parwati, N. N., Sari, N. K. A. I., & Suharta, I. G. P. (2025). Development Of Interactive E-Lkpd Based On Pbl To Improve Mathematical Problem Solving Skills In Trigonometry Material Tenth Grade High School Students. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 8(1), 001–013. <https://doi.org/10.17977/um038v8i12025p001>
- Purwasih, J. H. G., Rozakiyah, D. S., Sholicha, A. F., & Tayeb, A. M. (2023). Students' Knowledge, Perception, and Motivation on Preservice Teacher Profession Education Program (Vol. 1). Atlantis Press SARL. [https://doi.org/10.2991/978-2-38476-078-7\\_36](https://doi.org/10.2991/978-2-38476-078-7_36)
- Rahnev, D. (2025). A comprehensive assessment of current methods for measuring metacognition. *Nature Communications*, 16(1), 701.
- Rao, N. J., & Banerjee, S. (2023). Classroom Assessment in Higher Education. *Higher Education for the Future*, 10(1), 11–30. <https://doi.org/10.1177/23476311221143231>
- Sari, Y. P., Sinaga, M., & Hermendra, H. (2024). Pengembangan Modul Pembelajaran Teks Poster Bermuatan Budaya Melayu Riau Untuk Siswa Kelas VIII MTs. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 7(4), 238. <https://doi.org/10.17977/um038v7i42024p238>
- Sari, T. N. I., & Rakhmawati, R. (2024). Analysis the quality of critical thinking and creativity questions in high school biology subjects with the Rasch model. *Journal of Biology Education and Learning (JBEL)*, 5(2), 45–56.
- Schimmelfing, L. C., & Persky, A. M. (2020). Activating prior knowledge using multiple-choice question distractors. *Medical Education*, 54(10), 925–931. <https://doi.org/10.1111/medu.14162>
- Supendi, P., Daryani, A., & Safitri, D. (2023). Pendidikan Profesi Guru (PPG). *Cendekia: Jurnal Pendidikan Dan Pengajaran*, 1(5), 7–17.

Wahab, A. M., Badriah, D. L., & Surya, A. (2023). Item analysis of final examination questions for social studies in junior high schools through the ITEMAN program. *Journal of Education Research and Evaluation, 7(2)*, 101–113.