



Analysis of Tryout Test Item for Students of the Elementary School Teacher Professional Education Program

Titis Angga Rini*
Yuniawatika

Teacher Professional Education Program Study Program, Graduate School, Universitas Negeri Malang
*Corresponding author, E-mail: angga.rini.fip@um.ac.id,

Paper received: 20-9-2025; revised: 20-10-2025; accepted: 23-1-2026

Abstract

This research is motivated by the importance of the availability of valid and reliable test instruments in measuring the four main competencies of elementary school teachers, which include professional, pedagogical, personality, and social competencies according to educational professionalism standards in Indonesia. The purpose of this study is to analyze the quality of the competency test tryout instrument in the elementary teacher professional education program (PPG SD) through tests of differentiation, difficulty, validity, reliability, and effectiveness of the diverter, so that the feasibility of the question items used can be determined. The research method used a quantitative approach with the analysis of question items on 35 test items, including the validity test, the reliability of Cronbach's Alpha, as well as the analysis of the distribution of answers and the effectiveness of the trickster. The results of the study showed that some of the questions had sufficient quality with adequate differentiation, moderate difficulty, and a well-functioning distractor, but still many invalid items and ineffective deceptionists were found. The reliability value of the instrument was around 0.7, indicating a fairly good internal consistency, although there were still items with low correlations that needed to be revised. Thus, this test instrument has the potential to be used as a competency measurement tool for students of the PPG SD Program, but requires a thorough revision of several items to be more valid, reliable, and representative.

Keywords: analysis item; elementary school teacher; teacher competency test; teacher professional education program;

INTRODUCTION

The Elementary School Teacher Professional Education Program (PPG) is one of the government's strategic efforts in preparing prospective teachers who are professional, competent, and ready to face the demands of the world of basic education. PPG graduates are expected to have four main competencies, namely pedagogic, professional, social, and personality, all of which will be measured through the Teacher Professional Education Student Competency Test (UKMPPG) (Chang et al., 2025; Mardhatillah & Surjanti, 2023) Before taking the exam, students are usually given a competency tryout test, which functions as an exercise as well as a means of self-evaluation. Through tryouts, students can map their abilities, identify weaknesses, and prepare the right learning strategies. However, the tryout test instruments often have not gone through a thorough quality testing process, so their validity and reliability are still doubtful. This condition has the potential to reduce the function of tryouts as an effective and accurate means of preparation (Manggopa & Batmetan, 2024).

Tryout test instruments should be designed to meet measurement quality standards, especially validity and reliability aspects. Validity ensures that the instrument actually measures the competency being targeted, while reliability indicates the consistency of measurement

results when used repeatedly. If the instrument used does not meet these two important criteria, then the results of the tryout test risk producing biased and less reliable data. As a result, students do not get a clear picture of their competencies, while PPG program managers do not have a strong evaluation basis to improve learning (Yuniawatika et al., 2022) Therefore, testing tryout instruments is an urgent need so that test results can be used validly, accurately, and provide optimal benefits for both students and teacher professional education providers (Thompson et al., 2020).

In addition to serving as a measure of ability, the tryout test also has a strategic role in building students' mental readiness to face the national competency exam. Through the results of the tryout, students get feedback that can be used as a reference to improve the learning process. They can recognize weaknesses in mastering the material, then strengthen the aspects that are still weak before facing the real test. In this case, the tryout instrument not only functions as a means of simulation but also as a diagnostic instrument that supports improving the quality of learning (Atkins et al., 2022) However, these benefits can only be obtained optimally if the instrument used has gone through the process of item analysis, construct validity test, and adequate reliability measurement, so that the results provided truly reflect the condition of the students.

In the context of PPG Elementary School, the quality of tryout instruments has a very significant impact because the results of student competency tests have direct implications for graduation and the feasibility of their profession as teachers. Invalid instruments can cause students to fail, not because they are less competent, but because of the weakness of the measuring instruments used (Kumar et al., 2021) This kind of situation has the potential to harm students while reducing the credibility of the PPG program. Furthermore, this will also affect the quality of graduates who will later become teachers in elementary schools, which in turn affects the quality of education at the elementary level (Tias & Tongjean, 2023) Therefore, research that focuses on testing tryout test instruments is very important, both to ensure fairness in the evaluation process and to improve the quality of teacher professional education in an ongoing manner (Pratiwi & Oktaviani, 2025).

Based on this exposure, this study aims to test the competency tryout test instrument of PPG elementary school students through the analysis of the validity, reliability, and characteristics of the question items used. With systematic testing, it is hoped that this instrument will be able to function as a valid, consistent, and representative measurement tool for students' competencies. The novelty of this research lies in the systematic testing of the instrument so that it functions as a valid, consistent, and representative measurement tool for students' competencies. This test is needed because the implementation of PPG in Indonesia every year uses item test results as the main determinant of graduation (Pangesti et al., 2020). The results of the research are expected not only to make a practical contribution to PPG program managers in compiling or improving evaluation instruments, but also to make a theoretical contribution to the development of educational assessment science in Indonesia. Thus, this research has a strategic role in supporting the improvement of the quality of the learning process, as well as ensuring that PPG graduates are truly worthy and ready to carry out their profession as educators in elementary schools (Nurwataniah et al., 2022).

METHOD

This study uses a quantitative approach with the test instrument analysis method. The subject of the study is 150 students of the Elementary School Teacher Professional Education Program (PPG) who took the competency tryout test in the current semester with probability sampling technique. The research instrument is a questionnaire in the form of multiple-choice question tools that are arranged in accordance with the PPG student competency grid. Data was collected through the results of tryout tests, which were then statistically analyzed. The analysis includes calculating the differentiation of question items to determine the ability of the question to distinguish between high-ability and low-ability students, as well as the level of difficulty to identify the level of difficulty of each item. The data processing process is carried out using the help of statistical software so that the results are more accurate, systematic, and scientifically accountable. The research procedure is carried out in several stages in Figure 1.

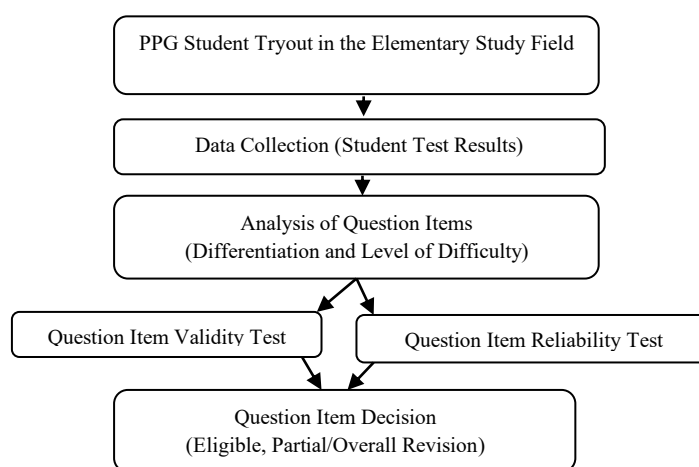


Figure 1. Research Flow

The analysis stage includes calculating the differentiation, difficulty, validity, and reliability. The differentiation and difficulty level were calculated using the question item analysis technique, while the validity of the instrument was tested through item-total correlation analysis. Reliability is calculated using Cronbach's Alpha coefficient to determine the internal consistency of the instrument. The results of this test are used to determine which question items are worth keeping, revising, or discarding. With this procedure, the research is expected to produce a tryout test instrument that meets the measurement quality criteria so that it can be used as a competency evaluation tool for PPG elementary school students in a valid and reliable manner.

RESULTS AND DISCUSSION

Elementary School PPG Competency Academic Test

The tryout instrument used in this study was designed to measure the four main competencies of teachers as required in professional standards of education. The four competencies include pedagogical, personality, social, and professional aspects, which are realized through learning management, personal example, communication skills, and extensive and in-depth mastery of the material (Rini et al., 2024). The four teacher competencies

measured in the PPG Elementary School tryout instrument are designed to be in accordance with the standards of teacher professionalism in Indonesia as described in Table 1 along with the proportion of question items. The first competency emphasizes the ability of teachers to manage student-centered learning, meaning that teachers are required to be creative in using methods, strategies, and approaches that allow students to actively build knowledge (Mardhatillah & Surjanti, 2023) The second competency emphasizes the personality side that is an example for students, including a steady attitude, noble character, and wisdom in making decisions. This aspect is strengthened by self-reflection on professional responsibilities and adherence to the code of ethics. The two complement each other because teachers are not only teachers, but also moral figures and role models for students (Widiyani et al., 2024)

Table 1. Competency Mapping of Tryout Test for PPG Elementary School Program

Graduate Competencies	Description	Number of Questions	Level of Questions		
			Easy	Moderate	Difficult
KL 1. Professional Competencies	Mastering and implementing student learning management with student-centered learning methods	11	1 (C2)	7 (C4)	3 (C5)
KL 2. Personality Competencies	Strong personality ability, noble character, wisdom, and authority, and become an example for students. These personality abilities are carried out through reflection in carrying out their responsibilities as teachers in accordance with the professional code of ethics and are student-oriented.	6	2 (C2)	3 (C4)	1 (C5)
KL 3. Social Competence	The ability of teachers to communicate and interact effectively and efficiently with students, fellow teachers, parents/guardians of students, and the surrounding community. The ability to communicate and interact effectively and efficiently is carried out in learning and self-development.	6	3 (C2)	3 (C4)	
AT 4 AM. Competence in Education	Mastering the subject matter broadly and in-depth to set learning objectives and organize student-centered learning knowledge content.	12	3 (C2)	6 (C4)	4 (C5)

The third and fourth competencies complete the profile of a professional teacher. Communication and interaction skills ensure that teachers can build harmonious relationships with students, colleagues, parents, and the community, thus creating a conducive learning environment. Effective communication is also key in delivering materials, giving direction, and building collaboration. Meanwhile, mastery of the material is the academic foundation that determines the quality of learning (Afriadi & Tola, 2022). Teachers who master content well can formulate clear learning objectives and arrange materials according to students' needs. Thus, these four competencies are not just separate indicators, but a complete series that ensures that PPG elementary school teachers have comprehensive professionalism in pedagogic, personality, social, and professional aspects (Chang et al., 2025).

Quality Test Results

Analysis of question items through differentiation, difficulty, and validity tests is an important stage to assess the quality of test instruments. Differentiation is used to see the ability of question items in distinguishing participants with high and low abilities, while difficulty level provides information on whether the questions are easy, medium, or difficult. Meanwhile, the validity of the question item shows the extent to which the question really measures the competence in question. By conducting this analysis, it can be determined which question items can be maintained, revised, or discarded so that the instruments used in measuring the competence of PPG elementary school students are truly quality and representative (Ariani, 2022).

The results of the analysis showed variations in the quality of the question items. In terms of differentiation, most of the questions are in the sufficient category, although there are still a number of questions with poor differentiation (e.g., Q2, Q9, Q15, Q18, Q25, Q28, and Q30), which indicate weaknesses in distinguishing students' abilities. In terms of difficulty, the majority of the questions are in the medium category, which is ideal for measurement, although there are items that are too easy (e.g., Q2, Q10, Q11, Q19, Q20, Q25, Q34) and some that are too difficult (Q9, Q15, Q17, Q18, Q28, Q32). Meanwhile, validity analysis shows that only a small percentage of items are valid, such as Q3, Q5, Q7, Q10, Q11, Q13, Q20, and Q23, while the majority of others are invalid. This indicates that although the instrument is generally quite functional, significant improvement is needed through revision or replacement of items in order for the instrument to optimally measure the targeted competencies. The details of the data from this test result can be observed in Table 2.

Next, the results of the Phase 1 reliability test with 35 questions showed a Cronbach's Alpha value of 0.701, which is in the reliability category, quite good. However, in the item-total correlation analysis, there are still several items with low correlations, even negative, such as items 15, 18, and 30, which indicate that the problem is inconsistent with the overall instrument. Therefore, a screening of question items that do not meet the criteria is carried out so that there are 22 items left in Phase 2. The results of the reanalysis in Stage 2 resulted in Cronbach's Alpha of 0.699, which is relatively stable compared to the previous stage, although it decreased slightly. This score still shows that the instrument has an acceptable enough internal consistency to be used in measuring student competence, although some low-correlated question items, such as numbers 8, 17, 27, and 28, still need to be considered for revision

Table 2. Test Results of Differentiation, Difficulty, and Test Validation

Item	Differentiation	Qualification	Difficulty Level	Qualification	Validity	Qualification
Q1	0.531	Enough	0.489	Moderate	0.297	Invalid
Q2	0.167	Signs	0.931	Easy	0.277	Invalid
Q3	0.592	Enough	0.349	Moderate	0.337716	Valid
Q4	0.371	Enough	0.332	Moderate	0.199996	Invalid
Q5	0.550	Enough	0.694	Moderate	0.377193	Valid
Q6	0.399	Enough	0.580	Moderate	0.232038	Invalid
Q7	0.312	Enough	0.846	Easy	0.320133	Valid
Q8	0.352	Enough	0.570	Moderate	0.164653	Invalid
Q9	0.139	Signs	0.103	Difficult	0.095924	Invalid
Q10	0.385	Enough	0.809	Easy	0.34501	Valid
Q11	0.342	Enough	0.820	Easy	0.331979	Valid
Q12	0.296	Signs	0.340	Moderate	0.121164	Invalid
Q13	0.526	Enough	0.717	Easy	0.386314	Valid
Q14	0.423	Enough	0.566	Moderate	0.251038	Invalid
Q15	0.102	Signs	0.204	Difficult	0.008402	Invalid
Q16	0.375	Enough	0.333	Moderate	0.177617	Invalid
Q17	0.339	Signs	0.251	Difficult	0.164464	Invalid
Q18	0.068	Signs	0.178	Difficult	-0.02941	Invalid
Q19	0.225	Signs	0.872	Easy	0.109686	Invalid
Q20	0.276	Signs	0.873	Easy	0.309222	Valid
Q21	0.272	Enough	0.728	Easy	0.296291	Invalid
Q22	0.350	Enough	0.459	Moderate	0.251491	Invalid
Q23	0.544	Enough	0.563	Moderate	0.306362	Valid
Q24	0.451	Enough	0.367	Moderate	0.232092	Invalid
Q25	0.529	Signs	0.932	Easy	0.292328	Invalid
Q26	0.172	Enough	0.449	Moderate	0.282472	Invalid
Q27	0.346	Enough	0.605	Moderate	0.163524	Invalid
Q28	0.329	Signs	0.137	Difficult	0.162232	Invalid
Q29	0.182	Signs	0.337	Moderate	0.11259	Invalid
Q30	0.064	Signs	0.920	Easy	-0.0393	Invalid
Q31	0.150	Enough	0.481	Moderate	0.208825	Invalid
Q32	0.405	Signs	0.248	Difficult	0.198141	Invalid
Q33	0.179	Enough	0.448	Moderate	0.067197	Invalid
Q34	0.523	Signs	0.879	Easy	0.292278	Invalid
Q35	0.209	Enough	0.489	Moderate	0.213744	Invalid

So that the reliability of the instrument increases in the next stage of development. The results of the Phase 1 reliability test with 35 questions showed a Cronbach's Alpha value of 0.701, which is in the reliability category, quite good. However, in the item-total correlation analysis, there are still several items with low correlations, even negative, such as items 15, 18, and 30, which indicate that the problem is inconsistent with the overall instrument. Therefore, a screening of question items that do not meet the criteria is carried out so that there are 22 items left in Phase 2. The results of the reanalysis in Stage 2 resulted in Cronbach's Alpha of 0.699, which is relatively stable compared to the previous stage, although it decreased slightly. This score still shows that the instrument has an acceptable enough internal consistency to be used in measuring student competence, although some low-correlated question items, such as numbers 8, 17, 27, and 28, still need to be considered for revision so that the reliability of the instrument increases in the next stage of development. Thus, it can be concluded that the process of eliminating question items does not significantly reduce reliability, but revision steps are still needed to strengthen the quality of the instrument.

Table 3. Results of Distractor Interpretation

Item	Answer Options					Distractor Interpretation					Conclusion
	A	B	C	D	E	A	B	C	D	E	
Q1	18	17	10	50	8	√	√	√	Key	√	Option to Use Without Revision
Q2	2	1	1	94	3	X	X	X	Key	X	Overall Revision of Options
Q3	14	17	11	25	35	√	√	√	√	Key	Option to Use Without Revision
Q4	34	33	14	15	7	Key	√	√	√	√	Option to Use Without Revision
Q5	16	4	2	9	70	√	X	X	√	Key	Partial Revision on: B, C
Q6	21	58	3	9	10	√	Key	X	√	√	Partial Revision on: C
Q7	1	9	1	86	5	X	√	X	Key	√	Partial Revision on: A, C
Q8	1	22	19	58	2	X	√	√	Key	X	Partial Revision on: A, E
Q9	65	8	10	9	9	√	√	Key	√	√	Option to Use Without Revision
Q10	1	7	3	82	8	X	√	X	Key	√	Partial Revision on: A, C
Q11	2	83	11	2	3	X	Key	√	X	X	Overall Revision of Options
Q12	34	8	11	14	33	Key	√	√	√	√	Option to Use Without Revision
Q13	8	13	72	1	7	√	√	Key	X	√	Partial Revision on: D
Q14	35	2	5	57	2	√	X	√	Key	X	Partial Revision on: B, E
Q15	25	6	21	47	2	√	√	Key	√	X	Partial Revision on: E
Q16	34	43	7	7	10	Key	√	√	√	√	Option to Use Without Revision
Q17	14	12	35	33	8	√	√	Key	√	√	Option to Use Without Revision
Q18	26	58	3	10	5	Key	√	X	√	√	Partial Revision on: C
Q19	8	7	28	58	0	√	√	√	√	Key	Option to Use Without Revision
Q20	2	8	88	1	3	X	√	Key	X	X	Overall Revision of Options
Q21	2	2	6	89	3	X	X	√	Key	X	Overall Revision of Options
Q22	6	74	17	2	3	√	Key	√	X	X	Partial Revision on: D, E
Q23	14	47	4	16	20	√	Key	X	√	√	Partial Revision on: C
Q24	66	0	21	15	0	√	X	√	Key	X	Partial Revision on: B, E
Q25	9	37	17	28	10	√	Key	√	√	√	Option to Use Without Revision
Q26	95	2	4	1	1	Key	X	X	X	X	Overall Revision of Options
Q27	0	32	49	0	21	Key	√	√	X	√	Partial Revision on: D
Q28	26	61	2	3	9	√	Key	X	X	√	Partial Revision on: C, D
Q29	6	33	22	14	27	√	√	√	Key	√	Option to Use Without Revision
Q30	34	18	28	9	13	Key	√	√	√	√	Option to Use Without Revision
Q31	2	4	2	1	93	X	X	X	X	Key	Overall Revision of Options
Q32	70	0	23	8	0	Key	X	√	√	X	Partial Revision on: B, E
Q33	19	0	47	35	0	√	X	√	√	Key	Partial Revision on: B
Q34	6	36	45	6	8	√	√	Key	√	√	Option to Use Without Revision
Q35	5	4	89	1	2	√	X	Key	X	X	Overall Revision of Options

The question item analysis test was also carried out by analyzing the distribution of answers, the effectiveness of the diverter (trickster), and the feasibility of the answer key. Through this analysis, the researcher can identify whether each answer option works as it should, i.e., the key is that the majority of participants can choose correctly while the distraction can attract some respondents who have not mastered the material (Rezigalla et al., 2024; Chauhan et al. 2023). Thus, this test serves as a diagnostic tool to ensure that the question items actually measure the student's abilities according to the expected competencies. The test results then provide an overview of the question items that are suitable for reuse, items that need partial improvement, and items that must be thoroughly revised. This analysis is very important because questions that are invalid or do not have an effective trickster can lower the overall reliability and validity of the test (Uba & Khairani, 2024). Therefore, the question item analysis test not only helps to improve the quality of the instrument but also ensures fairness, objectivity, and accuracy in measuring student learning outcomes. The results of the distractor test for this question are described in Table 3.

Based on the results of the analysis of the distribution of answers and the effectiveness of the distractor, most of the question items have shown good quality because the deceiver option (distractor) functions as it should, so no revision is needed. Question items such as Q1, Q3, Q4, Q9, Q12, Q16, Q17, Q19, Q25, Q29, Q30, and Q34 can be maintained without change because each option is chosen by the respondent, and the answer key can distinguish well. This shows that these items have met the eligibility criteria of the instrument, both in terms of content validity and deceptive appeal, so that they can be reused in learning evaluation. However, several questions require revision of both parts and all options. Partial revision is required for questions with uneven answer distribution or options that do not work at all, for example, in Q5, Q6, Q7, Q8, Q10, Q13, Q14, Q15, Q18, Q22, Q23, Q24, Q27, Q28, Q32, and Q33. Meanwhile, a thorough revision is needed on questions Q2, Q11, Q20, Q21, Q26, Q31, and Q35 because the majority of distractors are not selected at all, so their function as a decoy fails. This condition indicates weaknesses in both the wording of the questions, the clarity of the material, and the level of difficulty is too uneven, so that fundamental improvements are needed before they can be used in more valid measurements.

The results of the analysis of the PPG SD test instrument show that the quality of the questions is still diverse, both in terms of differentiation, difficulty level, and validity. According to the theory of educational evaluation, good question items must be able to distinguish participants with high and low abilities (Sharma, 2021), have a moderate level of difficulty, and be valid in measuring targeted competencies (Okonkwo, 2018). In this study, most of the items were in the category of sufficient in differentiation, but there were still many questions that fell into the bad category, even invalid. These findings, in line with test instruments, often require repeated revisions to strike a balance between reliability and validity (Lange et al., 1967). A fairly good reliability value ($\alpha = 0.701$ in stage 1 and $\alpha = 0.699$ in stage 2) indicates the internal consistency of the instrument, although there are still a number of points that need improvement. Values of α around 0.70 are widely cited as the threshold for "acceptable" reliability in research instruments, indicating that the items are sufficiently correlated to be considered consistent measures of the same concept (Yun et al, 2023). In addition, the results of the distractor analysis reinforce the idea that some trick options work well, but some questions require partial or thorough revision.

A diverter that the participant did not choose indicated that the option did not function as a deception, thus contradicting the principle of objective question construction that emphasizes the effectiveness of all answer options (Gierl et al, 2017). Studies show that non-functional distractors (options rarely or never chosen) signal that an item is poorly constructed and should be revised or removed (Rezigalla et al., 2024). A good distractor is chosen mainly by low-ability students; distractors attracting no one, or more high- than low-ability students, are considered poor and must be reviewed or replaced (Arhin et al., 2023). Thus, questions that have met the criteria can be reused because they have proven to be effective in measuring students' abilities, while weak items must be revised to make the instrument more representative (Uba & Khairani, 2024). The results of this study confirm that instrument evaluation is an important process in ensuring that the PPG test truly measures the four competencies of teachers accurately, fairly, and in accordance with educational professionalism standards in Indonesia (Kurniadi, 2025; Yuniawatika et al., 2022).

CONCLUSION

Based on the overall results of the analysis, it can be concluded that the PPG SD test instrument, designed to measure professional, pedagogical, personality, and social competencies, still needs improvement to be more optimal in assessing students' abilities. In terms of item quality, even though some of the questions have met the criteria with sufficient variability, moderate difficulty, and a well-functioning distress, there are still many items that are invalid or have ineffective tricks, so that they cannot distinguish participants optimally. The results of the reliability test showed fairly good internal consistency with a Cronbach's Alpha value above 0.69, indicating that the instrument was relatively stable, but there were several problems with low correlation that lowered the overall quality of the instrument. These findings reinforce the theory of educational evaluation that the quality of tests is not only determined by reliability, but also by the validity of the content and construction of the questions. Therefore, partial and thorough revision steps are very necessary, especially for items with poor differentiation, the level of difficulty is too easy or too difficult, and the validity is low. The expected improvement can truly represent the competence of teachers according to educational professionalism standards, as well as provide a fair, objective, and accurate measurement tool in assessing the readiness of PPG students to become professional educators in elementary schools.

ACKNOWLEDGEMENT

The author expresses his gratitude to the Teacher Professional Education Program Study Program, Graduate School, Universitas Negeri Malang for providing facilities and funding support so that the preparation of this article can be published.

REFERENCE

- Afriadi, B., & Tola, B. (2022). Evaluation Of The Implementation Of Teacher Professional Education in Indonesia. *International Education Trend Issues*, 1(1), 1–9. <https://doi.org/10.56442/ieti.v1i1.111>
- Arhin, D., Annan-Brew, R., & Owusuaah, R. (2023). Exploring the Quality of Multiple-Choice Question Type of Test Items in Information and Communication Technology Using Item Analysis. *E-Journal of Humanities, Arts and Social Sciences*. <https://doi.org/10.38159/ehass.2023414>
- Ariani, S. (2022). Analysing A Multiple-Choice Test To Find Out The Item Facility (IF), THE Item Discrimination (ID) and The Distractor Efficiency. *Inovasi Pendidikan*, 9(2). <https://doi.org/10.31869/ip.v9i2.3906>
- Atkins, D. L., Violette, L., Neimann, L., Tanner, M., Hoover, K., Rao, D., & Stekler, J. D. (2022). Provider perspectives on the use of motivational interviewing and problem-solving counseling paired with the point-of-care nucleic acid test for HIV care. *PLoS ONE*, 17(6 June). Scopus. <https://doi.org/10.1371/journal.pone.0270302>
- Chang, C.-F., Annisa, N., & Chen, K.-Z. (2025). Pre-service teacher professional education program (PPG) and Indonesian science teachers' TPACK development: A career-path comparative study. *Education and Information Technologies*, 30(7), 8689–8711. <https://doi.org/10.1007/s10639-024-13160-6>

- Chauhan, G. R., Chauhan, B. R., Vaza, J. V., et al. (2023). Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions. *Cureus*, 15(7), e42492. <https://doi.org/10.7759/cureus.42492>
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, 87(6), 1082–1116. <https://doi.org/10.3102/0034654317726529>
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85–S89. <https://doi.org/10.1016/j.mjafi.2020.11.007>
- Kurniadi, T. (2025). Enhancing Teacher Quality in Indonesia: A Systematic Literature Review of the PPG Prajabatan Program Perspectives. *Languae (Journal of Language and Education)*, 3(2), 121–127.
- Lange, A., Lehmann, I. J., & Mehrens, W. A. (1967). Using item analysis to improve tests. *Journal of Educational Measurement*, 4(2), 65-68. <https://doi.org/10.1111/j.1745-3984.1967.tb00572.x>
- Manggopa, H., & Batmetan, J. R. (2024). Evaluation of Students' Achievement in the Development of Innovative Learning Tools for PPG Computer Engineering and Informatics Students Using the Rasch Model. *International Journal of Information Technology and Education*, 3(2), 54–65. <https://doi.org/10.62711/ijite.v3i2.177>
- Mardhatillah, O., & Surjanti, J. (2023). Peningkatan Kompetensi Pedagogik dan Profesionalitas Guru di Indonesia Melalui Pendidikan Profesi Guru (PPG). *Jurnal Pendidikan Ekonomi Undiksha*, 15(1), 102–111. <https://doi.org/10.23887/jjpe.v15i1.65200>
- Nurwataniah, N., Hasan, S. H., & Susilana, R. (2022). Investigating the effectiveness of the teacher professional training program (PPG) in improving teacher competences. *European Online Journal of Natural and Social Sciences*, 11(4), pp-1390.
- Okonkwo, C. A. (2018). Evaluation of Validity of Computer Based Test Items in NOUN. *West African Journal of Open and Flexible Learning*, 6(2), 147-171. <https://wajofel.org/index.php/wajofel/article/view/18>
- Pangesti, F., Fauzan, F., & Risnawati, R. (2020). Kualitas butir soal try out uji pengetahuan dalam memprediksi tingkat kelulusan mahasiswa PPG. *Jurnal Pendidikan Profesi Guru*, 1(2), 91–98. <https://doi.org/10.22219/jppg.v1i2.13503>
- Pratiwi, S. S., & Oktaviani, H. I. (2025). Development Of Summative Assessments For PPG Students In The Sociology Program. *JKTP: Jurnal Kajian Teknologi Pendidikan*, 8(4), 379–388. <https://doi.org/10.17977/um038v8i42025p379-388>
- Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhoussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24(1), 445. <https://doi.org/10.1186/s12909-024-05433-y>
- Rini, T. A., Cahyanto, B., & Nurdianasari, N. (2024). Lesson Study-based Microteaching Design in Early Reading Instructional for Prospective Elementary School Teachers.

Sekolah Dasar: Kajian Teori Dan Praktik Pendidikan, 33(1), 59–72.
<https://doi.org/10.17977/um009v33i12024p59-72>

- Sharma, L. R. (2021). Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple Choice Questions of Speech Sounds of English. *International Research Journal of MMC*, 2(1), 15–28. <https://doi.org/10.3126/irjmmc.v2i1.35126>
- Thompson, V. L. S., Leahy, N., Ackermann, N., Bowen, D. J., & Goodman, M. S. (2020). Community partners' responses to items assessing stakeholder engagement: Cognitive response testing in measure development. *PLoS ONE*, 15(11 November). Scopus. <https://doi.org/10.1371/journal.pone.0241839>
- Tias, S. A., & Tongjean, W. (2023). Teacher Professional Education Program to Improve the Quality of Education: Lesson Learned from International Practice. In A. Mustofa, I. Widiyanah, B. K. Prahani, I. A. T. Rahayu, Moh. Mudzakkir, & C. D. M. Putri (Eds.), *Proceedings of the International Joint Conference on Arts and Humanities 2023 (IJCAH 2023)* (Vol. 785, pp. 1127–1134). Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-152-4_114
- Uba, A. R., & Khairani, A. Z. (2024). Assessing Item Reliability, Differential Item Functioning, and Wright Map Analysis of the GSP122 Test at a Public University in Nigeria. *Journal of Education For Sustainable Innovation*, 2(2), 107–120. <https://doi.org/10.56916/jesi.v2i2.934>
- Widiyani, T. P., Wijayanti, I., & Siswanto, J. (2024). Analisis Kompetensi Pedagogik Mahasiswa PPL PPG Prajabatan dalam Pembelajaran di Sekolah Dasar. *Ainara Journal (Jurnal Penelitian Dan PKM Bidang Ilmu Pendidikan)*, 5(2), 145–155. <https://doi.org/10.54371/ainj.v5i2.424>
- Yun, V., Ulang, N., & Husain, S. (2023). Measuring the Internal Consistency and Reliability of the Hierarchy of Controls in Preventing Infectious Diseases on Construction Sites: The Kuder-Richardson (KR-20) and Cronbach's Alpha. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. <https://doi.org/10.37934/araset.33.1.392405>
- Yuniawatika, Muksar, M., Angga Rini, T., & Alfian, M. (2022). Online Teacher Professional Education Program (PPG) Practices Through Moodle-Based E-learning. *2022 2nd International Conference on Information Technology and Education (ICIT&E)*, 60–64. <https://doi.org/10.1109/ICITE54466.2022.9759894>