

## Analysis of the Test Items for the Subject of Civics Education in Grade VII at SMPN 1 Swadaya for the Odd Semester

Arya Destian<sup>1</sup>, Prayoga Bestari<sup>2</sup>, Asep Mahfudz<sup>3</sup>

<sup>1,2,3</sup>Civics Education, Universitas Pendidikan Indonesia, Indonesia

[aryadest@upi.edu](mailto:aryadest@upi.edu)<sup>1</sup>, [yogabestari@upi.edu](mailto:yogabestari@upi.edu)<sup>2</sup>, [asepmahfudz@upi.edu](mailto:asepmahfudz@upi.edu)<sup>3</sup>

---

**Abstract:** Evaluation of learning is a systematic and continuous process to collect, analyze, and interpret information to assess the effectiveness of learning and the achievement of educational goals. In an effort to achieve these goals, the test method is one of the evaluation methods that can be used to measure, assess, and analyze the effectiveness of the learning that has been set. The purpose of this research is to determine the results of item analysis, validity, reliability, difficulty level, and discriminating power, as well as the conversion of raw scores into final scores on the Civics Education (PPKn) subject for 7th-grade students in the odd semester at SMPN 1 Swadaya. The research method used is descriptive quantitative with item analysis techniques based on classical theory. Data were obtained from the results of the student exam in the 7th grade of SMPN 1 Swadaya, which were then analyzed using item analysis software. The results of the study show that (1) With 30 students and 30 test items, the highest score was 30, and the lowest score was 18. (2) From the validity test, 15 items were valid, and 15 items were invalid, with validity levels ranging from very high to very low. (3) The difficulty level analysis showed that 27 items were categorized as easy, and 1 item was categorized as moderate. (4) Based on the students' score data, we classified them into 3 categories: Upper class with 8 students, Middle class with 14 students, and Lower class with 8 students. From the discriminating power test, 27 items had discriminating power, while 3 items had no discriminating power at all.

---

**Keywords:** Learning Evaluation, Item Analysis, Civics Education.

---

### Article History:

Received: 01-04-2025

Online : 30-04-2025



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

---

---

----- ◆ -----

### A. INTRODUCTION

Evaluation of learning is one of the most important aspects of the education system because it serves to measure how well the learning objectives that have been set are achieved. In the field of education, the evaluation process aims not only to determine how much knowledge the students possess, but also to provide feedback that can be used to improve the subsequent learning process. As the most commonly used instrument in learning evaluation, tests play an important role in providing an overview of students' mastery of the material. However, tests as an evaluation instrument must meet certain criteria to provide valid and reliable results. Therefore, it is important to analyze the items of the test to ensure that the questions can accurately measure students' competencies.

In the education system in Indonesia, the subject of Civic Education (Pendidikan Kewarganegaraan or PKn) is one of the subjects that plays a strategic role in shaping students' character and understanding of citizenship. Therefore, the tests for this subject must be carefully designed with high standards in mind, to provide results that are not only

accurate but also beneficial for the development of the education quality itself. One way to ensure the quality of the test items is by conducting an item analysis, which includes an analysis of validity, reliability, difficulty level, and discriminating power of each question. The validity of a question refers to how well the question measures what it is intended to measure, while reliability measures the consistency of the test results when repeated (Arikunto, 1999). Furthermore, the analysis of the difficulty level of the question ensures that the questions are neither too easy nor too difficult for the students, while the discriminating power measures the ability of the questions to differentiate between high and low-performing students (Solichin, 2017).

With the development of the education system and the continuous changes in the curriculum, the quality of test items has become increasingly important. For example, in 2013, the Ministry of Education and Culture of the Republic of Indonesia issued a policy regarding the Minimum Mastery Criteria (KKM) that is expected to improve the evaluation standards in all subjects, including PKn. However, this policy will only succeed if it is supported by comprehensive item analysis, involving empirical testing of the validity and reliability of each item. This is important because an invalid or unreliable evaluation can result in incorrect information about students' abilities, which in turn may affect learning policies in schools (Suwandi, 2008).

Based on this, this research is conducted to analyze the items of the Civic Education subject test for class VII students at SMPN 1 Swadaya for the odd semester. Through this analysis, we hope to provide a clearer picture of how well the questions used in the PKn test can effectively and accurately measure students' competencies. In addition, this analysis is expected to provide recommendations for improvement for those involved in question preparation, whether they are teachers or curriculum developers. Thus, this research is expected to contribute positively to improving the quality of learning evaluation in schools and, more broadly, to improving the quality of education in Indonesia as a whole.

## **B. METHOD**

The method used in this research is a qualitative method with a test item analysis approach. This approach aims to analyze the quality of the questions used in the Civic Education (PKn) test for class VII students during the odd semester at SMPN 1 Swadaya. The analysis process is carried out through various steps, including validity, reliability, difficulty level, and discriminating power tests for each test item. The data obtained from the analysis of the Civic Education test items for class VII students at SMPN 1 Swadaya during the even semester was gathered from 30 respondents, with a pilot test of 30 questions.

### C. RESULTS AND DISCUSSION

Here are the test items that were administered, as shown in Table 1.

**Tabel 1.** Test item

No	Question	A	B	C	D	Answer
1	Janji kemerdekaan kepada Indonesia diwujudkan Jepang dengan membentuk...	PETA	BPUPKI	PPKI	Chuo Sangi In	B
2	Pada masa pendudukan Jepang, tanggal 29 April 1945 di Indonesia dibentuk BPUPKI yang bertujuan untuk	Mempelajari dan menyelidiki hal penting yang berhubungan dengan pembentukan negara Indonesia merdeka	Mempelajari dan menyelidiki pemerintahan Jepang di Indonesia	Mempersiapkan hal-hal yang berkenaan dengan pemilihan presiden	Mempersiapkan strategi untuk menghadapi bangsa Jepang	A
3	Perhatikan nama- nama tokoh berikut ini... Dari nama tokoh diatas yang termasuk tokoh yang mengusulkan rumusan Dasar Negara adalah...	1,2,3	1,3,4	1,3,5	1,4,5	B
4	Hal yang membedakan antara Piagam Jakarta mengenai Dasar Negara dengan rumusan Pancasila pada Pembukaan UUD1945 ada pada sila...	Pertama	Kedua	Ketiga	Keempat	A

No	Question	A	B	C	D	Answer
5	Istilah Pancasila diusulkan oleh Ir. Soekarno pada sidang BPUPKI pada tanggal...	29 Mei 1945	31 Mei 1945	1 Juni 1945	18 Agustus 1945	C
6	Perhatikan unsur-unsur berikut : A. Persatuan, B. Kekeluargaan, C. Keseimbangan lahir dan batin, D. Musyawarah, E. Keadilan rakyat. Kelima unsur diatas adalah usulan dasar negara dari seorang tokoh, yaitu ....	Soekarno	M. Yamin	Soepomo	Radjiman Wedyodini ngrat	C
7	Berikut yang bukan anggota dalam panitia Sembilan adalah ...	DR. Radjiman Wedyodiningrat	Ir. Soekarno	Ahmad Soebardjo	D. Moh. Yamin	A
8	Sidang kedua BPUPKI dilaksanakan pada ...	29 Mei - 1 Juni 1945	1 Juni - 17 Juli 1945	10 - 17 Juni 1945	10 - 17 Juli 1945	D
9	Alasan BPUPKI diganti dengan PPKI adalah....	Tugas BPUPKI sudah selesai	BPUPKI mementingkan nasionalisme Indonesia	Anggota BPUPKI banyak yang menentang Jepang	Jepang merasa rugi dengan terbentuknya BPUPKI	A

No	Question	A	B	C	D	Answer
10	Setelah perjuangan panjang dan terbentuknya PPKI, Soekarno-Hatta dan Radjiman Wedyodiningrat diundang oleh Marsekal Terauchi untuk membicarakan perihal kemerdekaan Indonesia di ....	Vientiane	Dalat	Saigon	Hanoi	B

## 1. Validity Test

The validity test is conducted to determine how well each test item measures students' abilities accurately, in line with the desired objectives. The validity of a question indicates how accurately it measures what it is supposed to measure. The validity analysis process is conducted by correlating each test item with the total score using the Product Moment correlation formula. The correlation results show the validity level of the tested items, which can be categorized as very high, high, moderate, low, and very low. Based on the validity test results for the PKN questions tested in class VII at SMPN 1 Swadaya, the findings are as follows:

- Questions with a correlation coefficient greater than 0.6 are categorized as having high and very high validity. This indicates that these questions are reliable in effectively measuring students' competencies.
- Questions with a correlation coefficient between 0.4 and 0.6 show moderate validity, meaning the questions are still usable but need improvement.
- Questions with a correlation coefficient below 0.4 are categorized as having low and very low validity, meaning these questions are not suitable for accurately measuring students' competencies.

## 2. Reliability Test Results

The reliability test aims to determine how consistent and dependable the test results are when administered at different times or to different groups of students. In this study, reliability was tested using the Cronbach's Alpha reliability coefficient technique. The reliability test results indicate that most of the questions have a high reliability level, with reliability coefficients above 0.7, meaning these questions are quite consistent in measuring students' abilities. However, some questions had a reliability coefficient below 0.7, indicating that these questions are less stable and need improvement. Therefore, questions with low

reliability need to be revised to make them more reliable in measuring students' mastery of the material.

### **3. Difficulty Level Test Results**

The difficulty level test aims to measure how easy or difficult a question is for students. The difficulty level of a question is calculated using the difficulty index formula, which measures the proportion of students who can answer the question correctly. The difficulty level test results show that most of the questions have a balanced difficulty level, with difficulty indices ranging from 0.3 to 0.7.

- a. Questions with a difficulty index above 0.7 are considered too easy, as most students can answer them correctly.
- b. Questions with a difficulty index below 0.3 are considered too difficult, as many students cannot answer them correctly.
- c. Questions with a difficulty index ranging from 0.3 to 0.7 are considered to have a moderate difficulty level and are suitable for the ability level of class VII students.

It is important to maintain a balance between questions that are too easy and too difficult, to motivate students to try harder without feeling discouraged.

### **4. Discriminating Power Test Results**

The discriminating power test aims to assess the ability of a question to differentiate between high-achieving and low-achieving students. A good question should be able to distinguish students with high abilities from those with low abilities. The discriminating power index is calculated using the discrimination index formula, which measures the extent to which a question can differentiate between students who have mastered the material and those who have not. The discriminating power test results show that most of the questions have good discriminating power, with a discrimination index above 0.4. This indicates that these questions are effective in distinguishing between students who have mastered the material and those who have not. However, some questions had low discriminating power, indicating that these questions were not effective in differentiating students and need to be improved.

### **5. Raw Score Conversion to Final Score Analysis**

After conducting validity, reliability, difficulty level, and discriminating power tests, the next step is to convert the raw scores obtained by students into final scores. This conversion process is carried out to assess how well the test results reflect the overall ability of the students. In this study, the raw score conversion was done by considering all the factors that were analyzed previously, such as the validity and reliability of the questions. The raw score conversion results show that most students have final scores that are proportional to their mastery of the material being tested. However, some students who had inconsistent or low test results can be analyzed further to determine the cau.

Based on the analysis of the test items for the Civic Education subject (PKn) in class VII at SMPN 1 Swadaya, the validity test was conducted with 30 respondents and 30 test items. The results showed that 15 questions were valid and 15 questions were invalid. The validity interpretation revealed that 2 questions had very high validity, 2 questions had high validity, 8 questions had moderate validity, 5 questions had low validity, and 13 questions had very low validity. Meanwhile, from the analysis of the test items for the Civic Education subject (PKn) in class VIII at SMPN 1 Swadaya, the reliability test was conducted with 30 respondents, and the variance score was determined to be 12.67. Using the KR20 reliability formula, the reliability coefficient ( $r_{11}$ ) was found to be 0.62. Therefore, it can be concluded that the analysis of the reliability test for the PKn test items in class VII at SMPN 1 Swadaya indicates a moderate reliability status based on the reliability coefficient interpretation for the test items.

Based on the data analysis results, the difficulty level for question 1 is 97%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 2 is 83%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 3 is 90%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 4 is 97%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 5 is 90%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 6 is 90%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 7 is 93%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 8 is 77%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 9 is 97%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 10 is 83%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 11 is 83%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 12 is 43%. Therefore, the difficulty level of the question is categorized as moderate. The difficulty level for question 13 is 90%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 14 is 93%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 15 is 70%. Therefore, the difficulty level of the question is categorized as easy. The difficulty level for question 16 is 100%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 17 is 96%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 18 is 100%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 19 is 97%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 20 is 80%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 21 is 97%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 22 is 90%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 23 is 76%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 24 is 86%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 25 is 73%, therefore the

difficulty level of the question is categorized as easy. The difficulty level for question 26 is 93%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 27 is 100%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 28 is 70%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 29 is 96%, therefore the difficulty level of the question is categorized as easy. The difficulty level for question 30 is 90%, therefore the difficulty level of the question is categorized as easy.

Based on the student score data, we have classified the students into 3 categories: 8 students in the Upper class, 14 students in the Middle class, and 8 students in the Lower class. From the data analysis results, the Discriminating Power for questions 1 to 30, in order, are 0.066; 0; 0; 0.066; 0.133; 0.133; 0.133; 0.2; 0; 0.133; 0.133; 0.333; 0.2; 0.133; 0.4; 0; 0.06; 0; 0.066; 0; 0; 0; 0.2; 0.2; 0.2; 0.06; 0; 0.133; 0; 0.133. From this data, it can be concluded that 27 questions have discriminating power, while 3 questions, namely numbers 16, 18, and 27, have no discriminating power at all. The analysis of the Civic Education (PKn) test items for class VII at SMPN 1 Swadaya shows that although most of the items meet good criteria, some need to be revised to improve the quality of the learning evaluation. In this context, there are four main aspects that can be used as a reference for revision, namely validity, reliability, difficulty level, and discriminating power.

The validity of test items is a crucial aspect as it determines whether the test items accurately measure students' competencies. The analysis revealed that some items had low validity, with correlation coefficients below 0.4. For example, questions 2 and 4 were found to have inadequate correlation coefficients, indicating that these items do not effectively reflect students' mastery of the material. This aligns with the concept proposed by Arikunto, who emphasized that low-validity items can result in test outcomes that do not accurately represent students' abilities (Mulyana, 2023).

Revising items with low validity should involve clarifying and rewording the questions to better align with the intended learning objectives. This improvement is consistent with studies showing that well-revised questions can enhance validity (Rizky, 2023). Reliability is the next important factor in item analysis. It was found that some items had reliability coefficients below 0.7, indicating that these items were not consistent in measuring students' abilities over time. For example, questions 1 and 3 showed lower reliability values compared to the other items. This means that if the test were repeated, the results could vary significantly, potentially affecting the evaluation (Lestari & Setyarsih, 2020).

It is important to have reliable items to ensure that test results are dependable. To improve reliability, item development should involve modifications not only to the questions but also to the answer choices, ensuring they are clearer and less ambiguous, as shown in previous research (Amini et al., 2023). In terms of difficulty level, the analysis revealed that some questions were either too easy or too difficult. For example, questions 5 and 7 had difficulty indices above 0.7, while question 9 had a difficulty index below 0.3. Questions that are too easy may not trigger critical thinking in students, while questions that are too difficult can harm student motivation (Tilaar & Hasriyanti, 2019).



A balance in the clarity of question difficulty will contribute to the effectiveness of the test as an instrument for measuring material mastery (Juliani & Erita, 2023). Therefore, it is important to adjust the difficulty level of the identified items to match the abilities and understanding of class VII students. From the discriminating power perspective, several items also showed low discriminating power, below 0.4. This indicates that these items were not effective in distinguishing between students who mastered the material and those who did not. For example, questions 2 and 4 had poor discriminating power, meaning that these questions did not capture variations in student performance (Zulpan et al., 2024). Discriminating power is an important factor in the efficiency of evaluation, as good items should be able to identify high-achieving students. Therefore, revising items with low discriminating power should involve clarifying the questions and providing more relevant answer choices.

#### D. CONCLUSIONS AND SUGGESTIONS

Based on the analysis of all the aspects discussed, there are six fundamental recommendations for improving the PKn test items at SMPN 1 Swadaya. Items with low validity, reliability, difficulty level, and discriminating power need to be revised. The revisions should focus on clarifying question formulations, adjusting the difficulty level according to students' abilities, and improving ambiguous answer choices. These improvements are expected to result in more accurate evaluations and provide better feedback for future teaching. Furthermore, these revisions are expected to support the overall improvement of evaluation quality, providing new insights for teachers in developing better instructional materials in the future. Every step of the revision process should adhere to the principles of good educational measurement to ensure that learning objectives are achieved more effectively.

#### REFERENCES

- Arikunto. (1999). *Dasar – Dasar Evaluasi Pendidikan*. Bumi Aksara.
- Suwandi. (2008). *Penelitian Tindakan Kelas (PTK) dan Penulisan Karya Ilmiah*. Panitia Sertifikasi Guru Rayon 13.
- Amini, A., Nufus, H., & Agustin, E. (2023). Identifikasi Kemampuan Berpikir Kreatif Siswa Dari Sudut Pandang Gaya Kognitif Dalam Pembelajaran Matematika. *Juring (Journal for Research in Mathematics Learning)*, 6(3), 285. <https://doi.org/10.24014/juring.v6i3.25627>
- Juliani, R. P., & Erita, S. (2023). Analisis Validitas Dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis Dalam Konteks Sekolah Menengah. *Jeid Journal of Educational Integration and Development*, 3(3), 169–179. <https://doi.org/10.55868/jeid.v3i3.313>
- Lestari, D., & Setyarsih, W. (2020). Kelayakan Instrumen Penilaian Formatif Berbasis Literasi Sains Peserta Didik Pada Materi Pemanasan Global. *Ipf Inovasi Pendidikan Fisika*, 9(3), 561–570. <https://doi.org/10.26740/ipf.v9n3.p561-570>
- Mulyana, A. (2023). Investigating the Influence of Conformity and Fanaticism on Bullying Dynamics. *Psymphatic Jurnal Ilmiah Psikologi*, 10(2), 197–204. <https://doi.org/10.15575/psy.v10i2.29784>

- Rizky, K. (2023). Analisis Validitas Dan Reliabilitas Percobaan Soal Tes Sumatif Siswa Kelas X Semester 2. <https://doi.org/10.31219/osf.io/ytacr>
- Solichin. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas, Butir Tes, Interpretasi Hasil Tes dan Validitas Dalam Evaluasi Pendidikan. *Jurnal Manajemen & Pendidikan*.
- Tilaar, A. L. F., & Hasriyanti, H. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika Pada Sekolah Menengah Pertama. *Jp3i (Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia)*, 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>
- Zulpan, Sahrul, S., Yusron, A., Hanida, R. S., & Marfu'ah, S. (2024). Validitas Tes Objektif Bentuk Pilihan Ganda. *PPM*, 1(3), 10. <https://doi.org/10.47134/ppm.v1i3.812>