Analysis of Test Instrument Quality in Measuring Computational Thinking Skills in Seventh Grade Junior High School Ratio Material

Gustina Nandari[1], Ayu Kusumaningtyas[2], Meliyani Lutfiah[3], Agung Putra Wijaya[4*)], Chika Rahayu[5]
[1,2,3,4,5] Universitas Lampung
[*)]agung.wijaya@fkip.unila.ac.id

Abstract

The integration of computational thinking in mathematics learning is essential to support 21st-century skills; however, assessment practices in schools still inadequately measure students' thinking processes. This study aims to analyze the quality of a test instrument designed to measure junior high school students' computational thinking ability on ratio material. A quantitative descriptive-evaluative method was employed involving 64 seventh-grade students. The instrument consisted of five contextual essay items developed based on computational thinking indicators: decomposition, pattern recognition, abstraction, generalization, and algorithmic thinking. Item analysis was conducted to examine validity, reliability, difficulty index, and discrimination power. The results indicate that all items were valid and the instrument showed high reliability ($\alpha = 0.81$). The difficulty levels ranged from easy to moderate, while the discrimination power was categorized as sufficient to good. These findings demonstrate that the instrument is valid, reliable, and appropriate for measuring students' computational thinking ability in mathematics learning at the junior high school level.

Keywords: computational thinking, discriminating power, item analysis, level of difficulty, validity, reliability

Introduction

The development of digital technology in the 21st century requires the younger generation to have competencies that go beyond mere academic mastery, such as critical thinking, problem solving, and systematic information processing. One of the key skills that supports these demands is computational thinking (CT), which includes the ability to decompose, recognize patterns, abstract, and design algorithms to produce logical and efficient solutions. Since its comprehensive introduction by Wing (2006) as a fundamental skill for all individuals, CT has been increasingly integrated into various fields of learning, including mathematics, which has a strong affinity for logical reasoning and problem solving (Irawati and Hadi, 2025). However, the implementation of CT in learning has not been fully matched by the availability of high quality, valid, and reliable CT testing instruments, so that measuring students' CT abilities remains an issue that needs to be studied in depth.

In Indonesia, the integration of computational thinking (CT) into mathematics learning is gaining attention. Various studies have explored the development of CT-based

mathematics learning models, the application of interactive media, and the development of assessment instruments to measure students' CT abilities (Irawan et al., 2024). In addition, the application of CT in mathematics learning has been reported to improve students' problem-solving, reasoning, and higher order thinking skills (HOTS) (Kaswar and Nurjannah, 2024). However, most of these studies focus more on the aspects of implementation and learning impact, while studies that specifically examine the quality of CT test instruments, especially in terms of validity, reliability, and item characteristics, are still limited. Therefore, research that explicitly examines the quality of CT test instruments is needed so that the measurement of students' CT abilities in mathematics learning can be carried out accurately and reliably.

However, assessment practices in schools are still dominated by procedural questions that focus on final answers rather than the thinking process, so students' computational thinking (CT) skills are not measured optimally. In the context of assessment, CT components can be operationalized through questions that require students to break down complex problems into sub-problems (decomposition), identify regularities or patterns from the given data or situations (pattern recognition), filter important information and represent problems in a simple manner (abstraction), draw conclusions or general rules from the solutions obtained (generalization), and develop logical and systematic steps for solving problems (algorithms). Therefore, testing instruments with good psychometric qualities, including validity, reliability, difficulty level, and discriminating power, are needed so that each item truly measures aspects of CT accurately and reliably.

Based on this need, this study aims to evaluate the quality of test items developed to assess junior high school students' computational thinking abilities. The analysis includes validity testing, reliability, item difficulty index, and discrimination power. The findings of this study are expected to provide reliable assessment instruments aligned with 21st-century competency demands and serve as a reference for teachers and researchers in designing assessments that comprehensively capture students' computational thinking processes.

Method

This study used a quantitative approach with a descriptive-evaluative design, which is a study that aims to describe and evaluate the quality of test instruments based on trial data. The evaluation was conducted through an analysis of the characteristics of the

questions, including validity, reliability, difficulty level, and discriminating power, to assess the suitability of the instruments in measuring the computational thinking abilities of junior high school students. The developed instrument consists of five contextual mathematics essay questions on the subject of ratios, which are compiled based on computational thinking ability indicators, including problem decomposition, pattern recognition, abstraction, generalization, and algorithms. The developed questions are presented in the following figure.

1. Pemerintah kota sedang merancang peta wisata digital dan peta cetak untuk mempromosikan tempat-tempat wisata unggulan di daerahnya.Untuk versi digital, digunakan Peta A dengan skala 1 : 100, sedangkan versi cetaknya, Peta B, menggunakan skala 1 : 200 agar bisa menampilkan seluruh wilayah dalam satu halaman.Pada Peta A, taman kota yang menjadi pusat kegiatan festival budaya tampak memiliki luas 16 cm².

   Sebelum mencetak peta versi B, tim desain ingin memastikan agar proporsi gambar taman tetap akurat sesuai skala.Tanpa menghitung luas sebenarnya di lapangan, tentukan berapa luas taman tersebut akan tampak pada Peta B, dan jelaskan alasan matematisnya agar desain peta tetap proporsional!

2. Pada kegiatan Olimpiade Olahraga Sekolah, Siti dan Rina terpilih mewakili kelas mereka dalam lomba lari cepat.Mereka berdua sama-sama ingin menjadi juara, sehingga berlatih keras selama seminggu.

   Saat lomba berlangsung, Siti berhasil menempuh jarak 100 meter dalam waktu 20 detik, sedangkan Rina berlari 120 meter dalam waktu 30 detik pada sesi latihan terakhirnya. Pelatih ingin menilai siapa di antara keduanya yang memiliki kecepatan rata-rata lebih tinggi, agar bisa disiapkan strategi latihan yang berbeda.Tentukan siapa yang lebih cepat!

3. Sebuah proyek pembuatan taman sekolah dikerjakan oleh 4 orang pekerja dengan target selesai dalam 6 jam.Mereka bekerja dengan kecepatan yang sama dan semangat tinggi.

   Namun setelah 3 jam bekerja, salah satu pekerja jatuh sakit dan harus pulang. Agar taman selesai tepat waktu sebelum upacara peresmian, kepala sekolah harus memperkirakan berapa lama tambahan waktu yang dibutuhkan bila hanya 3 pekerja yang melanjutkan pekerjaan itu.Berapa jam tambahan yang dibutuhkan agar pekerjaan selesai. Uraikan langkah penyelesaian mu dengan memecah masalah menjadi beberapa tahap!

4. Dalam sebuah keluarga, Ayah, Ibu, dan Anak sedang merencanakan perjalanan wisata keluarga.Mereka menyadari bahwa umur mereka memiliki perbandingan 6 : 4 : 2.Dua tahun yang lalu, ketika mereka berfoto bersama di acara ulang tahun keluarga, jumlah umur mereka adalah 54 tahun.

   Sekarang mereka ingin menulis biodata keluarga untuk dokumen asuransi perjalanan, dan mereka harus mengisi umur masing-masing secara tepat Berdasarkan informasi yang diketahui, tentukan umur Ayah, Ibu, dan Anak sekarang, serta jelaskan langkah berpikir matematis untuk mendapatkannya!

5. Rudi sedang mengisi tangki air rumahnya yang besar. Tangki tersebut bisa diisi oleh dua keran, yaitu keran A dan keran B.Keran A dapat mengisi penuh tangki dalam 12 menit, sedangkan keran B dalam 8 menit.Untuk mempercepat, Rudi membuka kedua keran bersamaan selama 3 menit, lalu menutup keran A karena ingin menghemat air. Namun, tangki belum penuh.

   Rudi perlu menghitung berapa lama lagi keran B harus tetap dibuka agar tangki benar-benar terisi penuh.Tentukan lama waktu tambahannya!

Figure 1. Test Item

The instrument was tested on 64 seventh-grade students at SMP Negeri 1 Kota Agung Barat, who were selected using purposive sampling, considering that the class had studied material relevant to computational thinking ability indicators. The use of this technique enabled the collection of data that was relevant and in line with the objectives of the instrument evaluation. However, the research results have limitations in terms of generalization to a wider population, so the research findings need to be interpreted in the context of the subjects studied.

Data collection was conducted through a written test administered in a single session, with scoring guided by a rubric designed to maintain consistency among assessors. Each student's response was scored according to the indicators in each test item, producing quantitative data for further analysis.

Item analysis was performed to assess the instrument's quality based on four key aspects: validity, reliability, difficulty level, and discrimination power. Item validity was calculated using Pearson's Product-Moment correlation by comparing the obtained r_calculated and r_table values at a 5% significance level. Instrument reliability was analyzed using Cronbach's Alpha coefficient, with an instrument considered reliable if $\alpha \geq$

0.70. The difficulty level was determined by comparing the mean score of each item with the ideal maximum score, and items were categorized as difficult, moderate, or easy. The discrimination index was computed by comparing the mean scores of the upper and lower groups to evaluate each item's ability to differentiate between high- and low-ability students.

All analytical processes were conducted using Microsoft Excel to facilitate statistical computation. Through these procedures, a comprehensive overview of the instrument's quality was obtained, leading to conclusions about the appropriateness of the test items as valid evaluation tools for measuring junior high school students' computational thinking skills.

Results and Discussion

The quality analysis of the instrument was conducted on five test items measuring computational thinking skills, administered to 64 eighth-grade students. The analysis covered validity, reliability, difficulty level, and discrimination index.

1. Validity Test

The validity analysis was carried out using the criterion-validity formula through the Pearson Product-Moment correlation coefficient, as follows:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (1)$$

The test results were obtained using Microsoft Excel.

Table 1. Validity Test Results

Validity Test	**Question**				
	1	**2**	**3**	**4**	**5**
R Calculated	0,729	0,701	0,753	0,786	0,806
R Table	0,2465	0,2465	0,2465	0,2465	0,2465
Description	**Valid**	**Valid**	**Valid**	**Valid**	**Valid**

Based on the calculations in the table, all items in the computational thinking ability test instrument are valid because the calculated r value is greater than the table r value (0.2465). This shows that each item contributes significantly to the total score and is able to measure the targeted aspects of computational thinking ability. In accordance with Arikunto

(2013), the validity of an item is determined through the correlation between the item score and the total score, where an item is declared valid if its correlation value exceeds the r-table at a certain significance level. In addition to being statistically supported, the validity of this instrument is also substantively reinforced because each item has been designed based on clear CT indicators, namely problem decomposition, pattern recognition, abstraction, generalization, and algorithmic thinking, so that the items are not only empirically valid but also consistent with the theoretical construct being measured. Thus, the instrument used is suitable as a representative measure of computational thinking ability.

This result aligns with several studies conducted over the past decade. Research by El-Hamamsy et al. (2022) demonstrated that the Computational Thinking Test (cCTt) instrument possesses excellent psychometric properties, as analyses using classical test theory and item response theory confirmed its validity and reliability in measuring elementary school students' computational thinking skills. Furthermore, the study by Maksum et al. (2022) supported this conclusion, showing that their developed computational thinking assessment instrument was valid based on Aiken's V analysis, with each item achieving a high validity category. The consistency of these findings indicates that the instrument in this study meets recognized measurement standards in the literature and can be appropriately used to assess students' computational thinking abilities.

2. Reliability Test

Reliability analysis was performed using Cronbach's Alpha coefficient, because the instrument was in the form of essay questions with politomus scoring. This coefficient was used to measure the internal consistency of the instrument. The Cronbach's Alpha formula is expressed as follows:

$$\propto = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right) \tag{2}$$

where k denotes the number of items, σ_i^2 the variance of each item, and σ_t^2 the variance of the total score.

Table 2. Reliability Test Results

Reliability Test	Question				
	1	**2**	**3**	**4**	**5**
Variance	1,313	1,51	1,348	1,833	2,293
Sum of Variances	8,298				
Total Variances	23,59				
Reliability	**0,81**				

Based on the table, the instrument's reliability coefficient is 0.81, exceeding the recommended minimum threshold of 0.70. According to Arikunto (2013), an instrument is considered reliable if its reliability coefficient reaches or exceeds 0.70, indicating consistent data production across repeated uses. This reliability value of 0.81 confirms that the computational thinking test instrument in this study exhibits strong internal consistency and stability, making it a trustworthy representative measurement tool.

These findings align with various Indonesian studies over the past decade, which also demonstrate high reliability in computational thinking instruments. Inasari et al. (2023) developed a Rasch model-based computational thinking test instrument and found strong measurement consistency, deeming it suitable for instructional use. Another study by Novianto et al. (2021) showed that a computational thinking assessment instrument for elementary mathematics learning met feasibility criteria, with 86% validity and excellent reliability. The consistency across these national studies reinforces that the instrument in this research meets high measurement quality standards and can reliably assess students' computational thinking abilities.

3. Difficulty Level

The difficulty level of a question is a number indicating how easy or difficult it is for students to solve. This index is typically calculated by comparing the number of students who answered correctly with the total number of test participants (Masullah et al., 2024). The difficulty index formula used is :

$$TK = \frac{\bar{X}}{SMI} \tag{3}$$

Table 3. Difficulty Level Result

	Question				
	1	**2**	**3**	**4**	**5**
Average	3,359	2,828	2,281	2,234	2,266
SMI	4	4	4	4	4
Difficulty Level	**0,839**	**0,707**	**0,570**	**0,558**	**0,566**

The criteria for determining the difficulty level categories refer to Arikunto (2013),

Table 4. Difficulty Index Range Criteria

Difficulty Index Range	**Criteria**
0,00 - 0,30	Difficult
0,31 - 0,70	Moderate
0,71 - 1,00	Easy

Based on the difficulty index calculation results, it is known that item 1 has a difficulty index value of 0.839, which is in the range of 0.71–1.00, so it is categorized as an item with an easy level of difficulty. This value indicates that most students were able to answer the item correctly. This condition indicates that item 1 does not pose any significant difficulty for students, thus serving more as a measure of basic understanding of ratio and initial computational thinking concepts. Although items with an easy level of difficulty are still necessary in a test instrument, the proportion of items that are too easy needs to be controlled so that the instrument does not lose its ability to distinguish variations in student abilities.

Furthermore, item 2 has a difficulty index of 0.707, which is at the upper limit of the moderate category. This indicates that the difficulty level of this item is still within the ideal range, as it is neither too easy nor too difficult for students. Items in this category are generally able to measure students' conceptual understanding more deeply, while maintaining their discriminatory power. A similar condition is also shown by items 3, 4, and 5, which have difficulty indices of 0.570, 0.558, and 0.566, respectively. These values are in the range of 0.31–0.70, so all of these items are included in the moderate difficulty category.

The majority of questions in the moderate category indicate that the test instrument has been designed with a relatively balanced level of difficulty. Questions with a moderate level of difficulty are considered most ideal in learning evaluation because they are able to comprehensively measure students' abilities, from understanding concepts to applying computational thinking strategies in problem solving. With this level of difficulty, students

with high and low abilities have different opportunities in answering questions, so that the test results become more informative.

This classification is in line with the findings of various previous studies. Mustaqim and Sulisti (2024) stated that questions are categorized as easy if the proportion of correct answers exceeds 0.70, while the range of 0.30–0.70 indicates a moderate level of difficulty. They also emphasized that questions that are too easy have the potential to reduce the quality of evaluation because they are less able to distinguish students' abilities optimally. Therefore, a good test instrument should have a variety of difficulty levels with a predominance of moderate categories. Meanwhile, Harbit et al. (2024) assert that questions with a moderate level of difficulty tend to have better discriminating power and are more appropriate for use in learning evaluation because they are able to reveal differences in student abilities more effectively and objectively.

Thus, the results of the difficulty index calculation in this study show consistency with the item analysis standards commonly used in educational research in Indonesia. The dominance of items with a moderate level of difficulty and the presence of easy questions to reinforce basic understanding indicate that the test instrument developed is capable of measuring students' computational thinking skills in ratio material proportionally, balanced, and accurately. This finding reinforces that the test instrument used is suitable for application as an evaluation tool in the context of mathematics learning at the junior high school level.

4. Discrimination Index

The discrimination index is the ability of a test item to differentiate between high- and low-ability students, measured through the discrimination index that indicates the item's effectiveness in distinguishing student abilities (Solichin, 2017). According to Hadmar et al. (2024), items with high discrimination power accurately assess material mastery and enhance instrument quality, while those with low discrimination are less effective and require revision to maintain test validity and reliability. The formula used is :

$$DP = \frac{\bar{x}a - \bar{x}b}{SMi} \tag{4}$$

Discrimination Index Criteria (Arikunto, 2013),

Table 5. Discrimination Index Range Criteria

Discriminant Index Range	Criteria
0,7 < DP ≤ 1	Very Good
0,4 < DP ≤ 0,7	Good
0,2 < DP ≤ 0,4	Enough
0,0 < DP ≤ 0,2	Poor
DP ≤ 0	Very Poor

Test Results using Microsoft Excel:

Table 6. Results of Discrimination Power Test

	Question 1				
	1	**2**	**3**	**4**	**5**
Upper class average	3,90625	3,5625	2,90625	3,25	3,46875
Lower class average	2,8125	2,09375	1,65625	1,21875	1,0625
SMI	4	4	4	4	4
DI	0,273438	0,367188	0,3125	0,507813	0,601563
Criteria	**Enough**	**Enough**	**Enough**	**Good**	**Good**

Based on the table of discriminating power analysis results, it is known that item 1 has a discriminating power index of 0.27, which is classified as sufficient. This indicates that the item is able to distinguish between students with high and low abilities, although the level of discrimination produced is not yet optimal. A similar condition is also found in items 2 and 3, which have discrimination indices of 0.36 and 0.31, respectively. Both items are still in the adequate category, so they are still functionally suitable for use, but require improvement or refinement in terms of wording and level of difficulty in order to increase their discriminatory ability.

Meanwhile, items 4 and 5 showed higher discrimination indices, namely 0.50 and 0.60, respectively, which are classified as good. These discrimination indices indicate that both items are highly effective in distinguishing between students with high and low computational thinking abilities. Items with good discrimination indices generally have a balanced level of difficulty and are designed with clear indicators, enabling them to measure students' abilities more accurately.

These results are in line with the findings of various studies in Indonesia. Hadmar et al. (2024) stated that most test instruments with sufficient to good discriminatory power can function effectively in distinguishing students' abilities, especially in the context of learning

evaluation. In addition, Rachmawati and Pradana (2025) report that items with good discriminating power can accurately distinguish between students with high and low academic achievement, thereby improving the quality of evaluation results. Research by Saputri and Larasati (2023) reports that items with good discriminating power can accurately distinguish between students with high and low academic achievement, thereby improving the quality of evaluation results.

Thus, the results of the discriminating power analysis in this study indicate that most of the items have met the criteria for suitability as learning evaluation instruments. The test instruments used can be said to be effective in measuring and distinguishing students' computational thinking abilities in ratio material. However, items with sufficient discriminating power still need to be revised or refined so that the overall quality of the instruments can be improved and aligned with the educational evaluation standards applicable in Indonesia.

Conclusion and Suggestion

Based on the analysis of the five test items measuring computational thinking skills, it can be concluded that the developed instrument meets good quality criteria, as all items are valid (calculated r values exceeding the table r), the reliability coefficient of 0.81 falls in the high category, the difficulty levels range from easy to moderate, and the discrimination indices fall in the adequate to good categories, enabling effective differentiation between high- and low-ability students. This indicates that the instrument is suitable for use as an evaluation tool to measure junior high school students' computational thinking skills accurately and reliably, in line with the demands of 21st-century learning.

In light of these findings, it is recommended that items with only adequate discrimination be revised to improve the clarity of indicators and measurement precision, and that future studies increase the number of items and involve more diverse samples to enhance the comprehensiveness and generalizability of the instrument. Furthermore, teachers are encouraged to employ similar instruments in classroom assessment to evaluate students' thinking processes rather than focusing solely on final products, while technology-based assessment instruments may be developed as relevant alternatives for more authentic measurement of computational thinking in modern education.

References

Arikunto, S. (2013). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Rineka Cipta.

El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022). The competent computational thinking test: Development and validation of an unplugged computational thinking test for upper primary school. *Journal of Educational Computing Research*, *60*(7), 1818–1866. https://doi.org/10.1177/07356331221081753

Hadmar, S. S. A., Ali, A. M., & Yurfiah, Y. (2024). Analisis Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda Pada Mata Pelajaran IPA Di Sekolah Dasar. *Prosa: Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, *2*(3), 875–884. https://jurnal-umbuton.ac.id/index.php/Prosa/issue/view/284

Harbit, H., Samritin, S., & Natsir, S. R. (2024). Analisis Tingkat Kesukaran dan Daya Pembeda Soal Ulangan pada Mata Pelajaran Matematika di Sekolah Dasar. *Prosa: Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, *2*(1), 400–407. https://scholar.google.com/scholar?q=+intitle:%27Analisis%20Tingkat%20Kesukaran%20dan%20Daya%20Pembeda%20Soal%20Ulangan%20pada%20Mata%20Pelajaran%20Matematika%20di%20Sekolah%20Dasar%27

Inasari, L., Lidinillah, D. A. M., & Prehanto, A. (2023). Pengembangan instrumen tes computational thinking Siswa Sekolah Dasar melalui analisis RASCH model. *COLLASE (Creative of Learning Students Elementary Education)*, *6*(1), 102–110. https://pdfs.semanticscholar.org/2e3e/9ba2aa68f514af09b8641495b5e9fac9a0eb.pdf

Irawan, E., Rosjanuardi, R., & Prabawanto, S. (2024). Advancing computational thinking in mathematics education: a systematic review of indonesian research landscape. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *8*(1), 176–194. https://doi.org/10.31764/jtam.v8i1.17516

Irawati, L., & Hadi, M. S. (2025). Computataional Thinking dalam Pengembangan Berpikir Matematis di Sekolah Dasar. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, *8*(2), 2358–2364. https://doi.org/10.54371/jiip.v8i2.7106

Kaswar, A. B., & Nurjannah, N. (2024). Keefektifan computational thinking dalam meningkatkan kemampuan pemecahan masalah matematika siswa. *SIGMA: JURNAL PENDIDIKAN MATEMATIKA*, *16*(1), 109–120. https://doi.org/10.26618/sigma.v16i1.14574

Maksum, K., Ardiyaningrum, M., & Sukati, S. (2022). Pengembangan Instrumen Tes Keterampilan Berpikir Komputasi pada Pelajaran Matematika Sekolah Dasar (SD)/Madrasah Ibtidaâ€™ iyah (MI). *MODELING: Jurnal Program Studi PGMI*, *9*(1), 39–53. https://doi.org/10.69896/modeling.v9i1.1038

Masullah, B. D., Zuhry, L. H., Usman, L. H., & Maulana, L. M. G. (2024). Analisis butir soal ujian akhir semester mata pelajaran matematika smp negeri 6 praya timur. *Elips: Jurnal Pendidikan Matematika*, *5*(2), 152–161. https://doi.org/10.47650/elips.v5i2.1219

Mustaqim, M., & Sulisti, H. (2024). Analisis butir soal pas matematika peminatan: daya pembeda, tingkat kesukaran, dan kualitas pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, *3*(1), 44–56. https://doi.org/10.24260/add.v3i1.3011

Novianto, A., Maknun, I. L. Il, Aliyah, N. Y. N., & Khalifah, A. N. (n.d.). Pengembangan Instrumen Penilaian Computational Thinking Pada Pembelajaran Matematika SD. *Kalam Cendekia: Jurnal Ilmiah Kependidikan*, *13*(1).

https://doi.org/10.20961/jkc.v13i1.93072

Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: Validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, *13*(3), 273–284. https://doi.org/10.26740/jupe.v13n3.p273-284

Saputri, H. A. S., & Larasati, N. J. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *9*(5), 2986–2995. https://doi.org/10.36989/didaktik.v9i5.2268

Solichin, M. (2017). Analisis daya beda soal, taraf kesukaran, validitas butir tes, interpretasi hasil tes dan validitas ramalan dalam evaluasi pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, *2*(2), 192–213. https://doi.org/10.26594/dirasat.v2i2.879

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35.