Journal of Authentic Research
https://journal-center.litpam.com/index.php/jar/index

December 2025 Vol. 4, No. 2
e-ISSN: 2828-3724
pp. 2288-2304

# Smart Assessment: A Computer-Based English Assessment System For Junior High School Students in Merauke Integrated With Deep Learning

[1]*Survey Sijabat, [1]Abdul Rahim
[1] Universitas Musamus, Merauke, Indonesia

*Corresponding Author e-mail: surveysijabat@unmus.ac.id

*Received: Nopember 2025; Revised: Nopember 2025; Published: December 2025*

**Abstrak**

This study aims to develop and analyze a computer-based Smart Assessment system integrated with a deep learning pedagogical approach for assessing the English language skills of junior high school students in Merauke. The system was implemented using the Quizizz platform as a Computer-Based Test (CBT), incorporating three core principles of deep learning: meaningful learning, mindful learning, and joyful learning. The research employed the 4D development model (Define, Design, Develop, Disseminate) and involved 45 junior high school students during the field implementation phase. Smart Assessment was applied in regular English assessment activities, enabling students to complete tests digitally while teachers monitored performance and results in real time. Several challenges emerged during implementation, including limited internet connectivity, differences in students' digital literacy levels, and teachers' initial adaptation to technology-based assessment. Despite these constraints, the system proved effective in facilitating teachers to conduct assessments more efficiently and accurately, while providing immediate and automated feedback that supported students' reflective learning processes. Item quality analysis was conducted using Classical Test Theory (CTT) with the support of RStudio software. The results demonstrated very high instrument reliability (Cronbach's Alpha = 0.961). Most test items were classified as having moderate difficulty, with good to very good discrimination power, and all distractors functioned effectively. These findings confirm that the developed instrument meets acceptable psychometric standards for educational assessment. Implementation of Smart Assessment shows strong potential to support objective, efficient, and engaging English language assessment, while enhancing teachers' assessment practices and improving students' learning experiences. This system offers a practical and innovative solution for English language assessment in the context of digital and 21st-century education, particularly in remote and border regions.

**Keywords:** *smart assessment, classical test theory (CTT), computer-based test (CBT), deep learning*

**How to Cite:** Sijabat, S., & Rahim, A. (2025). Smart Assessment: A Computer-Based English Assessment System For Junior High School Students in Merauke Integrated With Deep Learning. *Journal of Authentic Research*, 4(2), 2288-2304. https://doi.org/10.36312/s7ax7530

. https://doi.org/10.36312/s7ax7530



## INTRODUCTION

Mastery of English plays a strategic role in facing the challenges of globalization and the development of a technology-based society in the era of Society 5.0. English not only serves as a means of international communication, but also as a key to accessing global knowledge and technology (Golzar et al., 2025; Rababah, 2025; Y. Wang, 2025; Zolfaghari et al., 2025). The demands of 21st-century competencies, which include critical thinking, creativity, communication, and collaboration (4Cs), further emphasize the importance of English proficiency starting from junior high school (Herlinawati et al., 2024; Nurhayati & Suhartini, 2025; Zhang et al., 2024).

English is positioned as one of the strategic subjects because it plays a role in equipping students with language skills that support cognitive, affective, and psychomotor development (Kaprawi, 2017; Pondelíková & Luprichová, 2025; Siahaan et al., 2025).

Advances in information technology have driven significant transformation in education assessment systems, from conventional paper-based methods to computer-based testing (CBT) (Caro et al., 2026; Rachayu et al., 2026). Computer-based assessment systems offer various advantages, including speed in data processing, accuracy in presenting results, and flexibility in applying various forms of tests, both objective and subjective (Karavidas et al., 2024; Tan, 2021; Q. Wang et al., 2023). CBT also supports transparency in assessment because it minimizes subjectivity, especially in objective questions. From a pedagogical perspective, CBT can increase student motivation to learn because it can be packaged in a more interesting, adaptive form that is in line with developments in digital technology.

The deep learning approach in education is not only understood as artificial intelligence technology, but also as a learning paradigm that emphasizes three main elements, namely meaningful learning, mindful learning, and joyful learning (Feriyanto & Anjariyah, 2024; Nafi'ah & Faruq, 2025; Suparti et al., 2025). In the context of English language assessment, meaningful learning is realized through an assessment system that is capable of providing comprehensive information about the development of students' competencies, so that the assessment results are not merely numbers, but also a mapping of strengths and weaknesses to support a more meaningful learning process. Furthermore, mindful learning is reflected in the provision of adaptive and reflective automatic feedback, which allows learners to have metacognitive awareness of their learning progress. Meanwhile, joyful learning is realized through interactive and enjoyable computer-based assessment designs, so that the evaluation process is no longer considered a burden, but rather part of a motivating learning experience. The integration of these three aspects makes deep learning a potential approach to realizing an assessment system that is innovative, objective, and relevant to the learning needs of the 21st century.

The quality of test items can be evaluated using Classical Test Theory (CTT), an analytical approach that is widely used because it is simple, easy to apply, and does not require complex statistical software (Moreta-Herrera et al., 2025; Rahim et al., 2025). Through CTT, a number of important indicators can be obtained to assess question quality, including difficulty index, discrimination index, and distractor effectiveness (Anselmi et al., 2025; C. Li et al., 2025a; Y. Li et al., 2025). With these advantages, CTT is considered highly relevant for application in the world of education. Various previous studies have extensively examined item analysis through the application of the Classical Test Theory approach. Previous research related to Classical Test Theory (CTT) has been used in analysis in various fields, especially to measure the reliability and validity of research instruments (Apino et al., 2024; Chen et al., 2024; Doğan Keskin et al., 2025; Grau-Gonzalez et al., 2024; Joseph & Fleary, 2024; C. Li et al., 2025b; H. Wang et al., 2025).

Schools in remote regions frequently encounter significant barriers to the successful implementation of technology-based assessments, such as Computer-Based Tests (CBT), due to limited internet connectivity, lack of sufficient hardware, and low levels of digital literacy among both teachers and students (Firdaus & Ritonga, 2024). Research has shown that these infrastructural and human resource constraints remain central obstacles to the adoption of technology in educational assessment in underdeveloped or geographically isolated areas, affecting institutional readiness and equity in access to digital learning environments (Nama & Tanggur, 2022). Studies in educational technology underscore that inadequate access to reliable networks and the necessary digital infrastructure hampers educational inclusion and the effective use of assessment technologies, particularly in rural settings where connectivity and device availability are inconsistent (Sulistyowati & Asriati, 2024). Furthermore, the transition to computer-based systems often demands additional training and capacity building for educators, who may not be adequately prepared to integrate such tools into their practice without sufficient professional development (Hulu, 2023). These challenges are compounded in regions like Merauke, where geographical isolation and infrastructure gaps can further widen the digital divide, making tailored solutions for local contexts crucial to realizing the full potential of CBT and related assessment innovations. Such insights highlight the need for adaptive, contextually grounded approaches to technology implementation that address both infrastructure and user competency to ensure equitable access and meaningful assessment outcomes.

Based on previous research, the main objective of this study is to design and implement a computer-based Smart Assessment for assessing the English language skills of junior high school students in Merauke with the integration of deep learning, which is realized through three main elements, namely Meaningful Learning, Mindful Learning, and Joyful Learning. This system is expected to provide a more objective, rapid, and comprehensive evaluation, covering both receptive and productive skills. In addition, this study aims to demonstrate how a deep learning approach based on the principles of meaningful learning, mindful learning, and joyful learning can improve the quality of assessment and the learning experience of students. The results of this assessment system development are expected to be an innovative solution to the limitations of conventional assessments, while also contributing to efforts to improve the quality of English education in Indonesia's border regions in the digital era and Society 5.0.

**Learning Assessment**

Assessment is one of the fundamental components for measuring the achievement of learning objectives and providing relevant information about the development of student competencies. Assessment in learning is a systematic process of collecting, analyzing, and interpreting information about student learning achievements (Dianti et al., 2025; Upiyani et al., 2023). Assessment not only serves to determine learning outcomes, but also as a basis for decision-making in improving the learning process (Hamid et al., 2024; Laila Laila et al., 2024; Natasya Lady Munaroh, 2024). Formative and summative functions in assessment, where formative assessment is used to monitor learning progress on an ongoing basis, while summative assessment is used to evaluate final achievement (Stufflebeam, 1972). Thus, these theories emphasize that assessment is an important instrument for ensuring the achievement of learning objectives.

Assessment theory emphasizes the principles of validity, reliability, practicality, and authenticity. Validity is a fundamental aspect that indicates the extent to which an instrument is capable of measuring the intended competency, while reliability emphasizes the consistency of assessment results (Bhisma Murti, 2011; Musyarofah et al., 2021). (Gronlund, 1982) emphasizes the importance of practicality, namely ease of administration, processing, and interpretation of assessment results. According to (Hughes, 2011) shows that assessment has a direct influence on the teaching and learning process. Well-designed assessments can have a positive impact on students' motivation and learning strategies. Thus, assessment theories emphasize that effective assessments must fulfill these principles in order to provide an objective, meaningful, and useful picture for improving the quality of learning.

**Quizizz! as an Assessment Tool in Learning**

Quizizz is a digital platform developed to support the assessment process in a more practical, flexible, and efficient manner (Marpaung, 2021). Through the features provided, teachers can compile evaluation instruments in various forms, such as multiple choice, true or false, and short answer questions, which can be accessed via computers and mobile devices (Asmara et al., 2022; Matlan & Maat, 2021). This characteristic is in line with technology-based assessment theories that emphasize efficiency, speed, and accuracy in obtaining learning achievement data. Effective assessment instruments must be able to provide an objective picture of the achievement of student competencies, and in this context, Quizizz provides an automatic result processing mechanism that can strengthen the validity and reliability of the assessment.

In addition, the use of Quizizz makes the assessment process more transparent and accessible. Real-time learning outcome reports allow teachers to comprehensively analyze student achievement, while helping students understand their level of mastery of the material (Degirmenci, 2021; Huei et al., 2021; Ilahi et al., 2025). This is in line with the basic principles of assessment, which emphasize practicality, clarity of interpretation of results, and a positive impact on the learning process. Presenting assessments in an engaging and interactive way can also increase student involvement and encourage them to perform at their best. Thus, Quizizz not only functions as a technology-based evaluation tool, but also as a means of supporting effective, relevant learning that is in line with the educational needs of the 21st century.

**Computer Based Test (CBT)**

Computer-Based Test (CBT) is a modern form of assessment that utilizes computer technology in the presentation of questions, management, and processing of test results (Khairi et al., 2025; Prasetyo et al., 2023; Sultan, 2023). CBT is an innovation in assessment systems that replaces paper-based tests, which tend to be more time-consuming and costly (Im et al., 2008; Piaw Chua, 2012). The main advantages of CBT lie in its efficiency, accuracy, and speed in processing test results (Abdul Gani et al., 2021; Efendi et al., 2021). In addition, CBT also allows questions to be presented in various interactive formats, such as text, images, audio, and video, so that students' skills can be measured more comprehensively. This is in line with the principles of assessment that emphasize the practicality, validity, and reliability of assessment instruments.

Furthermore, CBT has important implications for learning quality because it supports the implementation of more adaptive and flexible assessments (Manurung et al., 2019; Temitayo et al., 2013). In addition, CBT can also increase student motivation because it is more suited to the characteristics of the digital native generation who are accustomed to using technology. Thus, the implementation of CBT not only provides practical benefits in terms of efficiency, but also contributes to the realization of a more modern, transparent, and relevant assessment system for 21st century education needs.

## METHOD

This study uses a research and development approach with the aim of producing a computer-based English assessment system integrated with deep learning. The development model used adapts the 4D stages (Define, Design, Develop, Disseminate). In the Define stage, an analysis of English assessment needs in junior high schools was conducted. The Design stage included the design of the Smart Assessment system and the development of test instruments. The Develop stage involved the creation of a computer-based system and limited testing. The Disseminate stage was carried out by implementing the system on a wider scale in schools, while also evaluating its effectiveness in learning.

The research participants consisted of junior high school students in Merauke Regency. Participants were selected using purposive sampling, which involved selecting schools and classes that had implemented regular English language learning. The number of respondents in the trial involved approximately 45 students. In addition, English teachers were involved as test instrument validators and early users of the Smart Assessment system.

The main instrument used in this study was a multiple-choice test based on Quizizz. The test instrument was designed to measure English language skills in the aspect of reading comprehension based on deep learning, which was realized through three main elements, namely Meaningful Learning, Mindful Learning, and Joyful Learning. The test consists of 30 multiple-choice questions with four answer options, compiled based on the basic competency indicators applicable in the junior high school curriculum. The Quizizz platform was chosen because it has the advantage of providing game-based assessment, making it more interesting, interactive, and able to provide immediate results.

The validity and reliability of the instruments were tested to ensure that the tests used were able to measure English language proficiency accurately and consistently. Content validity was obtained through expert judgment by involving experts who assessed the suitability of the items with the competency indicators and the representativeness of the material. Furthermore, construct validity was analyzed using the Classical Test Theory (CTT) approach, taking into account the level of difficulty, discriminating power, and distractors of each item. The reliability of the instrument was tested by calculating Cronbach's Alpha coefficient to determine the internal consistency between items.

## RESULTS AND DISCUSSION

The Quizizz display in the assessment shows each question in a multiple-choice format with four answer options that students can choose from. Each question is presented digitally with a timer to manage the time allocated for completion, so that the evaluation process is more structured. The system automatically corrects answers, gives scores, and displays results in real time, both in the form of rankings and individual achievement analyses. The Quizizz display can be seen in Figure 1.

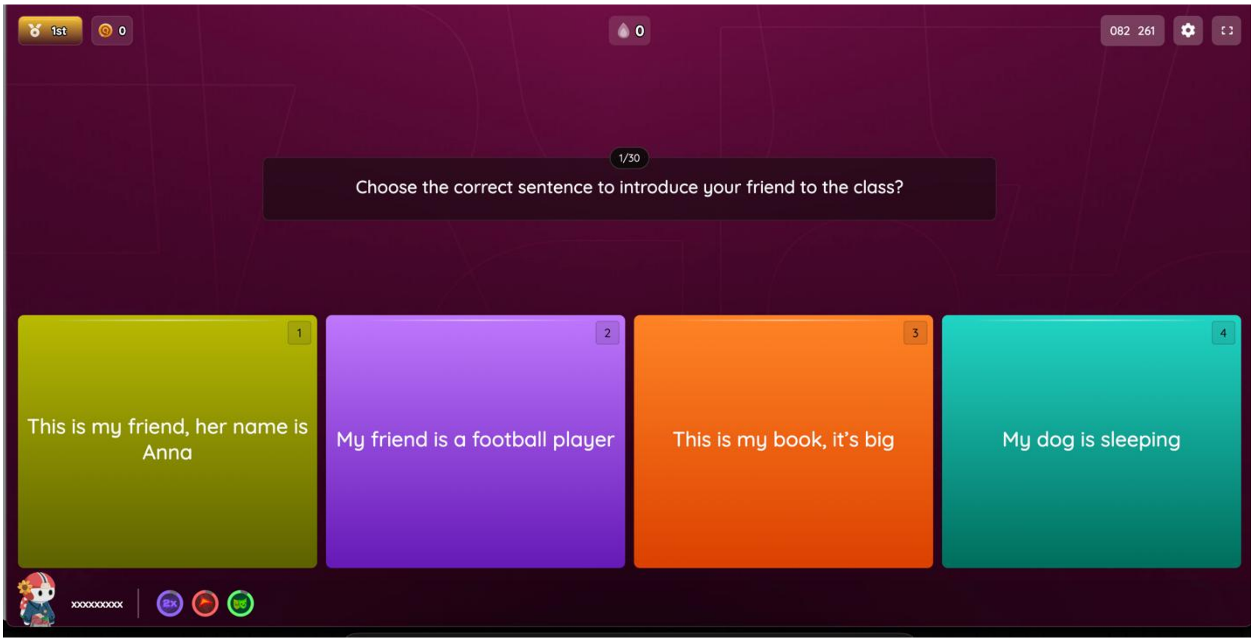



Figure 1. Quizizz display

The results of item analysis using the Classical Test Theory (CTT) approach with the help of RStudio software show that each item can be comprehensively evaluated through several key indicators, namely difficulty level, discriminating power, distractor effectiveness, and overall test reliability. Item analysis provides a more in-depth picture of the quality of each item, where the level of difficulty shows the extent to which the item is able to measure variations in test takers' abilities, while discriminating power identifies the item's ability to distinguish between high and low ability test takers. In addition, distractor effectiveness provides information on the extent to which incorrect answer choices function well in distracting participants who have not mastered the material, so that it can be determined whether each answer option works according to its role. The reliability calculation results using Cronbach's Alpha coefficient also show the level of internal consistency of the instrument used, which indicates that the test is reliable enough to be used in measuring learning outcomes objectively. The results of data analysis using Rstudio can be seen in Figure 2 and Figure 3.

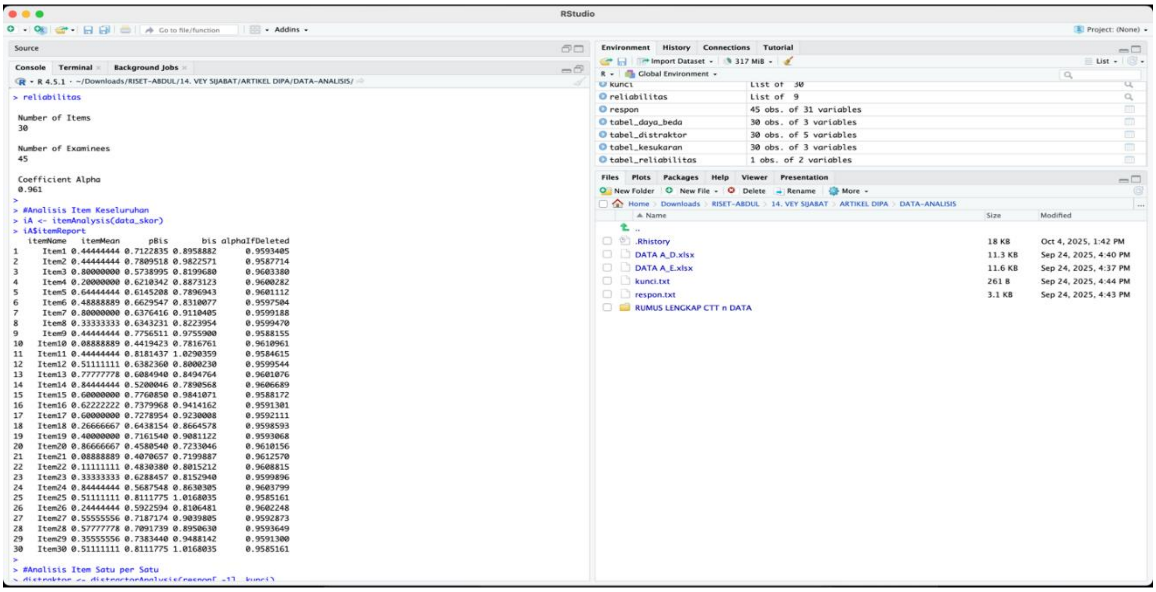



**Figure 2.** RStudio Analysis Display

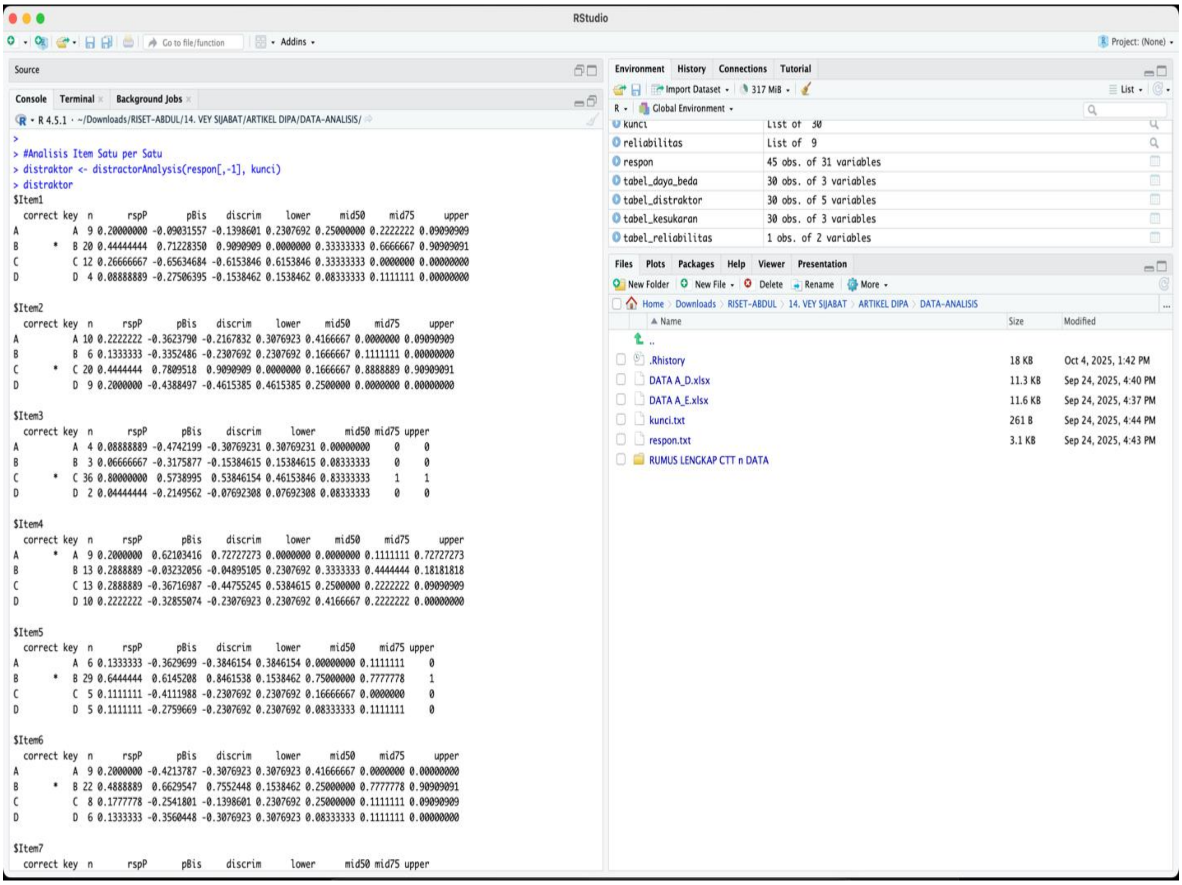



**Figure 3.** RStudio Analysis Display

Based on the results of reliability analysis using the Classical Test Theory (CTT) approach in RStudio, a Cronbach's Alpha coefficient value of 0.9610 was obtained, which is classified as very high. This value indicates that the test instrument has a very good level of internal consistency, meaning that the items compiled are able to measure the same construct in a stable and consistent manner. With this very high reliability, it can be concluded that the instrument is suitable for accurately and reliably measuring participants' abilities, and provides confidence that the variation in scores obtained is due to differences in participants' abilities rather than inconsistencies in the instrument. The reliability analysis results can be seen in Figure 4.

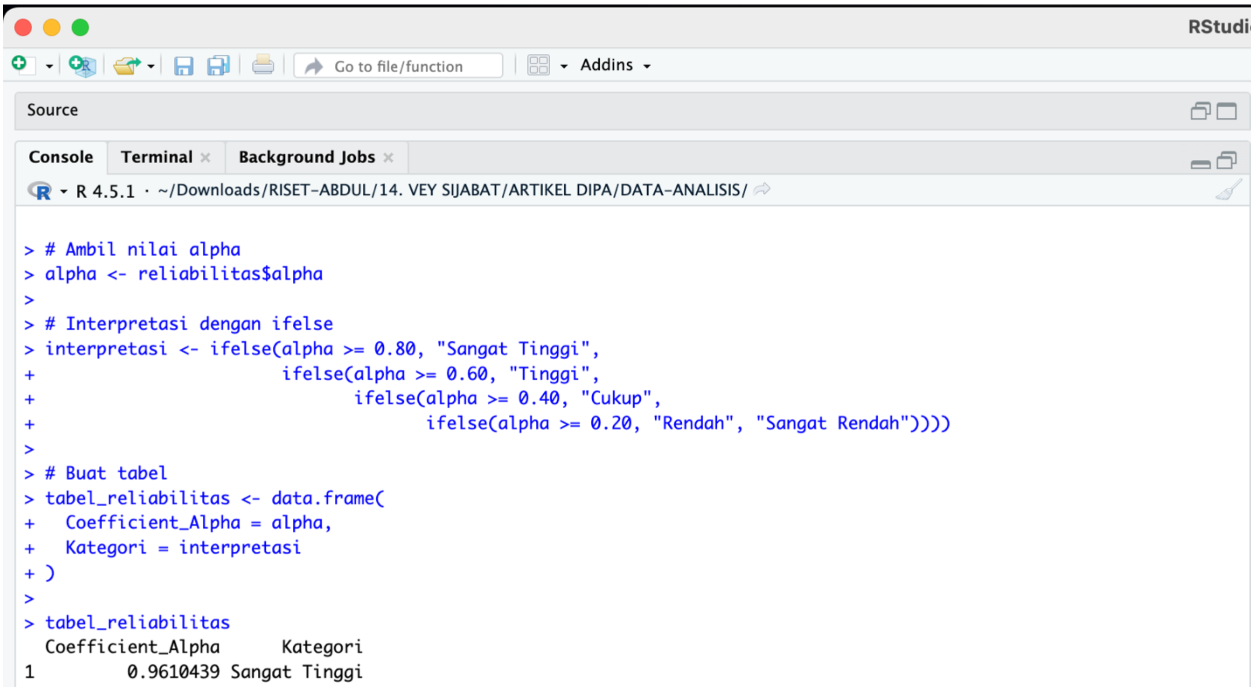
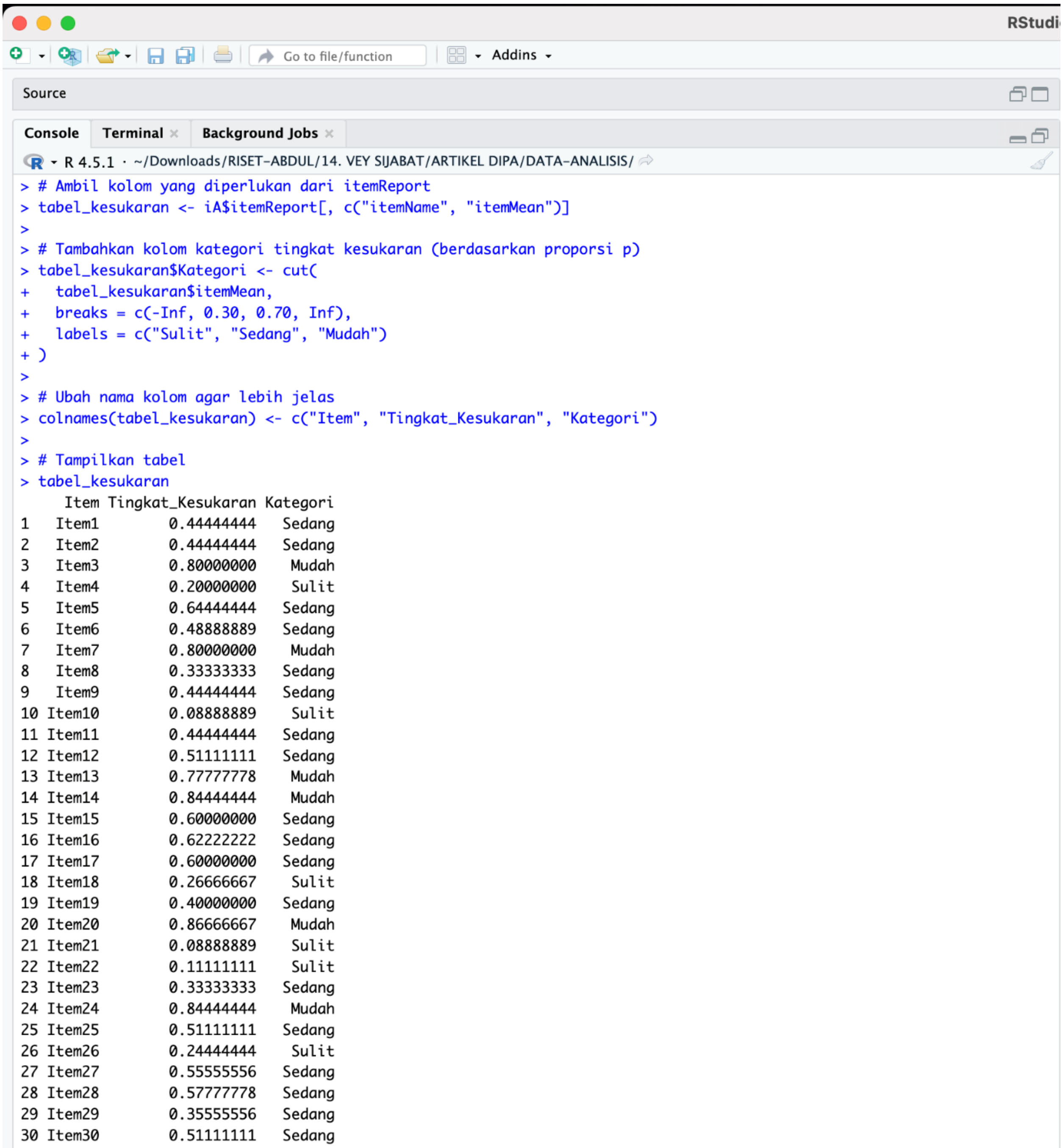



**Figure 4.** Reliability

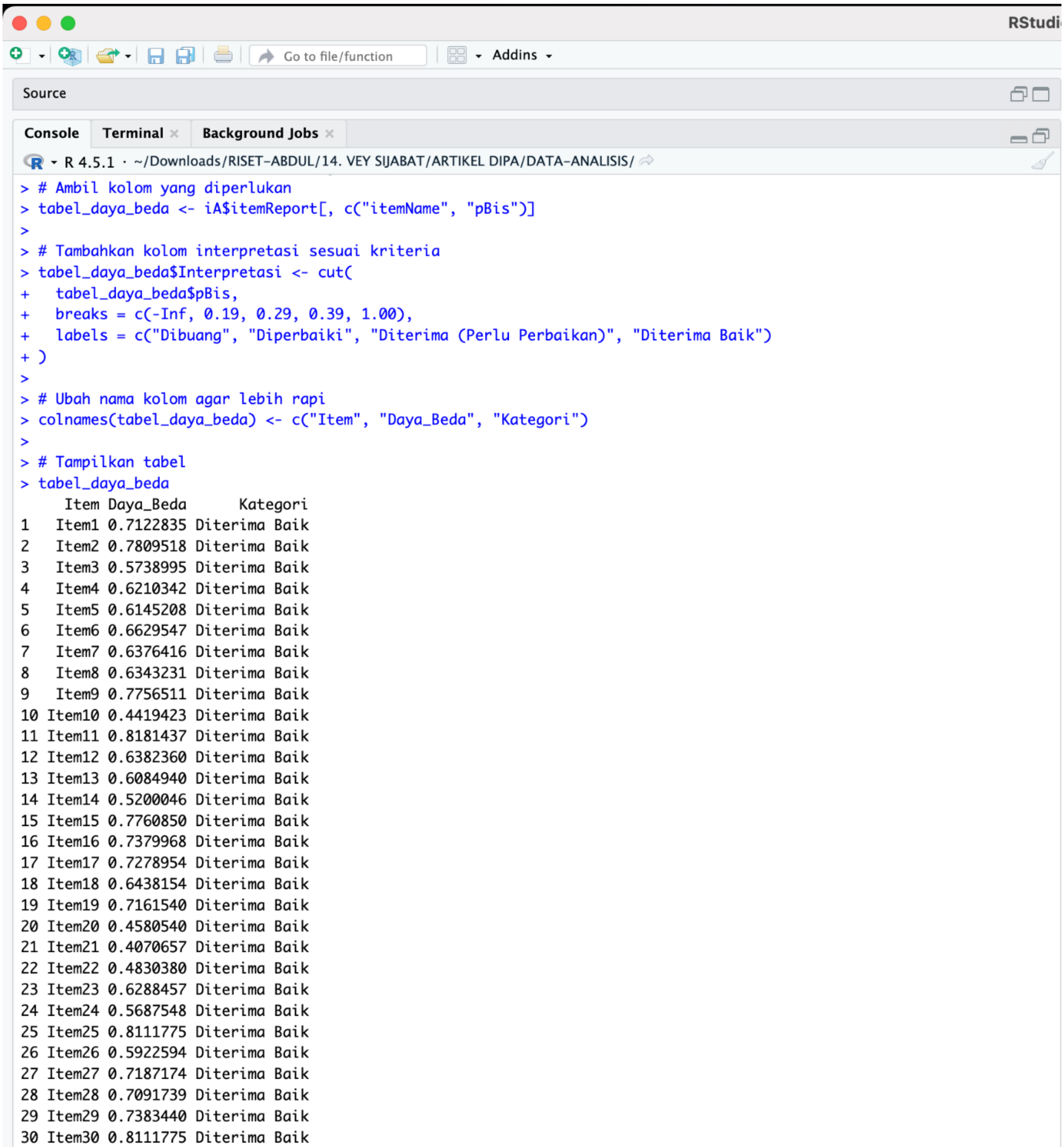
Based on the results of the analysis of the difficulty level of the questions displayed in the RStudio output, out of a total of 30 questions, 6 questions were classified as "Difficult." There were 18 questions in the 'Moderate' category and 6 questions in the "Easy" category. This composition shows that most of the items are at a medium level of difficulty, which is ideal in test development because it can provide a good spread of scores and increase the discriminating power between participants. Meanwhile, the existence of easy and difficult items in proportional numbers also strengthens the validity of the instrument by covering the variation in the ability levels of the students. The results of the analysis of the difficulty level of the items can be seen in Figure 5.



```
RStudio
Go to file/function   Addins
Source
Console   Terminal ×   Background Jobs ×
R 4.5.1 · ~/Downloads/RISET-ABDUL/14. VEY SIJABAT/ARTIKEL DIPA/DATA-ANALISIS/
> # Ambil kolom yang diperlukan dari itemReport
> tabel_kesukaran <- iA$itemReport[, c("itemName", "itemMean")]
>
> # Tambahkan kolom kategori tingkat kesukaran (berdasarkan proporsi p)
> tabel_kesukaran$Kategori <- cut(
+   tabel_kesukaran$itemMean,
+   breaks = c(-Inf, 0.30, 0.70, Inf),
+   labels = c("Sulit", "Sedang", "Mudah")
+ )
>
> # Ubah nama kolom agar lebih jelas
> colnames(tabel_kesukaran) <- c("Item", "Tingkat_Kesukaran", "Kategori")
>
> # Tampilkan tabel
> tabel_kesukaran
     Item Tingkat_Kesukaran Kategori
1   Item1        0.44444444   Sedang
2   Item2        0.44444444   Sedang
3   Item3        0.80000000    Mudah
4   Item4        0.20000000    Sulit
5   Item5        0.64444444   Sedang
6   Item6        0.48888889   Sedang
7   Item7        0.80000000    Mudah
8   Item8        0.33333333   Sedang
9   Item9        0.44444444   Sedang
10 Item10        0.08888889    Sulit
11 Item11        0.44444444   Sedang
12 Item12        0.51111111   Sedang
13 Item13        0.77777778    Mudah
14 Item14        0.84444444    Mudah
15 Item15        0.60000000   Sedang
16 Item16        0.62222222   Sedang
17 Item17        0.60000000   Sedang
18 Item18        0.26666667    Sulit
19 Item19        0.40000000   Sedang
20 Item20        0.86666667    Mudah
21 Item21        0.08888889    Sulit
22 Item22        0.11111111    Sulit
23 Item23        0.33333333   Sedang
24 Item24        0.84444444    Mudah
25 Item25        0.51111111   Sedang
26 Item26        0.24444444    Sulit
27 Item27        0.55555556   Sedang
28 Item28        0.57777778   Sedang
29 Item29        0.35555556   Sedang
30 Item30        0.51111111   Sedang
```

**Figure 5**. Level of Difficulty

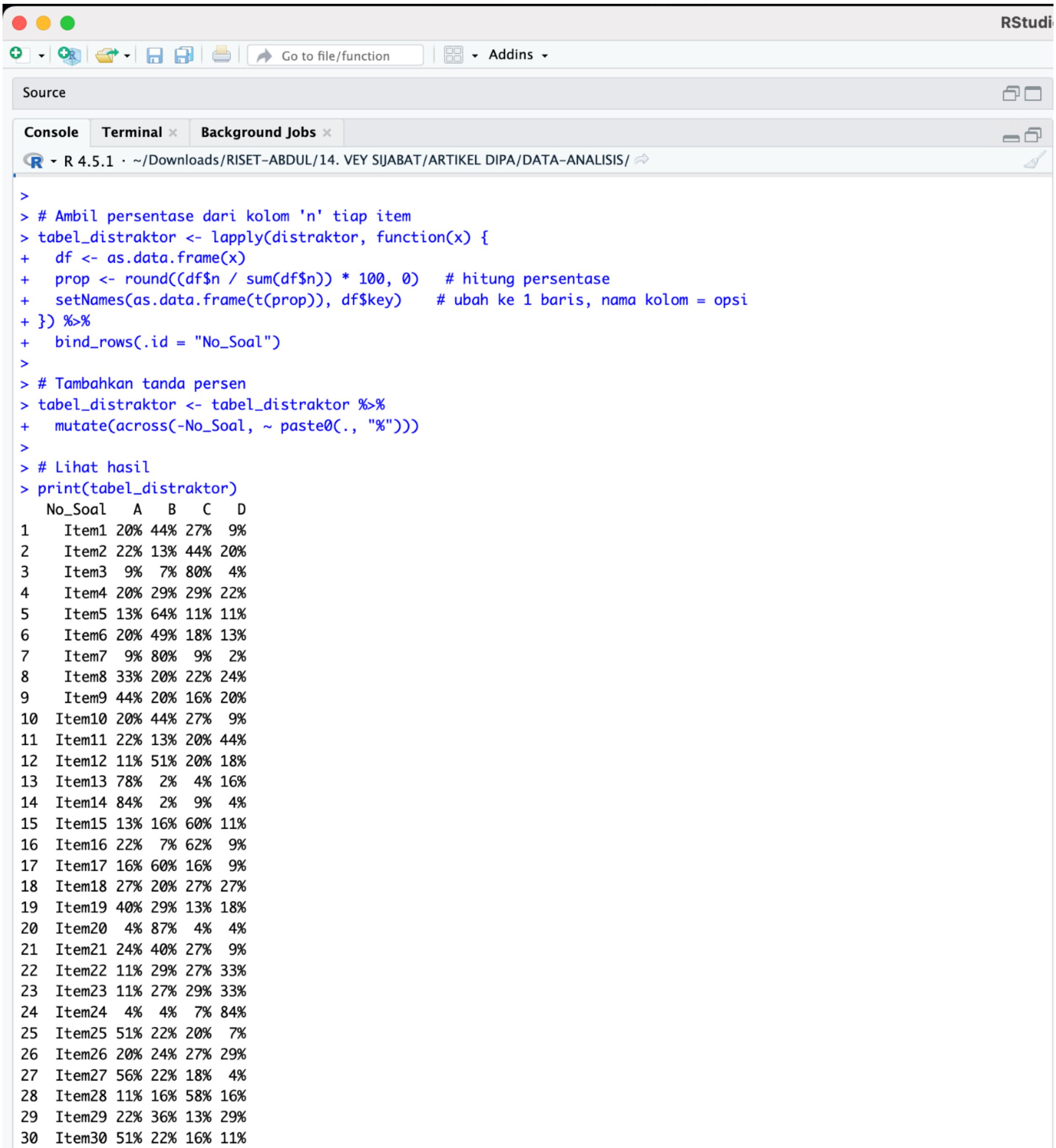
Based on the results of item discrimination analysis using the Classical Test Theory (CTT) approach through RStudio software, it was found that all 30 items were in the "Well Accepted" category, with discrimination values ranging from 0.40 to 0.81. This indicates that each item has a good ability to distinguish between participants with high and low abilities. No items were classified as "Rejected" or "Accepted with Improvement," so overall, the instrument can be said to have excellent item quality in terms of discrimination. These findings indicate that the test is able to measure participants' abilities effectively and proportionally in accordance with the measurement objectives. The results of the item discrimination analysis can be seen in Figure 6.



```
RStudio
Go to file/function   Addins
Source
Console  Terminal  Background Jobs
R 4.5.1 · ~/Downloads/RISET-ABDUL/14. VEY SIJABAT/ARTIKEL DIPA/DATA-ANALISIS/
> # Ambil kolom yang diperlukan
> tabel_daya_beda <- iA$itemReport[, c("itemName", "pBis")]
>
> # Tambahkan kolom interpretasi sesuai kriteria
> tabel_daya_beda$Interpretasi <- cut(
+   tabel_daya_beda$pBis,
+   breaks = c(-Inf, 0.19, 0.29, 0.39, 1.00),
+   labels = c("Dibuang", "Diperbaiki", "Diterima (Perlu Perbaikan)", "Diterima Baik")
+ )
>
> # Ubah nama kolom agar lebih rapi
> colnames(tabel_daya_beda) <- c("Item", "Daya_Beda", "Kategori")
>
> # Tampilkan tabel
> tabel_daya_beda
     Item Daya_Beda      Kategori
1   Item1 0.7122835 Diterima Baik
2   Item2 0.7809518 Diterima Baik
3   Item3 0.5738995 Diterima Baik
4   Item4 0.6210342 Diterima Baik
5   Item5 0.6145208 Diterima Baik
6   Item6 0.6629547 Diterima Baik
7   Item7 0.6376416 Diterima Baik
8   Item8 0.6343231 Diterima Baik
9   Item9 0.7756511 Diterima Baik
10 Item10 0.4419423 Diterima Baik
11 Item11 0.8181437 Diterima Baik
12 Item12 0.6382360 Diterima Baik
13 Item13 0.6084940 Diterima Baik
14 Item14 0.5200046 Diterima Baik
15 Item15 0.7760850 Diterima Baik
16 Item16 0.7379968 Diterima Baik
17 Item17 0.7278954 Diterima Baik
18 Item18 0.6438154 Diterima Baik
19 Item19 0.7161540 Diterima Baik
20 Item20 0.4580540 Diterima Baik
21 Item21 0.4070657 Diterima Baik
22 Item22 0.4830380 Diterima Baik
23 Item23 0.6288457 Diterima Baik
24 Item24 0.5687548 Diterima Baik
25 Item25 0.8111775 Diterima Baik
26 Item26 0.5922594 Diterima Baik
27 Item27 0.7187174 Diterima Baik
28 Item28 0.7091739 Diterima Baik
29 Item29 0.7383440 Diterima Baik
30 Item30 0.8111775 Diterima Baik
```

**Figure 6.** Item discrimination

Based on the results of the distractor analysis using RStudio, it is known that each item has a relatively proportional distribution of answer choices for options A, B, C, and D, with a percentage of choices ranging from 4% to 87%. This shows that all distractors function properly because each option is chosen by a number of respondents, so that no answer alternative is completely ignored. This condition indicates that each item is able to attract the attention of respondents in a balanced manner and does not cause bias towards a particular choice. Thus, it can be concluded that the distractors in this instrument are effective in measuring participants' understanding and support the validity of the items as a whole. The results of the distractor analysis can be seen in Figure 7.




```
>
> # Ambil persentase dari kolom 'n' tiap item
> tabel_distraktor <- lapply(distraktor, function(x) {
+   df <- as.data.frame(x)
+   prop <- round((df$n / sum(df$n)) * 100, 0)   # hitung persentase
+   setNames(as.data.frame(t(prop)), df$key)    # ubah ke 1 baris, nama kolom = opsi
+ }) %>%
+   bind_rows(.id = "No_Soal")
>
> # Tambahkan tanda persen
> tabel_distraktor <- tabel_distraktor %>%
+   mutate(across(-No_Soal, ~ paste0(., "%")))
>
> # Lihat hasil
> print(tabel_distraktor)
   No_Soal   A   B   C   D
1    Item1 20% 44% 27%  9%
2    Item2 22% 13% 44% 20%
3    Item3  9%  7% 80%  4%
4    Item4 20% 29% 29% 22%
5    Item5 13% 64% 11% 11%
6    Item6 20% 49% 18% 13%
7    Item7  9% 80%  9%  2%
8    Item8 33% 20% 22% 24%
9    Item9 44% 20% 16% 20%
10  Item10 20% 44% 27%  9%
11  Item11 22% 13% 20% 44%
12  Item12 11% 51% 20% 18%
13  Item13 78%  2%  4% 16%
14  Item14 84%  2%  9%  4%
15  Item15 13% 16% 60% 11%
16  Item16 22%  7% 62%  9%
17  Item17 16% 60% 16%  9%
18  Item18 27% 20% 27% 27%
19  Item19 40% 29% 13% 18%
20  Item20  4% 87%  4%  4%
21  Item21 24% 40% 27%  9%
22  Item22 11% 29% 27% 33%
23  Item23 11% 27% 29% 33%
24  Item24  4%  4%  7% 84%
25  Item25 51% 22% 20%  7%
26  Item26 20% 24% 27% 29%
27  Item27 56% 22% 18%  4%
28  Item28 11% 16% 58% 16%
29  Item29 22% 36% 13% 29%
30  Item30 51% 22% 16% 11%
```

**Figure 7.** Distractors

The results of this study reinforce the findings of various previous studies that confirm the effectiveness of the Classical Test Theory (CTT) approach in evaluating the quality of assessment instruments. The Cronbach's Alpha reliability value of 0.961 indicates that the instrument has very high internal consistency, in line with the results of the study (Apino et al., 2024; Chen et al., 2024) which shows that CTT is capable of producing stable reliability measures in various educational assessment contexts. The distribution of difficulty levels, which is dominated by the moderate category, is also in line with research findings (Rahim et al., 2025) which confirms that a balanced proportion of easy, medium, and difficult questions will improve the test's ability to comprehensively describe the variation in students' abilities. The discriminating power of all items is in the good to very good category, indicating that each item functions effectively in distinguishing between high- and low-ability participants, as found in (Doğan Keskin et al., 2025; C. Li et al., 2025) which emphasizes the importance of item discrimination in ensuring the construct validity of an instrument. The results of the distractor analysis show that all distractors function optimally, with a proportional distribution of answer choices among participants. This supports the research (Grau-Gonzalez et al., 2024; Moreta-Herrera et al., 2025) which explains that the effectiveness of distractors plays an important role in reducing bias and increasing the validity of test items. The results of this study indicate that the application of a computer-based Smart Assessment system integrated with CTT analysis is not only effective in producing valid and reliable instruments, but also in line with modern assessment trends that emphasize objectivity, efficiency, and integration with digital technology.

The results of item analysis using the Classical Test Theory (CTT) approach have important implications for assessment practices in schools, particularly in supporting teachers to develop more effective and objective test instruments. The analysis shows that the majority of test items fall into the moderate difficulty category, which is considered ideal for identifying differences in students’ ability levels. Items with moderate difficulty enable teachers to distinguish students who have mastered the material from those who still require additional support, thereby providing more accurate information for instructional decision-making. Furthermore, the good to very good discrimination indices indicate that the items function effectively in differentiating high- and low-performing students, which contributes to a fairer and more reliable evaluation process. From a practical perspective, these findings guide teachers in revising or retaining test items based on empirical evidence rather than subjective judgment. In addition, the results of the distractor effectiveness analysis, which show that all answer options function properly, help teachers design more plausible incorrect alternatives, reducing guessing behavior and increasing the validity of the assessment. By utilizing item analysis results from CTT, teachers can continuously refine their assessment instruments, ensuring that the questions used are aligned with learning objectives, accurately reflect students’ competencies, and support more meaningful feedback for improving teaching and learning processes.

## CONCLUSION

This study successfully developed a computer-based Smart Assessment system integrated with a Deep Learning approach through three main elements: Meaningful Learning, Mindful Learning, and Joyful Learning. This system is designed to support the English language learning evaluation process at the junior high school level in a more objective, efficient, and adaptive manner to the needs of students in the digital age. The results of the analysis using the Classical Test Theory (CTT) approach through RStudio software show that the assessment instrument has excellent quality, with high reliability ($\alpha = 0.961$), a proportional distribution of difficulty levels, strong item discrimination, and effective distractors. This confirms that the developed instrument is suitable for measuring students' English language proficiency in a valid and reliable manner.

## REFERENCES

Abdul Gani, H., Putra Pratama, M., & Wijaya, H. (2021). Development of an Android-based Computer Based Test (CBT) In Middle School. *Journal of Education Technology*, *5*(2), 272–281. https://ejournal.undiksha.ac.id/index.php/JET

Anselmi, P., Heller, J., Stefanutti, L., & Robusto, E. (2025). EDUCATIONAL AND CLINICAL ASSESSMENT WITHIN COMPETENCE-BASED KNOWLEDGE STRUCTURE THEORY. TPM - Testing, Psychometrics, Methodology in Applied Psychology, 32(1), 1–18. https://doi.org/10.69978/Tpmap.V32.I1.1

Apino, E., Istiyono, E., Retnawati, H., Widihastuti, W., & Hidayati, K. (2024). Development and calibration of an instrument measuring attitudes toward statistics using classical and modern test theory. Journal of Pedagogical Research, 8(3), 63–78. https://doi.org/10.33902/JPR.202427097

Asmara, C. H., Muhammad, R. N., & Almubarokah, Q. (2022). Effect of Online Learning Using Quizizz Application to Improve English Reading Skills of Higher Education Students. JELLT (Journal of English Language and Language Teaching), 6(1), 17–29. https://doi.org/10.36597/jellt.v6i1.11964

Bhisma Murti. (2011). VALIDITAS DAN RELIABILITAS PENGUKURAN.

Caro, M. F., Flórez, E. P., & Muñoz, I. C. (2026). A formal model for assessing the learning outcomes of academic programs. Evaluation and Program Planning, 114. https://doi.org/10.1016/j.evalprogplan.2025.102644

Chen, H., Wang, X., Bai, T., Cui, H., Shi, S., Li, Y., Xu, G.-Y., Li, H., & Shen, B. (2024). Reliability, validity, and simplification of the Chinese version of the Global Pain Scale in patients with rheumatoid arthritis. BMC Nursing, 23(1). https://doi.org/10.1186/s12912-023-01664-4

Degirmenci, R. (2021). The Use of Quizizz in Language Learning and Teaching from the Teachers' and Students' Perspectives: A Literature Review Article Info Abstract. In Language Education and Technology (LET Journal) (Vol. 1, Issue 1). http://langedutech.com

Dianti, K., Ulfah, M., Salam, A., Gunawan, G., & Luthfiyah, L. (2025). Analisis Asesmen Diagnostik, Formatif dan Sumatif Serta Implikasinya terhadap Efektivitas Sistem Evaluasi Pendidikan. Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI), 5(2), 555–565. https://doi.org/10.53299/jppi.v5i2.1234

Doğan Keskin, A., Öztürk, N., Kaytez, N., Aral, N., & Elibol, F. (2025). Validity and reliability study of the Turkish version of the autism spectrum knowledge scale. Acta Psychologica, 258. https://doi.org/10.1016/j.actpsy.2025.105204

Efendi, R., Lesmana, L. S., Putra, F., Yandani, E., & Wulandari, R. A. (2021). Design and Implementation of Computer Based Test (CBT) in vocational education. Journal of Physics: Conference Series, 1764(1). https://doi.org/10.1088/1742-6596/1764/1/012068

Feriyanto, F., & Anjariyah, D. (2024). Deep Learning Approach Through Meaningful, Mindful, and Joyful Learning: A Library Research. Electronic Journal of Education, Social Economics and Technology, 5(2), 208–212. https://doi.org/10.33122/ejeset.v5i2.321

Firdaus, K., & Ritonga, M. (2024). PERAN TEKNOLOGI DALAM MENGATASI KRISIS PENDIDIKAN DI DAERAH TERPENCIL. JURNAL KEPEMIMPINAN & PENGURUSAN SEKOLAH, 9(1). https://doi.org/10.34125/jkps.v9i1.427

Golzar, J., Sarwarzada, T., & Miri, M. A. (2025). Coping with Teacher Professional Identity Tensions: Using TPACK During Global Crisis Time. Journal of Language, Identity and Education, 24(5), 1208–1226. https://doi.org/10.1080/15348458.2023.2281952

Grau-Gonzalez, I. A., Villalba-Garzon, J. A., Torres-Cuellar, L., Puerto-Rojas, E. M., & Ortega, L. A. (2024). A psychometric analysis of the Early Trauma Inventory-Short Form in Colombia: CTT and Rasch model. Child Abuse and Neglect, 149. https://doi.org/10.1016/j.chiabu.2024.106689

Gronlund, J. (1982). Errors due to data reduction in single-breath method for measurement of pulmonary blood flow. Journal of Applied Physiology, 52(1), 104–108. https://doi.org/10.1152/jappl.1982.52.1.104

Hamid, S. F., Shopia, K., & Isnaniah, I. (2024). DEVELOPING TPACK BASED ASSESSMENT INSTRUMENT FOR TEACHING PRACTICE OF ENGLISH EDUCATION STUDENTS. Journal of English Language and Language Teaching, 8(2), 39–56. https://doi.org/10.36597/jellt.v8i2.18125

Herlinawati, H., Marwa, M., Ismail, N., Liza, L. O., & Situmorang, D. D. B. (2024). The integration of 21st century skills in the curriculum of education. Heliyon, 10(15). https://doi.org/10.1016/j.heliyon.2024.e35148

Huei, L. S., Yunus, M. M., & Hashim, H. (2021). Strategy to improve english vocabulary achievement during COVID-19 epidemic. Does quizizz help? Journal of Education and E-Learning Research, 8(2), 135–142. https://doi.org/10.20448/JOURNAL.509.2021.82.135.142

Hughes, G. (2011). Towards a personal best: a case for introducing ipsative assessment in higher education. Studies in Higher Education, 36(3), 353–367. https://doi.org/10.1080/03075079.2010.486859

Hulu, Y. (2023). Problematika Guru Dalam Pengembangan Teknologi dan Media Pembelajaran. ANTHOR: Education and Learning Journal, 2(6), 2023. https://doi.org/10.31004/anthor.v2i6.285

Ilahi, D. S. K., Zaini, F. M., Muhammad, B., Zakiyullah, A., & Sofa, A. R. (2025). Penilaian Pembelajaran PAI Berbasis Google Forms, Quizizz, dan Grade Scope: Strategi dan Implementasi Efektif di MA Zainul Hasan 1 Genggong. Indonesian Research Journal on Education Web Jurnal Indonesian Research Journal on Education, 5(2).

Im, E.-J., Lee, W.-K., Lee, Y.-C., Choe, B.-H., Chung, S.-K., Lee, T.-H., Cho, H., Sohn, J.-H., Won, D.-I., Kong, H.-H., Chang, B.-H., & Lee, J.-M. (2008). Development of Computer-Based Test (CBT) and Student Recognition Survey on CBT. Korean J Med Educ, 20(2), 145–154. https://doi.org/10.3946/kjme.2008.20.2.145

Joseph, P. L., & Fleary, S. A. (2024). Construct Validity and Measurement Invariance of the eHEALS in a Diverse US High School Sample. Journal of Media Psychology, 36(3), 188–198. https://doi.org/10.1027/1864-1105/a000399

Kaprawi, N. (2017). Obstacles facing learners in speaking english: Non-English Teachers' Perspective, Putra Batam School. Pertanika Journal of Social Sciences and Humanities, 25(April), 119–127. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85054663400&partnerID=40&md5=66c9a929687b3c605935adc55c4ba213

Karavidas, L., Skraparli, G., & Tsiatsos, T. (2024). A Comparison between Online Quizzes and Serious Games: The Case of Friend Me. Computers, 13(3). https://doi.org/10.3390/computers13030058

Khairi, M. U., Rokhmat, J., & Mustari, M. (2025). Penilaian Berbasis Digital Menggunakan Computer Based Test (CBT) di SMP NW Jerua. Jurnal Praktisi Administrasi Pendidikan), 9(2), 158–165. https://doi.org/10.29303/jpap.v9i2.1104

Laila Laila, Alawiyah Nabila, & Eka Widyanti. (2024). Konsep Dasar Evaluasi Pembelajaran. Jurnal Manajemen Dan Pendidikan Agama Islam, 2(5), 252–262. https://doi.org/10.61132/jmpai.v2i5.536

Li, C., Lin, Y., Tosun, B., Wang, P., Guo, H. Y., Ling, C. R., Qi, R., Luo, Q. Y., Wang, Y., Huang, F., Wang, J., Ma, S. H., Xu, D. F., Wu, S. Z., & Zhang, L. (2025a). Psychometric evaluation of the Chinese version of the BENEFITS-CCCSAT based on CTT and IRT: a cross-sectional design translation and validation study. Frontiers in Public Health, 13. https://doi.org/10.3389/fpubh.2025.1532709

Li, C., Lin, Y., Tosun, B., Wang, P., Guo, H. Y., Ling, C. R., Qi, R., Luo, Q. Y., Wang, Y., Huang, F., Wang, J., Ma, S. H., Xu, D. F., Wu, S. Z., & Zhang, L. (2025b). Psychometric evaluation of the Chinese version of the BENEFITS-CCCSAT based on CTT and IRT: a cross-sectional design translation and validation study. Frontiers in Public Health, 13. https://doi.org/10.3389/fpubh.2025.1532709

Li, Y., Zhao, X., Wang, J., Han, X., Wu, H., Li, X., Tan, F., & Guo, S. (2025). Developing and Validating a Knowledge Questionnaire of Blood Sampling via Central Vascular Access Devices. Journal of Infusion Nursing, 48(2), 142–150. https://doi.org/10.1097/NAN.0000000000000580

Manurung, A., Fildzah, A., & Rajagukguk, J. (2019, October 11). Computer Based Test (CBT) System for Student Academic Examination. https://doi.org/10.4108/eai.18-10-2018.2287186

Marpaung, S. (2021). PEMANFAATAN QUIZIZZ SEBAGAI MEDIA PENILAIAN PEMBELAJARAN BAHASA INGGRIS. 5(1), 43–48.

Matlan, S. J., & Maat, S. M. (2021). Penggunaan Aplikasi Quizizz Sebagai Alternatif Penilaian Formatif dalam Pengajaran dan Pembelajaran Matematik. Jurnal Dunia Pendidikan. https://doi.org/10.55057/jdpd.2021.3.4.18

Moreta-Herrera, R., Caycho-Rodríguez, T., Salinas, A., Jiménez-Borja, M., Gavilanes-Gómez, D., & Jiménez-Mosquera, C. J. (2025). Factorial Validity, Reliability, Measurement Invariance and the Graded Response Model for the COVID-19 Anxiety Scale in a Sample of Ecuadorians. Omega (United States), 90(3), 1078–1093. https://doi.org/10.1177/00302228221116515

Musyarofah, A., Abdurrahman Siddik Bangka Belitung, S., Syaikh Abdurrahman Siddik, & ABDUL RAHIM. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab, I(1), 34–44. https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB

Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing Deep Learning Approach in Primary Education: Integrating Mindful, Meaningful, and Joyful. Journal of Educational Research and Practice, 3(2), 225–237. https://doi.org/10.70376/jerp.v3i2.384

Nama, D. Y., & Tanggur, F. S. (2022). DISPARITAS MEDIA PEMBELAJARAN PADA ERA DIGITALISASI PENDIDIKAN DI WILAYAH PERBATASAN RI-RDTL (REFLEKSI PEMBELAJARAN ONLINE DAERAH PERBATASAN). Jurnal Pendidikan Teknologi Informasi (JUKANTI), 5(2).

Natasya Lady Munaroh. (2024). Asesmen dalam Pendidikan : Memahami Konsep,Fungsi dan Penerapannya. Dewantara : Jurnal Pendidikan Sosial Humaniora, 3(3), 281–297. https://doi.org/10.30640/dewantara.v3i3.2915

Nurhayati, I., & Suhartini, R. (2025). The influence of critical, creative, communication, collaboration thinking ability on the entrepreneurial ability and competency of Tefa vocational school teachers in East Java. International Journal of Innovative Research and Scientific Studies, 8(3), 4653–4668. https://doi.org/10.53894/ijirss.v8i3.7577

Piaw Chua, Y. (2012). Effects of computer-based testing on test performance and testing motivation. Computers in Human Behavior, 28(5), 1580–1586. https://doi.org/https://doi.org/10.1016/j.chb.2012.03.020

Pondelíková, I., & Luprichová, J. (2025). Exploring the Efficacy of Chatgpt in Enhancing Specialized Communication Skills in English Language Learning. Journal of Teaching English for Specific and Academic Purposes, 13(1), 25–40. https://doi.org/10.22190/JTESAP250125003P

Prasetyo, J. A., Rufi'i, R., & Wiyarno, Y. (2023). PENGEMBANGAN APLIKASI COMPUTER BASED TEST (CBT) BERBASIS WEB DENGAN JARINGAN NIRKABEL (WIRELESS NETWORK) PADA HASIL BELAJAR. JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika), 8(3), 1010–1021. https://doi.org/10.29100/jipi.v8i3.4379

Rababah, L. M. (2025). Attitudes of Engineering Students towards English Courses at Jadara University in Jordan. World Journal of English Language, 15(2), 240–249. https://doi.org/10.5430/wjel.v15n2p240

Rachayu, I., Riyanto, Y., & Dewi, U. (2026). Development of IMMARA learning model effectively improve digital literacy skills and creativity. Multidisciplinary Reviews, 9(3). https://doi.org/10.31893/multirev.2026116

Rahim, A., Olga Siba Sabon, Y., Nurhayati, Ayu Priyudahari, B., Fachrul Islam, R., & Amri, H. (2025). Analisis Butir Soal Pilihan Ganda dengan Pendekatan Classical Test Theory pada Ujian Akhir Semester Mata Pelajaran Matematika. Indo-MathEdu Intellectuals Journal, 6(4), 4763–4776. https://doi.org/10.54373/imeij.v6i4.3357

Siahaan, M. M., Sijabat, A., Samosir, H., Purba, R., & Terizla, R. F. (2025). Enhancing Hots and Entrepreneurial Competencies Through Avnet Academic Applications in English Learning. APTISI Transactions on Technopreneurship, 7(1), 206–216. https://doi.org/10.34306/att.v7i1.496

Stufflebeam, D. L. (1972). Benjamin S. Bloom, J. Thomas Hastinos, and George F. Madaus Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Company, 1971. Pp. x + 913 $12.50. Psychometrika, 37(4), 487–489. https://doi.org/DOI: 10.1007/BF02291223

Sulistyowati, C., & Asriati, N. (2024). PEMANFAATAN TEKNOLOGI UNTUK MENINGKATKAN EFEKTIVITAS PEMBELAJARAN DAN KETERLIBATAN BELAJAR DI ERA DIGITAL. Jurnal Ilmiah Pendidikan Citra Bakti, 11(4), 1176–1188. https://doi.org/10.38048/jipcb.v11i4.4542

Sultan, E. H. (2023). PENGEMBANGAN MODEL ASESMEN MENGGUNAKAN APLIKASI COMPUTER BASED TEST (CBT ) UNTUK PENINGKATAN HASIL BELAJAR SISWA SDN GUNUNGSARI 01 BATU. Jurnal Pendidikan Taman Widya Humaniora(JPTWH), 2(3), 1794–1814. https://jurnal.widyahumaniora.org/

Suparti, S., Milawati, M., & Kholiq, A. (2025). Dynamic Assessment as A Strategy to Enhance Deep Learning in Narrative Writing for Elementary School Students. Educational Process: International Journal, 17. https://doi.org/10.22521/edupij.2025.17.385

Tan, J. (2021). Research on Computer-Aided English Language Evaluation System. In J. Zhang & Z. Qian (Eds.), Journal of Physics: Conference Series (Vol. 1992, Issue 3). IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/1992/3/032103

Temitayo, F. M., Adebisi, A. A., & Alice, O. O. (2013). Computer-Based Test (Cbt) System For University Academic Enterprise Examination. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 2(8). www.ijstr.org

Upiyani, U., Cahyawati, I., Supendi, S. S., Bachtiar, M., & Hunainah, H. (2023). Alat Penilaian Pembelajaran: Implementasi Alat Analisis Untuk Tujuan Belajar. Jurnal Darussalam: Jurnal Pendidikan, Komunikasi Dan Pemikiran Hukum Islam, 1(XV), 221–232.

Wang, H., Xue, D., & Wang, X. (2025). Revision of the emotion and motivation self-regulation questionnaire in Chinese middle school students. BMC Psychology, 13(1). https://doi.org/10.1186/s40359-025-02462-0

Wang, Q., Mousavi, A., Lu, C., & Gao, Y. (2023). Examining adults' behavioral patterns in a sequence of problem solving tasks in technology-rich environments. Computers in Human Behavior, 147. https://doi.org/10.1016/j.chb.2023.107852

Wang, Y. (2025). Enhancing english oral translation through cross-modal learning and synchronous optimization. PLOS ONE, 20(8 August). https://doi.org/10.1371/journal.pone.0329381

Zhang, Y., Mu, G. M., & Hu, Y. (2024). Gauging 21st Century Competencies of Chinese Students: A Rural-Urban Comparative Perspective. Journal of Research on Educational Effectiveness, 17(2), 344–364. https://doi.org/10.1080/19345747.2023.2181245

Zolfaghari, Z., Karimian, Z., Zarifsanaiey, N., & Farahmandi, A. Y. (2025). Navigating challenges in medical english learning: leveraging technology and gamification for interactive education – a qualitative study. BMC Medical Education, 25(1). https://doi.org/10.1186/s12909-025-07511-1