KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya

P-ISSN: 2442-7632 | E-ISSN: 2442-9287 | Vol. 10, No. 1, April 2024, pp. 366 – 387

10.22219/kembara.v10i1.32974 | http://ejournal.umm.ac.id/index.php/kembara

Cognitive Diagnostic Assessment Instrument with Individual Feedback for Language Receptive Learning on Multimodal Text

(Asesmen Diagnostik Kognitif dengan Feedback Individual untuk Pembelajaran Reseptif Bahasa pada Teks Multimodal)

Giati Anisah
Universitas Nahdlatul Ulama Sunan Giri, Indonesia
giati@unugiri.ac.id

Midya Yuli Amreta
Universitas Nahdlatul Ulama Sunan Giri, Indonesia
midyaamreta2@gmail.com

***Corresponding author: Giati Anisah | email:** anisahgiati@gmail.com

Article History **Accepted: February 28, 2023** **Revised: April 22, 2024** **Available Online: April 30, 2024**

Abstract: Cognitive diagnostic assessment with individual feedback plays an important role in detecting students' cognitive strengths and weaknesses, providing directions for steps that teachers and students must take to follow up on this, as well as facilitating the clustering of students in preparing differentiated learning. This study aims to develop a cognitive diagnostic assessment instrument for receptive language learning on multimodal text with individualized feedback. This study is R&D research with an ADDIE design. The study involved grade 9 students with varying cognitive levels working through the questions, and the results were analyzed to develop individualized feedback for each student. Individualized feedback was then followed up by teachers and students in differentiated learning. We conducted interviews, questionnaires, and observations to gather data on the feasibility of the assessment instrument. Based on the results of the analysis, it can be concluded that a set of cognitive diagnostic assessment instruments along with individualized feedback produced in the study meets the requirements of validity, reliability, and readability, as well as applicability. In addition, this study shows that the cognitive diagnostic assessment with individualized feedback provides an overview of the receptive cognitive subcompetence of language in multimodal text that is mastered so that it facilitates the teacher in developing content, process, and target differentiated learning products for the next material. This research also contributes to addressing the long-standing criticism that language assessments fail to provide individualized feedback to link assessment to learning.

Keywords **cognitive diagnostic assessment, individual feedback, language receptive learning, multimodal text**

Abstrak: Asesmen diagnostik kognitif dengan umpan balik individual berperan penting dalam mendeteksi kekuatan dan kelemahan kognitif siswa, memberikan arahan langkah-langkah yang harus dilakukan guru dan siswa untuk menindaklanjutinya, serta memudahkan pengelompokan siswa dalam mempersiapkan pembelajaran yang berdiferensiasi. Penelitian ini bertujuan untuk mengembangkan instrumen asesmen diagnostik kognitif untuk pembelajaran bahasa reseptif pada teks multimodal dengan umpan balik individual. Penelitian ini merupakan penelitian R & D dengan desain ADDIE. Penelitian ini melibatkan siswa kelas 9 dengan tingkat kognitif yang bervariasi untuk mengerjakan soal-soal dan hasilnya dianalisis untuk mengembangkan umpan balik individual untuk setiap siswa. Umpan balik individual kemudian ditindaklanjuti oleh guru dan siswa dalam pembelajaran yang berbeda. Wawancara, kuesioner, dan observasi dilakukan untuk melengkapi data kelayakan instrumen penilaian. Berdasarkan hasil analisis, dapat disimpulkan bahwa seperangkat instrumen asesmen diagnostik kognitif beserta umpan balik individual yang dihasilkan dalam penelitian memenuhi persyaratan validitas, reliabilitas, dan keterbacaan, serta keterterapan. Selain itu penelitian ini menunjukkan bahwa asesmen diagnostik kognitif dengan umpan balik terindividualisasi memberikan gambaran tentang subkompetensi kognitif reseptif bahasa dalam teks multimodal yang dikuasai sehingga memudahkan pengajar dalam menyusun konten, proses, dan target produk pembelajaran berdiferensiasi untuk materi berikutnya. Penelitian ini juga berkontribusi dalam menjawab kritik yang sudah berlangsung lama bahwa penilaian bahasa gagal memberikan umpan balik yang bersifat individual untuk menghubungkan penilaian dengan pembelajaran.

Kata Kunci **asesmen diagnostic kognitif, feedback individual, pembelajaran reseptif bahasa, teks multimodal**

How to Cite Anisah, G., & Yuli Amreta, M. (2024). Cognitive Diagnostic Assessment Instrument with Individual Feedback for Language Receptive Learning on Multimodal Text. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *10*(1), 366-387. https://doi.org/10.22219/kembara.v10i1.32974

INTRODUCTION

All teachers would agree that the level of difficulty of the learning content, the learning process, and the target learning outcomes must be adjusted to the level of students' abilities and needs. Teachers, therefore, need information about the extent of competence that students have achieved, what factors influence this achievement, and what difficulties students face. Cognitive diagnostic assessment (CDA), an assessment that diagnoses students' cognitive abilities at the beginning or end of learning, can generate this information. CDA formulates the description of students' abilities in the form of feedback. The impact of this feedback varies greatly depending on the complexity of its utilization (Carless & Boud, 2018).

The implementation of the Merdeka Curriculum in 2022 initiated the promotion of CDA in Indonesia. Preliminary studies conducted in secondary schools in Bojonegoro Regency reveal that classroom learning practices often do not incorporate CDA. Even if it is done only through simple questions that are done classically, the feedback generated is also general and applies to all students in the class. The feedback is difficult to follow up on because the varied nature of achievements, difficulties, and inhibiting factors cannot be accommodated. Therefore, individualized feedback is needed to encourage more effective learning (Habsy et al., 2024; Misbah, 2022; Hui et al., 2021; Kramer et al., 2023; Laudel & Narciss, 2023).

Research conducted by Tang & Zhan (2021) has explored whether cognitive diagnostic feedback can promote learning and whether it is more effective than traditional feedback in promoting learning. Tang and Zhan used longitudinal cognitive diagnostic tests in a kind of experiment to compare the effects of three types of feedback: cognitive diagnostic feedback, true-false response feedback, and no feedback. The results can be concluded that cognitive diagnostic feedback can improve student learning and is more effective than true-false response feedback in improving comprehension, especially in more challenging knowledge areas. It is clear that CDA without feedback will reduce the function of CDA itself (Hattie & Clarke, 2018; Irons & Elkington, 2021). The difference between this study and Tang & Zhan (2021) lies in the types of feedback involved as research variables as well as the way feedback is treated. The research (Tang & Zhan, 2021) involved cognitive diagnostic feedback, true-false response feedback, and no feedback, and then compared them. This study creates cognitive diagnostic feedback to enhance language receptive learning through multimodal text analysis. The research equation lies in the use of cognitive-diagnostic feedback variables in learning.

Another study conducted by Zhang & Hyland (2022) identified student engagement with feedback. Zhang and Hyland integrated three types of feedback in learning: automatic, peer, and teacher feedback. It was found that the integrated feedback encouraged students' behavioral, affective, and cognitive engagement with the feedback provided. We need to tailor feedback to each individual student to increase their engagement with it. Of course, it requires detailed analysis of students' answers on CDA. The difference between this study and Zhang & Hyland (2022) is the integration of the types of feedback. This study focuses on feedback obtained from analyzing students' answers without involving other feedback. We use this to accurately detect students' abilities, enabling teachers and students to follow up appropriately.

Another study conducted by Prihatni et al., (2016) has discussed in detail the steps of developing CDA, which begin with the formulation of a learning continuum by meeting the eligibility requirements of content validity and reliability. Both requirements are important, but not perfect, without construct validity, readability, and applicability (Supriyadi et al., 2022; Larenas et al., 2021; Astalini et al., 2019). Therefore, this study fulfills all of these eligibility requirements.

In Indonesia, the lack of development of individualized feedback in the implementation of CDA is also indicated by the absence of the term individualized feedback in research on CDA in

2022–2023. For example, research conducted by (Putri, 2021; Hasna et al., 2023; Budyartati 2016; Prihatni et al., 2016; Huda et al., 2023; Suarni, 2023; Wulandari et al., 2023). Previous research has provided examples of CDA preparation and encouraged teachers to carry out diagnostic assessments to identify students' abilities, strengths, and weaknesses so that teachers can design learning tailored to students' abilities and characteristics. However, if CDA implementation fails to generate feedback, this goal will remain unachieved. For example, in the study Budyartati (2016), CDA was used to detect misconceptions, but it was not stated what misconceptions were experienced by students, so the recommendations to overcome these misconceptions were difficult to explain.

In the language receptive domain, the lack of research on developing feedback leads to claims that cognitive diagnostic tests are hard to use to give personalized feedback. This is because the range of answers shows how complex students' language learning levels are, which is hard to figure out (Min & He, 2022). The receptive language competencies identified in this study include listening and reading when viewing (Kementerian Pendidikan, Kebudayaan, 2022). The two abilities have a close relationship because they both use the decoding process (Diakidoy et al., 2019). This research focuses on receptive language in multimodal texts, namely texts that use different modes at the same time and contain verbal and visual semiotic sources for various communication purposes (Al Fajri, 2020; Pérez-González, 2019; Dressman, 2019). CDA for receptive ability in multimodal texts can measure competence in life in the 21st century, which includes competence in understanding the advanced level, critical thinking, creative thinking, communicating, and collaborating (Imamyartha et al., 2019; Redlo, 2021; Astuti et al., 2019; Musaad & Suparman, 2023).

This study presents a cognitive diagnostic assessment by first mapping the receptive language sub-competencies, including explication, interpretation, application, and perspective, which are cognitive aspects of comprehension design (Wiggins & Mac Tighe, 2005; Kokotsaki et al., 2016). The RASCH model analyzes the questions related to each sub-competency directly after students' answers appear. Additionally, the competency construct organizes the CDA, making it suitable for classroom administration, easy to use, and targeted (Sun & Hwang, 2023; Jang & Sinclair, 2021; Astalini et al., 2019).

RASCH model analysis is a type of analysis that is being widely used due to its foresight in producing objective and detailed analysis in educational science research because it uses probability functions (Arijanty, 2014; Aryadoust et al., 2021; Chan et al., 2021; Müller, 2020; Muslihin et al., 2022; Karlimah, 2022). RASCH is able to present person parameters (ability) and problem parameters (difficulty). A logarithmic scale, known as logits, expresses the function. In addition, the RASCH model analysis is also able to detect guessing answers (Andrich & Marais, 2019; Stemler & Naples, 2021; Parmaningsih & Saputro, 2021).

This study aims to develop a cognitive diagnostic assessment with individualized feedback for language receptive learning on multimodal texts that meets the requirements of validity, reliability, readability, and applicability so that it is feasible for use in school learning. More specifically, this study reveals the success of CDA in generating individualized feedback in the language receptive domain, which has been difficult to do. In order to achieve this goal, this study uses the research and development method with the ADDIE development model (analysis, design, development, implementation, and evaluation) (Molenda, 2015). A series of structured processes can increase the efficiency and consistency of the quality of the instrument to be produced.

The results of this study contribute to making it easier for teachers to detect students' cognitive strengths and weaknesses so that they can develop differentiated learning content, processes, and product targets for the next material. In addition, this research also contributes to answering the long-standing criticism that language assessment fails to provide individualized feedback to link assessment with learning.

METHOD

This study aims to develop a cognitive diagnostic assessment instrument for receptive language learning on multimodal text with individualized feedback that is valid, reliable, easy to read,

and easy to implement so that it is useful for grouping students specifically and encouraging students to make learning improvements. This research is R&D research with an ADDIE design, which consists of analysis, design, development, implementation, and evaluation stages. The ADDIE model consists of the stages of analysis, design, development, implementation, and evaluation. We can use this model because we evaluate each stage before moving on to the next. A slight error at one stage will affect both. Thus, we will minimize errors in fulfilling the CDA eligibility requirements.

Figure 1
Stages of ADDIE Development Research

This research expert judgement involved two experts as data sources, namely Indonesian language learning specialists and assessment experts. Both are lecturers at Nahdlatul Ulama Sunan Giri University, and they were selected based on their expertise and doctoral qualifications.

In addition, this study involved students at MTS Abu Darrin Bojonegoro. We used purposive sampling to select the research participants. This research was intended for Phase D, so the teacher recommended Grade 9 be the participant. The selected students had varied abilities, namely low, medium, and high, which were recommended by the teacher based on the teacher's assessment document. The ability level selection of students was crucial as it aligned with the research objectives. We also did this to ensure that the resulting CDA instrument can adapt to the diversity of students in the real world (Saleh Alharbi et al., 2021).

Data collection was done through documentation, interviews, questionnaires, and tests. The instrument was structured based on its purpose and the stage at which it is used. Specifically, we arranged the questionnaire instrument according to indicators that demonstrate its validity, readability, and applicability. Experts validated all instruments before use.

We carried out the documentation at the analysis stage, specifically to analyze the learning outcomes in Phase D of the Merdeka Curriculum's listening and reading components. We also used documentation to collect student score data on the listening and reading elements. We conducted interviews with teachers to learn about their previous assessments and the challenges they encountered when preparing the CDA. Questionnaires were used during the expert validation test, small-scale trial, and field test. Tests were conducted to determine the effectiveness of the CDA questions produced.

Qualitative data from interviews was analyzed using three stages: data reduction, data presentation, and conclusion (Huberman et al., 2014). We carry out data reduction by selecting the suitability of the collected qualitative data. After that, the data was presented in tabular form to make it easier to analyze. Finally, data collection was carried out.

The scores of content validity, construct validity, facial validity, applicability, and readability of experts, teachers, and students were analyzed using classical analysis, according to (Arikunto, 2019). Sumintono & Widhiarso, (2015) used RASCH modeling to analyze the validity and reliability of the questions. The classical analysis was carried out with the following formula:

Data processing formula per item.

$$P = \frac{X}{Xi} x\ 100\%$$

Information:
P : percentage
X : respondent's answer in one item
Xi : ideal score in one item
100: constant

Overall data processing formula

$$P = \frac{\sum X}{\sum Xi} x\ 100\%$$

Information:
P : percentage
X : total answers of respondents in one item
Xi : total ideal score in one item
100: constant

Eligibility criteria for assessment instruments using classical analysis can be seen in Table 1 below.

Table 1
Eligibility Criteria for Assessment Instruments
(Classical Analysis)

Test results	Qualification	Follow-up
85%-10%	Very Qualified	Implemented
75%-84%	Qualified	Implemented
56%-74%	Quite Qualified	Revised
>55%	Less Qualified	Revised

Based on Table 1 above, if the results of calculating the score of content validity, construct validity, facial validity, readability, and applicability reach a score of 75% and above, it was declared qualified and very qualified if it reaches 85% and above so that it can proceed to the field test implementation stage. However, if the score is 74% or below, it was quite qualified, and less qualified if it is less than 55%, so it needed to be revised again.

We used RASCH modeling to analyze the students' answers to CDA questions. This analysis was carried out to obtain two results. First, it was used to calculate the validity and reliability of the product. The second step involved obtaining a profile of students' responses, which we then formulated as individual feedback. The analysis employed RASCH modeling to determine validity, specifically using an item measure that assessed the outfit means square (MNSQ), Z-standart outfit score (ZSTD), and point measure correlation score (Pt Mean Corr). Reliability can be seen through summary statistics by analyzing the person reliability, item reliability score, and Cronbach alpha score and then comparing them with reliability criteria.

a. This validity test was based on the validity requirements of the question items, according to Sumintono & Widhiarso, (2015).
b. Outfit means-square score (output MNSQ): 0,5 < MNSQ < 1,5
c. Score of outfit Z-Standart (Outfit ZSTD): -2,0 <ZSTD< +2,0
d. Poin measure correlation (PT-Measure Corr) = 0,4; PT-Measure Corr <0,85.

The reliability prerequisite criteria can be seen in Table 2 dan 3 below.

Table 2
Person Reliability and Item Reliability Criteria
(Sumintono, B., & Widhiarso, 2015)

Person Score / Item reliability	Information
<0,67	Weak
0,67-0,80	Enough
0,81-0,90	Good
0,91-0,94	Very good
>0,94	Special

Tabel 3
Alpha Cronbach Criteria

Alpha Cronbach Score	Information
<0,50	Very Bad
0,50-0,60	Bad
0,61-0,70	Enough
0,71-0,80	Good
>0,80	Very Good

We used the Winstep application to analyze the results of students' answers to the CDA questions using RASCH modeling. RASCH modeling, which included item measure, person measure, scalogram, and person-wright map, generates student answer profiles. Item measure was used to determine the difficulty of question items; person measure to determine student ability; scalogram made it easy for us to analyze, provided explanations, and predicted simultaneously individual abilities and item difficulty; and person wright map to map student ability clustering. The entire analysis results were formulated in the form of individual feedback, which was then reported to teachers and students.

RESULT AND DISCUSSION

In the digital era, students' capacity for reading and listening to multimodal texts is important. To maximize this capacity, teachers need to prepare appropriate lessons. For this reason, teachers need information on the extent of competence achieved by students, what factors affect that achievement, and what difficulties students face (Leighton & Gierl, 2007). CDA can generate such information.

Cognitive Diagnostic Assessment for Language Receptive Learning on Multimodal Text

CDA is designed to measure students' specific knowledge structure and skill processes to generate information about their cognitive strengths and weaknesses. The field's need for more relevant information for learning and the push to change assessment design led to the emergence of CDA. In order to produce such important information, CDA needs to fulfil four eligibility requirements, namely validity, reliability, readability, and applicability.

Validity

Validity demonstrates the instrument's reliability in measuring the intended outcome (Anisah, 2018). Validity is the main requirement of an assessment instrument, before any other requirement. This was because if the instrument was invalid, it would be difficult to meet other eligibility requirements. Content validity, construct validity, facial validity, and question validity were just a few of the diverse types of validity that are satisfied. The results of the analysis of content validity, construct validity, face validity, and question validity can be seen in Table 4 below.

Table 4
Expert Validation Test Results

Data Source	Criteria	%
Assessment Expert	Content Validity	91,6%
	Facial Validity	95%
	readability	92%
Language Learning Expert	Facial Validitas	92%
	Content Validity	91,6%
	readability	92%

The validity of the content reaches 91.6%, meaning that the material tested was in accordance with the demands of the curriculum. The validity of the content indicated how far the assessment instrument can measure the mastery of student competencies (Anisah, 2018; Himawan & Nurgiyantoro, 2022). A high percentage of content validity signified that the tested material's content aligns with the demands of the relevant curriculum. Content validity was the main requirement before testing the validity of others in the assessment instrument development process (Shrotryia & Dhanda, 2019). The high validity of the content was achieved because the assessment instrument was in accordance with the learning outcomes in the curriculum and the learning objectives, the question indicators were in accordance with the learning objectives, the indicators were in accordance with the question items, the scope of test material was in line with the scope of competence in the curriculum, and the suitability of the proportion on the test was in accordance with the proportion in the curriculum. The content's high validity was also supported by curriculum analysis conducted at the start of the research, specifically at the analysis stage. Curriculum analysis provided a true picture of the types of texts and skills that the curriculum demands. Curriculum analysis also generated target users.

In this study, the CDA instrument was intended for Phase D students, namely students in grades 7-8-9. We chose this phase because the learning outcomes of the Independent Curriculum in Phase D required the use of multimodal texts. In addition, another reason was that at the age of Phase D (12–15 years), intellectually, students had been able to perform moral reasoning, mathematical logic reasoning, and social transmission, where knowledge comes from receptive activity (Mauliya, 2019).

In addition to content validity, another validity tested was construct validity. Expert tests yielded a construct validity of 90% for the developed cognitive diagnostic assessment instrument, indicating its ability to generate characteristics of the measured competence (Anisah & Amreta, 2023). So it can be said that the CDA instrument was able to measure receptive language competence in multimodal texts.

This study focused on receptive skills, specifically the ability to listen and read multimodal texts, as students frequently encounter this type of text in the digital era. Reading skills are very helpful for students to find information that can be processed into new ideas or findings, ideas, and references in developing writing products (Beauty et al., 2023; Hendaryan & Noviadi, 2023). In general, listening is a process of receiving oral symbols with full attention, understanding, appreciation, and interpretation to obtain information, capture content, or messages and understand the meaning of communication that has been conveyed by the speaker through spoken language (Susanti, 2019).

Multimodal texts refer to texts with the purpose of communicating by using different modes at the same time (Van Leeuwen, 2015). Multimodal contains verbal and visual semiotic resources that can be used to realize various types of texts with the desired communicative purpose (Al Fajri, 2020). CDA for receptive ability in multimodal texts can measure 21st century competencies, which include competencies of advanced comprehension, critical thinking, creative thinking, communication, and collaboration.

The learning objectives for each element are broken down by taxonomy Wiggin and McTighe (2005), which includes explanation, interpretation, application, and perspective. The four sub-competencies are nonhierarchical. Indeed, in the taxonomy of Wiggins and McTighe, there are two more subcompetencies, namely empathy and self-knowledge. However, both competencies are more

likely to involve emotions, so they are not included in instrument development. The assessment instruments aim to concentrate on cognitive abilities.

According to Wiggins & McTighe's taxonomy (Wiggins & McTighe, 2005), the Independent Curriculum uses subcompetencies as forms of understanding in its learning outcomes. In constructivism, the concept of "understanding" refers to the process of building knowledge through real experience. Understanding is not static, but evolves and changes constantly as students construct new experiences that modify previous understandings. According to constructivist theory, the ability to understand is a union of various subcompetencies. It is different when referring to Bloom's Taxonomy, which places the ability to understand at level C2 (Wilson, 2016).

Explanation is the competence to describe an idea in one's own words, build relationships, demonstrate work, explain reasons, explain a theory, and use data (Jain & Wallace, 2019; Taylor & Noë, 2021). Interpretation deals with translating text, artwork, or situations. Interpretation also means interpreting an idea, feeling, or work from one medium to another. Application means using knowledge, skills, and understanding of something in a real situation or a simulation (resembling reality). Perspective relates to the competence to see things from a different point of view. Students can explain the other side of a situation, see the big picture, see the assumptions underlying a thing, and give criticism.

Another validity measure was facial validity. The facial validity of CDA instruments, according to assessment experts, was 95%, while according to defense experts, it was 92%, so an average of 93% was obtained. This indicated that the CDA instrument's face aligned with the measured competency (Johnson, 2021). The instrument adopted a multiple-choice format, presenting multimodal text at the outset before instructing students on how to solve the problem. In some questions, students were asked to read or watch relevant posters to compare, found relevance, found new information, or conclude.

To diagnose students' abilities in all sub-competencies in 3 types of multimodal texts (expalanation texts, exposition texts, and discussion texts), 24 questions were developed, both for listening and reading competencies, so that the total number was 48 multiple-choice questions. The multiple-choice form was chosen because it is easier to administer in the application of CDA instruments. However, it is undeniable that the form of multiple-choice questions has a weakness, namely the existence of guesswork answers. However, it can still be detected using Scalogram analysis on RASCH modeling.

Before students do the assessment, they are asked to listen to or watch videos related to scientific or social themes. Students adapt the theme to a real-world context they are familiar with and understand. It is in accordance with the learning objectives. The questions developed require students to think critically when perceiving the information contained in the text they listen to or read. Students' capacity in critical thinking in accordance with the demands of the twenty-first century (Hesse et al., 2015; Gravemeijer et al., 2017; Safi'i et al., 2021; Kamaruddin et al., 2023). Students can achieve this if they are accustomed to answering questions at a high level and capable of providing explanations or justifications for their responses.

Students are required to describe ideas and provide reasons in explanation-related questions. This is in accordance with the essence of explanation, which is to describe an idea in one's own words, build relationships, explain reason, and explain a theory (Agnafia, 2019). Interpretation skills are related to comprehensiveness and expression of meaning from various kinds of experiences, situations, data, events, decisions, conferences, beliefs, legal procedures, or criteria, so that in problems related to interpretation, students are asked to interpret the text and categorize information in the text. In application-related questions, students are required to connect the text's content to various real-world scenarios and other relevant sources. In perspective-related questions, students must draw conclusions by comparing the text's information with other sources.

We hope that by working through the entire series of questions, we can accurately describe the profile of students' abilities in the competencies tested. However, in the implementation of this CDA, schools must have adequate facilities and infrastructure because it is necessary to play

multimodal texts in the form of audiovisuals. The researchers acknowledge that not all schools, particularly those lacking infrastructure, can use the produced CDA instrument.

Another test to determine validity was the question validity test using RASCH modeling, as seen from the scores of MNSQ, ZSRD, and Pt Mean Corr. The RASCH modeling test results declared the 48 developed questions valid. Although not all questions meet all three criteria, if at least two criteria were met, then the question was declared valid (Sumintono & Widhiarso, 2015).

The analysis, which included both expert test analysis and tests using RASCH modeling, indicated that 48 questions from the cognitive diagnostic assessment fell into the valid category. The implication of the overall validity score was that the CDA instrument developed was in accordance with the material in the curriculum, in accordance with the construct of the measured competence, in accordance with the face of the competence to be measured, and was able to measure what it was supposed to measure.

Reliability

Reliability is defined as the consistency and reliability of measurements (Wahyuni et al., 2020). In this study, reliability was calculated using RASCH modeling. Table 5 below displayed the results of the reliability analysis using the RASCH model.

Table 5
Results of the RASCH Modeling Question Item Reliability Test

Person reliability	Item Reliability	Alpha Cronbach
0,86	0,67	0,69
Good	Enough	Enough

In RASCH modeling, reliability was indicated by the scores of person reliability, item reliability, and the alpha cronbach score (Sumintono & Widhiarso, 2015). The test results using RASCH modeling obtained a person reliability result of 0.86, which was included in the good category. The item reliability score of 0.67 fell into the adequate category. Cornbach's alpha score was 0.69, which means enough. Based on the test results, the reliability of the assessment instruments developed was in the good category in terms of student consistency in answering, but it was quite good when viewed from the perspective of the perspective of the questions. Cronbach's alpha score was quite good, showing the interaction between students (persons) and the question items as a whole was quite sufficient because the students' answers to each question item were quite consistent (Azizah & Wahyuningsih, 2020; Pratama, 2020).

Overall, the questions generated in this study could produce credible diagnostics. In addition, questions could make reliable mastery/non-mastery classifications at the level of competence and subcompetence. Therefore, we can assert that this cognitive diagnostic assessment served as a tool for low-risk learning decisions, such as allocating students' abilities within the class and categorizing students for remedial or enrichment courses.

A range of studies have explored the use of diagnostic language assessment to provide targeted feedback for language learners. Wang (2023) and Toprak & Cakir (2021) both emphasize the importance of this approach, with Wang focusing on spoken language assessment and Hirschi on L2 pronunciation. Both studies highlight the potential for automatic models to outperform human experts in providing feedback. Huilin (2013) further supports this, suggesting that a cognitive diagnostic approach can accurately diagnose language skills and promote individualized language teaching. Wang (2023) builds on this by demonstrating the usefulness of cognitive diagnostic feedback in a large-scale Spanish proficiency test, showing that it can assess reading skills more accurately and provide valuable feedback for academic improvement.

Research conducted by Yeh (2022) found that diagnostic language tests can be a valuable tool for independent language learners, providing insight into their strengths and weaknesses. The study revealed that these learners often struggle with vocabulary and have a strong desire to improve their

communication, speaking, and listening skills. The findings suggest that teacher interventions, learning strategies, and self-evaluation skills are essential for the early development of self-directed language learning.

Readability

The ease of understanding words and sentences is known as readability. We select letters, sentences, and text to meet the readability requirements (Anisah & Amreta, 2023). Table 6 below displays the results of the readability analysis.

Table 6
Result of Readability Analysis

Data Source	%
Assessment Expert	92%
Language Learning Expert	92%
Student	76,8%
Teacher	83,30%
Average	**86,025%**

Based on the test results, the CDA instrument had a readability of 86,025%, which means that all parts of the text were easy to understand, could be read at maximum speed, and had a text length and video duration according to student ability.

This high percentage could not be separated from the selective selection of letters, words, and sentences and adjusted to the ability of students. The font used was Bookman Old Size 12. This letter made it easier for students to read books quickly because it used a serif typeface to speed up eye movements. The American Psychological Association (APA) recommended this letter in text writing (Perea, 2013). The sentences used vary, namely declarative, interrogative, and imperative sentences. These three types of sentences were appropriate for assessment instruments. The use of declarative sentences lied in each part of the question that functioned to explain something, both in the text listened to or read and in the stem of the question. Stem questions employed interrogative sentences to elicit answers from students. For passages that instructed students to act on the problem's stem, we used imperative sentences.

The text chosen to be listened to and read was one of the types of explanation, exposition, and discussion. The text's theme was scientific and socio-cultural. We can state, based on the presented evidence, that cognitive diagnostic assessment instruments with individual feedback for language receptive learning in multimodal texts, which used detectable and understandable words, sentences, and texts, were highly qualified.

Applicability

Applicability is the degree to which CDA instruments are easy to use in the field. Table 7 below displays the results of the applicability analysis.

Table 7
Result of Aplicability Analysis

Data Source	%
Student	83,03%
Teacher	87,50%
Average	**85,04%**

Teachers and students' assessments showed that the CDA instrument had an applicability of 85.4%, indicating that the difficulty level of the questions aligned with students' abilities, the number of questions aligned with the allotted time, and the instrument was easy to administer.

As discussed in the sub-item measure, the questions tested had the appropriate proportion between very difficult, difficult, medium, and easy questions. This had been appropriate to measure the abilities of students with low, medium, and high abilities.

The assessment time was in accordance with the allocation of time provided, which was 40 minutes for 24 listening questions and 40 minutes for 24 reading questions. So answering 48 cognitive diagnostic assessment questions required two hours of lessons or one meeting. This duration was suitable for the start of a series of materials that encompass multiple learning objectives concerning language receptiveness in multimodal texts.

Assessment instruments were easy to administer. Assessment instruments were distributed using ICT Nearpod, where students could listen to and read texts as well as do questions in the application. Distributing questions in paper form prevented this from happening. The NearPod application was easy for students to use because the tools are intuitive and in accordance with the characteristics of digital natives. In addition, teachers could also monitor the progress of students' answers while they were working on the questions.

By working on the entire series of questions, it was expected that the profile of students' abilities in the competencies tested would be truly described. However, in implementing this CDA, schools must have adequate facilities and infrastructure because multimodal text playback in the form of audiovisuals is required. Therefore, the researcher acknowledged that not all schools, particularly those lacking infrastructure, could use the resulting CDA instrument.

In producing feedback, teachers need to analyze using the RASCH model to produce individualized feedback, teachers must analyze the RASCH model. Teachers could only provide feedback once they have thoroughly analyzed the students' answers. This was also one of the limitations of the study, where feedback could be accessed immediately after answering the questions.

Individual Feedback

Although research on feedback has grown rapidly over the past three decades, little attention has been paid to it by Indonesian researchers. Feedback will provide an in-depth understanding of individual differences in their engagement and the factors that influence them (Zheng et al., 2023). The study generated individualized feedback for each student who took the test on their cognitive strengths and weaknesses. However, this study has yet to identify the factors that influence these cognitive strengths and weaknesses.

Individual feedback was developed based on the results of analyzing students' answers using RASCH modeling. Therefore, the teacher can only provide feedback once she has analyzed the students' answers. This is also one of the limitations of research, where feedback cannot be accessed immediately after doing the problem.

To formulate feedback, students' answers to 48 multiple-choice questions were analyzed using RASCH modeling, specifically item measure, person measure, scalogram, and person:wright map. These four things provided a profile of each student's answers regarding language-receptive competence in multimodal texts.

Firstly, we conducted an analysis of the measurement items. This analysis serves to determine the level of difficulty of the question items. This was done at the beginning before the researcher mapped the student's ability. If the student was able to correctly answer the difficult questions, then he/ she was categorized as a high-ability student, and vice versa. The categories of item difficulty test results can be seen in Table 8 as follows.

Table 8
Question Item Difficulty Category

Score Measure	Information
>1,08	Very Difficult
1,08-0,00	Difficult
-1,08-0,00	Medium
<-1,08	Easy

Based on Table 8, if the item measure test results showed a value of > 1.08, then the question was categorized as a very difficult question; if the value was 1.08–0.00, then it was in the difficult category. The value -1.08–0.00 was in the medium category, and <1.08 was in the easy category. The results of the item measure test on listening and reading multimodal texts can be seen in Figures 2 and 3.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	27	70	1.36	.26	1.22	2.26	1.24	1.86	.07	.34	54.3	67.1	simakeksplanasi03
16	33	70	.96	.26	.95	-.56	.93	-.64	.41	.35	70.0	65.1	simakeksposisi08
21	33	70	.96	.26	1.04	.49	1.06	.61	.30	.35	61.4	65.1	simakdiskusi05
15	35	70	.83	.26	.89	-1.31	.87	-1.37	.49	.35	67.1	65.0	simakeksposisi07
13	36	70	.77	.26	1.03	.33	1.04	.46	.32	.35	65.7	65.0	simakeksposisi05
1	40	70	.51	.26	1.10	1.07	1.13	1.18	.23	.35	62.9	65.9	simakeksplanasi01
7	40	70	.51	.26	.83	-2.00	.79	-2.08	.57	.35	74.3	65.9	simakeksplanasi07
8	42	70	.37	.26	1.16	1.65	1.16	1.36	.16	.35	57.1	66.9	simakeksplanasi08
9	44	70	.23	.26	1.14	1.35	1.15	1.15	.18	.34	58.6	68.2	simakeksposisi01
6	45	70	.16	.27	1.00	.06	.96	-.25	.35	.34	65.7	68.9	simakeksplanasi06
24	45	70	.16	.27	1.10	.99	1.12	.92	.22	.34	65.7	68.9	simakdiskusi08
4	47	70	.02	.27	.98	-.14	.98	-.06	.36	.34	74.3	70.5	simakeksplanasi04
19	48	70	-.05	.27	.97	-.24	.86	-.89	.40	.33	67.1	71.3	simakdiskusi03
11	50	70	-.20	.28	.99	.00	1.00	.03	.33	.33	74.3	73.0	simakeksposisi03
2	52	70	-.37	.29	1.02	.18	1.02	.15	.29	.32	77.1	75.0	simakeksplanasi02
17	52	70	-.37	.29	.97	-.16	.95	-.21	.35	.32	74.3	75.0	simakdiskusi01
5	53	70	-.45	.29	1.08	.58	1.19	.93	.19	.31	71.4	76.1	simakeksplanasi05
10	53	70	-.45	.29	.99	-.04	1.07	.39	.30	.31	80.0	76.1	simakeksposisi02
22	54	70	-.54	.30	.98	-.07	1.04	.23	.31	.31	81.4	77.2	simakdiskusi06
12	55	70	-.63	.31	.95	-.27	.84	-.64	.38	.30	78.6	78.5	simakeksposisi04
23	55	70	-.63	.31	.79	-1.33	.64	-1.67	.58	.30	78.6	78.5	simakdiskusi07
14	57	70	-.83	.32	1.09	.52	.97	-.04	.21	.28	81.4	81.4	simakeksposisi06
18	59	70	-1.05	.34	.83	-.77	.60	-1.42	.51	.27	84.3	84.2	simakdiskusi02
20	61	70	-1.30	.37	.89	-.38	.61	-1.14	.43	.25	87.1	87.1	simakdiskusi04

Figure 2
Item Measure Results Listening Questions

Based on Figure 2, it can be seen that the most difficult listening question was the one that occupies the top row in the entry number column, which is question 3. There were only 27 students who answered correctly (total score) out of 70 (total count), so the measure score was only 1.36 and was categorized as very difficult. The next difficult question was number 16, 21, 15, and 13. The easiest question was the question at the bottom row in the entry number column, which was question number 20. There were 61 students who answered correctly out of a total of 70 students, so the measurement score was -1.30. Figure 3 presents the results of the item measure test for the reading problem.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
23	29	70	1.06	.26	.90	-1.17	.87	-1.23	.46	.33	70.0	65.5	membacadiskusi07
10	31	70	.93	.25	1.05	.64	1.05	.52	.26	.33	64.3	64.4	membacaeksposisi02
8	33	70	.80	.25	1.09	1.18	1.10	1.06	.21	.33	64.3	63.7	membacaeksplanasi08
21	35	70	.67	.25	1.07	.85	1.07	.80	.24	.33	58.6	63.7	membacadiskusi05
18	37	70	.54	.25	1.05	.65	1.06	.64	.26	.33	62.9	64.0	membacadiskusi02
4	38	70	.48	.25	1.04	.49	1.07	.76	.27	.33	61.4	64.3	membacaeksplanasi04
13	38	70	.48	.25	1.02	.26	1.03	.38	.30	.33	67.1	64.3	membacaeksposisi05
15	38	70	.48	.25	1.11	1.31	1.11	1.12	.19	.33	55.7	64.3	membacaeksposisi07
2	39	70	.41	.25	.94	-.74	.95	-.48	.40	.33	68.6	64.6	membacaeksplanasi02
6	39	70	.41	.25	1.18	2.12	1.24	2.30	.08	.33	57.1	64.6	membacaeksplanasi06
24	43	70	.15	.26	1.03	.33	.99	-.03	.30	.32	58.6	66.7	membacadiskusi08
3	44	70	.08	.26	1.05	.50	1.07	.61	.26	.32	65.7	67.4	membacaeksplanasi03
16	47	70	-.13	.27	.95	-.40	.91	-.63	.39	.31	70.0	69.8	membacaeksposisi08
5	48	70	-.20	.27	.97	-.23	.91	-.54	.36	.31	72.9	70.7	membacaeksplanasi05
7	48	70	-.20	.27	.82	-1.69	.72	-1.99	.57	.31	72.9	70.7	membacaeksplanasi07
11	49	70	-.27	.27	1.01	.15	.96	-.21	.31	.31	68.6	71.6	membacaeksposisi03
19	49	70	-.27	.27	1.04	.38	.94	-.36	.29	.31	65.7	71.6	membacadiskusi03
1	50	70	-.35	.28	.88	-.95	.79	-1.29	.48	.30	78.6	72.6	membacaeksplanasi01
20	52	70	-.51	.29	.95	-.33	.96	-.17	.35	.29	78.6	74.8	membacadiskusi04
17	54	70	-.68	.30	.97	-.17	1.02	.15	.31	.28	81.4	77.3	membacadiskusi01
12	55	70	-.77	.30	.82	-1.13	.67	-1.53	.54	.28	78.6	78.5	membacaeksposisi04
14	55	70	-.77	.30	.98	-.09	.94	-.21	.31	.28	78.6	78.5	membacaeksposisi06
22	55	70	-.77	.30	1.05	.37	1.13	.60	.19	.28	78.6	78.5	membacadiskusi06
9	62	70	-1.57	.39	1.06	.32	.97	.04	.15	.22	88.6	88.6	membacaeksposisi01

Figure 3
Item Measure Results Reading Problem

The most difficult listening question, according to the item measure test, was question number 3. There were only 27 students who answered correctly out of 70 students, so the measure score of only 1.36, which meant that it was categorized as very difficult. In the reading question, the most difficult question was question number 23, which was a question about reading discussions. There were only 29 students who answered correctly out of 70 students, so the measure score of 1.06 meant it was classified as a difficult question.

The purpose of question point analysis was to provide students with diagnostic information and produce quality questions. Quality questions were questions that can provide precise information so that students who have mastered the material can be known and those who have not (Magdalena et al., 2021). The difficulty of the questions in the resulting CDA had a proportional level of difficulty, consisting of very difficult, difficult, medium, and easy questions.

In addition to the difficulty of the question items, another thing that was analyzed to produce individual feedback was the person's measure or individual ability. Individual ability levels detail each respondent's logit information, ranging from highest to lowest ability (Sumintono & Widhiarso, 2015). The logit score indicated the students' ability to solve the problem. The higher the logit means that students can do the questions well. The results of the person measure analysis can be seen in Figures 4 and 5.

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
3	21	24	2.11	.63	.92	-.03	.91	.03	.31	.22	87.5	87.5	003
10	21	24	2.11	.63	.81	-.33	.51	-.83	.55	.22	87.5	87.5	010
38	21	24	2.11	.63	.88	-.13	.59	-.63	.45	.22	87.5	87.5	038
44	21	24	2.11	.63	.88	-.13	.59	-.63	.45	.22	87.5	87.5	044
2	20	24	1.75	.56	1.00	.11	.92	-.02	.26	.24	83.3	83.3	002
9	20	24	1.75	.56	.93	-.08	.88	-.11	.34	.24	83.3	83.3	009
22	20	24	1.75	.56	1.00	.11	.95	.05	.26	.24	83.3	83.3	022
25	20	24	1.75	.56	.91	-.14	.89	-.09	.36	.24	83.3	83.3	025
31	20	24	1.75	.56	1.01	.16	.97	.09	.24	.24	83.3	83.3	031
45	20	24	1.75	.56	.76	-.64	.53	-1.05	.63	.24	83.3	83.3	045
56	20	24	1.75	.56	1.01	.15	1.00	.16	.22	.24	83.3	83.3	056
57	20	24	1.75	.56	1.05	.27	1.11	.38	.13	.24	83.3	83.3	057
59	20	24	1.75	.56	1.01	.15	1.00	.16	.22	.24	83.3	83.3	059
65	20	24	1.75	.56	1.01	.15	1.00	.16	.22	.24	83.3	83.3	065
66	20	24	1.75	.56	1.05	.27	1.11	.38	.13	.24	83.3	83.3	066
68	20	24	1.75	.56	1.01	.15	1.00	.16	.22	.24	83.3	83.3	068
21	19	24	1.46	.52	1.00	.08	.89	-.15	.30	.26	79.2	79.1	021
58	19	24	1.46	.52	.92	-.19	.75	-.56	.42	.26	79.2	79.1	058
60	19	24	1.46	.52	.97	-.02	.85	-.28	.34	.26	79.2	79.1	060
67	19	24	1.46	.52	.92	-.19	.75	-.56	.42	.26	79.2	79.1	067
69	19	24	1.46	.52	.97	-.02	.85	-.28	.34	.26	79.2	79.1	069
23	18	24	1.21	.49	.95	-.14	.81	-.53	.40	.28	70.8	75.3	023
24	18	24	1.21	.49	1.06	.33	1.02	.16	.21	.28	70.8	75.3	024
30	18	24	1.21	.49	1.06	.33	1.02	.16	.21	.28	70.8	75.3	030
47	18	24	1.21	.49	1.03	.22	1.03	.20	.23	.28	79.2	75.3	047
49	18	24	1.21	.49	1.05	.29	1.07	.30	.20	.28	79.2	75.3	049
50	18	24	1.21	.49	1.00	.08	.93	-.12	.30	.28	79.2	75.3	050
12	17	24	.98	.47	1.06	.35	1.10	.46	.19	.29	66.7	71.6	012
14	17	24	.98	.47	1.09	.49	1.17	.69	.14	.29	66.7	71.6	014

Figure 4
Person Measure Test Results Listening Questions

Based on Figure 4, it can be seen that the student with the highest ability is the student who was on the top row in the entry number column, that was, student number 3. He correctly answered 21 questions out of the 24 listening questions presented, so the person's score was 2.11. Students with high ability and the same grades were students numbers 10, 38, and 44.

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
45	21	24	2.09	.63	.91	-.07	.66	-.46	.39	.18	87.5	87.5	045
3	20	24	1.74	.56	.90	-.17	.72	-.51	.40	.21	83.3	83.3	003
9	20	24	1.74	.56	.90	-.17	.72	-.51	.40	.21	83.3	83.3	009
10	20	24	1.74	.56	.96	.00	.78	-.35	.32	.21	83.3	83.3	010
21	20	24	1.74	.56	1.03	.20	1.06	.28	.15	.21	83.3	83.3	021
30	20	24	1.74	.56	.96	.00	.78	-.36	.32	.21	83.3	83.3	030
38	20	24	1.74	.56	.96	.01	1.03	.22	.24	.21	83.3	83.3	038
44	20	24	1.74	.56	.96	.01	.93	.01	.26	.21	83.3	83.3	044
56	20	24	1.74	.56	.90	-.17	.69	-.59	.41	.21	83.3	83.3	056
59	20	24	1.74	.56	.90	-.17	.69	-.59	.41	.21	83.3	83.3	059
65	20	24	1.74	.56	.90	-.17	.69	-.59	.41	.21	83.3	83.3	065
68	20	24	1.74	.56	.90	-.17	.69	-.59	.41	.21	83.3	83.3	068
2	19	24	1.45	.52	.98	.01	.95	.00	.27	.23	79.2	79.1	002
24	19	24	1.45	.52	.98	.01	.95	.00	.27	.23	79.2	79.1	024
25	19	24	1.45	.52	1.07	.35	1.14	.49	.10	.23	79.2	79.1	025
22	18	24	1.20	.49	1.19	.85	1.36	1.12	-.10	.25	75.0	75.0	022
60	18	24	1.20	.49	1.06	.32	1.06	.28	.16	.25	75.0	75.0	060
69	18	24	1.20	.49	1.06	.32	1.06	.28	.16	.25	75.0	75.0	069
7	17	24	.97	.47	1.04	.25	.98	.02	.22	.26	66.7	71.0	007
11	17	24	.97	.47	1.02	.19	1.00	.09	.23	.26	66.7	71.0	011
12	17	24	.97	.47	.93	-.28	.94	-.15	.35	.26	75.0	71.0	012
14	17	24	.97	.47	.94	-.24	.94	-.15	.34	.26	75.0	71.0	014
33	17	24	.97	.47	.80	-1.04	.69	-1.20	.59	.26	75.0	71.0	033
37	17	24	.97	.47	.89	-.51	.87	-.40	.42	.26	75.0	71.0	037
57	17	24	.97	.47	1.05	.34	.97	-.01	.21	.26	66.7	71.0	057
66	17	24	.97	.47	1.05	.34	.97	-.01	.21	.26	66.7	71.0	066

Figure 5
Person Measure Test Results Reading Questions

Based on Figure 5, it can be seen that the student with the highest ability to read questions was student number 45. He answered correctly 21 questions out of 24 listening questions presented, so the score of the person was 2.09. Students with high ability below student number 45 were students number 3, 9, and 10.

Not only did it detect the sequence of students' abilities, but the feedback was also combined with scalogram results. Testing using scalograms was used to identify error responses, predict scores on missing data, find out respondents' abilities not only based on correct answers but also identified the origin of guesses, and label careless students (Sumintono & Widhiarso, 2015). With scalogram analysis, it can be detected whether students were really able to solve problems or were just guessing. Students from guessing can be detected by inconsistencies in answers, for example he was correct in answering one difficult question and wrong in another difficult question, the student was also wrong in answering inquiries that are categorized as easy.

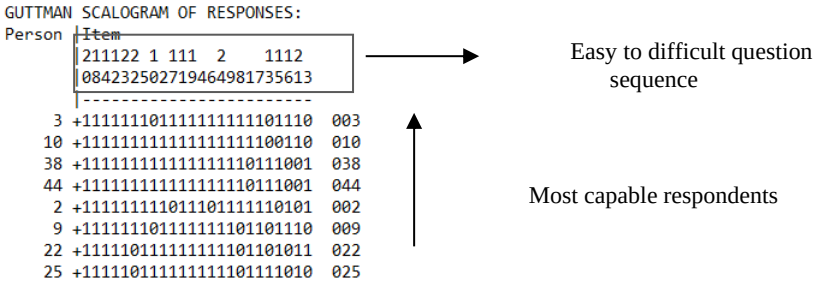

Figure 6
Scalogram Analysis Results Listening Questions

Figure 6 provides information that the most capable respondents sequentially were students with numbers 3, 10, 38, and 44. Students with numbers 38 and 44 had the same ability because they had the same answers. While student number 3, although he was the most capable student, had careless potential because he answered incorrectly on question number 10, which was in the category of questions with moderate difficulty, even though he was able to answer correctly on questions in the difficult category.

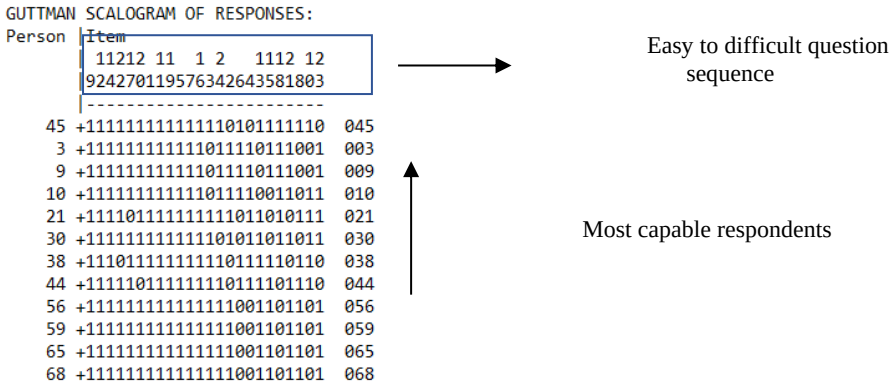

Figure 7
Scalogram Analysis Result Reading Question

Figure 7 can indicate that the most capable respondents for sequential reading competence were students with sequence numbers 045,003,009,010, and 021. Students with sequence numbers 3 and 9 had the same ability because they had the same answers. Meanwhile, student number 21 had careless potential because he answered incorrectly on question number 22, which was categorized as a question with moderate difficulty, even though he was able to answer correctly on difficult questions.

In addition to item measure, person measure, and scalogram, RASCH modeling can also analyze the distribution of respondents' abilities with the same scale, namely person: wright map. The Person:Wright map delineates groups of learner abilities that could serve as tools for differentiation in learning. The results of the person-right map analysis can be seen in Pictures 8 and 9 below.

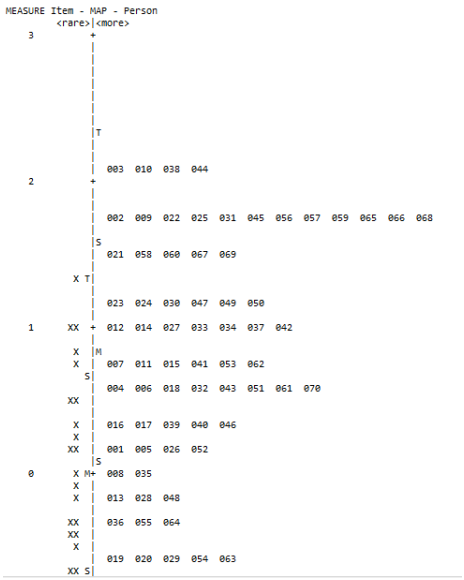

Figure 8
Person Map Results of Listening Question Answers

Based on Figure 8, it can be seen that there was a laterization of students' abilities from the highest to the lowest. The group of students with high listening competence were students with sequence number 003,010,038,044. Followed by groups that had abilities below, which were students numbers 002, 009, 022, 025, 031, 045, 057, 059, 065, 066, and 068. While the lowest group of student abilities were students numbers 019,020, 029,054, and 063.

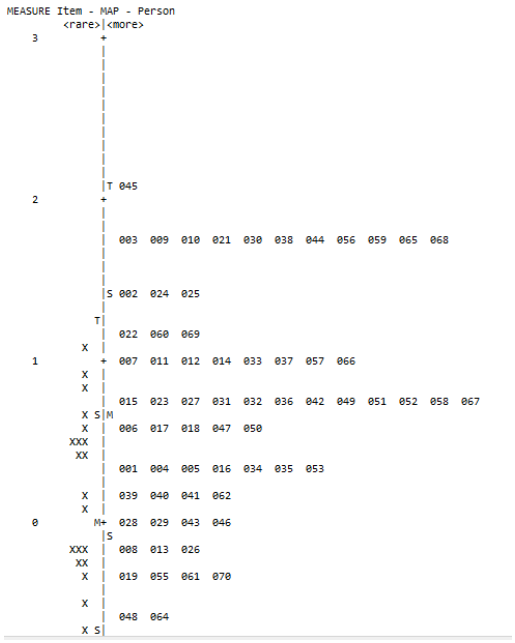

Figure 9
Person Map of Reading Question Answer Result

Figure 9 illustrates the progression of students' reading abilities from the highest to the lowest. The student with the highest ability in competence was the student with sequence number 045. He did not have a group because other students were not able to match his ability. Followed by groups that have the ability below, namely students numbers 003,009, 010,021, 030, 044, 056, 059, 065, and 068, while the lowest group of student abilities were students numbers 048 and 064.

Research conducted by Tang & Zhan (2021) proved that feedback provided based on cognitive diagnostic assessments can improve student learning and is more effective than true-false feedback on student answers. This is especially true in the field of higher-order thinking skills (HOTS). This study assumed that the more information provided in feedback, the more benefits students can get from correcting and improving their learning abilities. Therefore, in this study, students were not only shown the ability to answer questions but also at what point they are weak, at what point they are strong, and their position in the classroom.

Based on the assessment results, diagnostic information can be given in the form of feedback on what students should do from a cognitive perspective (Jang et al., 2015; Kim, 2015). Remedial teaching can utilize this individual feedback as a foundation. This feedback provides sufficient detail to assess students' proficiency in each sub-competency, enabling teachers and students to monitor progress (Sawaki & Koizumi, 2017). To facilitate user interpretation of feedback, researchers translate diagnostic results into qualitative descriptions in as few technical words as possible. Researchers use simple language, assuming even the lowest ability students will be able to understand the feedback given. Researchers use positive, active sentences to create feedback descriptors.

Feedback will provide an in-depth understanding of individual differences in their engagement and the factors that influence it (Zheng et al., 2023). This research has indeed produced individualized feedback for each student who took the test in the form of the students' cognitive strengths and weaknesses. However, this research has not yet identified the factors that affect these cognitive strengths and weaknesses. So, the researchers recommend further researchers develop information technology-based CDA that can broadcast feedback directly as soon as students finish working on the questions. Further research expects feedback to identify student weaknesses and strengths, as well as to describe the factors that influence them.

CONCLUSION

Based on a series of studies conducted, this study has produced a cognitive diagnostic assessment instrument that met the requirements of validity, reliability, readability, and applicability and was ready to be implemented. Thus, it can be claimed that the CDA instrument produced was able to measure cognitive ability and produce individualized feedback for language receptive ability on multimodal text in sub-competencies, namely explanation, interpretation, application, and perspective. The results of this study played an important role in detecting students' cognitive strengths and weaknesses and providing direction for steps that teachers and students must take to follow up. In addition, it also facilitated student grouping, selecting content, process, and product targets for differentiated learning of the next material. This study also contributes to addressing the long-standing criticism that language assessments fail to provide individualized feedback to link assessment to learning, as there has been no research addressing CDA with individualized feedback in the language domain. However, it needs to be recognized that in its application, the resulting CDA instrument requires adequate infrastructure support, so it cannot be applied in schools that are lacking. Feedback cannot be formulated quickly because teachers need to do an analysis first. This study has not yet detected factors that affect cognitive strengths and weaknesses. So that researchers recommend further researchers develop information technology-based CDA that can broadcast feedback directly as soon as students finish working on the problem. Further research expects feedback to identify student weaknesses and strengths, as well as to describe the factors influencing them.

DECLARATIONS

Author contribution	:	Giati Anisah leads and is responsible for all research projects implemented in Bojonegoro District. She also wrote the manuscript and collaborated with the second author. Midya Yuli Amreta participated in field data collection and analysis. Both authors approved the final manuscript.
Funding statement	:	This research was funded by the Directorate of Research, Technology, and Community Service of the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia in 2023.
Conflict of interest	:	Both authors declare that they have no competing interests.
Ethics Approval	:	The authors agree to have this article published in KEMBARA in 2024.
Additional information	:	No additional information is available for this paper.

REFERENCE

Agnafia, D. N. (2019). Analisis Kemampuan Berpikir Kritis Siswa dalam Pembelajaran Biologi. *Florea: Jurnal Biologi dan Pembelajarannya*, *6*(1), 45–53. http://doi.org/10.25273/florea.v6i1.4369

Al Fajri, T. A. (2020). Pentingnya Penggunaan Pendekatan Multimodal dalam Pembelajaran. *Waskita: Jurnal Pendidikan Nilai dan Pembangunan Karakter*, *2*(1), 57–72. https://doi.org/10.21776/ub.waskita.2018.002.01.5

Andrich, D., & Marais, I. (2019). A Course in Rasch Measurement Theory. In *Measuring in the educational, social and health sciences* (Vol. 41). Springer. https://link.springer.com/book/10.1007/978-981-13-7496-8

Anisah, G. (2018). Validitas Instrumen Asesmen Menulis Cerpen Terintegrasi Pendidikan Karakter. *AL-Ulya: Jurnal Pendidikan Islam*, *3*(1), 15–25. https://doi.org/10.32665/ulya.v3i1.695

Anisah, G., & Amreta, M. Y. (2023). Pengembangan Instrument Assessment as Learning Berbasis Projek untuk Pembelajaran Menyimak dan Berbicara Monologis Dialogis Bermuatan Karakter Bhineka Tunggal Ika. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *9*(1), 76–88. https://doi.org/10.22219/kembara.v9i1.22701

Arijanty, A. (2014). Model Rasch sebagai Kerangka Acuan Penyusunan Alat Ukur. *Jurnal Pendidikan dan Kebudayaan*, *20*(1), 109–123. https://doi.org/10.24832/jpnk.v20i1.130

Arikunto, S. (2019). *Prosedur Penelitian*. Rineka Cipta.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A Comprehensive Review of Rasch Measurement in Language Assessment: Recommendations and Guidelines for Research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Astalini, A., Darmaji, D., Kurniawan, W., Anwar, K., & Kurniawan, D. (2019). Effectivenes of Using e-Module and e-Assessment. *International Journal of Interactive Mobile Technologies*, *13*(9), 21–39. https://doi.org/10.3991/ijim.v13i09.11016

Astuti, A. P., Aziz, A., Sumarti, S. S., & Bharati, D. A. L. (2019). Preparing 21st Century Teachers: Implementation of 4C Character's Pre-Service Teacher Through Teaching Practice. *Journal of Physics: Conference Series*, *1233*(1), 12109. https://doi.org/10.17509/ijal.v9i2.20218

Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model RASCH untuk Analisis Instrumen Tes pada Mata Kuliah Matematika Aktuaria. *Jurnal Pendidikan Matematika (JUPITEK)*, *3*(1), 45–50. https://doi.org/10.30598/jupitekvol3iss1pp45-50

Beauty, M., Rahmawati, L. E., Markhamah, M., & Tawandorloh, K.-A. (2023). Evaluasi CIPP Pembelajaran Keterampilan Membaca di Program Studi Pendidikan Bahasa dan Sastra Indonesia: Analisis proses dan produk. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *9*(1), 1–14. https://doi.org/10.22219/kembara.v9i1.23516

Budyartati, S. (2016). Tes Kognitif Diagnostik untuk Mendeteksi Kesulitan Belajar Siswa Sekolah Dasar. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, *1*(1). https://doi.org/10.33830/penaanda.v1i2.6202

Carless, D., & Boud, D. (2018). The Development of Student Feedback Literacy: Enabling Uptake of Feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chan, S.-W., Looi, C.-K., & Sumintono, B. (2021). Assessing Computational Thinking Abilities Among Singapore Secondary Students: A Rasch Model Measurement Analysis. *Journal of Computers in Education*, *8*(1), 213–236. https://doi.org/10.1007/s40692-020-00177-2

Diakidoy, I.-A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2019). The Relationship Between Listening and Reading Comprehension of Different Types of Text at Increasing Grade Levels. *Reading Psychology*, *26*(1), 55–80. https://doi.org/10.1080/02702710590910584

Dressman, M. (2019). Multimodality and Language Learning. In *The Handbook of Informal Language learning*. Wiley Online Library. https://doi.org/10.1002/9781119472384.ch3

Ermiyanto, E., Asroa, I., & Ilyas, A. (2023). Asesmen Diagnostik Gaya Belajar Siswa Kelas VII di SMPN 4 Padang Panjang. *MANAZHIM*, *5*(1), 166–177. https://doi.org/10.36088/manazhim.v5i1.2845

Gravemeijer, K., Stephan, M., Julie, C., Lin, F., &, & Ohtani, M. (2017). What Mathematics Education May Prepare Students for the Society of the Future?. *International Journal of Science and Mathematics Education, 15*(1), 105–123. https://doi.org/10.1007/s10763-017-9814-6

Habsy, B. A., Azizah, N. H., Viola, N. P., & Mahendra, W. (2024). Peran Guru dalam Menciptakan Lingkungan Pembelajaran yang Efektif dengan Cara Pengelolaan Kelas yang Menarik. *TSAQOFAH*, *4*(2), 545–566. https://doi.org/10.58578/tsaqofah.v4i2.2305

Hasna, S., Azizah, M., & others. (2023). Implementasi Asesmen Diagnostik Non Kognitif Siswa Kelas 3 SD Negeri Gayamsari 02 Kota Semarang. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *9*(2), 6037–6049. https://doi.org/10.36989/didaktik.v9i2.1390

Hattie, J., & Clarke, S. (2018). *Visible Learning: Feedback*. Routledge.

Hendaryan, & Noviadi, A. (2023). The Role of Metacognition Strategies (Metacomprehension) and Inferential Ability in Improving Reading Comprehension Ability. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *9*(2), 363–375. https://doi.org/10.22219/kembara.v9i2.26097

Hesse, F., Care, E., Buder, J., Sassenberg, K., & G., & P. (2015). A Framework for Teachable

Collaborative Problem Solving Skills. In *Assessment and teaching of 21st century skills* (pp. 37–56). https://doi.org/10.1007/978-94-017-9395-7_2

Hilman, I., Akmal, R., & Nugraha, F. (2023). Analisis Gaya Belajar Peserta Didik Melalui Assessment Diagnostik Non Kognitif pada Pembelajaran Diferensiasi di Sekolah Dasar. *Naturalistic: Jurnal Kajian dan Penelitian Pendidikan dan Pembelajaran*, *8*(1), 161–167. https://doi.org/10.35568/naturalistic.v8i1.3911

Himawan, R., & Nurgiyantoro, B. (2022). Analisis Butir soal Latihan Penilaian Akhir Semester Ganjil Mata Pelajaran Bahasa Indonesia Kelas VIII SMPN 1 Bambanglipuro Bantul Menggunakan Program ITEMAN. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *8*(1), 160–180. https://doi.org/10.22219/kembara.v8i1.20530

Huda, A. A. S., Alamsyah, A., Selvia, S., & Sangadah, N. (2023). Asesmen Diagnostik Kognitif pada Mata Pelajaran PAI Kelas 7 di SMPN 3 Lembang. *Al'Ulum Jurnal Pendidikan Islam*, *3*(2), 213–224. https://doi.org/10.54090/alulum.298

Hui, L., de Bruin, A. B. H., Donkers, J., & van Merriënboer, J. J. G. (2021). Does Individual Performance Feedback Increase the use of Retrieval Practice? *Educational Psychology Review*, *33*(1), 1–23. https://doi.org/10.1007/s10648-021-09604-x

Huilin, C. L. (2013). *Cognitive Diagnostic Approach — A New Stage of Language Testing Research*. https://api.semanticscholar.org/CorpusID:63374199

Imamyartha, D., Fitriyah, S. M., Tasnim, Z., Puspa, A., Fardhani, A. E., Wahjuningsih, E., Sundari, S., Hudori, R. F., & Arya, B. (2019). The Efficacy of 4Cs-based Reading to Foster 21st-Century Learning Competencies. *Indonesian Journal of Applied Linguistics*, *9*(2), 275–285. https://doi.org/doi: 10.17509/ijal.v9i2.20218

Irons, A., & Elkington, S. (2021). *Enhancing Learning Through Formative Assessment and Feedback*. Routledge.

Jain, S., & Wallace, B. C. (2019). Attention is Not Explanation. *ArXiv*. https://doi.org/https://doi.org/10.48550/arXiv.1908.04626

Jang, E. E., Dunlop, M., Park, G., & Van Der Boom, E. H. (2015). How do Young Students with Different Profiles of Reading Skill Mastery, Perceived Ability, and Goal Orientation Respond to Holistic Diagnostic Feedback? *Language Testing*, *32*(3), 359–383. https://doi.org/10.1177/0265532215570924

Jang, E. E., & Sinclair, J. (2021). Diagnostic Assessment in Language Classrooms. *The Routledge Handbook of Language Testing*, 187–205. https://doi.org/10.4324/9781003220756-15

Johnson, E. (2021). Face Validity. In *Encyclopedia of Autism Spectrum Disorders*. Springer. https://doi.org/10.1007/978-3-319-91280-6_308

Kamaruddin, A., Patmasari, A., Whanchit, W., & Suriaman, A. (2023). The Effect of IBL (Inquiry-Based Learning) Model on EFL Students' Critical Thinking Skills. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *9*(1), 15-23.. https://doi.org/10.22219/kembara.v9i1.22766

Karlimah, K. (2022). How Does RASCH Modeling Reveal Difficulty and Suitability Level the Fraction Test Question? *Jurnal Elemen*, *8*(1), 66–76. https://doi.org/10.29408/jel.v8i1.4170

Kementerian Pendidikan, Kebudayaan, R. dan T. (2022). *Capaian Pembelajaran Kurikulum Merdeka 2022*. Kementerian Pendidikan, Kebudayaan, Riset Dan Teknologi. https://www.kemdikbud.go.id/main/blog/2021/12/kemendikbudristek-dorong-sekolah-memahami-opsi-kurikulum-prototipe-untuk-pulihkan-pembelajaran

Kim, A.-Y. (2015). Exploring Ways to Provide Diagnostic Feedback with an ESL Placement Test: Cognitive Diagnostic Assessment of l2 Reading Ability. *Language Testing*, *32*(2), 227–258. https://doi.org/10.1177/0265532214558457

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-Based Learning: A Review of the Literature. *Improving Schools*, *19*(3), 267–277. https://doi.org/10.1177/1365480216659733

Kramer, A.-W., Schaaf, J. V, & Huizenga, H. M. (2023). How Much do You Want to Learn? High-School Students' Willingness to Invest Effort in Valenced Feedback-Learning Tasks. *Learning*

and Individual Differences, *108*, 102375. https://doi.org/10.1016/j.lindif.2023.102375

Larenas, C. D., Díaz, A. J., Orellana, Y. R., & Villalón, M. J. S. (2021). Exploring the Principles of English Assessment Instruments. *Ensaio*, *29*(111), 23-37. https://doi.org/10.1590/S0104-40362021000290285l

Laudel, H., & Narciss, S. (2023). The Effects of Internal Feedback and Self-Compassion on The Perception of Negative Feedback and Post-Feedback Learning Behavior. *Studies in Educational Evaluation*, *77*(1), 101237. https://doi.org/10.1016/j.stueduc.2023.101237

Leighton, J., & Gierl, M. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press.

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *Bintang*, *3*(3), 198–214. http://dx.doi.org/10.24042/ajpm.v6i2.50

Mauliya, A. (2019). Perkembangan Kognitif pada Peserta Didik SMP Menurut Jean Piaget. *ScienceEdu*, *2*(2), 86–91. https://doi.org/10.19184/se.v2i2.15059

Miles, Matthew B., Huberman, A. Michael. dan Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE.

Min, S., & He, L. (2022). Developing Individualized Feedback for Listening Assessment: Combining Standard Setting and Cognitive Diagnostic Assessment Approaches. *Language Testing*, *39*(1), 90–116. https://doi.org/10.1177/0265532221995475

Misbah, S. (2022). Penerapan Metode Umpan Balik (Feed Back Partner) untuk Meningkatkan Prestasi Belajar Siswa pada Mata Pelajaran Bahasa Indonesia Materi Struktur dan Kebahasaan Teks Anekdot Kelas X IPS-2 SMAN 4 Kota Bima Semester 1 Tahun Pelajaran 2020/2021. *Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI)*, *2*(2), 143–154. https://doi.org/10.53299/jppi.v2i2.219

Molenda, M. (2015). In Search of the Elusive ADDIE Model. *Performance Improvement*, *54*(2), 40–43. https://www.researchgate.net/profile/Michael-Molenda/publication/251405713_In_search_of_the_elusive_ADDIE_model/links/5b5f098ba6fdccf0b200e5b2/In-search-of-the-elusive-ADDIE-model.pdf

Müller, M. (2020). Item Fit Statistics for Rasch Analysis: Can we Trust Them? *Journal of Statistical Distributions and Applications*, *7*(1), 1–12. https://doi.org/10.1186/s40488-020-00108-7

Musaad, F., & Suparman, S. (2023). Pengembangan e-Modul Berbasis Problem Based Learning untuk Memacu Kemampuan Berfikir Kritis Abad-21. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(3), 3162–3171. http://dx.doi.org/10.24127/ajpm.v12i3.6119

Muslihin, H. Y., Suryana, D., Suherman, U., Dahlan, T. H., & others. (2022). Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using RASCH model. *International Journal of Instruction*, *15*(2), 207–222.

Parmaningsih, T. J., & Saputro, D. R. S. (2021). RASCH Analysis on Item Response Theory: Review of Model Suitability. *AIP Conference Proceedings*, *2326*(1), 13-27. https://doi.org/10.1063/5.0040305

Perea, M. (2013). Why Does the APA Recommend the Use of Serif Fonts? *Psicothema*, *25*(1), 13–17. https://doi.org/10.7334/psicothema2012.141

Pérez-González, L. (2019). Multimodality. In *Routledge Encyclopedia of Translation Studies* (pp. 346–351). Routledge.

Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru melalui Pendekatan Item Response Theory (IRT) Model RASCH. *Tarbawy: Jurnal Pendidikan Islam*, *7*(1), 61–70. https://doi.org/10.21776/ub.waskita.2018.002.01.5

Prihatni, Y., Kumaidi, K., & Mundilarto, M. (2016). Pengembangan Instrumen Diagnostik Kognitif pada Mata Pelajaran IPA di SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*(1), 111–125. https://doi.org/10.21831/pep.v20i1.7524

Putri, A. O. (2021). Pengembangan Tes Diagnostik Berbentuk Uraian Berdasarkan Ranah Kognitif untuk Mengetahui Letak Kesulitan Belajar Siswa SMP pada Materi Bangun Ruang. *Jurnal*

Pusaka, *10*(1), 40–54. https://doi.org/10.35897/ps.v10i1.597

Redlo, J. M. (2021). Barriers in Teaching the Four C's of 21st Century Competencies: Dismantling the Obstacles. In *Handbook of research on barriers for teaching 21st-Century competencies and the impact of digitalization* (pp. 1–14). IGI Global. https://doi.org/10.4018/978-1-7998-6967-2.ch001

Safi'i, I., Tarmini, W., & Wahdini, L. (2021). Critical Thinking in Evaluation Instruments at BSE Indonesian Language. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, *7*(2), 232–240. https://doi.org/10.22219/kembara.v7i2.17300

Saleh Alharbi, A., Abdullah Alhebshi, A., & Meccawy, Z. (2021). EFL Students' and Teachers' Perceptions of Google forms as a Digital Formative Assessment Tool in Saudi Secondary Schools. *Arab World English Journal (AWEJ) Special Issue on CALL*, *7*(2), 140–154. https://dx.doi.org/10.24093/awej/call7.10

Sawaki, Y., & Koizumi, R. (2017). Providing Test Performance Feedback that Bridges Assessment and Instruction: The Case of Two Standardized English Language Tests in Japan. *Language Assessment Quarterly*, *14*(3), 234–256. https://doi.org/10.1080/15434303.2017.1348504

Shrotryia, V. K., & Dhanda, U. (2019). Content Validity of Assessment Instrument for Employee Engagement. *SAGE Open*, *9*(1), 234–245. https://doi.org/10.1177/2158244018821751

Stemler, S. E., & Naples, A. (2021). Rasch Measurement V. Item Response Theory: Knowing When to Cross The Line. *Practical Assessment, Research & Evaluation*, *26*, 11.

Suarni. (2023). Implementasi Asesmen Diagnostik dalam Kurikulum Merdeka pada Mata Pelajaran Bahasa Inggris Kelas XI/Fase F di Madrasah Aliyah Negeri Kota Sorong. *JKIP*, *1*(4), 263–270. https://doi.org/10.61116/jkip.v1i4.188

Sumintono, B., & Widhiarso, W. (2015). *RASCH Modeling Application in Educational Assessment*. Trimkom Publising Home.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan*. Trim Komunikata.

Sun, H., & Hwang, Y. (2023). Enhancing EFL learners' English Reading Performance Through the Diagnostic Feedback of CDA. *Korean Journal of English Language and Linguistics*, *23*(July), 535–553. https://doi.org/10.15738/kjell.23..202307.535

Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan Instrumen Asesmen Diagnostik untuk Persiapan Kurikulum Merdeka. *Journal of Community Empowerment*, *2*(2), 67–73. https://doi.org/10.15294/jce.v2i2.61886

Susanti, E. (2019). *Keterampilan Menyimak*. Rajawali Pers.

Tang, F., & Zhan, P. (2021). Does Diagnostic Feedback Promote Learning? Evidence from a Longitudinal Cognitive Diagnostic Assessment. *AERA Open*, *7*. https://doi.org/10.1177/23328584211060804

Taylor, C., & Noë, A. (2021). *The Explanation of Behaviour*. Routledge.

Toprak, T. E., & Cakir, A. (2021). Examining The L2 Reading Comprehension Ability of Adult ELLs: Developing a Diagnostic Test within the Cognitive Diagnostic Assessment framework. *Language Testing*, *38*(1), 106–131. https://doi.org/10.1177/0265532220941470

Van Leeuwen, T. (2015). Multimodality. *The Handbook of Discourse Analysis*, 447–465. https://doi.org/10.1002/9781118584194.ch21

Wahyuni, L. D., Gumela, G., & Maulana, H. (2020). Interrater Reliability: Comparison of Essay's Tests and Scoring Rubrics. *Journal of Physics: Conference Series*, *1*(1), 1–6. https://doi.org/10.1088/1742-6596/1933/1/012081

Wang, M. (2023). Bridging Assessment and Learning: A Cognitive Diagnostic Analysis of a Large-Scale Spanish Proficiency Test. *Porta Linguarum*, *2023*(40), 9–24. https://doi.org/10.30827/portalin.vi40.15930

Wiggins, G., & McTighe, J. (2005). Understanding by Design (expanded 2nd ed.). In *Association for Supervision and Curriculum Development*.

Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy Revised: Understanding the New Version of Bloom's Taxonomy. *Quincycollege.Edu*, 2–4.

Wulandari, G. A. P. T. W., Putrayasa, I. B., & Martha, I. N. (2023). Efektivitas Asesmen Diagnostik dalam Pembelajaran Berdiferensiasi pada Pelajaran Bahasa Indonesia. *Nusantara: Jurnal Pendidikan Indonesia*, *3*(3), 433-448. https://doi.org/10.14421/njpi.2023.v3i3-5

Yeh, H. W. (2022). Effects of Diagnostic Tests on Self-Directed Language Learning Development. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *12*(1), 1-15. https://doi.org/10.4018/ijcallt.291105

Zhang, Z. V., & Hyland, K. (2022). Fostering Student Engagement With Feedback: An Integrated Approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zheng, Y., Yu, S., & Liu, Z. (2023). Understanding Individual Differences in Lower-Proficiency Students' Engagement with Teacher Written Corrective Feedback. *Teaching in Higher Education, 28*(2), 301-321. https://doi.org/10.1080/13562517.2020.1806225