ONLINE CLASSROOM-BASED READING ASSESSMENT: COMPREHENSION AND PRACTICE DEVELOPMENT

Eko SUHARTOYO
ORCID: 0000-0002-4881-3997
Faculty of Letters
Universitas Negeri Malang
Malang, INDONESIA

Dr. Rida AFRILYASANTI
ORCID: 0000-0002-0049-1490
Faculty of Letters
Universitas Negeri Malang
Malang, INDONESIA

Dr. Nur MUKMINATIEN
0000-0002-9031-5174
Faculty of Letters
Universitas Negeri Malang
Malang, INDONESIA

ABSTRACT

In this paper, we investigated the impact of an online classroom-based reading assessment on implementing practices in reading instruction among 30 EFL learners in an intermediate reading course at a public university in East Java, Indonesia. Our study aimed to develop an online classroom-based reading assessment and evaluate its efficacy in measuring students' reading skills. The online classroom-based assessment, characterized as systematic, comprehensive, instructional, diagnostic, and adaptive, yielded positive outcomes, as evidenced by specific trends in the qualitative analysis. While limitations were considered, the study's findings showcase the practical implications of incorporating online classroom-based assessment in monitoring student performance and evaluating teaching quality. The results emphasize the importance of online classroom-based reading assessment as a robust evaluation tool for aligning learning goals and needs. Overall, our study contributes valuable insights into the benefits of online classroom-based reading assessments and their potential to enhance traditional assessment methods in English as a Foreign Language (EFL) settings in this technologically advanced era.

Keywords: Classroom-based assessment, EFL classroom, online assessment, reading assessment, technology inclusion.

INTRODUCTION

Reading is a fundamental skill that significantly impacts all facets of life, including academic achievement, employment efficiency, and social relationships. Proficiency in reading involves meeting cognitive demands, encompassing the comprehension of explicit and implicit meanings, content organization, and retention of relevant information (Hoi, 2021). With dual objectives of fostering literary experiences and facilitating knowledge acquisition, reading skills form the bedrock of language learning environments (Duke & Carlisle, 2011; Murphy et al., 2009).

Within the context of language proficiency enhancement, evaluating reading abilities is of utmost importance, especially in English as a Foreign Language (EFL) classes (Ekpang et al., 2021). Contrary to the widespread misunderstanding that assessment is just used to evaluate student performance, its more

considerable function includes monitoring progress, matching teaching methods with learning goals, and adapting instructional materials to different levels of competence (Hoi, 2021). Based on Vygotskian ideas that emphasize the importance of language in social interaction, classroom-based evaluation plays a crucial role in evaluating and promoting literacy development (Vygotsky, 1981; Guthrie & Wigfield, 2000; Maplethorpe et al., 2022).

Recognizing the importance of classroom assessment in shaping students' learning paths, educators need to provide top-notch evaluations, particularly in reading. In addition to their inherent academic significance, proficient reading abilities broadly impact both professional and social contexts. Moreover, the changing media and information environment, driven by technological progress, highlight the crucial importance of reading in English as a Foreign Language (EFL) settings, requiring web-based English language teaching and readily available online evaluation tools (Raine, 2018).

Although widely recognized standardized reading assessments like the TOEFL and IELTS reading examinations are familiar (Khabir et al., 2022), their practical use in classrooms often exposes a lack of teachers' understanding of reading assessment procedures (Sawaki et al., 2009; Tengberg, 2018). The heavy dependence on standardized tests and a lack of recognition for teacher-created assessments in the classroom highlights the necessity for empirical investigation into reading assessments that focus on learning in the classroom setting.

The increasing need for online reading assessment in classroom settings necessitates developing customized methodologies aligned with the unique characteristics and requirements of English as a Foreign Language (EFL) classes. Indonesian EFL environments lack systematic, comprehensive, and diagnostic classroom-based learning-oriented reading evaluations. This study aims to fill this void by creating reading assessment tools similar to standardized instruments to evaluate comprehension abilities, identify reading difficulties, and assess the complexity of instructional materials in Indonesian EFL classes at the intermediate competency level. Our investigation aims to improve reading assessment methods and instructional methodologies in the Indonesian English as a Foreign Language (EFL) context.

LITERATURE REVIEW

Reading Test

The development of reading tests is a critical aspect of educational assessment, and several studies have emphasized the importance of their validity and reliability. Mullis et al. (2016) underline the importance of creating instruments that accurately measure intended aspects and produce consistent, replicable results. Ozdemir and Akyol (2019) further stress the need for reliable, validated tests to reflect students' learning outcomes accurately.

Despite this, traditional test development methods often lead to assessments that fail to gauge student achievement accurately. Hanafi (2016) points out that many teachers develop tests without proper guidelines, leading to inadequate measurement of learning outcomes. Similarly, Brown (2004) identifies that a lack of understanding of reading abilities and the reading process can result in flawed test development.

Selecting appropriate skills to assess and measure varied reading abilities is key to effective test development. Nisa and Helmanda (2019) argue that factors like reading materials, test formats, and types of reading should be considered to create comprehensive and valid assessments. This critical analysis underscores the importance of teachers engaging with past theories and methodologies in test development to ensure substantial achievement of validity, reliability, and accuracy. It also suggests that incorporating authentic classroom-based assessments could better align with learning goals and student needs.

Keenan et al. (2008) emphasize that comprehension is the primary goal of reading, and the ability to measure and understand it is increasingly demanded. Reading tests typically involve students responding to questions about texts that traditionally focus on single-source texts from printed materials. However, as Grabe (2009) notes, profound reading activities like integrating background knowledge and forming inferences are necessary for fully understanding text information. Agota (2020) identifies two primary purposes of reading tests: ensuring equitable grading across abilities and serving as a benchmark for evaluating teaching practices. Diagnostic exams, in particular, assess various reading sub-skills and identify strengths and weaknesses.

Recent research advocates for reading assessments based on classroom evidence (Andronova, 2021). According to Snow (2002), reading comprehension is deeply embedded in a sociocultural context involving the reader, task, and text, significantly influencing comprehension outcomes. Yu (2021) suggests that considering the intricacy of tests is crucial when examining reading comprehension to have a more profound appreciation of these associations. Additionally, readability assessments using computational methods are increasingly used to estimate text difficulty.

This study intends to reinforce using classroom-based assessment as an effective tool for diagnosing student comprehension and selecting appropriate instructional materials. Educational institutions may find the results of such assessments crucial for improving academic quality. The findings will also contribute to future research on classroom-based assessments in various language skills.

Types of Reading Test

Reading tests should encompass diverse tasks that address different reading strategies, such as scanning, skimming, and comprehensive reading. The first type of task generally assesses students' overall understanding of the reading material. The second focuses on analyzing specific components and the content of the text, while the third type evaluates the ability to comprehend detailed text information (Andronova, 2021; Walvoord, 1998). Nisa and Helmanda (2019) elaborate on various reading types that should be incorporated into reading tests.

According to Nisa and Helmanda (2019), reading materials can be classified into selective, participative, and comprehensive categories. Selective reading includes tasks involving graphics, true/false questions, and short response queries. Participatory reading involves interactive elements with the text, such as anecdotes and short stories. Comprehensive reading refers to longer texts like professional articles and essays.

Reading comprehension tests can be objective-based or essay-based (Nisa & Helmanda, 2019). Objective tests may include formats like true/false, multiple-choice, and match-up questions. In contrast, essay-based tests require test-takers to write responses, often evaluating their ability to defend answers using their own thoughts.

While the transition to online testing in reading comprehension offers advantages such as flexibility and instant feedback, it also presents several challenges and limitations. Technical issues are a primary concern, as online tests rely on stable internet connections and functional computer systems. Technical difficulties during testing can disrupt the process and affect student performance (Kostaki & Karayianni, 2022). A further significant obstacle is the disparity in technology. Unequal accessibility to essential technology among learners results in discrepancies in test-taking possibilities (Johnson et al., 2016). This lack of access can disproportionately affect students from lower socio-economic backgrounds.

Security concerns are heightened in online testing environments. The potential for cheating and unauthorized assistance increases due to the remote nature of these tests (Newton & Essex, 2023). Additionally, online tests often lack the immediate human interaction and support available in traditional settings, which can be crucial for some students, especially in understanding and interpreting test questions (Barrot et al., 2021).

Adapting to different learning styles is also more challenging in online formats. Traditional methods may cater more effectively to diverse learning preferences, thus providing a more inclusive assessment environment (Santiago et al., 2021). Online tests might also struggle to assess higher-order thinking skills as effectively as traditional essay-based questions (Fensham & Bellocchi, 2013; FitzPatrick et al., 2015; Koksal & Ulum, 2018). While online tests can provide immediate feedback, the quality and depth of this feedback may be limited compared to traditional classroom settings (Lu et al., 2021). This limitation could impact the learning and improvement opportunities typically derived from thorough and personalized feedback.

Therefore, while online reading tests offer notable advantages in efficiency and scalability, educators and test developers must consider and address these challenges to ensure fairness and effectiveness in assessment (Andronova, 2021; Nisa & Helmanda, 2019). A balanced approach to reading test administration is crucial, one that leverages the benefits of technology while mitigating its limitations. This understanding is essential for developing effective and equitable reading assessments in ELT contexts.

The Construction of Reading Test

Teachers need to consider factors such as validity and reliability when creating a test. To be valid, a test fulfills the course goals and objectives, test purposes, and comprehension processes such that they do not work in isolation from one another or the students' learning context (Mullis et al., 2016). The constructed instruments must be capable of measuring the aspects to be assessed (Fatimah et al., 2018). Furthermore, reliability means consistency. The test is considered reliable if the same test results are consistently obtained using the same strategies under the same conditions. A reliable test, however, is not necessarily valid. Although the results are replicable, they are not always accurate. Hence, developing a reading test as a formal assessment instrument necessitates several crucial decisions, and testing must be reliable and valid (Ozdemir & Akyol, 2019).

However, most teachers commonly develop tests without considering test development guidelines, resulting in inaccurate and invalid assessments that cannot gauge students' achievement (Hanafi, 2016). Consequently, there is a widespread occurrence of exams that are not valid or trustworthy and that do not effectively assess students' learning results. In his 2004 publication, Brown outlines several phases of the test development process. The creator should decide which skills to examine to begin the test development process. This assessment might be based on a logical analysis of the requisite reading abilities or a psychological understanding of the reading process (or both). Both factor analysis and experimentation can be used to evaluate the independence of different skills from one another.

Second, while developing a reading test, the test developer must determine how to measure the varied reading skills. Because test results are frequently misinterpreted by readers (and even educational specialists), test creators must also specify how and when their tests should be used (Schwartz, 1984). There are several techniques for measuring each reading skill, and test developers typically pick one or the other based only on personal choice. Reading theories are generally beneficial in selecting measuring techniques because they provide crucial test variable specifications. The next stage in developing a test is determining how the test will be graded. Several distinct scores have been used, and a test can produce multiple types of results.

Furthermore, teachers can administer tests online or offline using paper and pencil. However, in this technologically advanced era and response to the characteristics of today's digital native students, online tests are preferable. Students' Z-generation characteristics have led them to choose online-based assessment over offline assessment, and it is vital to meet students' reading learning preferences to perform better in reading comprehension (Ming & Aziz, 2019). Additionally, incorporating technology in classroom practices resulted in a favorable perspective of technological integration among instructors and students (Cahyono et al., 2023). Online tests are also more effective than paper and pencil tests because they run more smoothly in emergencies such as the COVID-19 pandemic (Sudarwati et al., 2021). Besides, online tests have several advantages over paper and pencil tests. It allows for adaptive assessment phases in which the test can be conducted (Spivey & McMillan, 2014), automatically randomizing the order of questions and answers for multiple-choice and matching questions to reduce the possibility of student cheating and providing instant feedback and reports.

As a result, in this study, we developed an online classroom-based test for EFL reading comprehension. As previously stated, the test for EFL reading comprehension is considered necessary because classroom-based learning-oriented assessments tailored to EFL classroom-specific features and needs and can be easily implemented in classroom instruction are still in their infancy. Furthermore, a few educational institutions still administer standardized competency examinations that can be used as diagnostic reading assessments.

To summarize, this research project was intended to produce an online classroom-based test for students in the English Language Education major's intermediate reading class. The informational and literary reading texts are accommodated in the content. The product of this study is a set of classroom-based learning-oriented tests for reading skills, which contain 40 multiple-choice test items. In brief, the product is expected to benefit teachers and students in the intermediate reading course by demonstrating how classroom-based learning-oriented assessment is built and functions to assess students' reading proficiency vigorously, diagnosing students' competency level and areas for teaching improvement, and contributing to the advancement of assessment theory in English Language Teaching (ELT).

The results of a classroom-based reading test formulation and tryouts for second-year university students participating in an intermediate reading course are examined in this study. It intends to create reading test instruments similar to standardized assessment tools for assessing students' comprehension, diagnosing reading difficulties, and defining instructional material difficulty levels for usage at the intermediate competency level in the Indonesian EFL setting. This research seeks to tackle the following inquiries:

1. What are the findings of the classroom-based reading test validation?
2. How are the results of the developed classroom-based reading test tryouts?
3. What are the ramifications for the tryout results?

METHOD

This study employed a two-stage approach encompassing research and development (R&D) for constructing an online classroom-based reading assessment. The methodology drew upon Borg and Gall's (1983) design, which was integrated with McGrath's (2016) design and eventually integrated with McGrath's (2016) stages, as illustrated in Figure 1. The R&D process was divided into three major phases: 1) exploration, 2) development, and 3) validation and implementation.
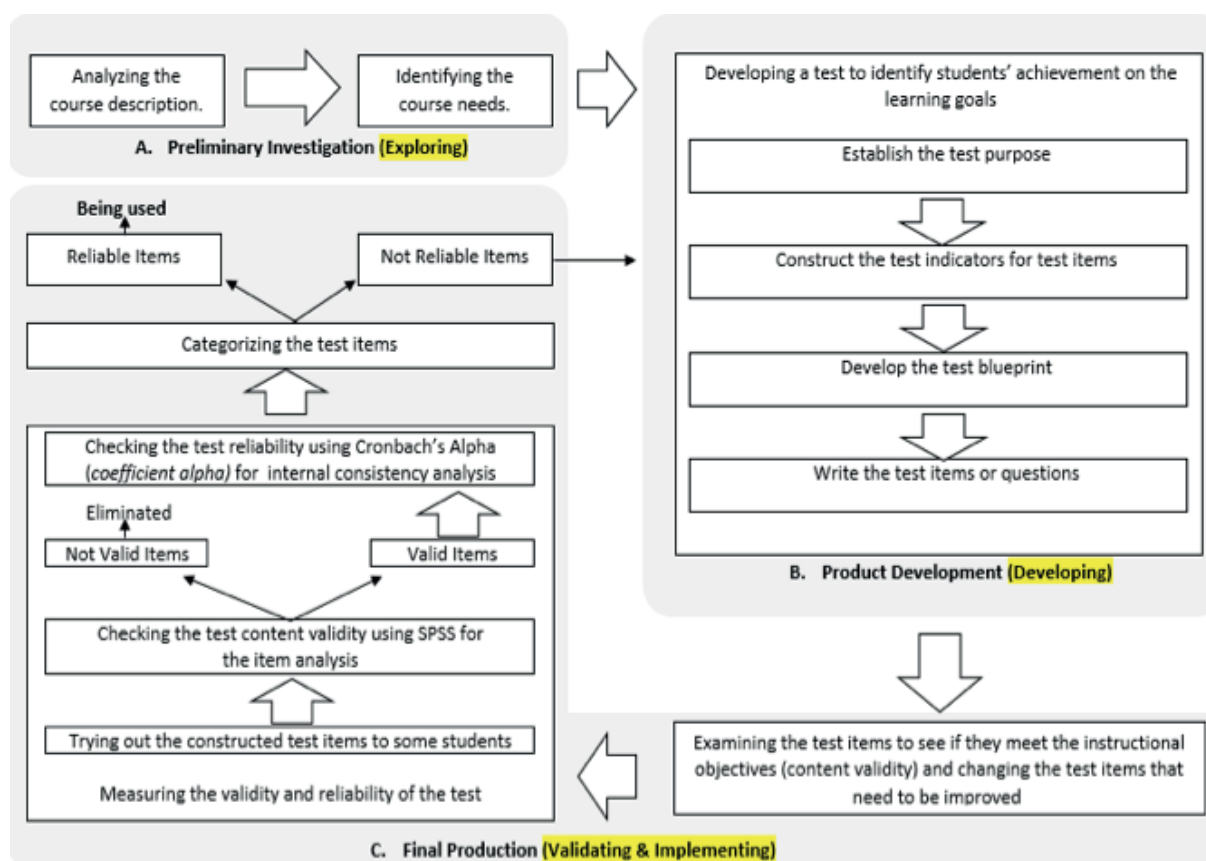

Figure 1. Material Development Procedure

The first stage entails doing a preliminary inquiry or exploration. During the initial inquiry phase, we comprehensively evaluated the English Language Education Study Program's intermediate reading course at an East Java public institution. This process involves the examination of course descriptions, conversations with instructors, and a comprehensive evaluation of pertinent literature. We systematically analyzed the course description to identify explicit objectives and competencies related to reading as outlined in the curriculum. Subsequently, we proceeded to conduct interviews with the teachers. We engaged in discussions with experienced instructors of the intermediate reading course. These discussions generated subjective

insights regarding teachers' evaluations of students' reading ability, challenging areas, and their expected educational goals. Afterward, an extensive review of pertinent literature was undertaken to discover practical approaches for assessing reading proficiency in English as a Foreign Language (EFL) contexts. This measure ensured compliance with established industry norms.

The second phase entailed product development, which began with identifying the test's objective. The assessment's diagnostic nature was acknowledged, leading to further processes of identifying test signs, creating a test blueprint, and composing test items. We generate and develop test metrics based on the findings from our investigation. Five essential indicators were devised, which include detecting keywords and themes, making inferences, interpreting dictions and figures of speech, assessing culture-bound diction, and examining text organization and idea development. Subsequently, a test blueprint was produced by delineating the framework and allocating test items according to the discovered indications. In addition, we developed test items that were aligned with the specified indicators and test plan.

The last phase involved the process of validating and implementing. The validation method encompassed face validity tests and content validity and reliability assessments. During face validity testing, we evaluated whether the test items aligned with instructional objectives, if the reading texts were suitable, and if the things followed proper syntax and vocabulary usage. Items that did not pass face validity testing were revised as required. Concurrently, regarding content validity and reliability, the examination was given to 30 students in an intermediate reading class of 18 females and 12 men. The Quizizz program was utilized for this purpose. This program was selected based on its user-friendly interface, efficient management capabilities, and rapid score analysis (Rahmad et al., 2019).

Additionally, it was picked for its high level of user involvement and accuracy (Zhao, 2019). The material's validity and reliability were evaluated through internal consistency analysis. Statistical analyses were used to identify items that were both valid and reliable. In contrast, known test item development techniques were followed to reconstruct things that were found to be invalid and unreliable.

This rigorous approach aimed to create an online reading assessment for intermediate English as a Foreign Language (EFL) students in Indonesia that is both relevant to their culture and successful in measuring their skills.

FINDINGS AND DISCUSSION

An analysis of the course description was performed in the preliminary study to determine the student's learning objectives in reading and the test goals. The study's findings were then used to guide the construction of the test blueprint. The expert validation testing and tryout were carried out as the blueprint was built. The following parts describe the results of the course description analysis, expert validation testing, and tryout.

Course Description Analysis

We selected an intermediate reading course offered as part of the English Language Education major at a public university in East Java, Indonesia. The course was chosen because it covers literary and informative text categories commonly found on standardized reading tests. In addition, the course text types serve two specific functions of reading: to engage with literature (Duke & Carlisle, 2011) and to gain and apply knowledge (Murphy et al., 2009). The test was then generated, which included several steps, including a) deciding the test purpose, b) analyzing the test goals according to the course outline to construct the test blueprint, which includes the instructional objectives, test indicators, and the targeted learning outcomes, c) creating the test items based on the blueprint, d) reviewing and revising the constructed test items, and e) getting expert validation for the test items.

The findings indicate that seven competencies should be developed as instructional indicators for students' reading ability. These seven abilities are identifying keywords and underlying themes, making inferences, interpreting dictions, figures of speech, and culture-bound expressions, examining juxtaposition, examining structure and development of ideas, classifying text types, and appraising short stories. However, only five of those seven skill categories could be further developed into reading tests since categorizing text genres and appraising short stories are intended to be utilized in teaching and assessed through students' performance as a part of productive skill performance. As demonstrated in Table 1, the intermediate reading course focuses on two primary purposes that comprise the majority of reading undertaken by intermediate language level students: reading for literary experiences (Duke & Carlisle, 2011) and reading to gain and apply knowledge (Murphy et al., 2009).

Table 1. The Course Goals, and Instructional Objectives

Course Goals

The students are able to read and comprehend popular articles and short stories and develop their reading strategies by identifying keywords, making inferences, analyzing figures of speech, analyzing dictions, analyzing juxtaposition, analyzing the organization, and development of ideas, identifying text types, and appreciating short stories

Reading Purposes

- Literary experience
- Acquire and use information

Instructional Objectives

At the end of the lesson, the students are able to:

1. identify keywords and the underlying theme,
2. make inferences from the content of a text,
3. interpret dictions, the figure of speech, and culture-bound dictions,
4. analyze juxtaposition in a text,
5. analyze organization and development of ideas (the development of the plot of a text)

It should be underlined that reading goals and purposes and instructional objectives reflecting comprehension processes in integration do not operate independently (Mullis et al., 2016).

Test Blueprint Development

Table 2 displays the indicators for the intermediate reading test and the construction of the test blueprint by referring to the course goal and instructional objectives. Based on the blueprint, we developed 40 items of classroom-based learning-oriented tests for an intermediate reading course as a product. The test consists of forty items, each with ten test item indicators and a time restriction of 60 minutes. Additionally, there are five alternative responses for each item. Because the test items are intended to evaluate students' higher-order thinking skills, they span comprehension, analysis, and application skills. Furthermore, each item scores 2.5 for the correct answer and 0 for the incorrect answer.

Table 2. The Test Blueprint

Instructional Objectives	Indicator of Test Items	Material	Test Format	Bloom's Taxonomy Classification	Number of Test Items & Item Weight	Question Number
At the end of the lesson, the students are able to:	Given a short story, students are able to:	Literary text in the form of short story	Multiple choice			
identify keywords and the underlying theme	identify the underlying theme of the story			comprehension	3 (2.5 points each)	1, 2, & 31
make inferences from the content of a text	make inferences from what the characters say or do and by how the author portrays them			comprehension	3 (2.5 points each)	8, 32, & 35
interpret dictions, figure of speech, and culture-bound dictions	interpret dictions of rhetorical questions and part of speech of the text			application	4 (2.5 points each)	4, 5, 34, & 36
analyze juxtaposition in a text	analyze juxtaposition in a text			analysis	5 (2.5 points each)	3, 6, 7, 33, & 39
analyze organization and development of ideas (the development of the plot of a text)	analyze organization and development of ideas (the development of the plot of a text)			analysis	5 (2.5 points each)	9, 10, 37, 38, & 40
At the end of the lesson, the students are able to:	Given an expository text, students are able to:	Informational text in the form of expository text	Multiple choice			
identify keywords and the underlying theme	identify underlying theme of a particular paragraph in the text			comprehension	4 (2.5 points each)	11, 17, 21 & 26
make inferences from the content of a text	make inferences from particular words, phrases, or sentences in the text			comprehension	5 (2.5 points each)	13, 15, 19, 22, & 24
interpret dictions, figure of speech, and culture-bound dictions	interpret dictions, illustrations, graphics, and the figure of speech of the text			application	3 (2.5 points each)	12, 14, & 18, 27 & 28
analyze juxtaposition in a text	analyze juxtaposition in a text			analysis	3 (2.5 points each)	20, 23, & 25
analyzes organization and development of ideas (the development of the plot of a text)	analyze organization and development of ideas of the text			analysis	3 (2.5 points each)	16, 29, & 30
Total Items					40 items	

Expert Validation Test Results

Validity is one of the factors to consider when developing a test. The developed test should be capable of measuring the aspect to be assessed (Fatimah et al., 2018). Two EFL instructors from Indonesian institutions were involved in the validation process. Experts provided analyses, comments, and recommendations for improving the developed classroom-based reading test items. Experts appraised three aspects: materials, construction, and rhetorical devices. The experts' validation results are demonstrated in Table 3.

Table 3. The expert Validation Results

Aspects	Expert 1	Expert 2
Materials	The texts used were appropriate for the course outline.	The texts used were appropriate for the course outline and student level.
Test Construction	The composition of the distractors (concerning sentence length) in several items required revision.	The order of the items should be adjusted depending on the difficulty level and location of the clue for the answers in the text.
Grammar/ language structure	Some items' sentence construction required to be simplified.	The phrasal and clausal complexity of Reading 4 should be simplified.

In addition to the validation results, experts recommend using the online method for some considerations, such as the possibility to randomize the item order and answer choices, the ability to review and change answers before submitting the test, the more appealing visuals and program features, as well as the immediate report for the test results and feedback. The experts' recommendation to use an online method for administering the test is consistent with previous research (e.g., Ming & Aziz, 2019; Raine, 2018; Spivey & McMillan, 2014; Sudarwati et al., 2021), which found that technological tools in assessment provide beneficial impacts for both students and teachers in terms of effectiveness and efficiency.

The test that was created has been deemed to meet the requirements for both content and language. This was achieved by adjusting the test items based on expert feedback and addressing the specific areas noted in the feedback.

Trial Results

After having expert validation, a tryout was conducted online using the Quizizz application. The Quizizz assessment medium was chosen because it provides simple test development, control or supervision, an accurate final score, and quick analysis (Rahmad et al., 2019; Zhao, 2019). Furthermore, the features included in Quizizz applications are easily accessible and usable by students with no prior experience. Thirty students from the intermediate reading class were chosen to participate in the tryout. The validity and reliability tests and the difficulty level of the test items were accessed. The validity of the test is determined through the processing of the data using the SPSS version 25 program. Table 4 shows the first trial result.

Table 4. First Trial

1	Moderate	0.080	Not significant	Revised
2	Easy	0.367	Significant	Used
3	Moderate	0.445	Significant	Used
4	Moderate	0.504	Very significant	Used
5	Moderate	0.465	Very significant	Used
6	Moderate	0.525	Very significant	Used
7	Moderate	0.493	Very significant	Used
8	Moderate	0.627	Very significant	Used
9	Easy	0.445	Significant	Used
10	Moderate	0.464	Very significant	Used
11	Moderate	0.585	Very significant	Used
12	Moderate	0.617	Very significant	Used
13	Moderate	0.719	Very significant	Used
14	Easy	0.897	Very significant	Used
15	Moderate	0.250	Not significant	Revised
16	Moderate	0.687	Very significant	Used
17	Moderate	0.686	Very significant	Used

18	Moderate	0.412	Significant	Used
19	Moderate	0.421	Significant	Used
20	Moderate	0.593	Very significant	Used
21	Moderate	0.297	Not significant	Revised
22	Moderate	0.729	Very significant	Used
23	Moderate	0.618	Very significant	Used
24	Moderate	0.522	Very significant	Used
25	Moderate	0.352	Not significant	Revised
26	Easy	0.706	Very significant	Used
27	Moderate	0.922	Very significant	Used
28	Moderate	0.947	Very significant	Used
29	Moderate	0.653	Very significant	Used
30	Easy	0.717	Very significant	Used
31	Moderate	0.705	Very significant	Used
32	Moderate	0.636	Very significant	Used
33	Moderate	0.563	Very significant	Used
34	Easy	0.811	Very significant	Used
35	Moderate	0.728	Very significant	Used
36	Moderate	0.725	Very significant	Used
37	Moderate	0.862	Very significant	Used
38	Easy	0.463	Very significant	Used
39	Moderate	0.571	Very significant	Used
40	Moderate	0.605	Very significant	Used

Table 4 shows four test items categorized as low validity since the correlation scores are more or equal to the reliability of Table > 0.361 (from 30 participants). Those questions were number 1, 15, 21, and 25. Since a test must be reliable and valid (Ozdemirl & Akyol, 2019), the four question numbers were eliminated and replaced with new questions based on the indicators specified in the blueprint. Meanwhile, the other significant test items proceeded with reliability analysis. The result of the reliability analysis is 0.955, which is in the high category. This high-reliability rating denotes that the test results are more consistent or reliable.

Following the initial tryout, we adjusted the test items based on the results of the validity and reliability tests. We not only revised the invalid items but also altered the sequence of the items and modified the linguistic structure based on the test complexity (Yu, 2021). Structuring the sequence of the presentation of the test items from easy to difficult is vital to provide the test takers a pleasant impression when they first access the test and not frustrate the students at the beginning of the test. Then, we had the second tryout. Table 5 displays the results of the second tryout.

Table 5. Second Trial

1	Moderate	0.401	Significant	Used
2	Moderate	0.384	Significant	Used
3	Moderate	0.464	Very significant	Used
4	Difficult	0.502	Very significant	Used
5	Moderate	0.458	Significant	Used
6	Moderate	0.531	Very significant	Used
7	Moderate	0.485	Very significant	Used
8	Moderate	0.633	Very significant	Used

9	Moderate	0.440	Significant	Used
10	Moderate	0.471	Very significant	Used
11	Moderate	0.590	Very significant	Used
12	Moderate	0.614	Very significant	Used
13	Moderate	0.723	Very significant	Used
14	Difficult	0.892	Very significant	Used
15	Moderate	0.386	Significant	Used
16	Moderate	0.694	Very significant	Used
17	Difficult	0.682	Very significant	Used
18	Moderate	0.418	Significant	Used
19	Moderate	0.414	Significant	Used
20	Moderate	0.602	Very significant	Used
21	Moderate	0.375	Significant	Used
22	Moderate	0.725	Very significant	Used
23	Moderate	0.619	Very significant	Used
24	Moderate	0.502	Very significant	Used
25	Moderate	0.417	Significant	Used
26	Moderate	0.708	Very significant	Used
27	Moderate	0.919	Very significant	Used
28	Difficult	0.940	Very significant	Used
29	Moderate	0.656	Very significant	Used
30	Moderate	0.711	Very significant	Used
31	Moderate	0.705	Very significant	Used
32	Moderate	0.639	Very significant	Used
33	Moderate	0.548	Very significant	Used
34	Moderate	0.814	Very significant	Used
35	Moderate	0.728	Very significant	Used
36	Moderate	0.727	Very significant	Used
37	Difficult	0.859	Very significant	Used
38	Moderate	0.468	Very significant	Used
39	Moderate	0.568	Very significant	Used
40	Moderate	0.608	Very significant	Used

Table 5 shows that all the item questions were valid, and most of them were categorized as a very significant category. In detail, 21 test items are classified as very substantial, and the rest are significant since all the correlation scores are more or equal to the reliability Table > 0.361 (from 30 participants). Furthermore, the reliability result of the analysis is 0.955, which is in the high category. This means that all items are reliable. The higher the reliability rating, the more consistent or reliable the test results. This study demonstrated that the classroom-based learning-oriented test for intermediate reading generated with the Quizizz application was reliable and valid. Therefore, all test items can be used in the test. Distractors' consistency and complexity are essential in tasks that assess misunderstandings and critical thinking. The quantity and quality of mistakes students make while responding to test items are critical for developing the assessment blueprint and information about students' comprehension and ability to accomplish the instructional objectives.

Reflection and Implications: Towards the Development of Students' Comprehension and the Practice

As an online classroom-based assessment, we accommodated two reading purposes in the test items mandated by the course outlines: reading for literary experiences (Duke & Carlisle, 2011) and reading to

acquire and use information (Murphy et al., 2009). Furthermore, the reading materials included literary and informational text types (Nisa and Helmanda, 2019) that incorporated social context (Snow, 2002). As a result, teachers can immediately assess whether or not their students have met the learning objectives. The findings also demonstrate that the created examination successfully fulfills the primary purpose of an evaluation, which is to enhance teaching and learning procedures by providing feedback on the outcomes.

Moreover, constructivism acknowledges the student's active role in the learning process. Classroom-based reading assessment effectively combines teaching and assessment by aligning the curriculum with students' contexts, fostering their development (Yang & Qian, 2022). Considering that learning activities, including assessment and the thinking that goes with them, are deeply ingrained in specific cultural practices and norms, it is argued that applying classroom-based assessment principles to students' learning needs and contexts is beneficial for assessment teachers, practitioners, and assessment scholars. It is required for the ongoing development of frameworks for classroom-based evaluation.

Interestingly, even though the exam has been seen to be valid and reliable, some students earned low points. This might be because, due to the COVID-19 pandemic, some students did not participate in profound reading activities during their online learning (Grabe, 2009), lowering their reading skills. Furthermore, some students may be unwilling to practice their reading comprehension independently at home while participating in remote learning during the pandemic. To confirm this conclusion, we conducted informal interviews with students about their development of reading abilities throughout the pandemic. Their answers highlighted their lack of independent reading comprehension learning at home. Students emphasized that teachers' support is required when reading. Therefore, they do not do it effectively when they are left to learn independently at home.

According to the findings, we also propose in this study that relevant, contextual, and classroom-based assessment be utilized to accurately exhibit students' reading skills (Andronova, 2021) because it meets the instructional objectives. Furthermore, analyzing data from students' scores can confirm the effectiveness of classroom-based learning-oriented reading assessment, and it is known that proper classroom-based learning-oriented reading assessment can benefit teachers and students. Classroom-based learning-oriented reading assessment not only measures students' performance and progress in their learning but also analyzes the effectiveness of the instructions, including teaching strategies or methods.

Furthermore, both students and teachers regarded the use of technology in test delivery positively because having an online test delivery method allows them to access and finish the test efficiently (Raine, 2018), especially during emergencies such as the Covid-19 pandemics (Sudarwati et al., 2021). Furthermore, the teachers and students who took the test also discussed the advantages of using an online test delivery technique. According to teachers and students, online tests provide flexible testing schedules during which the test was administered (Spivey & McMillan, 2014). Furthermore, the teachers emphasize that automatically randomizing the order of questions and answers for Quizizz's multiple-choice option reduces the chance of student cheating while simultaneously providing fast feedback and reports. As a result, it can be stated that the online classroom-based reading test using the Quizziz application is practical for testing students' reading comprehension.

CONCLUSION

Ultimately, the generated online classroom-based reading test conducted using the Quizizz program proves to be a reliable assessment of students' reading skills. This may be ascribed to the fact that it is highly relevant to the context, fitting perfectly with the demands of students and the goals of instruction. As a result, it produces precise depictions of their comprehension and proficiency in the curriculum. The test effectively examines the literary aspects and information acquisition in reading comprehension. It provides diagnostic insights into students' strengths and weaknesses in different reading sub-skills. The exam satisfies the urgent need for a systematic, thorough, and instructional diagnostic reading evaluation by meeting these requirements.

Moreover, the study emphasizes the broader influence of efficient assessments conducted within the classroom. These evaluations reveal insights into individual students' performance and development and offer helpful feedback on the effectiveness of teaching strategies and approaches. This dual functionality is

crucial in continuously enhancing education, benefiting students, teachers, practitioners, and academics. Adopting online classroom-based assessment features customized to students' language and cultural settings is essential to advance assessment frameworks, as learning and assessment are closely intertwined with these contexts. The online text form also correlates with the students' technologically dependent characteristics and is beneficial regarding efficacy and efficiency.

Although the research has produced significant benefits, it lacks constraints. Further research is required to explore the intricate complexity of the online classroom-based test, driven by the emphasis on validity, reliability, and diagnostic value. Gaining a deep understanding of the complexities in designing tests for intermediate reading comprehension in a classroom setting would provide a more thorough viewpoint, enhancing the ongoing discussion on successful assessment methods. Thus, this study is a foundation for future research efforts to improve and progress classroom-based evaluation systems.

BIODATA AND CONTACT ADDRESSES OF AUTHORS

Eko SUHARTOYO is one of faculty members at English Language Education, Faculty of Teacher Training and Education, Universitas Islam Malang, Malang, East Java, Indonesia. He is currently a Ph.D candidate at English Language Education, Faculty of Letters, Universitas Negeri Malang, Malang, East Java, Indonesia. His doctoral study is funded by Indonesian Education Scholarship (Beasiswa Pendidikan Indonesia), from the Ministry of Education and Culture of the Republic of Indonesia (Kemendikbudristek) and Indonesia Endowment Fund for Education (LPDP). His research interests are mainly in the area of critical thinking in writing, teaching writing and feedback provision, and ICT based Learning.

Eko SUHARTOYO
English Language Education, Faculty of Letters, Universitas Negeri Malang
Address: Jl. Semarang 5, 65145, Malang, Indonesia
Phone: +6281232752666
E-mail: eko.suhartoyo.2102219@students.um.ac.id; suhartoyoeko@unisma.ac.id

Dr. Rida AFRILYASANTI is an English teacher at SMA Negeri Taruna Nala, East Java, and holds a Ph.D. in English Language Education from Universitas Negeri Malang, Indonesia. She participated in a student exchange program at Humboldt State University, California, during her undergraduate studies. She also completed a language enrichment program at the Regional Language Center (RELC) in Singapore as part of her master's degree. Dr. Afrilyasanti has authored national English textbooks for Indonesian schools and a globally published handbook on teaching critical media literacy. Her research interests include technological pedagogy, pedagogical innovation, digital cultures, English language teaching, and media literacy.

Rida AFRILYASANTI
English Language Education, Faculty of Letters, Universitas Negeri Malang
Address: Jl. Semarang 5, 65145, Malang, Indonesia
Phone: +6281235778278
E-mail: rida.afrilyasanti.2102219@students.um.ac.id; afrilyasanti@gmail.com

Dr. Nur MUKMINATIEN is a Professor in English Language Teaching at Universitas Negeri Malang, Indonesia. Her publications are in the area of teacher education, teaching methods, and language education. Her expertise is Teaching English as a Foreign Language (TEFL), assessment in EFL Writing, and English language teaching methodology (TEFL).

Nur MUKMINATIEN
English Language Education, Faculty of Letters, Universitas Negeri Malang
Address: Jl. Semarang 5, Malang 65145, Indonesia
Phone: +62811303138
E-mail: nur.mukminatien.fs@um.ac.id

REFERENCES

Agota, S. (2020). *Reading English for academic purposes: The reading processes of first-year EFL learner BA student.* [Unpublished doctoral dissertation]. Eotvos Lorand University.

Andronova, E.S. (2021, April 9-10). *Assessment techniques for ESP students' reading skills* [Conference session]. *Selected papers from assessment in teaching foreign language: assessments, challenges, and solutions international conference* (pp. 55-62). Tomkhet. https://conference.fledu.uz/wp-content/uploads/sites/6/2021/11/horijiy-tillarni-oqitishda-baholash-1.pdf#page=55.

Barrot, J. S., Llenares, I. I., & Del Rosario, L. S. (2021). Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines. *Education and Information Technologies*, *26*(6), 7321–7338. https://doi.org/10.1007/s10639-021-10589-x

Borg, W.R. & Gall, M.D. (1983). *Educational research: An introduction (5th Eds.).* Longman.

Brown, H.D. (2004). *Language assessment: Principles and classroom practices.* Pearson Education.

Cahyono, B. Y., Istiqomah, F., Fitriyah, I., & Gozali, I. (2023). EFL teachers' voice on their preferred strategies in teaching EFL writing during the pandemic: Investigating the role of technology. *Turkish Online Journal of Distance Education 24*(3), 330-350. https://doi.org/10.17718/tojde.1175925

Duke, N. K. & Carlisle, J. (2011). The development of comprehension. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P.P. Afflerbach (Eds.). *Handbook of reading research* (Vol. IV, pp.199-228). Routledge.

Ekpang, J.N., Ghevolor, A., & Bassey, V. (2021). Literacy, oral language, and rhetoric. *Journal of Languages, Linguistics and Literary Studies*, *10*(1), 9-17.

Fatimah, A.W.N., Suryani, N., & Yamtinah, S. (2018). The development of critical thinking test based on higher-order thinking PISA version in the historical learning at senior high school. *International Journal of Multicultural and Multireligious Understanding*, *5*(2), 136-144. https://doi.org/10.18415/ijmmu.v5i2.132

Fensham, P. J., & Bellocchi, A. (2013). Higher order thinking in chemistry curriculum and its assessment. *Thinking Skills and Creativity*, *10*, 250–264. doi:10.1016/j.tsc.2013.06.003

FitzPatrick, B., Hawboldt, J., Doyle, D., & Genge, T. (2015). Alignment of learning objectives and assessments in therapeutics courses to foster higher-order thinking. *American Journal of Pharmaceutical Education*, *79*(1), 10. doi:10.5688/ajpe79110

Fulcher, G., & Davidson, F. (2012). Introduction. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 1-19). Routledge.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, Vol. 3, pp. 403–422). Lawrence Erlbaum Associates Publishers.

Hanafi, A. (2016). Developing reading comprehension test for the first semester students of English department. *Journal of English Language, Literature and Teaching*, *1*(1), 57-74.

Hoi, V. N. (2021). *Constructing a validity argument for a locally developed test of English reading proficiency.* [Unpublished Doctoral dissertation]. Queensland University of Technology.

Johnson, A. M., Jacovina, M. E., Russell, D. E., & Soto, C. M. (2016). Challenges and solutions when using technologies in the classroom. In S. A. Crossley & D. S. McNamara (Eds.) *Adaptive educational technologies for literacy instruction* (pp. 13-29). New York: Taylor & Francis. Published with acknowledgment of federal support.

Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*(3), 281-300. https://doi.org/10.1080/10888430802132279

Khabir, M., Jabbari, A. A., Fazilatfar, A. M., & Razmi, M. H. (2022). Flipped presentation of IELTS reading: Impacts of grit, autonomy, and reading achievement in an EFL context. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ)*, *23*(2), 178-197.

Koksal, D. ,& Ulum, O.G. (2018). Language assessment through Bloom's taxonomy. *Journal of Language and Linguistic Studies*, *14*(2), 76-88.

Kostaki, D. & Karayianni, I. (2022). Houston, we have a pandemic: Technical difficulties, distractions and online students engagement. *Student Engagement in Higher Education Journal, 4*(2), 105-127, Retrieved from https://sehej.raise-network.com/raise/article/view/1063

Lu, X., Sales, A., & Heffernan, N. T. (2021). Immediate versus delayed feedback on online learning: Do people's instincts really conflict with reality? *Journal of Higher Education Theory and Practice*, *21*(16). https://doi.org/10.33423/jhetp.v21i16.4925

Maplethorpe, L., Kim, H., Hunte, M. R., Vincett, M., & Jang, E. E. (2022). Student-generated questions in literacy education and assessment. *Journal of Literacy Research, 54*(1), 74-97. https://doi.org/10.1177/1086296X221076436

McGrath, I. (2016). *Materials evaluation and design for language teaching (2nd Eds.).* Edinburgh University Press.

Ming, T. S., & Aziz, N. B. A. (2019). Using eye tracking to investigate Malaysian ESL students' reading processes and preferences. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ)*, *20*(3), 92-116.

Mullis, I. V.S., Martin, M.O., & Sainsbury, M. (2016). *PIRLS 2016 Reading Framework (2nd Eds.).* TIMSS & PIRLS International Study Center.

Murphy,P.K., Wilkinson, I.A.G., Soter, A.O., Hennesssey, M.N., & Alexander, J.F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740-764.

Newton, P.M., & Essex, K. (2023). How common is cheating in online exams and did it increase during the Covid-19 pandemic? A systematic review. *Journal of Academic Ethics* https://doi.org/10.1007/s10805-023-09485-5

Nisa, R & Helmanda, C.M. (2019, September 18-19). *The classification of reading comprehension test items.* Proceedings of the 2nd English education international conference (EEIC) in conjunction with the 9th annual international conference (AIC), *2*(2019), 137-141. http://jurnal.unsyiah.ac.id/EEIC/article/view/15250

Ozdemirl, E.C. & Akyol, H. (2019). The development of a reading comprehension test. *Universal Journal of Educational Research, 7*(2), 563-570. https://doi.org/10.13189/ujer.2019.070229

Rahmad, N., Lestari, A., Musa, L. A. D., & Sugilar, H. (2019, July). Quizizz online digital system assessment tools. In *2019 IEEE 5th International Conference on Wireless and Telematics (ICWT)* (pp. 1-4). IEEE.

Raine, P. (2018). Developing web-based English learning applications: Principle practice. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ)*, *19*(2), 125-138.

Santiago, C. J., Ulanday, M. L., Centeno, Z. J., Bayla, M. C., & Callanta, J. (2021). Flexible learning adaptabilities in the new normal: E-learning resources, digital meeting platforms, online learning systems and learning engagement. *Asian Journal of Distance Education*, 16(2). Retrieved from https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/580

Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL internet-based test. *Language Testing*, *26*(1), 5-30.

Schwartz, S. (1984). *Measuring Reading Competence.* Plenum Press.

Setyarini, S. (2019, May 15). *Teachers' understanding in constructing higher order thinking-based assessment: voice from English teachers' experience*. In Proceeding of the 4th Asian Education Symposium (AES 2019) (pp. 39-42). https://doi.org/10.2991/assehr.k.200513.009

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Education.

Spivey, M. F., & McMillan, J. J. (2014). Classroom versus online assessment. *Journal of Education for Business, 89*(8), 450–456. https://doi.org/10.1080/08832323.2014.937676

Sudarwati, E., Fatimah, Astuti, Y., & Ubaidillah, F. (2021). Developing online learning assessment instrument for English sentence structure course during Covid-19 pandemic. *Langkawi: Journal of the Association for Arabic and English, 7*(2), 170-181. http://dx.doi.org/10.31332/Ikw

Tengberg, M. (2018). Validation of sub-constructs in reading comprehension tests using teachers' classification of cognitive targets. *Language Assessment Quarterly*, *15*(2), 169-183.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J.V. Wertsch (Eds.), *The concept of activity in Soviet psychology* (pp. 144-188). Sharpe.

Walvoord B.E., & Anderson V.J. (1998). *Effective grading: A tool for learning and assessment*. Jossey-Bass.

Yang, Y., & Qian, D. D. (2022). Enhancing EFL learners' reading proficiency through dynamic assessment. *Language Assessment Quarterly, 20*(1). 20-43. https://doi.org/10.1080/15434303.2022.2132160

Yu, X. (2021). Text complexity of reading comprehension passages in the national matriculation English test in China: The development from 1996 to 2020. *International Journal of Language Testing, 11*(2), 142-167.

Zhao, F. (2019). Using Quizizz to integrate fun multiplayer activity in the accounting classroom. *International Journal of Higher Education, 8*(1), 37-43. https://doi.org/10.5430/ijhe.v8n1p37