The Analysis of Reliability an English Subject Final Test and the Factors Affecting Result at the Tenth Grade of MAN 1 Kebumen in the Academic Year of 2021/2022

Setiya Cahya Ningsih[1], Abdul Ngafif[2], Ismawati Ike Nugraheni[3]
{ setyacahyan1898@gmail.com[1], abdulngafif@gmail.com[2], ikewardhana@gmail.com }

English Language Education, Universitas Muhammadiyah Purworejo, Indonesia[123]

Abstract. The purpose of this study was to discover the reliability of the English final test and to describe the factors that affect the reliability test results of students at MAN 1 Kebumen. This research has a mixed method. This research involved a mix of quantitative and qualitative approaches. The instruments used by researchers to obtain data were documentation, closed questionnaires, and open interviews. the result of the final English Reliability test was 0.821. This means that the reliability of the English summative test is in a very high classification (0.80-1.000). Judging from the results of the questionnaire, the factors that affect the reliability of the test are related to the variety of people, the composition of the test items, the variation in the administration of test instructions, and the variation in the test environment. From the results of interviews, the factors of test reliability are mastery of learning materials, namely preparation for studying before facing exams, strategies for facing exams, physical condition of students, clarity of exam instructions, environmental conditions of exams and time allocation.

Keywords: English Subject, Final Test, Reliability.

1. Introduction

English is the most widely spoken language in the world, and it is used for communication in many areas of life, such as politics, social media, education, economics, and culture [1]. Furthermore, a large number of nations, including Singapore, South Africa, Belize, the Philippines, Hong Kong, Ireland, Canada, Nigeria, and others, have adopted English as one of their official languages. International organizations including the United Nations and the International Olympic Committee also recognize it as an official language English is the common tongue. English is the most widely studied of the seven languages learned worldwide, studied by 1.5 billion people, and the official language of instruction in 42 countries (the others being French, Mandarin, Spanish, and German). Italian and Japanese are two of the 839 languages that are spoken in 60 countries across the world. 4. No country has disputed that English is currently the de facto language used throughout the ASEAN area since 2009 [2].

Teachers have to work hard to make sure their students have good English skills. All learning processes and content that the teacher has given to students and whether students can understand what is being taught. After giving the material, the teacher should give a test to see how well the students understand it.

A test is a tool for measuring a person's skill, knowledge, or performance in a particular area [3]. Test, on the other hand, a test is typically defined as a tool or instrument of measurement used to collect data on a certain attribute or characteristic of a person or group. The researcher concludes that a test is one of the evaluation techniques used by teachers and consists of a series of tasks or things that must be answered by students. How well students do on an assessment determines how well it is. Assessment is a way to obtain information from students' learning processes to achieve learning goals. assessment is a comprehensive assessment and involves team members to determine weaknesses and strengths and the results of the decision can be used for educational services that children need as a basis for developing learning designs. The assessment can determine the success of a learning program [4]. The reason for giving exams to students is to determine students' knowledge and give appropriate grades as well as evaluate less effective teaching. This can also provide information to students about mastery of learning material [5].

The teacher's evaluation is utilized to determine the extent to which the pupils' attitudes have improved, in accordance with the National Standard of Education and Regulation of Education and Culture Number 23 of 2016. It is completed in accordance with the processes for process maintenance, teaching and learning evaluation, learning improvement, and subsequent improvement. Assessments are used by the teacher to accomplish test objectives, which include monitoring student progress and assigning grades sets [6]. According to Brown, there are two types of tests: standardized tests and tests created by teachers. A test that is developed by the teacher utilizing the lesson plan and curriculum that they have been using all semester is known as a teacher-made test. Its goal is to assess the degree to which pupils, following the teacher's instruction, meet the curriculum's objectives. As a result, while deciding which questions to ask, educators must use logic and reason. This test is typically administered as part of daily, formative, and summative assessments. There are numerous test-related processes that are frequently employed in the test program. The assessment consists of two parts, namely test and non-test. Several types of tests are written tests, oral tests, practical tests, and assignments. Therefore, the test is one of the instruments needed to evaluate. Minimum test quality standards must be met in order for the test to be valid and reliable. It is important to determine the validity and reliability of an assessment. The objective of these examinations is to generate an assessment that is of high quality. The reliability test is a test that proves that, despite the fact that the assessment is tested repeatedly, it will still yield scores that are relatively consistent.

As a measurement stage, tests on Final Semester Examination questions are the starting point for measuring student abilities. A tool that can be used to measure something using predetermined rules is called test [7]. Based on its preparation, there are two types, namely first, tests carried out by teachers, tests that are not standardized, and tests carried out by individuals who meet the requirements. The effectiveness of this type of test depends on the teacher's skills and abilities in designing the test. The second is a standardized test with carefully defined administration and scoring procedures to ensure uniformity to achieve broader goals (http://www.britannica.com). In this sense, standardized tests are tests conducted by certain educational institutions or departments. Teacher-made tests have two types, namely subjective tests and objective tests. Subjective tests have two types are description tests or essays and short

answer tests. Meanwhile, objective tests have 3 types of objective tests, including: true false (TrueFalse), which is a test that asks students to determine the answer from one of the choices, true or false, matching (Matching).

Although it was first restricted to higher education levels, English has long been an essential component of Indonesian education. As a result, the curriculum for teaching English has changed as the language's demand and popularity increased. In the Independent Curriculum, as decreed by Permendikbudristek No. 262/M/2022, English is one of the compulsory subjects and optional subjects at the national secondary education level, namely: Junior High School (SMP) and High School (SMA), while English is only local content at the school level the basis of Indonesia [8]. The teaching of English as a local or extracurricular activity in elementary schools, shows that education levels cannot be separated from education in Indonesia. English lessons are compulsory subjects in all majors at MAN 1 Kebumen. One of them which is class 10 MIPA the time and schedule of English lessons is ninety minutes per week. There are three tests: one at the end of the material, one in the middle, and one at the end. The Final Semester Exam is one that will be seen.

In this research, the researcher intends to analyze the reliability results and describe the factors that influence the reliability of the MAN 1 Kebumen English test. Based on the research background, researchers have two objectives that they want to accomplish through this research. The aim is to analyze the reliability of the reliability results and analyze the factors that influence the reliability of the MAN 1 Kebumen English test.

Tests are an important element in the learning process. Test are one of the evaluation techniques used by teachers and consist of a series of tasks or things that students must answer. A test is a tool to measure a person's skills, knowledge, or performance in a particular area. It is an instrument for assessing students' skills or knowledge to make educational decisions [9]. This means that tests are a useful tool for assessing a person's achievements and abilities [10]. The function of tests is to determine the effectiveness of teaching, assess student abilities, evaluate student self-evaluation, evaluate learning programs, materials and textbooks [11]. Therefore, tests are important to make education better. From there we can determine what we will do with students, teaching and learning methods, teachers, and curriculum.

A test needs to meet the requirements of a quality test. It needs to be legitimate and dependable [12]. Three qualities make a test good: validity, reliability, and practicality. Gamper further classifies a variety of testing criteria into five categories: practicality, validity, backwash, authenticity, and reliability. According to these sources, a test's quality can be assessed using particular metrics. Giving students a high-quality exam gives them the opportunity to learn, and the results can enhance instruction and learning while also having an impact on students' grades [13].

Reliability is one of the most essential aspects of exams in general and language tests in particular [14]. The degree to which test scores are free of measurement error is the definition of reliability [15]. Reliability as "measurement consistency" In other words, a language exam is dependable if whatever it assesses is measured consistently [16]. In actuality, an unreliable test is worthless. A measure is considered reliable if a person's score on the same test given twice is similar. It means that identical measurements get the same results [17]. It is trustworthy if the

instructor administers the same tests to the same candidates on many occasions and the results are comparable [18].

Test-retest reliability as a measurement of stability over time. The degree to which a measurement of a phenomenon yields steady and predictable findings is known as reliability [19]. That is, A test is considered more dependable if the scores of students who take it in different places or under different conditions are similar [18].

According to [20], in this short book on applied statistics in education, test/instrument reliability is influenced by four factors, the first is the wording of test items. Generally in a test, the easy test items are placed in the beginning and followed by the difficult ones. Difficult items if placed in the beginning may induce nervousness among the learners. This affects their performance and in turn influences reliability. The second is Variations in the test instructions, this factor explains the unclear instructions for carrying out the test, error-question, by the test maker so that it makes it difficult for students to determine the correct answer and even students can answer wrongly even though they know the correct answer. The third is Variations in the test environment, means that affect the learner due to the environment may temporarily be profound on the day of the test and in turn increase or decrease the reliability coefficient significantly. such as temperature, humidity, noise, and comfortable seating. The last is Variations in a person, this factor arise from the physical and psychological conditionstaking a test due to illness, pain, medication, tiredness, or stress.

Similar to what was quoted in Heaton stated that the five factors that influence test reliability are first, the breadth of the material sample selected [21]. The larger the, the more likely it is that the test as a whole will be reliable. This means that the more test questions there are, the more reliable the test is. Therefore, objective tests are preferred, which allow a wide range of coverage. The second is the dilemma in administering the test, if test conditions fluctuate from one administration to another then reliability will be affected. The third is personal factors, referring to individual aspects of students' physical and mental health, such as weariness, bad health, lack of motivation or interest, anxiety, and melancholy. The fourth is test instructions; a clear set of instructions will increase the reliability of the test findings. The final one is assessment freezing, which refers to how subjectivity in evaluation can result in inconsistent assessment and inaccurate measurements.

The similarity of this research and the previous research is that all research analyzes the quality of English tests made by teachers. Furthermore, the similarities between the two studies lie in their objectives. The difference between this research and previous research is that this research focuses on analyzing factors that influence test reliability, different from previous studies which examined the validity and reliability of English multiple choice tests in midterm, final and midterm exams.

2. Method

Definition of mixed research methods (Mixed Methods) is research that involves a mixture of quantitative and qualitative approaches. In general, mixed methods research is done

when both quantitative and qualitative data are present. There are two types of data: qualitative data (such as open-ended questionnaires and interviews) and quantitative data (such as scores or numbers that can be statistically analyzed) [22]. The quantitative data used in this study is the outcome of the tenth grade English final exam. While the questionnaire and interview produced the qualitative data. In this study, the researcher used documentation tools to describe the reliability of the English final test given to students. The documentation method involves searching for information about something in books, newspapers, magazines, transcripts, notes, and soon. Researchers provide in-depth details regarding the factors influencing the reliability of teacher English assessments in this study. Fifty multiple-choice examinations created by the MAN 1 English teacher served as the study's focus. Students in the class served as the research subjects. The test for the final semester exam is a multiple-choice test. The researcher asked the teacher for permission to take test result data carried out by the teacher himself. After receiving permission, the researcher then called the teacher to give the students' work papers to be analyzed.

In analyzing the reliability of English subject final tests, then enter the scores into SPSS, calculate the reliability results using Cronbach's Alpha, categorize the reliability results, explain the reliability results, then draw conclusions. Researchers used closed questionnaires and in-depth interviews to find out and describe the factors that influence the reliability test results at MAN 1 Kebumen. In providing a closed questionnaire, the researcher asked permission from the English teacher and principal at each school to obtain data using questionnaires and interviews. In analyzing interviews, researchers see and hear interview recordings, identify them, categorize and explain the factors that influence reliability tests, then draw conclusions. The researcher gave instructions on how to complete the questionnaire in class before handing it to the students. After that, the researcher sent the link to the Google Form, requested the students to complete it, and downloaded the findings. When conducting a questionnaire analysis, researchers review the responses, classify and explain that affect test reliability, and then make conclusion.

3. Findings

In this section, the researcher discussed the result of the research. The researcher used documentation, questionnaire and interview as the instruments of the research. The documentation instrument is used to analyze the reliability of English final test of tenth grade of MAN 1 Kebumen.

Table.1 The Number of Students
Case Processing Summary

	N	%
Cases Valid	29	100.0
Excluded[a]	0	.0
Total	29	100.0

Table.2 The Result of Cornbach's Alpha
Reliability Statistics

Cronbach's Alpha	N of Items
.821	50

From table 1 and table 2, it gives information that the number of students doing the test are 29 students. Then, from the output of SPSS 16 computation available on table 2, it can be known that the result of Cronbach's Alpha is 0.821 with the total number of items (questions) 50.

Table.3 The Value of r Product Moment

N	5	1
2	0.3	0.4

The english fina l test at tenth grade of MAN 1 Kebumen in the academic year of 2021/2022 is reliable 1%. According to the table above, the coefficient of reliability is higher than the value of r product moment (0.821 > 0.467).

According to Sugiyono (2016) by seeing the categories of coefficient in interpretation of r-value, the coefficient of reliability of English test at tenth grade in MAN 1 Kebumen is high. The interval coefficient of reliability (0.80 – 1000) is very high.

4. Discussion

Discussion is the most important part in this research. To know the factors of affecting reliability of the test, the researcher was using questionnaire and interview. To arrange the statement of the questionnaire and the question of interview, the researcher was using theory from Sutton (2019). The theory is the basis for compiling questionnaires and interviews. The second instrument is questionnaire. These two instruments were used to fit out each other's result in order to get the more valid data. The questionnaire is consisted of 20 items statements. Here is the result of data collected from student's :

Table 1. The Factors Affecting Reliability of a Test

No	Statement	Strongly Disagree	Disagree	Agree	Strongly Agree
	Variation in a person				
1	When I am sick I cannot do the test well	0%	6.9%	58.6%	34.5%
2	When I am tired, I cannot do the test well.	3.4%	13.8%	58.6%	24.1%
3	When I am sad, I cannot do the test well	6.9%	24%	51.7%	17.2%

4	When I am anxious, I cannot do the test well	13.8%	10.3%	51.7%	24.1%
5	When I am angry, I cannot do the test well.	10.3%	44.8%	24.1%	20.7%
	The arrangement of test items				
6	I cannot do the test well when the test item is too much	6.9%	31%	48.3%	13.8%
7	I cannot do the test well when the test is too long	20.7%	48.3%	24.1%	6.9%
8	I cannot do the test well ehen the test more than one answer	10.3%	48.3%	27.6%	13.8%
9	I cannot do the test well in a very limited time	3.4%	10.3%	44.8%	41.4%
10	I cannot do the test well when the font size is too	10.3%	17.2%	48.3%	24.1%
11	I cannot do the test well when the background of google/papper	10.3%	20.7%	41.4%	27.6%
	Variation of administrator test instructions				
12	When my teacher gives unclear instructions in the test, it will influence my score	10.3%	20.7%	44.8%	24.1%
13	When my teacher gives unclear scoring criteria , it will influence my score.	3.4%	20.7%	51.7%	24.1%
14	When my teacher gives ambiguous queston, it will influence my score	3.4%	24.1%	48.3%	24.1%
	Variation in test environment				
15	I cannot do the test well when the weather is too hot	3.4%	24.1%	55.3%	17.2%
16	I cannot do the test well when the internet access is not good	0%	20.7%	48.8%	34.5%
17	I cannot do the test well when there is too much noise	3.4%	6.9%	51.7%	37.9%
18	I cannot do the test well when the place is not comfortable	6.9%	13.8%	48.3%	31%
19	When my teacher treats students differently, it will influence my score	10.3%	34.5%	41.4%	13.8%
20	When my teacher dislikes me, it will influence my score	6.9%	37.9%	41.4%	13.8%

Based on the result of table 1 above, the researcher can find the percentage of questionnaire score of factors affecting the reliability in English final test:

a. Statement 1 asks the students about the effect of sick in doing the test. Based on the result of questionnaire, it can be seen that the percentage of questionnaire score Strongly Disagree (SD) is 0%, Disagree (D) is 6.9%, Agree (A) is 58.8%, and Strongly Agree (SA) is 34.5%. The percentage is dominantly Agree (58.8%). So, it can be concluded that there is the effect of sick in doing the test.

b. Statement 2 asks the students about the relationship of tired in doing the test. Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 3.4%, Disagree (D) is 13.8%, Agree (A) is 58.6%, and Strongly Agree (SA) is 24.1%. The percentage is dominantly Agree (58.6%). So, it can be concluded that there is effect of tired in doing the test.

c. Statement 3 asks the students about the effect of sad in doing the test.
Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 6.9%, Disagree (D) is 24%, Agree (A) is 51.7%, and Strongly Agree (SA) is 17.2%. The percentage is dominantly Agree (51.7%). So, it can be concluded that there is a sad effect on doing the test.

d. Statement 4 asks the students about is there any relationship between being anxious and doing the test.
Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 13.8%, Disagree (D) is 10.3%, Agree (A) is 51.7%, and Strongly Agree (SA) is 23.3%. The percentage is dominantly Agree (51.7%). It means that anxious affects in taking the test.

e. Statement 5 asks the students about the relationship feeling angry and doing the test. Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 44.8%, Agree (A) is 24.1%, and Strongly Agree (SA) is 20.7%. The percentage that has the highest score is Disagree (44.8%). Thus, only several students (24.1%) agree that anger has effect on doing the test.

f. Statement 6 asks the students about influence of too much test items. Based on the result, the percentage of score Strongly Disagree (SD) is 6.9%, Disagree (D) is 31%, Agree (A) is 48.3% and Strongly Agree (SA) is 13.8%. The percentage is dominantly Agree (48.3%). It can be concluded that the test items are too much is affecting the result.

g. Statement 7 asks the students about the effect of the test time is too long on test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 20.1%, Disagree (D) is 48.3%, Agree (A) is 24.1%, and Strongly Agree (SA) is 13.8%. The percentage is dominantly Disagree (48.3%). Only several students agree that if the time of the test is too long, so it is affecting the result of the test. Thus, still can be concluded that the time of the test is too long can affect the result of the test.

h. Statement 8 asks asks the students about the effect of the test has many answers on test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 17.2%, Agree (A) is 48.3%, and Strongly Agree (SA) is 24.1%. The percentage is dominantly Disagree (48.3%). But still several students agree that if the test has any answers, it will be affecting the result of the test.

i. Statement 9 asks asks the students about the effect of limited time on test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 3.4%, Disagree (D) is 10.3%, Agree (A) is 44.8%, and Strongly Agree (SA) is 41.4%. The percentage is dominantly Agree (44.8%). Thus, it may be said that the limited time of the test is affecting the result of the test.

j. Statement 10 asks the students about the effect of the font size is too small on test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 17.2%, Agree (A) is 48.3%, and Strongly Agree (SA) is 24.1%. The percentage is dominantly Agree (48.3%). It can be concluded that small size font is affecting the result of the test.

k. Statement 11 asks the students about the effect of quality of printer ink is low on test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 20.7%, Agree (A) is 41.4%, and Strongly Agree (SA) is 27.6%. The percentage is dominantly Agree (41.4%). It can be concluded that the low quality of printer ink is affecting the result.

l. Statement 12 asks students about the relationship of unclear instructions with influence of scoring. Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 20.7%, Agree (A) is 44.8%, and Strongly Agree (SA) is 24.1%. The percentage that has the highest score is Agree (44.8%). It means that unclear instructions will influence the students score.

m. Statement 13 asks the students about the unclear scoring criteria and influence of scoring. Based on the result, the percentage of score Strongly Disagree (SD) is 3.4%, Disagree (D) is 20.7%, Agree (A) is 51.7% and Strongly Agree (SA) is 24.1%. The highest percentage is Agree (51.7%). It can be concluded that unclear scoring criteria is affecting the students score.

n. Statement 14 asks the students about the ambiguous questions and influence of scoring. Based on the result, the percentage of score Strongly Disagree (SD) is 3.4%, Disagree (D) is 24.1%, Agree (A) is 48.3%, and Strongly Agree (SA) is 24.1%. The highest percentage is Agree (48.3%). It means that ambiguous questions are affecting the students score.

o. Statement 15 asks the students about the effect of the hot weather. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 3.4%, Disagree (D) is 24.1%, Agree (A) is 55.2%, and Strongly Agree (SA) is 17.2%. the percentage is dominantly Agree (55.2%). It means that the hot weather is affecting the result of the test.

p. Statement 16 asks the students about the effect of bad internet access of test result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 0%, Disagree (D) is 20.7%, Agree (A) is 48.8%, and Strongly Agree (SA) is 34.5%. The percentage is dominantly Agree (48.8%). It means that the bad internet access is affecting the result of the test.

q. Statement 17 asks the students about the effect of noisy class result. Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 3.4%, Disagree

(D) is 6.9%, Agree (A) is 51.7%, and Strongly Agree (SA) is 37.9%. The percentage is dominantly Agree (51.7%). It means that the noisy classroom is affecting the test result.

r. Statement 18 asks the students about the effect of uncomfortable place for testing on test result.Based on the result, it can be seen the percentage of score Strongly Disagree (SD) is 6.9%, Disagree (D) is 13.8%, Agree (A) is 48.3%, and Strongly Agree (SA) is 31%. The percentage is dominantly Agree (48.3%). It means that the uncomfortable place for testing is affecting the result of the test.

s. Statement 19 asks students about the unfair treatment by the teacher with influence of scoring.Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 10.3%, Disagree (D) is 34.5%, Agree (A) is 41.4% and Strongly Agree (SA) is 13.8%. The percentage is dominantly Agree (41.4%). It can be concluded that several students still agree that unfair treatment by teacher is influence the score.

t. Statement 20 asks students about the correlation between the teacher who does not like her/him and it is infecting on her/his score.

Based on the result, it can be seen that the percentage of score Strongly Disagree (SD) is 6.9%, Disagree (D) is 37.9%, Agree is 41.4% and Strongly Agree is 13.8%. The highest percentage is Agree (41.4%). It means that if the teacher dislikes the students, it will be affecting the score of the student.

According to the data were collected from questionnaire, the factors that affecting reliability of English test at tenth grade of MAN 1 Kebumen are as follows :

a. **Variations in a person.**

Factors that influence student-related reliability include physical and psychological conditions. can be physically healthy but psychologically unstable because of anger, sadness about something. Besides that students also lack strategies in working on questions. As a result, it interferes with concentration and causes difficulties for students (Sutton, 2019)

Based on the result of questionnaire, tenth grade students of MAN 1 Kebumen agree that sick (58,8%), tired (58,6%), sad (58,6%), anxious (51,7%), and angry (51.7%) are affecting on taking the test. Thus, students cannot do the test well. It can be concluded that physical and psychological condition such as, sick, tired, sad, anxious and angry are affecting the reliability of test.

b. **The arrangement of test item**

This factor talking about the arrangement of the items, the placement of the questions in the easy and difficult categories. Placing difficult questions at the beginning, makes students nervous, or makes students think mentally it is very difficult and even makes students focus on difficult questions so they spend a lot of time [20].

In this factor is talking about the media tools used such as paper, clear writing or not and test arrangements. Based on the result of the questionnaire, the tenth grade students of MAN 1 Kebumen agree that if the test item is too much (48.3%), the test is too long (48.3%), the test has more than one answer (48.3%), the test well in a very limited time (44.8%), and the teacher

gives ambiguous questions (48.3%), the test successfully when Google's history (41.4%) influences their grade. It may therefore be affecting the test's reliability. Based on the explanation above, the arrangement of the test as a medium and tool when carrying out the test, paper / google background, writing, too many questions, too many questions, questions that are too long all affect the reliability of the test.

c. Variation of administrator instructions

According to [20] the clarity of test instructions is a very important factor and affects the reliability results. the ambiguity of the test can result in student confusion in determining answers. such as multiple answers, instructions on how to fill in answers, ambiguous questions, and unclear scoring criteria. If the test items are not paraphrased carefully examinees/learners face difficulty in understanding them. This leads to confusion and learners may answer incorrectly despite knowledge of the correct answer.

Based on the result of the questionnaire, the students agree that if the teacher gives unclear instruction (44.8%), the teacher gives unclear scoring criteria (51.7%), and the teacher gives ambiguous questions (48.3%) are influence their score. Therefore, it can be affecting the reliability of the test.

The conclusion is human error, namely the test maker himself. Examples are giving unclear instructions, giving unclear assessment criteria, and giving ambiguous questions.

d. Variation in test environment

According to [20] the environment contributes to influencing reliability and can significantly increase or decrease the coefficient. such as temperature, humidity, atmosphere, noise, media, and teacher subjectivity make students uncomfortable.

Based on the questionnaire result, the students agree that, the weather is too hot (55.3%), the internet access is not good (48.8%), the classroom is noisy (51.7%), the place is not comfortable (48.3%), the teacher treats student differently (41.4%) and the teacher dislike me, it will influence my score, can affect the students cannot do the test well.

According to the explanation above, The reliability of a test can be affected by factors such as noisy environment, uncomfortable classroom, hot weather, long test items, and poor internet access.

Based on the results of the interviews, the factors that influenced the results of the final English test for class X MAN 1 Kebumen were, first, the lack of maximum study preparation before the exam. most of the students got low grades. They stated the importance of study preparation. Based on the results of the interviews, they were not optimal in preparing for learning. so that they do not master the learning material to be tested.

Second mastery of learning material by students. of the first factor affects the second factor. students who are less prepared to learn and not optimal, have an effect on the mastery of learning material. almost all students stated that they did not master the material well and were not confident in taking exams. so students have difficulty in doing the test. Thus, it means that student problems such as a lack of strategies for facing tests can be a factor affecting reliability.

The three conditions of students during the exam. Physical and mental psychic is very important in the exam. the majority of students stated that being physically unwell had a negative impact on doing exams, causing them to be unable to think properly and to be unable

to focus because of the pain. This means that the condition of students affects the reliability of the test.

Fourth, the variety of instructions. Students stated that if the instructions and test questions were clear, it would affect their grades. Clear instructions help students understand the questions, so they can answer questions more easily. If the instructions are not clear, it will be difficult for students to answer the questions. We can conclude that the quality of the instructions has an effect on the test's dependability.

The fifth variations of the test environment. Based on the results of interviews, class conditions during exams affect student concentration. if the class is not quiet, noisy then it is difficult to focus on thinking. conversely if the class is quiet and comfortable then students are not disturbed while taking the exam. thus the environmental conditions when the test can affect the results of the reliability test.

The sixth time allocations. Students' time availability significantly impacts test reliability. Mastery of learning materials and preparation for exams are key factors influencing test reliability. strategies for facing tests, physical condition of students, clarity of exam instructions, environmental conditions of exams and time allocation.

Based on the explanation of the results of the analysis the researcher explained that there were several differences between this study and previous related findings. Septia and Mir'atul did not focus on reliability tests but focused on content validity. While Ummul got quite reliable results as well as this research. The difference in this study focuses on the results and factors that affect reliability. Researchers used questionnaires, closed questionnaires, and open interviews to analyze the factors influencing reliability.

5. Conclusion

Based on data analysis and data interpretation, in the previous chapter the researcher concluded that the reliability of English final test results was 0.821%, which means that the test made by the English teacher of each school is reliable because it is in a high level classification (0.60-0.799). Judging from the results of the questionnaire, the factors that influence test reliability are variation in person, variation of administrator instructions, the wording of test item, and variation in test environment. The teacher should pay attention to the quality of the English test to be tested on students, and the teacher must also evaluate the test after completion. Because as teachers, we must provide a better and better teaching and learning process. One of them is by making a good quality test. Therefore students should prepare themselves and develop study strategies to face the English test, and students need to ensure that they are in good health for the exam.

6. References

[1] S. Nasution, "Penggunaan Bahasa Inggris Secara Maksimal Untuk Menyongsong Masyarakat Ekonomi Asean (Mea) 2025," *Bisnis-Net J. Ekon. dan Bisnis*, vol. 1, no. 1, 2018.

[2] I. Iriance, "Bahasa Inggris sebagai bahasa lingua franca dan posisi kemampuan Bahasa Inggris masyarakat Indonesia diantara anggota MEA," in *Prosiding Industrial Research workshop and national seminar*, 2018, pp. 776–783.

[3] M. Pangaribuan, "The Effect of Reciprocal Teaching Technique on Students ' A bility in Reading Comprehension at English Department of FKIP UHN Pematangsiantar," *JETAFL (Journal English Teach. as a Foreign Lang.*, vol. 5, no. 1, pp. 26–36, 2019.

[4] Y. Apsari and A. Haryudin, "THE ANALYSIS OF ENGLISH LECTURERSâ€™ CLASSROOM-BASED READING ASSESSMENTS TO IMPROVE STUDENTSâ€™ READING COMPREHENSION," *Eltin J. J. English Lang. Teach. Indones.*, vol. 5, no. 1, pp. 35–44, 2017.

[5] I. H. Wenno, D. Tuhurima, and Y. Manoppo, "How to Create a Good Test," *J. Pendidik. Profesi Guru Indones.*, vol. 1, no. 1, pp. 11–20, 2021.

[6] A. Setiyawan, "Faktor-Faktor yang Mempengaruhi Reliabilitas Tes," *AN NUR J. Stud. Islam*, vol. 6, no. 2, 2014.

[7] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta, 2012.

[8] dan T. Kementerian Pendidikan, Kebudayaan, Riset, "Merdeka Belajar."

[9] N. Hartati and H. P. S. Yogi, "Item analysis for a better quality test," *English Lang. Focus*, vol. 2, no. 1, pp. 59–70, 2019.

[10] A. Fazlina, "An analysis of college entrance test," *English Educ. J.*, vol. 9, no. 2, pp. 192–215, 2018.

[11] F. H. Saragih, "Testing and assessment in English language instruction," *J. Bahas Unimed*, vol. 27, no. 1, p. 74656, 2016.

[12] D. Furwana, "Validity and reliability of teacher-made English summative test at second grade of vocational high school 2 Palopo," *Lang. Circ. J. Lang. Lit.*, vol. 13, no. 2, 2019.

[13] U. Khaer, "Analyzing Validity and Reliability of English Summative Test Made by The English Teacher at Second Grade of Vocational High School 2 Palopo." State Islamic Institute Palopo, 2018.

[14] R. E. Chiedu and H. D. Omenogor, "The concept of reliability in language testing: issues and solutions," *J. Resour. Distinct.*, vol. 8, no. 1, pp. 1–9, 2014.

[15] H. Moses, R. N., & Yamat, "Testing the Validity and Reliability of a Waiting Skill Assessment," *Int. J. Acad. Res. Bussiness Soc. Sci.*, vol. 11, no. 4, 2021.

[16] C. Chapelle, *Reliability in Language Assessment. In The Encyclopedia of Applied Linguistics.* Oxford: Blackwell/wiley, 2013. [Online]. Available: http://lib.dr.iastate.edu/engl_pubs/207

[17] H. Taherdoost, "Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research," *How to test Valid. a Quest. a Res. (August 10, 2016)*, 2016.

[18] A. A. Muhammad, "Validity, reliability and practicality of the first certification in English (FCE) and the Business Language Testing Service (BULATS)," *IDEAS J. English Lang. Teach. Learn. Linguist. Lit.*, vol. 6, no. 2, 2018.

[19] R. C. Rosaroso, "Using reliability measures in test validation," *Eur. Sci. J.*, vol. 11, no. 18, 2015.

[20] G. W. Sutton, *Applied Statistics Concepts For Counselors*. Sunflower, 2019.
[21] I. Tosuncuoglu, "Importance of Assessment in ELT.," *J. Educ. Train. Stud.*, vol. 6, no. 9, pp. 163–167, 2018.
[22] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2014.