JoTELL *Journal of Teaching English, Linguistics, and Literature* published by English Education Study Program, Faculty of Languages and Arts, Universitas Negeri Manado, Vol. 4 No. 4, pp. 2302-2330
Note: Tahoma, 10 pt, Leave the Vol and No Blank.

Improving Students' Reading Skills Through Group Work in SMP Negeri 2 Tondano

NOFLIN PILENDATU, NOLDY PALENKAHU, NIHTA LIANDO

Universitas Negeri Manado
Correspondance author: tessanoflin11@gmail.com

Received: 26 Maret 2025
Accepted: 15 April 2025
Published: 24 April 2025

Abstract : This study investigated the effectiveness of group work in improving students' reading comprehension at SMP N 2 Tondano. The research was motivated by the observation that many students struggled with reading English texts and demonstrated low comprehension abilities. Using a pre-experimental quantitative design, the researcher administered a pre-test and post-test to a sample of the available 18 out of 30 seventh-grade students due to their tardiness. The intervention involved group work activities that aimed to foster collaboration and active participation during reading lessons. Data were collected through multiple choice, matching, and completion tests, and were analyzed for reliability and validity. The findings showed a significant improvement in students' reading comprehension after the implementation of group work, as indicated by higher mean scores in the post-test compared to the pre-test. Statistical analysis using a paired-samples t-test confirmed that this improvement was statistically significant. The results suggested that group work was an effective strategy for enhancing reading skills among junior high school students. However, the study also found that some students continued to face challenges, indicating the need for differentiated instruction and ongoing support. Overall, the research highlighted the value of collaborative learning in the English classroom and recommended the integration of group work strategies to improve reading comprehension.

Keywords: ***Reading comprehension, group work, collaborative learning, improving reading comprehension.***

INTRODUCTION

Education means to achieve justice, both individual justice and social justice. According to (Plato, 2008), individual justice can be obtained when each individual

develops his or her ability to the fullest. In this sense, justice means excellence. For the Greeks and Plato, excellence is virtue.

Learning English as a foreign language allows you to fully appreciate the culture and its context of a country, it widens your understanding. By understanding a person's culture, for example, you can avoid situations which could cause misunderstandings. Being able to understand other cultures and through learning a foreign language you can bridge the gap between cultures, which would be a powerful tool in today's modern world.

In global world the importance of English cannot be denied and ignored since English is the most common language spoken everywhere. People often talk about English as a global language. With more than 350 million people around the world speaking English as a first language and more than 430 million speaking it as a second language, there are English speakers in most countries around the world.

For the teaching of English to be successful, the four skills, reading, listening, speaking and writing, should be integrated in an effective way. When a teacher makes use of activities that have been specially designed to incorporate several language skills simultaneously, such as reading, writing, listening, and writing (Liando et al., 2021), they provide their students with situations that allow for well-rounded development and progress in all areas of language learning.

Reading was the ability to understand the material being read. In the area where English was considered as a foreign language, teaching reading skill became a vital point for the English teachers. Moreover, it is also supported by (Karisi et al., 2021) whom stated that reading is a process of getting a meaning from a text. Many experts asserted that reading was viewed as an active skill rather than a passive skill, which involves constant brain raking activity (Iwai, 2016) Some important reading skill strategies are making connections, visualization, organization, determining important information, asking question and monitoring comprehension (Moos, 2010). Additionally, (Wood, 1991) confirms that reading improved vocabulary, and a larger

vocabulary improved reading skill. Furthermore, reading could also improve grammar mastery which in result readers to read sensibly.

In addition, reading is a very important need for students, so there is a need for efforts to pleasure in reading (Tatipang et al., 2021). According to (Urquhart & Weir, 2014) reading skill can be described as "a cognitive ability which a person is able to use when interacting with the text". Moreover, reading is a crucial academic skill, especially for students, as it forms the foundation for acquiring knowledge and developing critical thinking. Therefore, creating a sense of pleasure and motivation in reading is essential (Tatipang et al., 2021). In line with (Grabe & Stoller, 2019), reading encourages learners to engage deeply with the material by making predictions, asking questions, and summarizing information while reading. This technique fosters critical thinking and better retention.

Reading comprehension is a skill to understand the information contained in a reading text. It is in line with (Lendo et al., 2021) whom stated that reading comprehension is when someone is able to comprehend the meaning of what is being read. Based on the statement, reading is an important skill that someone must have when it comes to reaching full understanding of a certain text. Where reading comprehension also aims to not only read for fun but to enable someone to fully understand the meaning and the context behind a reading text (Kasiha et al., 2022). It can be concluded that reading comprehension is a required skill, especially for students in order to obtain information and fully understand about the information given in the text according to certain context.

Research by (Iwai, 2016) and (Moos, 2010) shows that using reading strategies like making connections and visualizing content significantly improves students' comprehension, especially in EFL contexts. Moreover, collaborative learning methods such as group work have been found to increase student motivation and comprehension in reading tasks (Ghaith, 2003).

There are many people or learners who find English reading, difficult to understand the reading and have less interest towards it. Problem in reading is a

problem all learners who learn English foreign language and it also seem to be a problem for students at SMP N 2 TONDANO.

After the researcher conducted teaching practice Based on PPL 2 at SMP N 2 TONDANO, the researcher found most of learners who still lack of attention in English teaching process, especially in reading class. Most students lacked attention during the English teaching process, especially in reading classes. They have many problems in reading such as the students read very slowly, they faced on problem reading as they could not do well to understand the reading text and their lack of knowledge that is related to the text. Students struggled with reading fluently and understanding texts, often reading very slowly and showing low motivation. In addition, they were doubted to read a text loudly since they were not familiar in reading fluently and always boring if they surface with the reading material.

This kind of problem was also identified by several researchers, (Ahmada, 2020) identified that the reading comprehension of the first semester class in English Education Department of Islamic Institute of Darussalam Blokagung Banyuwangi in Academic Year 2018/2019 was low. One also used group work strategy in order to increase their reading comprehension. Furthermore, (Damanik & Herman, 2021) also found similar problem where they tried to find out the influence of Question Answer Relationship Strategy (QARS) on students' reading comprehension ability in SMP YPK Pematangsiantar. The result indicated a significant improvement in the experimental group's post-test scores compared to their pre-test scores, suggesting that QARS effectively enhanced reading comprehension.

For solving the problem in reading, the teacher is required to be innovative in choosing suitable technique. They also must be able to create pleasant atmosphere in the classroom in order to motivate students in reading English text. During this time, many ways that have been applied by teacher to improve students' reading comprehension like skimming and scanning method, and also reading aloud. But those are not enough to increase students' ability. The students still need another way to touch their memory or mind. In this case, the researcher offers to conduct a

research through group work using descriptive text as the text type where according to (Bachtiar & Kurniawan, 2005), descriptive text is a text, which says what a person or a thing is like. Its purpose is to describe and reveal a particular person, place, or thing.

Based on explanation above, the researcher was interested in doing research about "Improving Student Reading Skill through Group Work at SMP N 2 Tondano.

RESEARCH METHOD

The method used in this research is quantitative. Quantitative research is a type of research where the data was structured numerically Matthew & Ross in (Tumengkol et al., 2022). The type of research was pre-experimental research where the subject of the research only consisted of one group (Sugiharto et al., 2022). In pre-experimental research study use pre-test and post-test design. Pre-test is a test that given to students before the trial treatment was given to students or before the material to find out the students learning achievement. While post-test is a test that given after the material to see if student achievement can improve (Kasiha, Hampp & Kumajas, 2022). According to (Hatch & Farhady, 1982) 'Pre-test and post-test designs are similar to one occasion case study". So there are 2 tests, T1 is pre-test and T2 is Post Test.

This research was conducted in class SMP N 2 Tondano, which is located at Jl. Gn. Agung Rinegetan, Tondano Barat, Minahasa, Sulawesi Utara in the odd Semester 2023/2024. The sample of this research was the student in class VII, which consisted in 30 students.

The instrument in this research was a test. The test was containing 25 questions in the form of multiple choice, 5 match making questions, and 5 fill in the blank questions. The test is given based on the material. The test is given in the pre-test and post-test.

Data was collected through pre-test and post-test in the form of student worksheet (LKS) to see and compare scores on T1 and T2.

The data collection technique used by the writer in this research used a Pre-Test and Post-Test adapted from Ngadiso 2013 in (Dadara & Sudiyono, 2021). The steps (procedures) for data collection to be carried out are as follows:

1. Pre-Test

Pre-Test were carried out before treatment. The pre-test was conducted to determine the students' reading skill before using group work strategy or group work assignment. This part of data collection was where writer provide question sheets that students did with a total of 35 questions divided into 25 multiple-choice numbers, 5 match making numbers, and 5 completion numbers based on the reading material.

2. Treatment

   The next step is giving treatment, in this case the writer used the group work strategy, to determine students' reading skill achievement after using group work assignment.

3. Post-Test

The last step was giving the same test as the pre-test to determine the difference after giving the treatment.

In analysing the data, the researcher used SPSS. The formula used, according to (George & Mallery, 2003) are as follows:

1. Cronbach's Alpha:

   If the value of Cronbach's Alpha is > 0.90, the questionnaire is reliable.

   If it is between 0.70 and 0.90, the reliability is high.

   If it is between 0.50 and 0.70, the reliability is moderate.

2. Product Moment Validity Test:

If the Pearson correlation coefficient ≥ the critical value (two-tailed test with sig. 0.05), the instrument or items are significantly correlated with the total score and are considered valid.

3. Cronbach's Alpha Reliability Test:
   If the value of Cronbach's Alpha is > 0.60, the questionnaire or survey is considered reliable or consistent.

$$\underline{x} = \frac{\sum \quad x}{n}$$

FINDINGS AND DISCUSSION

In this chapter, the researcher presented the results of the data analysis conducted on the pre-test and post-test instruments. However, from the listed 30 students, only 18 of them were able to be taken as the samples, as some of them rarely came to school. The analysis aimed to evaluate the validity and reliability of the instruments, as well as to assess the students' performance and progress in reading skills. The findings included statistical summaries, correlation analyses, and reliability tests that provided insight into the effectiveness of the test items and the overall measurement quality. These results served as the foundation for interpreting the impact of the instructional intervention and for drawing conclusions about the students' learning outcomes.

Pre-Test Result

Pre-Test Mean Score & Cronbach's Alpha

Before conducting the treatments, the researcher administrated the pretest to identify the ability of the students before they got the treatments. As the result, the mean score of pretest was 37.30, the highest score was 68.57 with 24 correct answers; and the lowest score was 14.29 with only 5 correct answers. There were only 4 students who scored above 60, 1 student score between 40 and 60 and the rest were below 40. From the pre-test score it can be seen that the students possessed low ability in reading comprehension. Looking at the fact that the average was only 37.30 out of 100 means that there is a significant lack of understanding of

the material prior to the treatment. Furthermore, the large gap between the highest and lowest scores (68.57 and 14.29) indicates a wide range of pre-existing reading abilities in the classroom, with most of the students struggling significantly. The score of each student can be seen in the table below along with SPSS result of the mean.

Table 1: Students' Pre-Test Scores

No	Students' Initial	Total Correct Answer (n)	Score ((n)/35)x100
1	**A.T.**	24	68.57143
2	**N.W.**	13	37.14286
3	**G.P.**	15	42.85714
4	**K.M.**	12	34.28571
5	**R.M.**	23	65.71429
6	**R.M**	11	31.42857
7	**J.M**	11	31.42857
8	**K.S**	12	34.28571
9	**G.D.**	7	20
10	**C.E**	9	25.71429
11	**J.M.**	5	14.28571
12	**Y.L**	10	28.57143
13	**O.T.**	6	17.14286
14	**A.K**	13	37.14286
15	**S.K.**	10	28.57143
16	**O.P.**	21	60
17	**R.E.**	22	62.85714
18	**M.M**	11	31.42857
	Pre-Test Mean Score $\underline{x} = \frac{\sum x}{n}$		37.3016

Picture 1. Pre-Test mean Score

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ScorePerStudent	18	14.29	68.57	37.3016	16.49976
Valid N (listwise)	18				

From the data in the pre-test, it can be observed that the students' reading abilities varied widely before the treatment was administered. The minimum score recorded was 14.29, indicating that some students answered only 5 out of 35 questions correctly, reflecting a very low initial level of reading comprehension. On the other hand, the maximum score was 68.57, with the highest-performing student correctly answering 24 questions, showing that a few students already had a relatively good grasp of the material. The mean score of 37.30 suggests that, on average, students were able to answer about one-third of the questions correctly, which points to an overall low to moderate level of reading proficiency across the group.

Furthermore, the standard deviation of 16.50 indicates a considerable spread in the students' scores around the mean. This high variability reveals that while some students performed fairly well, many others struggled significantly, creating a broad range of abilities within the class. Such a wide distribution emphasizes the need for targeted instructional strategies that can address the diverse learning needs of the students. These statistics collectively highlight the importance of the forthcoming treatment to improve reading comprehension skills and to help raise the lower-performing students closer to the average, while also challenging those who are already performing well. Furthermore, the reliability of the pre-test can be seen from the images below.

Pre-Test Reliability Test

Picture 2: Pre-Test Reliability Test

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.835	.831	34

Picture 3: Question 1 was not reliable due to all students answered correctly

Warnings

Each of the following component variables has zero variance and is removed from the scale: Q1

The reliability analysis of the pre-test instrument yielded a Cronbach's Alpha of 0.835, indicating a high level of internal consistency among the test items. This suggests that the instrument reliably measures the students' reading ability as a whole. The Cronbach's Alpha based on standardized items was similarly high at 0.831, confirming the consistency of the results.

Although the instrument originally contained 35 items, the reliability analysis included only 34 items. This is because Question 1 (Q1) exhibited zero variance, with all students answering it correctly. An item with zero variance does not differentiate between students' abilities, as it provides no variability in responses. Consequently, statistical software like SPSS automatically excludes such items from the Cronbach's Alpha calculation since they do not contribute to the measurement of internal consistency.

Despite this exclusion in the analysis, Q1 is retained in the instrument because it serves as a baseline question that all students can answer, possibly to build confidence or assess fundamental knowledge. Its presence does not negatively impact the overall reliability of the test, as evidenced by the strong Cronbach's Alpha value.

In summary, the pre-test instrument is reliable and appropriate for assessing students' reading skills. The exclusion of Q1 from the reliability calculation is a statistical necessity due to its lack of variability, but it remains an integral part of the test content.

Pre-Test Pearson Correlation

Pearson Correlation Analysis is a statistical method used to measure the strength and direction of the linear relationship between individual test items and the total test score. In the context of a pre-test, this analysis helps determine the validity of each item-whether it effectively measures the intended construct and contributes meaningfully to the overall assessment. Items with higher Pearson correlation coefficients and statistically significant p-values (typically $p < 0.05$) indicate strong and reliable relationships with the total score, suggesting that these items are valid indicators of student ability or knowledge. Conversely, items with low or non-significant correlations may not align well with the construct and could require revision or removal. This process ensures that the test is both reliable and valid, providing accurate measurement of the targeted skills or knowledge areas. Below is the table for Pre-Test Pearson Correlation:

Table 2: Pre-Test Pearson Correlation

Correlation	**Pearson Correlation Score**	**Probability (Sig. 2 tailed)**	**Conclusion**
Q1 with total	.a	-	No variability since all students answered correctly
Q2 with total	.429	.075	Not valid and not statistically significant
Q3 with total	.587*	.011	Valid and statistically significant

Q4 with total	.597**	.009	Valid and statistically significant
Q5 with total	.196	.436	Not valid and not statistically significant
Q6 with total	.474*	.047	Valid and statistically significant
Q7 with total	.474*	.047	Valid and statistically significant
Q8 with total	.155	.540	Not valid and not statistically significant
Q9 with total	.066	.794	Not valid and not statistically significant
Q10 with total	.259	.299	Not valid and not statistically significant
Q11 with total	.060	.813	Not valid and not statistically significant
Q12 with total	.492*	.038	Valid and statistically significant
Q13 with total	.536*	.022	Valid and statistically significant
Q14 with total	.281	.258	Not valid and not statistically significant
Q15 with total	-.713**	.001	Valid and statistically significant
Q16 with total	.311	.208	Not valid and not statistically significant
Q17 with total	.500*	.035	Valid and statistically significant

Q18 with total	-.066	.793	Not valid and not statistically significant
Q19 with total	.784**	.000	Valid and statistically significant
Q20 with total	.668**	.002	Valid and statistically significant
Q21 with total	.090	.723	Not valid and not statistically significant
Q22 with total	.364	.138	Not valid and not statistically significant
Q23 with total	.757**	.000	Valid and statistically significant
Q24 with total	.613**	.007	Valid and statistically significant
Q25 with total	.066	.794	Not valid and not statistically significant
Q26 with total	-.111	.662	Not valid and not statistically significant
Q27 with total	.601**	.008	Valid and statistically significant
Q28 with total	.770**	.000	Valid and statistically significant
Q29 with total	.563*	.015	Valid and statistically significant
Q30 with total	.591**	.010	Valid and statistically significant
Q31 with total	.474*	.047	Valid and statistically significant

Q32 with total	.713**	.001	Valid and statistically significant
Q33 with total	.595**	.009	Valid and statistically significant
Q34 with total	.595**	.009	Valid and statistically significant
Q35 with total	.473*	.047	Valid and statistically significant

The Pearson correlation analysis between each test item and the total test score provides valuable insights into the validity and effectiveness of the items in measuring the intended construct. Most of the items demonstrate moderate to strong positive correlations with the total score, accompanied by statistically significant p-values (less than 0.05). This indicates that these items are valid and contribute meaningfully to the overall assessment. For example, items such as Q3, Q4, Q6, Q7, and several others show correlation coefficients ranging roughly from 0.47 to 0.78, highlighting their strong relationship with the total test score. These items effectively discriminate between students with different levels of understanding or ability, thereby enhancing the reliability and validity of the test.

One notable exception is item Q15, which exhibits a strong negative correlation (-0.713) that is highly significant ($p = 0.001$). This suggests that the item is reverse-coded or measures the construct in an inverse manner relative to the total score. Such items are often intentionally included in assessments to reduce response bias and ensure that students are carefully considering each question. Despite the negative sign, the strong correlation and statistical significance confirm that Q15 is a valid and important part of the test.

On the other hand, several items, including Q2, Q5, Q8, Q9, and others, show low correlation coefficients and non-significant p-values (greater than 0.05). These items do not demonstrate a meaningful relationship with the total score and are

therefore considered not valid. Their poor performance may stem from ambiguous wording, misalignment with the construct being measured, or inappropriate difficulty levels. Such items should be reviewed carefully, revised, or potentially removed to improve the overall quality of the test.

Regarding Q1, all students answered this item correctly, resulting in no variability in responses. Because there is no variation, the correlation coefficient and significance value cannot be calculated. While this item does not contribute to differentiating between students' abilities, it may serve as an introductory or easy question to build confidence or assess basic knowledge.

In summary, the majority of the test items are valid and statistically significant, supporting the overall construct validity of the assessment. However, a subset of items lacks sufficient correlation with the total score and should be reconsidered to enhance the test's effectiveness. Including reverse-coded items like Q15 adds balance and helps control for response bias. Moving forward, it is recommended to revise or remove invalid items and possibly conduct further analyses, such as factor analysis and reliability testing, to strengthen the instrument's psychometric properties.

Post-Test Result

Post-Test Mean Score & Cronbach's Alpha

After getting the pre-test scores, the researcher began the treatment by explaining what a group work strategy is, what its stages are, and the benefit to improve students' reading skills. The students then were given reading texts to be read carefully, then asked to consciously recall its content. Furthermore, they were instructed to create and present summaries using their own words to check for comprehension. The treatments were team evaluation, peer evaluation and self-evaluation of individual participation.

In the first meeting after the pre-test, the students were introduced to group work and the concept of descriptive text, then, they were divided into small groups and assigned to read a reading passage. In the second meeting, the researcher

assigned a new passage for the group discussion followed by role-playing activities in pairs to practice comprehension and communication, followed by reflecting the lesson. In the third meeting, students individually completed specific tasks such as identifying main ideas, summarizing, and vocabulary recognition, then shared their findings within groups to collaboratively summarize the passage. The treatment concluded with group presentations and discussions of challenges faced during the activities, fostering both comprehension and collaborative skills.

The researcher then proceed with the post-test after the treatment to know whether the treatment increased students' reading comprehension or not. From the result, it was found that the overall reading comprehension of the students increased. However, if the data were to be analyzed individually, the result showed that, even if the overall score increased, there were some declining scores the student made in the post-test. The scores along with the reliability statistics can be seen below:

Table 3: Post-Test Mean Score

No	Students' Initial	Total Correct Answer (n)	Score ((n)/35)x100
1	**A.T.**	24	68.57143
2	**N.W.**	13	37.14286
3	**G.P.**	15	42.85714
4	**K.M.**	12	34.28571
5	**R.M.**	23	65.71429
6	**R.M**	11	31.42857
7	**J.M**	11	31.42857
8	**K.S**	12	34.28571
9	**G.D.**	7	20
10	**C.E**	9	25.71429
11	**J.M.**	5	14.28571
12	**Y.L**	10	28.57143

13	**O.T.**	6	17.14286
14	**A.K**	13	37.14286
15	**S.K.**	10	28.57143
16	**O.P.**	21	60
17	**R.E.**	22	62.85714
18	**M.M**	11	31.42857
	Post-Test Mean Score		
	$\underline{x} = \frac{\sum \quad x}{n}$		45.2381

Picture 4: Post-Test mean Score

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Score_Post_Test	18	11.43	77.14	45.2381	22.60895
Valid N (listwise)	18				

As the result, the mean score of post-test was 45.238, the highest score was 77.14 with 27 correct answers; and the lowest score was 11.43 with only 4 correct answers. There were 6 students who scored above 60, 3 student score between 40 and 60 and the rest were still below 40. From the post-test score it can be seen that the students still possessed low ability in reading comprehension. Although the post-test results indicate that students still exhibited relatively low reading comprehension ability overall, there was a noticeable improvement compared to the pre-test scores. The increase in the average score from 37.30 in the pre-test to 45.238 in the post-test suggests that the instructional treatment had a positive effect on students' reading skills.

The standard deviation of the post-test scores was 22.61, which reflects considerable variability in student performance. This relatively high standard deviation indicates that while some students made significant progress, others continued to struggle with reading comprehension. The wide spread of scores is consistent with

the range observed in the pre-test, where the highest and lowest scores were 68.57 and 14.29, respectively, demonstrating diverse levels of reading ability within the classroom.

Looking at the pre-test data, the average score of only 37.30 out of 100 highlights a significant lack of understanding of the material prior to the treatment. The large gap between the highest and lowest pre-test scores further confirms the presence of varied reading competencies among the students, with many facing considerable difficulties. Furthermore, the reliability of the pre-test can be seen from the images below.

Post-Test Reliability Test

Picture 5: Post-Test Reliability Score

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.915	.917	35

The reliability analysis of the post-test instrument showed a Cronbach's Alpha coefficient of 0.915, indicating an excellent level of internal consistency among the 35 test items. Similarly, the Cronbach's Alpha based on standardized items was 0.917, which further confirms the stability and reliability of the test results. These high values suggest that the post-test items consistently measure the students' reading comprehension ability, making the instrument a reliable tool for assessing their progress after the instructional intervention.

Post-Test Pearson Correlation

Table 4: Post-Test Pearson Correlation

Correlation	**Pearson Correlation Score**	**Probability (Sig. 2 tailed)**	**Conclusion**
Q1 with total	.475*	.046	Valid and statistically significant
Q2 with total	.373	127	Not valid and not statistically significant
Q3 with total	.591**	.010	Valid and statistically significant
Q4 with total	.669**	.002	Valid and statistically significant
Q5 with total	.494*	.037	Valid and statistically significant
Q6 with total	.481	.043	Valid and statistically significant
Q7 with total	.436	.070	Not valid and not statistically significant
Q8 with total	.111	.660	Not valid and not statistically significant
Q9 with total	.628**	.005	Valid and statistically significant
Q10 with total	.514*	.029	Valid and statistically significant
Q11 with total	.414	.088	Not valid and not statistically significant
Q12 with total	.649	.004	Valid and statistically significant

Q13 with total	.780**	.000	Valid and statistically significant
Q14 with total	.340	.168	Not valid and not statistically significant
Q15 with total	-.397	.102	Not valid and not statistically significant
Q16 with total	.324	.169	Not valid and not statistically significant
Q17 with total	.542	.020	Valid and statistically significant
Q18 with total	.146	.564	Not valid and not statistically significant
Q19 with total	.737**	.000	Valid and statistically significant
Q20 with total	.562*	.015	Valid and statistically significant
Q21 with total	.414	.088	Not valid and not statistically significant
Q22 with total	.301	.225	Not valid and not statistically significant
Q23 with total	.585*	.011	Valid and statistically significant
Q24 with total	.707**	.001	Valid and statistically significant
Q25 with total	.417	.085	Not valid and not statistically significant
Q26 with total	.307	.215	Not valid and not statistically significant

Q27 with total	.601**	.008	Valid and statistically significant
Q28 with total	.804**	.000	Valid and statistically significant
Q29 with total	.707**	.001	Valid and statistically significant
Q30 with total	.834**	.000	Valid and statistically significant
Q31 with total	.755**	.000	Valid and statistically significant
Q32 with total	.858**	.000	Valid and statistically significant
Q33 with total	.513*	.029	Valid and statistically significant
Q34 with total	.591**	.010	Valid and statistically significant
Q35 with total	.611**	.007	Valid and statistically significant

The Pearson correlation analysis examined the relationship between each post-test item and the total test score to evaluate the validity of the individual items in measuring students' reading comprehension ability. Out of the 35 items, 23 showed statistically significant positive correlations with the total score, with coefficients ranging from moderate (.475 for Q1) to very strong (.858 for Q32). These significant correlations, with p-values less than .05, indicate that these items effectively distinguish between students with higher and lower reading comprehension skills. Items such as Q13 (.780), Q28 (.804), Q30 (.834), and Q32 (.858) demonstrated particularly strong correlations, suggesting they are highly aligned with the overall test construct and serve as excellent indicators of student ability. The prevalence of

many valid items with strong correlations supports the reliability and construct validity of the post-test instrument, confirming that these items contribute meaningfully to the overall assessment of reading comprehension after the instructional intervention.

On the other hand, 12 items exhibited non-significant correlations with the total test score, indicating that they did not align well with the test's overall construct. For example, Q2 (.373, p = .127), Q7 (.436, p = .070), Q8 (.111, p = .660), and Q15 (-.397, p = .102) showed weak or even negative correlations. The negative correlation for Q15 is particularly concerning, as it suggests this item may assess a different skill or contain scoring or content errors. Items with low or non-significant correlations may fail to differentiate effectively among students' abilities and could detract from the test's validity. Such items warrant careful review to identify potential issues such as ambiguous wording, misalignment with the tested skills, or inappropriate difficulty levels that result in uniform responses.

Several items showed moderate but statistically significant correlations, including Q1, Q5, Q6, Q10, Q17, Q23, and Q33. These items contribute positively to the test and help assess various aspects of reading comprehension, although they may not be as strong indicators as some others. Overall, the pattern of correlations suggests that the post-test is a generally valid instrument, with most items effectively measuring the intended construct. This finding aligns with the high reliability coefficient (Cronbach's Alpha = .915), indicating that the test items work well together to assess students' reading comprehension. However, the presence of several items with weak or negative correlations highlights areas for improvement. Revising or removing these items could enhance the test's validity and reliability further.

Comparison of Pre-Test and Post-Test Result

This section presents a comparative analysis of the students' performance on the pre-test and post-test to evaluate the effectiveness of the instructional intervention. The table below summarizes key descriptive statistics, including the minimum and maximum scores, mean scores, and standard deviations for both tests. Additionally, the results of a paired-samples t-test are included to determine whether the observed differences in scores are statistically significant. This comparison provides a clear picture of the students' progress in reading comprehension skills, highlighting any improvements achieved after the treatment.

The effectiveness of the instructional intervention was evaluated by comparing students' reading comprehension scores before and after the treatment using paired-samples t-test analysis. This method assesses whether the mean difference between the pre-test and post-test scores is statistically significant, indicating real improvement in student performance. The following tables present the descriptive statistics, correlation between the tests, and the results of the paired-samples t-test.

Picture 6: Paired Samples Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Score_Post_Test	45.2381	18	22.60895	5.32898
	ScorePerStudent	37.3016	18	16.49976	3.88903

The Paired Sample Statistics table shows that the average post-test score (45.23) was higher than the pre-test score (37.30), suggesting an improvement in students' reading comprehension after the intervention. The standard deviations indicate some variability in scores, with the post-test scores showing a wider spread.

Picture 7: Paired Samples Correlation

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Score_Post_Test & ScorePerStudent	18	.913	.000

The Paired Samples Correlation table reveals a very strong positive correlation (.913) between pre-test and post-test scores, which is statistically significant (p = .000). This indicates that students' relative performance remained consistent across the two tests, supporting the reliability of the measurements.

Picture 8: Paired Samples Test

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Score_Post_Test - ScorePerStudent	7.93651	10.11604	2.38437	2.90592	12.96710	3.329	17	.004

The Paired Sample Test table presents the results of the paired t-test, which confirms that the mean increase of 7.93651 points in test scores is statistically significant (t = 3.329, df = 17, p = .004). The 95% confidence interval (2.91 to 12.97) further supports that the true mean difference is positive and meaningful. These results demonstrate that the instructional intervention had a significant positive effect on students' reading comprehension abilities.

The comparison between the pre-test and post-test scores clearly indicates that the instructional intervention had a positive impact on students' reading comprehension abilities. The increase in the mean score from 37.30 on the pre-test to 45.24 on the post-test reflects a noticeable improvement in overall performance. This gain of approximately 7.94 points suggests that students were able to enhance their skills as a result of the teaching methods or materials used during the intervention period.

The strong correlation coefficient of .913 between the pre-test and post-test scores further strengthens the validity of this comparison. Such a high correlation implies that the test reliably measured students' reading comprehension consistently

across both testing occasions. It also suggests that while students improved overall, their relative rankings remained stable; those who performed well initially generally continued to do so after the intervention, and vice versa. This consistency is important because it indicates the test's ability to discriminate among different levels of student ability both before and after instruction.

The paired-samples t-test results provide robust statistical evidence that the observed improvement is not due to random chance. With a t-value of 3.329 and a p-value of .004, the data show a statistically significant difference between the pre-test and post-test means at the 0.05 significance level. The confidence interval for the mean difference (2.91 to 12.97) does not include zero, reinforcing the conclusion that the intervention had a meaningful effect. This significance confirms that the instructional program was effective in enhancing students' reading comprehension skills within the sample tested.

However, it is worth noting that the standard deviation of the post-test scores (22.61) was higher than that of the pre-test scores (16.50), indicating increased variability in student performance after the intervention. This could suggest that while many students improved, the degree of improvement varied widely among individuals. Some students may have made substantial gains, while others improved only slightly or not at all. This variability highlights the importance of considering differentiated instructional strategies to support all learners effectively.

In summary, the statistical analysis of the pre-test and post-test scores demonstrates that the instructional intervention was successful in improving students' reading comprehension on average. The results provide both descriptive and inferential evidence supporting the effectiveness of the teaching approach. Future research might explore factors influencing the variability in student gains and investigate ways to maximize learning outcomes for all students.

Discussion

The analysis of students' reading comprehension, as reflected in both the pre-test and post-test results, reveals important insights into their abilities and the impact of the instructional intervention. Initially, the pre-test scores demonstrated that the overall level of reading comprehension among students was quite low, with an average score of 37.30 out of 100. Only a small number of students managed to score above 60, while the majority fell below 40, indicating widespread difficulties in understanding reading materials. The considerable gap between the highest (68.57) and lowest (14.29) pre-test scores, along with a relatively high standard deviation of 16.50, further highlighted the diversity of reading skills within the classroom. This wide range suggested that while a few students already possessed a moderate grasp of reading comprehension, most struggled significantly, emphasizing the need for targeted instructional support.

Following the implementation of group work strategies and collaborative reading activities, the post-test results showed a noticeable improvement in students' reading comprehension. The mean post-test score rose to 45.24, and the highest score increased to 77.14, indicating that some students made substantial progress. The number of students scoring above 60 also increased, reflecting a positive shift in overall performance. However, the post-test also revealed that reading comprehension remained a challenge for many, as several students continued to score below 40. The standard deviation for the post-test was 22.61, even higher than in the pre-test, suggesting increased variability in student outcomes. This indicates that while the intervention was effective for some, others continued to face difficulties, possibly due to differences in learning pace, engagement, or prior knowledge.

Statistical analysis using a paired-samples t-test confirmed that the improvement in mean scores from pre-test to post-test was statistically significant ($t = 3.329$, $df = 17$, $p = .004$). This result demonstrates that the instructional intervention had a meaningful effect on students' reading comprehension, moving the average performance upward. The strong correlation between pre-test and post-test

scores (r = .913, p = .000) indicates that students' relative performance remained consistent, with those who initially performed well generally maintaining their advantage, while those who struggled continued to do so, albeit with some improvement.

The reliability and validity analyses further support the quality of the assessment instruments. Both the pre-test and post-test showed high Cronbach's Alpha values (.835 and .915, respectively), indicating strong internal consistency and reliable measurement of reading comprehension skills. The majority of test items in both assessments demonstrated significant positive correlations with the total score, confirming their validity as indicators of students' abilities. However, a few items in both the pre-test and post-test did not correlate well with the overall score, suggesting areas for future revision to enhance the assessments' effectiveness.

Is Group Work method effective in improving Students' Reading Skill at SMP N 2 Tondano? Based on the findings above, the answer is yes. The use of group work strategies led to a statistically significant improvement in students' reading comprehension, as evidenced by the increase in average test scores, the rise in the number of students scoring above 60, and the results of the t-test analysis (p = .004). Although not all students reached the same level of improvement, the intervention proved effective in fostering collaborative learning and enhancing reading skills, particularly for those who were moderately performing or struggling at the beginning.

Based on the findings of this study, it can be concluded that the students' reading comprehension ability was initially low, as indicated by the pre-test mean score of 37.30 out of 100, with most students scoring below 40. The wide range of pre-test scores and high standard deviation reflected significant variability in students' reading skills, highlighting the presence of both struggling and moderately proficient readers in the classroom. After the implementation of group work strategies and collaborative reading activities, there was a noticeable improvement in students' reading comprehension, as shown by the increase in the post-test mean score to 45.24. The results of the paired-samples t-test confirmed that this improvement was

statistically significant (t = 3.329, df = 17, p = .004), indicating that the instructional intervention had a meaningful positive effect on students' reading comprehension.

Furthermore, both the pre-test and post-test instruments demonstrated high reliability, with Cronbach's Alpha values of 0.835 and 0.915 respectively, ensuring the consistency and accuracy of the assessments. The majority of test items in both instruments were valid, showing significant positive correlations with the total score, although a few items did not perform as well and may require revision. Despite the overall improvement, the post-test results still revealed considerable variability in student performance, with some students continuing to score below 40. This suggests that while the intervention was effective for many, additional support may be needed for those who continue to face challenges in reading comprehension.

REFERENCES

Ahmada, A. (2020). USING GROUP WORK STRATEGY TO IMPROVE STUDENTS'READING COMPREHENSION SKILL. *Journey: Journal of English Language and Pedagogy*, *3*(1), 14–18.

Bachtiar, B. M., & Kurniawan, C. (2005). Lets Talk. *Bandung: Pakar Raya*.

Dadara, A., & Sudiyono, S. (2021). THE EFFECTIVNESS TEACHING READING COMPREHENSION USING GROUP WORK DISCUSSION METHOD. *PROJECT (Professional Journal of English Education)*, *4*(4), 599–603.

Damanik, A. S., & Herman, H. (2021). Improving Students' Reading Comprehension Through Question Answer Relationship Strategy (Qars). *Inovish Journal*, *6*(1), 84–101.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: answers to selected exercises. *A Simple Guide and Reference*, *63*(1), 1461–1470.

Ghaith, G. (2003). Effects of the learning together model of cooperative learning on English as a foreign language reading achievement, academic self-esteem, and feelings of school alienation. *Bilingual Research Journal*, *27*(3), 451–474.

Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading*. Routledge.

Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics.*

Iwai, Y. (2016). Promoting strategic readers: Insights of preservice teachers' understanding of metacognitive reading strategies. *International Journal for the Scholarship of Teaching and Learning*, *10*(1), 4.

Karisi, Y., Pelenkahu, N., & Maru, M. G. (2021). STUDENTS'PERCEPTION OF THE USE OF YOUTUBE IN TRANSLATION CLASS. *SoCul: International Journal of Research in Social Cultural Issues*, *1*(2), 126–138.

Kasiha, G., Hampp, P., & Kumayas, T. (2022). THE USE OF SMALL GROUP DISCUSSION TECHNIQUE TO IMPROVE STUDENTS'READING COMPREHENSION. *JoTELL: Journal of Teaching English, Linguistics, and Literature*, *1*(6), 687–696.

Lendo, N., Liando, N., & Olii, S. (2021). An Analysis of readability of reading texts on English national examination on junior high school. *Journal of English Culture, Language, Literature and Education*, *9*(2), 128–143.

Liando, N. V. F., Pelenkahu, N., & Mongkaren, S. (2021). STUDENTS AND PARENTS'PERCEPTIONS TOWARD ENGLISH ONLINE LEARNING DURING CORONA VIRUS PANDEMIC. *Jurnal Pendidikan Bahasa Inggris Undiksha*, *9*(1), 91–97.

Liando, N. V. (2010). STUDENTS'VS. TEACHERS'PERSPECTIVES ON BEST TEACHER CHARACTERISTICS IN EFL CLASSROOMS. *TEFLIN journal*, *21*(2), 118.

Liando, N. V. F., & Tatipang, D. P. (2024). Enlightened minds: Navigating the nexus of artificial intelligence and educational modernization. *Penerbit Tahta Media*.

Liando, N. V., & Lumettu, R. (2017). Students' Personal Initiative towards Their Speaking Performance. *International education studies*, *10*(8), 21-28.

Liando, N. V., Dallyono, R., Tatipang, D. P., & Lengkoan, F. (2023). Among English, Indonesian and local language: Translanguaging practices in an Indonesian EFL classroom. *Indonesian Journal of Applied Linguistics*, *13*(1), 204-216.

Moos. (2010). *Using visualisation, questioning, and other metacognitive strategies to boost EFL reading comprehension*.

Plato. (2008). *The Republic*. Trans. C.D.C. Reeve. Indianapolis: Hackett.

Sugiharto, P. A., Santoso, Y. I., & Shofyana, M. H. (2022). Teaching english pronunciation using tongue twister. *Acitya: Journal of Teaching and Education*, *4*(1), 189–197.

Tatipang, D., Oroh, E. Z., & Liando, N. V. F. (2021). THE APPLICATION OF MIND MAPPING TECHNIQUE TO INCREASE STUDENTS'READING COMPREHENSION AT THE SEVENTH GRADE OF SMP. *Kompetensi*, *1*(03), 389–397.

Tumengkol, M., Posumah, J., & Liando, N. (2022). STUDENTS'PERCEPTION ON THE USE OF MOVIE FOR VOCABULARY MASTERY. *JoTELL: Journal of Teaching English, Linguistics, and Literature*, *1*(11), 1250–1261.

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.

Wood, N. V. (1991). *Strategies for College Reading and Thinking*. McGraw-Hill.