

CHAPTER III

RESEARCH METHODOLOGY

3.1 Introduction

Chapter II has elaborated the information about the theoretical foundations in line with the topic of the study.

Chapter III discusses the research methodology that the researcher used in this study. It consists of five main sections: the research design, the research site and participants, the data collection techniques, the data analysis techniques and the hypothesis testing. This chapter will be closed with the concluding remarks.

3.2 Research Design

In relation to the nature and the purposes of the study which have been elaborated in Chapter 1, this study employed a quasi experimental design in the form of nonequivalent (pre-test and post-test) control-group design. It was chosen for three main reasons. First, this study involved two groups, control and experimental groups, in which only the experimental group received the treatment (Creswell, 2009, p. 160). Second, the study employed two tests in the forms of reading comprehension tests in collecting the data. They were pre-test which was conducted before the treatment and post-test which was conducted after the treatment (Nunan, 1992, p. 41; Emilia, 2000, p. 3). Third, different from the true experimental design, quasi experimental which was used in this study did not require an individual random selection of the students as the participants of the research (Hatch and Lazaraton, 1991, p. 95; Moore, 2008).

Besides, in order to confirm the research results which were elaborated quantitatively from a quasi experimental design in the form of nonequivalent (pre-test and post-test) control-group design, this study also employed classroom observations which were elaborated qualitatively. It was conducted to observe the effectiveness of the implementation of the NHT with teacher's scaffolding in teaching reading comprehension as well as to see the students' responses to it.

3.3 Research Site and the Participants

This study took place at one senior high school in Riau for about five weeks. It was from the first week of March to the first week of April 2014. The research site was chosen for three reasons. First, it was accessible, feasible and suitable for this study in terms of time, mobility and skills (McMillan and Schumacher, 2001, p. 432). Second, the researcher had an easy access to do the research in the site since the researcher was a teacher there. Third, the site fulfilled the requirements of a place to conduct research under the aims of the research.

Moreover, the participants of the research were two classes of the second grade students in this school. These include XI natural science 1 which was categorized as the control group and XI natural science 2 which was categorized as the experimental group. The control group consisted of 26 students comprising 10 males and 16 females. Meanwhile, the experimental group contained 28 students, including 15 males and 13 females. There were two considerations for choosing them as the participants of the research which were related to the historical factors of the internal validity and the pre-existing ability of the participants (Hatch and Farhady, 1982, p. 7; McMilan and Schumacher, 2001, p. 186-187; Mackey and Gass, 2005, p. 109-110). First, related to the historical factors, the participants were chosen because they had relatively the same language learning background and the experience in the school with the same teacher and the same learning materials.

Second, the researcher assumed that they have relatively adequate and similar pre-existing ability or language knowledge background which was proved through two sources of data. The first source was the data from the English teacher who taught in the site who asserted that both groups had the same pre existing ability based on the students' previous reading comprehension scores. The second source of data was obtained from the statistical data of the normality test and the independent t-test of the pre-test results in both groups. These two tests were used to ensure that both groups were homogeneous.

The normality test was conducted to see whether the data which were obtained from the pre-tests are normally distributed. If the significance of the normality test is same or higher than alpha (.05), it indicates that the data are normally distributed so that the parametric statistic including t-test can be used (Freyadefk, 2013). In this study, the normality test result of the pre-tests were higher than .05 (.067 for control group pretest and .706 for experimental group pre-test). It proves that those pre-tests results were normally distributed and the independent t-test can be used in order to see whether both groups were homogeneous. When sig. (2-tailed) is lower than alpha (.05), it indicates that there is difference between the two mean scores in 5% significance. Meanwhile, when sig. (2-tailed) is lower than alpha (.01), it indicates that there is difference between the two mean scores in 1% significance. Then, when sig. (2-tailed) is higher than alpha (.05), it indicates that there is no difference between the two mean scores (Widhiarso, n.d.). In this study, the result of the t-test shows that t_{obtain} was 1.844 (df=52). It was lower than t table which was 2.000. Then, the significance of the probability value was .71. It was higher than alpha .05. Therefore, it implies that there was no difference between the pre-test results of both groups and proves that both groups were homogeneous (see Appendix 16).

Furthermore, regarding the process of recruiting the research participants, there were two main steps in it. First, the researcher asked for the agreement to the headmaster of the school and the participants to conduct the research in the school. Moreover, before conducting the research, the researcher asked the participants to sign a consent form which indicated their agreement to be the participants of the research. After that, the researcher told the participants about the research aims, what they had to do, what data would be obtained from them, and how the research was conducted in their classes (see Appendix 5). The consent form was adopted from Emilia (2005).

3.4 Data Collection Techniques

This section discusses two main parts, including the instrumentation and the procedures of data collection techniques. Each of which will be discussed below.

3.4.1 Instrumentation

There were three instruments which were used to collect the data to answer both research questions (see Section 1.3 in detail). These included classroom observations, reading comprehension tests, and a questionnaire.

First, the classroom observations were intended to find out whether the NHT with teacher's scaffolding was effective to teach reading comprehension as well as to investigate the students' responses to the use of this method (see Section 1.3). They were conducted in the experimental group for seven meetings. Moreover, the classroom observations were undertaken for two reasons. Firstly, they were used to observe what the teacher and the students did and said in the classroom as a mean of adding knowledge and information regarding language learning (Nunan, 1992, p. 91; Darlington and Scott, 2002, p. 74), particularly about the use of the NHT with teacher's scaffolding in teaching reading comprehension. Secondly, they also provided valuable supplemental information which was more useful than formal testing because they can be carried out obtrusively (Westwood, 2008, p. 73).

Second, reading comprehension tests which consisted of the pre-test and post-test were intended to collect the data in order to answer the first research question and as supporting instrument to answer the second one. They were used to examine the effectiveness of the NHT with teacher's scaffolding which was implemented in teaching reading comprehension (Emilia, 2011, p. 12) as well as to see the students' responses to its implementation.

They were administered to twenty six students in the control group and 28 students in the experimental group. The pre-test was administered on March 5th

2014 before conducting the teaching program. Meanwhile, the post-test was administered on April 5th 2014 at the end of the program.

Furthermore, the tests were conducted in the form of multiple-choice tests which were considered as the most common test used in testing reading comprehension (Magliano, Millis, Ozuru, and McNamara, 2007, p. 117). There were four fundamental reasons for using multiple choice tests.

First, the multiple choice test is the most widely used test by teachers (Miller et al., 2009, p. 196). Second, they are the most popular and common test which are used in the educational assessment, including for testing students' reading comprehension (Alderson, 2000, p. 211; Cheung and Bucat, 2002, p. 1; Harmer, 2008, p. 382; Brown and Abeywickrama, 2010, p. 234). Third, they were easy to mark (Harmer, 2001, p. 323; Harmer, 2008, p. 382). Fourth, they can be perfectly reliable (Hughes, 2003, p. 75).

The third instrument was questionnaire. It was used as the main instrument to collect the data to answer the second research question and as a supporting instrument to answer the first one. It was administered to find out whether the NHT with teacher's scaffolding was effective to teach reading comprehension as well as to investigate the students' responses to the use of this method. It was administered on April 6th 2014 at the end of the teaching session in the experimental group to 25 out of 28 students because three of them were absent in the day.

Furthermore, the questionnaire was utilized for five reasons. First, it was appropriate to use the questionnaire since one of the purposes of the research is to elicit the participants' responses to the implementation of the NHT with teacher's scaffolding. It was done by presenting the questionnaire items with series of statements to which they react by selecting the existing answers (Brown, 2001 in Mackey and Gass, 2005, p. 92; Cohen, Manion, and Morrison, 2007, p. 502). Second, the questionnaire was used as it allowed to collect a huge amount of information in a short time (Dornyei and Taguchi, 2010, p. 6).

Third, the questionnaire was very versatile. It can be used successfully with a variety of people in a variety of situations targeting a variety of topics (Dornyei and Taguchi, 2010, p. 6). Fourth, the use of questionnaire could reduce the bias and could improve the reliability of gaining information related to someone's response about certain case. It could even tap into attitudes that the respondents are not completely aware of. A well-constructed questionnaire can reduce the bias of interview effects and increase the consistency and reliability of the results (Bryman in Dornyei and Taguchi, 2010, p. 6). Fifth, by using the questionnaire, the researcher was allowed to gather information in which the participants are able to report about themselves, such as their beliefs, motivations, and reactions to learning activity and classroom instruction (Mackey and Gass, 2005).

Besides, the closed-ended questionnaire was used in this study in the form of Likert scale. It was originally devised by Rensis Likert in 1932. A close-ended questionnaire was used for two reasons. First, it was appropriate to use a close-ended questionnaire since it can discover the opinion, perception, and strength of feelings or attitudes towards the given statements or a series of statements related to social phenomena (Bell, 2005, p. 142; Sugiyono, 2012, p. 134).

Second, Likert scale was simple, versatile, and reliable in which it consisted of the statements which included value or direction, and the respondents indicated agreement and disagreement with the statements (Mc. Millan and Schumacher, 2001, p. 262; Dornyei and Taguchi, 2010, p. 27).

Moreover, in this study, Likert scale was employed in four options with the scores starting from strongly agree (4 point), agree (3 point), disagree (2 point), and strongly disagree (1 point). The respondents were required to choose the scale that they think suit to their mind (McMillan and Schumacher, 2001, p. 262; Dornyei and Taguchi, 2010, p. 27; Sugiyono, 2012, p. 135).

Furthermore, eventough those three instruments were conducted separately, but they were intended to answer both research questions regarding an investigation of whether the NHT with teacher's scaffolding was effective to teach

reading comprehension as well as the the students' responses to the use of the method.

3.4.2 Procedures

The research was conducted in three procedures. The first procedure was the administration of the classroom observations. There were two steps of it.

The first step was the construction of the observations' guidelines. It was constructed by referring to the guidelines adopted from several sources. They were related to four indicators to reveal the effectiveness of teaching reading comprehension through the NHT with teacher's scaffolding (see Section 2.4.4 in detail).

The first indicator included the teacher's plan, management and scaffolding. These points were to do with the curriculum materials, pedagogical approach or reading program he/she uses, explanation and assistance of learning activities and materials (Cooper and McIntyre, 1996, p. 1; Hammond: 2001, p. 2-3; Richard Allington, 2002 in Archer, 2004, p. 14). It was done by observing how the teacher planned the teaching and learning process, how the teaching program run, and how she explained and assisted or gave scaffolding to the students.

The second indicator was students' participation in learning process (Mohr, 1998 in Archer, 2004, p. 14; Yeh, 2004 in Japar, 2011). It was done by seeing how the students responded the teacher's explanation, participated on their own group, joined in answering the questions in quiz among the groups.

The third indicator was the students' enthusiasm in learning process (Yeh, 2004 in Japar, 2011). It was elicited by investigating whether the students were enthusiastic in learning process, and how they showed their enthusiasm. Then, the fourth indicator was the students' learning achievement (Cooper and McIntyre, 1996, p. 1). This section focused on the students' learning achievement which can be seen from their individual test mean scores at every end of meeting (see Appendix 7).

The second step was conducting the classroom observations. In this research, the researcher took a role as the participant observer. By being the participant observer, it enabled the researcher, as far as it was possible, to yield the valuable data, to observe the changes over time, to share the same experiences as the subjects, to understand the students' act in the way they do, to participate in classroom activity of an individual, group, to listen, observe, ask and understand the activities concerned, and 'to see things as those involved see things' (Bell, 2005, p. 186-187; Denscombe, 1998 in Bell, 2005, p. 17). Moreover, in carrying out the classroom observations, the researcher also took note regarding the guidelines as the source of data to be analyzed.

The second procedure was the administration of the pre-test and post-test of reading comprehension tests. The tests were used to measure the participants' comprehension of hortatory exposition. It was one of the materials taught in the second semester in the grade two of the senior high school level. There were three steps in conducting the tests.

The first step was constructing the test items. The test items were constructed by the researcher herself. This is because teacher-made tests are often more effective than published tests (Westwood, 2008, p. 77). In this study, the tests were constructed by adopting the hortatory exposition texts from several sources (Kristono, 2007; Cendikia, Dj, Hanifah, and Saspida, 2013; Liliawati, 2013; Priharini, Yuliani, and Arini, 2013; Pujarsono, 2013). They consisted of 30 questions for each test (pre-test and post-test) which covered three reading comprehension levels: literal, interpretive/inferential, and evaluative/applied/critical reading/metacognitive (Mohamad, 1999; Block, Rodgers, and Johnson, 2004, p. 3; King, 2007, p. 267; Dije, 2009; Cuesta College, n.d.; "3-level model," 2014).

Moreover, the questions covered six passing criteria of reading skill for the senior high school level as attached in the Regulatory Ministries of Education and Culture (*Permendikbud*) number 97 year 2013 which is organized by Organization of Education National Standard (*BSNP*, 2013) about National Examination

passing criteria, and that of number 23 year 2006 which is organized by Organization of Education National Standard (*BSNP*, 2006) about content standard for the senior high school level, and 2013/2014 National Examination prediction. The criteria were deciding the general idea of the text, the specific/implied/detail information in the text, the purpose of the text, the meaning of the words/phrases/sentences in the text, the reference of words in the text, and the main idea of the paragraphs in the text. The reason of selecting this text was that it is one of the text types involving in the National Examination for the senior high school level as mentioned in the statements of passing standard (*SKL*) for reading skill for the senior high school level which is attached in Organization of Education National Standard (*BSNP*, 2006).

The second step was constructing and conducting try out of tests. The try out of the pre-test and post-test were conducted to a class out of the research participants. Each of them consisted of forty questions. The try out results were analyzed by using Anatest V4. Then, the analysis results were used as the consideration for the researcher to reconstruct the better items for the real version of the pre-test and post-test. They consisted of only thirty questions. They were obtained after the elimination of the try out questions which comprise 25% of difficult questions, 50% of average questions, and 25% of easy questions.

Besides, in order to ensure that there was no significant difference in the test items between the pre test and post test, the try out results were analyzed statistically through the normality test and the paired sample t-test. It was calculated by using SPSS V16.

First, the normality tests results reveal that the significance (P value) of the pre-test was .078, and that of the post-test was .080. Since the results of the significance (P value) were higher than alpha .05, it implies that the data were normally distributed.

Second, the paired sample t-tests result of the pre-test and post-test try out show that t_{obtain} was 1.587 ($df = 23$). It was lower than t table which was 2.069. Then, the significance (P value) (2-tailed) was .126. It was higher than .05. Since

t_{obtain} was lower than t table (2.069), and the P value was higher than alpha (.05), it can be concluded that there was no significant difference in the try out results between the pre-test and post-test (see Appendix 15).

The third step was the construction and administration of the real version of the pre-test and post-test items. The real version of the tests were obtained from the analyses of try out results. The pre-test and post-test items were basically conducted in the same way, with the same indicators and difficulty level. They were used to find out whether the NHT with teacher's scaffolding was effective to teach reading comprehension and to investigate the students' responses to it.

The third procedure was the administration of the questionnaire. There were three steps in conducting it. The first step was the construction of Likert Scale questionnaire. The items of the questionnaire were developed based on the guidelines related to the theory about the advantages of the NHT proposed by Ibrahim (2000) as cited by Suprasetyo (2009, p. 4) (see Section 2.4.2).

Moreover, to enhance the students' consistency in giving their responses, the questionnaire items included both positive and negative statements (Dornyei and Taguchi, 2010, p. 43). In this study, there were sixteen statements in both positive and negative forms related to the implementation of the NHT. The questionnaire items were written in *Bahasa Indonesia* in order to make it more comprehensible for the students to answer.

The second step was piloting the questionnaire items. Before the questionnaire items were administered to the research participants, it was tried out or being piloted first. The point of carrying out a pilot study was to test, often to revise, and then to finalize the test items. It was an important mean of assessing the feasibility and usefulness of the data collection methods and making any necessary revisions before they were tested to the research participants (Alderson, Clapham, and Wall, 1995, p. 74; Mackey and Gass, 2005, p. 43). It was done to enhance the content validity, reliability, and the practicability of the questionnaire items, to ensure that the items were representative sample of the test content, to convince that the items were in comprehensible level, and to avoid

misunderstanding to the students (Brown, 2005, p. 221; Cohen, Manion, & Morrison, 2007, p. 341; Dornyei and Taguchi, 2010, p. 117). Thus, in this pilot step, a group of the students, who were out of the sample of the research but have similar characteristic, read the questionnaire items and showed their feedback of whether the questionnaire items were already comprehensible or still caused ambiguity which indicated their capability in answering the questionnaire items and their reaction toward those items. Therefore, the findings of this step were used as the consideration for the researcher to produce the better questionnaire items to be used for this research.

The third step was the construction and administration of the real version of the questionnaire items. The items were obtained from the revision results of the pilot step. It was administered to 25 students in the experimental group at the end of the teaching program.

Table 3.1 below shows the data collection procedure in detail.

Table 3.1

The Data Collection Techniques Procedure

Phase	Time	Activities
Pre stage	March 3 rd 2014	Administering try out of the pre-test and post-test
	March 5 th 2014	Administering the pre-test in both groups
While stage	March 6 th – April 4 th 2014	Treatment (7 meetings)
Post stage	April 5 th 2014	Administering the post-test in both groups
	April 6 th 2014	Administering the questionnaire items

3.5 Data Analysis Techniques

As explained in Section 3.4, there were three data collection techniques in the research. These cover the classroom observations, the tests, and the questionnaire. Each of which will be discussed below.

3.5.1 Classroom observations

The data from the classroom observations were in the forms of researcher's field notes of what was going on the process of teaching and learning process in reading class in which the NHT was implemented (see Appendix 17). The data which were obtained from the field notes were condensed, analyzed, and described quantitatively as presented in chapter IV. The analysis was done based on the theory of the effectiveness of the NHT with teacher's scaffolding (Coffey: n.d.; Cooper and McIntyre, 1996, p. 1; Hammond: 2001, p. 2-3; Gillies and Ashman, 2003; Johnson and Johnson, 2003; Richard Allington, 2002 in Archer, 2004, p. 14; Westwood, 2008, p. 68; Ibrahim, 2000 in Suprasetyo, 2009, p. 4; Yeh, 2004 in Japar, 2011; Smith, 1996 in Barkley et al., 2012, p. 7; Barkley et.al., 2012, p. 8 (see 2.44 in detail).

3.5.2 Reading Comprehension Tests

This subchapter points out two main points. The first point is the test analysis which was used in the study. In this study, the tests were analyzed by using t-test. It was the most widely used statistical test to compare two mean scores in which can also be used with very small sample sizes (Hatch and Farhady, 1982, p. 108; Mc. Millan and Schumacher, 2001, p. 368). In this study, the researcher calculated the t-test through a software called IBM SPSS statistics version 16. This software is an innovative statistical computer program that can compute almost any statistical calculation and make the calculation analysis easier and simpler to conduct (Larson-Hall, 2010; Privitera, 2012). It is also a great tool to create graphs to help gain a better understanding of the data (Carlson and Winquist, 2014). However, the biggest challenge in using this software is making

sure to enter the data correctly. Entering even a single value incorrectly can alter the data analyses. It is important to always double-check the data to make sure that the correct values have been entered.

Through this application, the researcher got the statistical output to be analyzed in order to answer the research questions of this study. After doing the calculation, the researcher will be able to see the accepted hypothesis as the basis to take conclusion of the study.

However, before deciding to use t-test to compare two means of the scores, it was needed to calculate the normality test in order to find out whether the data were normally distributed, and the parametric statistic in the form of the t-test could be used. In this study, Shapiro-Wilk normality test was used because the number of the research participants were less than 50 students (Widhiarso, n.d.).

The second point is the types of the t-test which was used in the study. There were two types of the t-test which were conducted in analyzing the data from tests. They were paired sample t-test and independent sample t-test (Hatch and Farhady, 1982, p. 122-125; Sugiyono, 2012, p. 121-137).

The paired sample t-test was used to see the students' mean scores in the pre-test and post-test in each group. Moreover, the independent sample t-test was used to compare the progress of the students' reading comprehension ability in both groups. It was calculated by comparing the means scores of the progress scores which were obtained from the two groups of the students. The result indicates which one of those two groups made more significant progress and eventually answered the research questions whether the NHT with teacher's scaffolding was effective to teach reading comprehension. Besides, that would also automatically reflect the students' responses to the implementation of the NHT.

Furthermore, there were seven main terms which were revealed in the output of SPSS V16 as the results of the normality test and the t-test. These include the significance (sig.) of the normality test, the mean score, the standard deviation correlation and its significance, t, df (degree of freedom), the sig. (2

tailed), and Lavene's test as well as the choice of both equal variances assumed or equal variances not assumed.

The first term is the significance (sig.) of the normality test. When it is the same as or higher than alpha (.05), it indicates that the data are normally distributed so that the parametric statistic including the t-test can be used (Freyadefk, 2013).

The second term is the mean score. the mean score is calculated by adding all the values and divided them by the number of values in the set (Woods, Fletcher, and Hughes, 1986, p. 29; Shayib, 2013, p. 32). In this case, it was used to find the students' average scores in their reading comprehension.

The third term is the standard deviation which is one of the most important statistical measurement which indicates the amount by which values in the data set differ from the mean (Woods, Fletcher, and Hughes, 1986, p. 43). It shows the range of students' scores between the highest and the lowest one.

The fourth term is the correlation. It indicates that two variables are correlated. The correlation point ranges from +1 to -1. +1 indicates that there is a positive correlation between the variables. On the other hand, -1 indicates that there is a negative correlation between the variables. Moreover, its significance refers to the significance of the correlation. If the significance (sig.) was higher than alpha (.05), it indicates that there is no correlation between the variables. Meanwhile, when the significance (sig.) is lower than alpha (.05), it indicates that there is a correlation between the variables (Widhiarso, n.d.).

The fifth term is the t_{obtain} . It indicates the value of t which is shown in the t-test calculation. This t_{obtain} is used to reveal which hypothesis to be accepted and rejected by comparing it with the critical value in t_{table} . When t_{obtain} is higher than critical value in t_{table} , it means that the null hypothesis is rejected and the alternative hypothesis is accepted. On the contrary, when t_{obtain} is lower than critical value in t_{table} , it means that the null hypothesis is accepted and the alternative hypothesis is rejected (Hatch and Farhady, 1982, p. 110). In addition, the degree of freedom (df) indicates the number of the data minus 1. It affects the way to see the t table.

The sixth term is the significance (2 tailed). It is used to calculate the 2 tailed statistical data. It affects the way to decide acceptance and rejection of the hypothesis. When sig. (2-tailed) is lower than alpha (.05), it indicates that there is a difference between the two mean scores in 5% significance. Meanwhile, when sig. (2-tailed) is lower than alpha (.01), it indicates that there is a difference between the two mean scores in 1% significance. Then, when sig. (2-tailed) is higher than alpha (.05), it indicates that there is no difference between the two mean scores (Widhiarso, n.d.)

The seventh term is the Lavene's test and the choice of both equal variances assumed and equal variances not assumed. Lavene's test is a kind of homogeneous test which indicates whether the data are homogeneous or not. Moreover, the significance (sig.) value of Lavene's test value indicates that whether the raw of equal variances assumed or equal variances not assumed to be used. When the significance of the value is higher than alpha (.05), it indicates that the data are homogeneous and the equal variances assumed raw is used. On the contrary, when the significance of value is lower than alpha (.05), it indicates that the data are not homogeneous and the equal variances not assumed raw is used (Widhiarso, n.d.; Tyrrell, 2009, p. 80).

3.5.3 Questionnaire

The questionnaire which consisted of four options was completed by number of scores: strongly agree (4 point), agree (3 point), disagree (2 point), and strongly disagree (1 point) for positive word items. While the scores for negative word items were: strongly agree (1 point), agree (2 point), disagree (3 point), and strongly disagree (4 point). Then, the data from the questionnaire was analyzed in three steps. The first step was computing the total number and the percentage of participants who gave responses for those four options in positive and negative statements (16 statements in total). The second step was calculating the mean of every single statement (both negative and positive). The third step was recapitulating the mean of all statement (both positive and negative). In this stage, the mean of positive statement and that of negative statement was available. Hence, the findings could be computed statistically and interpreted accordingly.

3.6 Hypotesis Testing

There were two assumptions which were proposed in this study. Null Hypotesis (H_0) was accepted if there was no significant difference in the progress of the students' tests results between the control and experimental groups. Meanwhile, the alternative hypotesis (H_1) was accepted if there was significant

difference in the progress of the students' tests results between the control and experimental groups. To meet the assumption above, the statistical analysis of the t-tests were applied in this study.

3.7 Concluding Remarks

This chapter has revealed the research methodology. It contains the research design, the research sites and participants, the data collection techniques, the data analysis techniques, and the hypothesis testing.

Chapter IV will cover the research findings and discussion of the research. It will be presented in three main sections: the effectiveness of Number Heads Together in teaching reading comprehension, students' responses to the implementation of Number Heads Together in teaching reading comprehension, and concluding remarks.