CHAPTER III: RESEARCH METHOD

The previous chapter has described the theoretical background of the study. It discusses literature review about reading comprehension, comprehension levels, STAD Cooperative Learning (STAD CL) and Direct Instruction (DI) along with the principles, procedures and their weaknesses.

This chapter provides the information about the methodology of the research including the research questions, the setting and participants of the study, the design, data collection methods and data analysis. The conclusion ends this chapter.

3.1 Research Setting and Participants

This research was undertaken in a junior high school in Serang Regency, Banten Province. This site made the research more feasible and suitable in terms of time, mobility and skills (McMillan and Schumacher 2001 p. 432). Firstly, it was located in the same town as the researcher. This could ease the researcher in managing time and cost in running the study. Secondly, based on the researcher' previous observation, the English teacher had the necessary skill to implement the teaching procedures. Thirdly, the researcher had good access to the site, since the researcher was one of teaching personnel in the site.

The participants of the study were 64 students of eight graders which were divided into two groups each of which consist of 32 students. They were selected based on the historical factors and pre-existing ability (Hatch and Farhady, 1982 p. 7; Mc.Millan and Schumacher, 2001 pp. 186-7). In terms of historical factors, to follow Hatch and Farhady, the students in both groups had relatively the same language learning experience. Previously, they had been studying English for more than 1.5 year the same teacher and the same materials. No one of them had ever joined any English course. In such a time span and experience, they were assumed to have adequate knowledge background which is beneficial for their reading (Alderson, 1999). This condition helped the researcher assumed that the participants had similar prior knowledge (Hatch and Farhady, 1982 p.7).

In addition, they were also chosen based on the homogeneity of pre-existing reading comprehension ability. Among the three groups joining pretest, two groups who

were statistically homogenous, no significant difference on their pretest scores. In other words, the two groups have approximately similar pre-existing reading comprehension ability (Hatch and Larazaton, 1991 pp. 261-3). The selected groups, then, were assigned as experimental and control groups by flipping a coin (Hatch and Farhady, 1982 p. 19).

3.2 Research Design

This study employed triangulation mixed method design (Creswell, 2003 p. 213; 2008 pp. 558-9 see also Mc.Millan and Schumacher, 2001 p. 542; Darlington and Scott, 2002 p. 123). The design was the combination between two designs: quasi-experimental and survey designs (Hatch and Farhady, 1982; Mc.Millan and Schumacher, 2001 pp. 32-4). The quasi-experimental design was employed since this study compared the causal effect of the independent variables: STAD CL and DI, on the dependent one: reading comprehension levels (Mc.Millan and Schumacher, 2001 pp. 342; Hatch and Farhady, 1982; Vellutino, and Schatschneider, 2011 p. 157). With this design, this study was more practical to do as it was not nec essary to assign students randomly (Hatch and Farhady, 1982 pp. 24 see also Vellutino, and Schatschneider, 2011 p. 169; Cresswell, 2008 p. 313). During the teaching programs (see Chapter IV), an observation was conducted to provide additional evidence for the study (Cowie in Heigham and Crocker, 2009, p. 168) and to have better understanding of what was going on in the implementation of the teaching programs. Meanwhile, the survey design, wherein the researcher elicited the students' opinion about the effects of STAD CL or Di on their comprehension levels, was carried out after the programs to provide better understanding of the findings (Creswell, 2008 p. 552) and to elicit students' opinion about the effect of the teaching programs on their comprehension levels.

The representation for the designs is presented on the following table.

Table 3.2: Research Design

Time

| Pretest | Select control (Intact) group | STAD Cooperative Learning Instruction + observation | Post Test | Question-naire | Interview |
| | Select experimental (Intact) group | Direct Instruction + observation | | | |

(Adapted from Hatch and Larazaton, 1991 p. 89 and Cresswell, 2008, p. 314)

Table 3.2 above shows a pretest (see Section 3.3.1) preceded the research, then experimental and control groups were selected on the basis of the pretest. After that, the teaching programs (see chapter IV) were conducted. After the programs ended, a post test (see Section 3.3.1), questionnaire (see Section 3.3.2) and interview (see Section 3.3.4) were administered.

3.3. Data Collection Methods

This section discusses data collection methods including pretest and post-tests, observation, questionnaire and interview.

3.3.1 Pretest and Posttests of Reading Comprehension

As mentioned in the research design (see Section 3.1), to gather the data about the students' achievement in reading comprehension, a pencil and paper tests (Mc.Millan and Schumacher, 2001 pp. 250-7) were administered. The tests were pretest, posttest and quizzes. The pretest (see Appendix 1a and 1b for the items) was aimed to see and compare the existing ability of the groups involved in the study. It was done before the teaching programs were conducted. The posttest (see Appendix 2a and 2b) was conducted after all teaching programs were done. It was conducted to see the effect of the teaching programs on the students' comprehension levels. In addition, in the end of every treatment, a quiz (see Appendix 3a,b for the example of the items) was also administered to assess the students' comprehension of the text learned.

In the tests, multiple choice items were chosen for several considerations. Firstly, they are suitable items for testing reading as receptive skill (Brown, 2005 p. 47). Secondly, they are selected due to its objectivity and its practicality. Multiple choice items are more objective since there will be less scoring mistakes than other test formats (Brown, 29). They are also easy to administer and can be scored quickly (Brown, 2005 p. 47). It can be administered and scored in shorter time than that of other formats. It has simplicity in analyzing the test validity and the test reliability (Brown, 2001 p. 386; 2004 p. 20).

Some test items were taken from available reading tests and some others were constructed by the researcher. They were adopted from the items of previous national examinations (Balitbang – Kemendiknas, UN, 2011) or its try out. The selection and the construction of the items were based on the comprehension levels to measure and their difficulty levels.

The items were segmented into three sections, according to comprehension levels: literal, inferential and evaluative (see Section 2.1.5). For literal level, the test items required students to scan for specific information; read for detailed explicit information; guess the meaning of unfamiliar word from context, and identify or locate information. For inferential comprehension, test items were related to identifying or reading for main ideas and pronominal reference; reading for implicit meaning; outlining logical organization of text and drawing inferences in the content. For evaluative comprehension, test items were to assess students' ability in recognizing a writer's purpose, attitude, tone and emotion in the text and identifying the characteristics of characters; and analyzing the (generic) structure of the text (see Section 2.1.5 for detail).

The test consisted of thirty items with two genres of texts: recount and narrative (see Martin and Rose, 2008; Anderson and Anderson, 1997). These genres were also as the material in the teaching programs (see Chapter IV). They were chosen since they had potential to challenge all levels of students' comprehension, according to Barret Taxonomy of reading comprehension levels (www.campbellps.det.wa.edu.au; Clymer 1968 in Hudson, 2007 and Pettit and Cockriel, 1974; Berry, 2005). The pretest items included two recounts, about unforgettable experience and holiday, and two narratives about folktale and legend. The posttest ones were also composed of five texts: two recounts about lost dog and unforgettable experience, and three narrative texts about fables and legend.

To enhance the validity of the test items, some efforts were made. To maintain the content validity (Hatch and Farhady, 1982 pp. 250-5; Hughes, 2003 p. 26: Brown, 2005 p. 221), the construction of the test items were referred to reading comprehension levels, basic competence of the content standard (Depdiknas, 2006) as well as graduation criteria of English national examination of Junior High School (Depdiknas, 2011). To support the face validity, the items were constructed in such a way that they had similar length of options; the same number of options and similar organization (Hughes, 2003 p. 33). At last, to uphold the criterion validity (Hughes, 2003 p. 27), the results of the (pre- and post) tests were compared with those of the previous tests.

To ensure the reliability and the difficulty levels, the test items were tried-out and subsequently modified (Hatch and Farhady, 1982 p. 253). The items were tried out to another group of students not involved in the study. This try-out was intended to see bad items or distracters and to gain the ideal levels of difficulty, containing about 25% of easy,

50% of medium and 25% of difficult items (Hughes, 2003; Brown, 2005; Arikunto, 2009, p.207). There were 40 items on the try-out but only 30 items were selected for the tests. The selection was based on difficulty and comprehension levels. Further discussion of the try-out of the items is discussed in Chapter IV Section 4.2.2.

3.3.2 Class Observation

Observation was carried out to provide additional evidence for the study (Cowie in Heigham and Crocker, 2009, p. 168) and to get to know first-hand information about social process in a naturally occurring context (Silverman, 1993). In this case, the researcher acted as a participant observer (Creswell, 2001 p. 222; see also Cowie in Heigham and Crocker, 2009 p. 166) wherein he got involved in helping the teacher plan the teaching programs, prepare reading texts, implement the procedures of the programs and evaluate them. This observation aimed to have better understanding of what was going on in the implementation of the teaching programs and how the teacher as well as students participated in the activities.

To provide sharper focus of data collection (Nunan, 1992 p. 98), in this study, the researcher applied structured observation wherein specific categories to observe had been determined previously (Mc.Millan and Schumacher, 2001 pp.40-1; see also Nunan, 1992 p. 96). The categories included the principles and the procedures of the teaching programs. The observation sheet for STAD CL group (see appendix 4a) were adapted from Furtwengler (1992; see also Leighton in Cooper, 1990 pp. 330-331). It contained some indicators of effective cooperative learning related to classroom organization, classroom management, presentation content, group facilitation, monitoring and lesson summary. In the meantime, the observation sheet for DI group (see appendix 4b) were adapted from observational guideline of direct instruction from *education.byu.edu*. It included direct instruction procedures consisting of the indicators of opening, modeling, guided practice, independent practice and closing phases.

When the observation was being conducted, the researcher also made field notes (see Appendix 5a and 5b, for the example). The notes about what the teacher as well as the students did in the teaching programs.

3.3.3 Questionnaire

After all the teaching programs were done, a questionnaire (see Appendix 6a - 6d) was employed. It was chosen to keep the objectivity of the response and the

anonymity of respondents (Mc.Millan and Schumacher, 2001 p.257). In this study, the questionnaire surveyed the students' opinion about the effect of STAD CL or DI on their comprehension levels. It was aimed to provide a numeric information, data or trend about the students' opinions on the teaching programs (Mc.Millan and Schumacher, 2001 p. 304; Cresswell, 2008 p. 61; 2002 pp. 154; 1994 pp. 127-130). The questionnaires were distributed to both groups after the posttest.

The questionnaire was developed following the guideline from Oppenheim (1982); Barbie (1998) in Mc Millan and Schumacher (2001 p. 258) and Dornyei (2002). The items were arranged in Likert Scale, in the four gradations starting from 'strongly agree', 'agree, 'disagree' and 'strongly disagree' (Mc.Millan and Schumacher, 2001 p.262; Dornyei, 2002 p. 27; Creswell, 2008 p. 176;). The respondents were only required to check the place on the scale that best reflect their belief or opinion about the statements (Mc.Millan and Schumacher, 2001 p.262).

The questionnaire contained 18 statements about the reading skills related to the three levels comprehension: literal, inferential and evaluative (see Appendix 6a and 6b). They consisted of nine positive and nine negative ones to avoid a response set in which the respondents mark only on side of the scale (Dornyei, 2002 p. 43). They were randomly arranged to make sure the students' consistency, real perception, and to prevent guessing (Sugiyono, 2005; Oppenheim, 1982). They were written in Bahasa Indonesia to make it more comprehensible for the students and to avoid misunderstanding.

Further, to enhance the effectiveness and comprehensibility of questionnaire, it was validated, coded and pilot-tested (Mc Millan and Schumacher, 2001 p. 267; Dornyei, 2002 pp. 53-55; Creswell, 2008 p. 402;). Before being tried out, it was validated by having others to read and to give feedback whether it was vague or incomprehensible. It was also coded by giving the initial of comprehension levels, the number of statements of each level and positive (+) or negative (-) symbols. It was also pilot-tested to a group of students not involved in the study to see their capability of completing the questionnaire and their reaction.

In the pilot testing, it was found that generally all statements in the questionnaire were comprehensible for the students. But, the term '*pembelajaran langsung'* (Direct Instruction) needed describing since the students were not familiar with the terms. Then, to follow Juliandi (2007), the result of the pilot test was statistically analyzed via Microsoft Excel (see Appendix 7). It was to get the correlation value (r) in Pearson-Product Moment

Correlation. It was found that most statements (73%) were statistically valid to use. And, in terms of internal consistency, the score of split-half reliability (Brown, 2005, p. 177; Hughes, 2003, p. 40) between positive and negative statements was 0.97. It meant that the questionnaire had good reliability. The score was close to the perfect reliability, that is, one (1.00).

3.3.4 Interview

In addition to the previous data collections, interview was conducted to verify, extend, elaborate or explain the data collected (Creswell, 2008 p. 552; McMillan and Schumacher, 2001 p. 444). In this case, it was intended to elicit the students' opinions about the effects of STAD CL or DI on their learning reading strategies. It was done after the analysis of the posttest score.

The interview was a directive (Richard, 2003 p. 51) or structured one (Nunan, 1992 p. 149) wherein specific topics/agenda were totally predetermined (see also McMillan and Schumacher, 2001 p. 444). Its form was focused group interview. It was to encourage a sense of divergent opinion (Krueger and Morgan, 1993 p. 17-18 in Emilia 2005 p. 245), to collect shared understanding from several individuals of students, to take the pressure off the participants (Darlington and Scott, 2002 p. 62) and to provide welcoming condition so that they were not hesitant to talk (Creswell, 2008 p. 226). Twelve students from each group were purposefully chosen (Creswell, 2008 p. 214; McMillan and Schumacher, 2001 pp. 400-1) from low, middle and high achievers in the post tests.

Interview guideline (see Appendix 8a and 8b) was used when interviewing. It consisted of some questions to elicit students' opinions about the effect of the teaching programs. The questions were related to the advantages of STAD cooperative learning (STAD CL) or Direct Instruction (DI) for their learning and their comprehension levels. The questions were segmented into three parts: introduction, core part and winding down or closing (Ritchie and Lewis, 2003 p. 114). In the introduction, to follow Ritchie and Lewis, easy, opening questions were asked. The core part included deeper and more specific questions. Finally, suggestive questions were delivered in the closing.

The number of questions in the group interviews was at least 16 questions. They consisted of two opening questions, thirteen core questions and one closing questions. The opening ones were related to students' personal identity and achievement background. The

core ones were related to their impression and experience in implementing the instructions especially the difference between the STAD CL or DI models and those that the students had experienced before, the facilitation of the models to their learning and to their comprehension levels. The closing ones consisted of a question relevant to students' suggestions on the application of the learning models in the future.

3.4 Data Analysis

In this section, data analyses are discussed. The analyses include the four data collection techniques employed: tests, class observation, questionnaire and interview.

3.4.1 Pretest and Posttest

ANATES V.4 (version 4) and SPSS V.15 software were utilized to analyze the result of the pretest and the posttest. ANATES V.4 (KarnoTo and Wibisono, 2003) could automatically analyze the reliability, the level of difficulty, discrimination index and distracters' quality of the items. It helped researcher to identify the quality of the test items quickly and appropriately. The second one, SPSS (Nie and Hull,1968) could easily analyzed quantitative data just by clicking the menu. It was to statistically analyze the difference of means among data through independent t-test and non-parametric test.

The procedures of analyzing the data from the tests were as follows.

1. Data Entry and the Analysis via ANATES software

   Firstly, the coded names of participants, the answers and the key answers were entered into ANATES V.4. Then, the correct answers were scored, one correct answer got one score. The sum of correct answers was divided by the number of the items to have percentage point. Finally, the reliability, the level of difficulty, discrimination index and distracters' quality of the items were analyzed automatically, just by clicking the menu.

2. Independent t-test analysis via SPSS V.15

   The total scores obtained from ANATES V.4 were then statistically analyzed through the SPSS V.15 software. This analysis was intended to see which teaching program, STAD CL or DI, is more effective in improving the students' comprehension in general. It compared the mean scores of total scores of both groups. Since they were two independent groups, independent t-tests were chosen (Hatch and Farhady, 1982 pp. 111-4, Hatch and Lazaraton, 1991 pp. 258-262; Kinnear and Gray, 1994 p. 86). It was considered appropriate because the data were interval and were assumed to have

normal distribution since the participants were more than 30 students (Hatch and Farhady, 1982 p. 98).

3. SPSS independent t-test on every level of Comprehension

Further, to see the improvement of the scores of both STAD CL and DI groups on every level of comprehension, SPSS descriptive statistics and independent t-test were carried out. For the comparison, the students' answers on the raw data (*data mentah*) of ANATES.V4 were categorized into three levels of comprehension: literal, inferential and evaluative, each of which consisted of 10 items. Such number of items was appropriate enough to have descriptive statistics and inferential one (Kinnear and Gray, 1994 p. 83). This analysis was intended to compare the significance of the difference between the mean score both groups on every level of comprehension. Since the data was the same as mentioned previously, independent t-test was run. Then, the result of the analysis was described adequately.

3.4.2 Class Observation

The data from observation consisted of researcher's field notes (Creswell, 2008 p. 224) and the result of reading comprehension quizzes. The analysis of the data was conducted during and after the data collection (Miles and Huberman, 1994 p. 10). The obtained data then were analyzed and described adequately, all of which will be presented in Chapter IV together with the teaching programs. Specifically the data comprised mainly of the teacher's activities and the students' response on the procedure of the teaching programs. The data obtained were presented in a condensed body of information (Emilia, 2005 p. 86) (see Appendix 5a and 5 b) which will be discussed in Chapter IV, together with the teaching programs.

3.4.3 Questionnaire

The data from the questionnaire were analyzed in several steps. Firstly, the questions or the statements in the questionnaires were categorized into the central themes (Alwasilah, 2000 p. 160; Cresswell, 2008 pp. 251): the levels of comprehensions. Then, simple computation and percentage were conducted. The results were tabulated, statistically analyzed and interpreted adequately.

In analyzing the result of the questionnaire, the scores between the positive and negative statements were reversed (Dornyei, 2002 p.43). For positive statements, to follow Dornyei, 'strongly agree' was scored four and 'strongly disagree' was scored one,

meanwhile for negative ones, 'strongly agree' was scored one and 'strongly disagree' was scored four (see also Creswell, 2008 p. 184). All respondents' checks were presented in recapitulation table (see Appendix 9a and 9b). In the end, SPSS.V15 analysis: non parametric Wilcoxon signed rank was done. The result of the analysis will be presented in Chapter V.

3.4.4   Interview

The data from interview were analyzed through several steps. Firstly, recorded data were transcribed. In transcribing the data, the participants were coded to protect their privacy (Silverman, 1993). To follow Miles and Hubberman (1994 pp. 10-11), then, inappropriate or non relevant data were reduced. After that, the data were categorized into central themes which became the main concern of this study (Cresswell, 2008 pp. 251-261); they are the participants' impression on STAD CL or DI and their effects on their comprehension levels. The obtained data were presented in the condensed body of information (Emilia, 2005 p. 86) (see Appendices 16a and 16b).

3.5   Conclusion of the Chapter

This chapter has provided the information about the methodology of the research including the setting, the participants and the designs of the study. It also discusses the data collection techniques and analyses for the study including tests, observation, questionnaire and interview.

The following chapter discusses the teaching procedures as the teaching programs for the study. It describes the steps of STAD CL and Direct Instruction (DI) in teaching reading comprehension together with the result of observation.