CHAPTER III
RESEARCH METHODOLOGY

This chapter highlights the research methodology which consists of the research design, subject of the study, population and sample, research instrument, research procedure, and data analysis. The rationales underlying this research are also described in the subchapter.

3.1.   Research design

The main design of this research was a mixed-method with a sequential explanatory (Hamied, 2017). The explanatory design was meant to find out "how" the outcome occurred. Quantitative data was collected through quasi-experiment. After the researcher analyzed the quantitative data, the researcher elaborated on the data collection with qualitative data gathered. In other words, the research did not only stop on the statistical analysis but also elaborated on the possibilities beyond the number interpreted from the statistics.

The samples were grouped into three: Group1, group2, and group3. Group1 used GBL with inquiry strategy. Group2 used GBL with collaborative learning strategy. Group3 used task-based learning and lecturing without GBL. Teaching method1 emphasized individual task during game-based instruction. Teaching method2 emphasized collaborative group task during game-based instruction. Teaching method3 emphasized collaborative writing without the game-based learning framework. While group1 and group2 were using game-based learning framework, group3 did not use game-based learning framework.

By the nature of inquiry learning, the open-ended question in group1 encouraged interaction in teacher-student form, or what so-called Ping-Pong discussion by Duke (McNeill & Pimentel, 2010). The direction of ping-pong discussion is between a teacher and student. The approach in the inquiry group was teacher-directed which means that the teacher asks questions according to the context of teaching, how the context is investigated, and what to be presented (Dobber, Zwart, Tanis, & van Oert, 2017). Teacher-directed inquiry positions the teacher to be the center that mediates the flow of discussion, though, the inquiry

process occurred in the middle of teaching. In this case, the teacher acted as the mediator to solve the problem in the game.
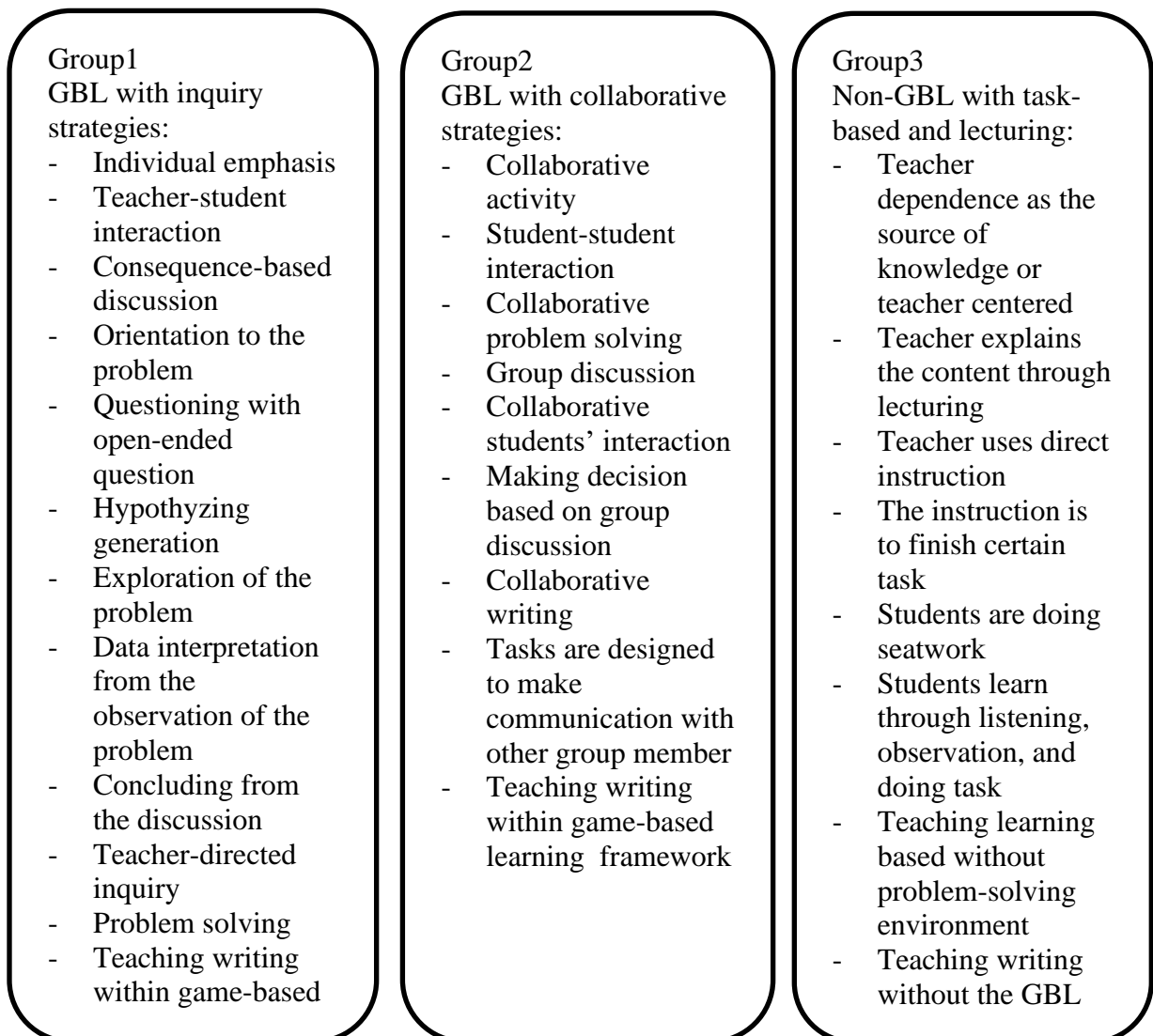
Group1 GBL with inquiry strategies:	Group2 GBL with collaborative strategies:	Group3 Non-GBL with task-based and lecturing:
- Individual emphasis - Teacher-student interaction - Consequence-based discussion - Orientation to the problem - Questioning with open-ended question - Hypothyzing generation - Exploration of the problem - Data interpretation from the observation of the problem - Concluding from the discussion - Teacher-directed inquiry - Problem solving - Teaching writing within game-based	- Collaborative activity - Student-student interaction - Collaborative problem solving - Group discussion - Collaborative students' interaction - Making decision based on group discussion - Collaborative writing - Tasks are designed to make communication with other group member - Teaching writing within game-based learning framework	- Teacher dependence as the source of knowledge or teacher centered - Teacher explains the content through lecturing - Teacher uses direct instruction - The instruction is to finish certain task - Students are doing seatwork - Students learn through listening, observation, and doing task - Teaching learning based without problem-solving environment - Teaching writing without the GBL

Figure 3.1-1 The characteristic of each group

3.2. Subjects of the study

The subjects of this research consisted of 37 students at a private college education in Cilegon, with the distribution of 14 subjects in group1, 11 subjects in group2, and 12 in group3. All subjects were in higher education students in a private college in Cilegon. Their average age was at 18 years old, late teenager. Their mother tongue was Indonesian, which means non-native speaker to English.

3.3. Population and sample

Sampling technique for this research was purposive sampling. The number of samples was chosen by the researcher as well as the institution where this research was conducted. All samples were the higher education students in the second semester in a college in Cilegon.

At first, there was a total of 48 students as samples. Moreover, the subjects were divided into three groups based on the pre-test result: group 1 had 16 students, group 2 had 16 students, and group 3 had 16 students. Group 1 represented game-based learning with inquiry strategies, group 2 represented game-based learning with collaborative strategies, and group 3 was the control group.

To make sure that the samples were homogenous, a test was conducted to measure the students writing score before grouping procedure. Writing test was conducted by writing a story after watching a riddle clip. The instruction was as follow:

> "Write a story from the riddle clip. Use your language abilities to develop your story. Feel free to modify, cut, or add the content of your writing. Use your own language ability in narrating your story but be sure to write in the past tense. The length of the writing is expected more than 100 words written"
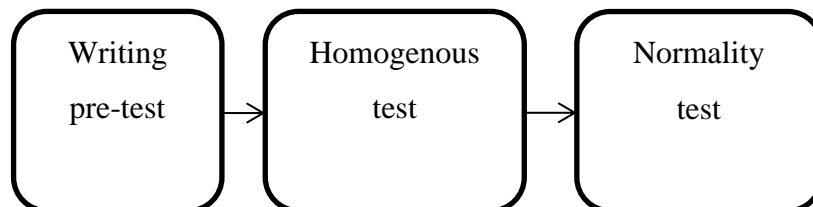
Writing pre-test → Homogenous test → Normality test

Figure 3.3-1 Sampling scheme procedure

After homogenous and normality test, the number of samples was reduced. Group 1 had 14 students, group 2 had 11 students, and group 3 had 12 students. The total of the samples were 37 students. The outliers in the groups were 11 students and indexed. The indexed students were not counted for post-test, questionnaire, and interview.

The samples in game-based learning groups had to fill in a questionnaire for evaluating their learning experience. For the interview, six representative participants were chosen from the low and high scorer after the post-test.

3.3.1. Homogenous of the sample groups

Obtaining an equal group is necessarily essential for experimental research. Although it was impossible to have 'the same' sample for the research, social research had accepted that 'equal' was the factor for conducting experimental research. Hence, in order to obtain equal group member, the researcher conducted homogenous tests of each group and set the samples into equal groups. Therefore, the researcher performed the homogenous test with ANOVA statistics measurement. The homogenous test was to compare the mean of each group. The result from the homogenous test was indicating the significant difference in the experiment. The method of calculation with ANOVA homogenous meets the requirement for interval data.

The statistics were designed to reject the hypothesis which was expecting for H1 to be rejected. The rejected H1 means that there was no difference between the groups, which mean that they were equal groups. Equal groups in this research mean that the groups had the same ability in developing a narrative text. Afterward, from the result of the homogenous test, the students were distributed into GBL with collaborative strategies, GBL with inquiry strategies and the control group without GBL.

To find equal groups, the homogenous test was conducted with the level of significance at 0.05% and two-tailed direction. Whereas, the hypothesis was:

H1: There is no difference between the groups' mean variable.

ANOVA

Result					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.410	2	1.205	.310	.736
Within Groups	132.334	34	3.892		
Total	134.743	36			

Based on the calculation, the significance value of the data showed at $p = 0.73$. Compared with the level of significance at 0.05, the p-value was higher than the level of significance ($p = 0.73 < 0.05$), which means the alternative hypothesis was rejected, and the null hypothesis was accepted. In conclusion, there was no

statistical significance difference between groups as determined by one-way ANOVA ($F(2,34) = 0.310$).

It means that there was no difference between the means of the three groups' variable. Thus, it also means that the three groups' ability in writing narrative was equally distributed based on the mean comparison.

3.3.2. Normality of the samples

The normality of the data had to be conducted to measure the distribution of students in a group. It was meant to reduce the outliers, so the data in a group were equal. An equal group in this research means that the group member had nearly the same ability between one to another. The number of outliers contributes significantly to the normality of the data. Outliers like a high achiever, or low achiever should be distributed evenly among the three groups to create normal data distribution, or removed.

A test that was conducted to measure the normality for this research was Kolmogorov-Smirnov since the data was interval.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group1	.122	14	.200[*]	.978	14	.958

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Table 3.3-1 Normality test result from group 1

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
Group2	.114	11	.200[*]	.943	11	.558

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Table 3.3-2 Normality test result from group 2

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group3	.176	12	.200[*]	.927	12	.347

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Table 3.3-3 Normality test result from group 3

The normality data of group 1, group 2, and group 3 were 0.200. The result showed the normality was above > 0.05. In conclusion, the distribution of the data was normal. In other words, all data in each group had been tested and normally distributed.

Then, the researcher set group 1 as the game-based learning with inquiry learning strategies, group 2 as the game-based learning with collaborative learning strategies, and group 3 as the control group with task-based learning and lecturing.

3.4. Research instrument

The instruments of this research included pre- and post-test, video recording, questionnaire, and interviews. Pre- and post-test was designed to answer research question number one. Observation was designed to answer research question number two. Questionnaire and interview were designed to answer research question number three.

3.4.1. Pre- and post-test

At the pre-test stage, the samples were required to write a narrative text. The students were expected to write a short narrative text after watching a riddle clip. The clip was validated through vocabulary measurement in vocab profile and vocabulary length at lextutor.ca. The scoring instrument was prepared for the narrative beginner writer. The rubric was adapted from NCTA corporation (2004). The criteria were scaled from the original according to the need of scoring. The rater to score the narrative text was two with the same background. The raters were tested for the reliability with Interrater Coefficient Correlation before scoring the pre-test.

Homogenous and normality test was run from the result of the pre-test. ANOVA homogenous test was conducted with a level of significance at 0.05. The direction of the test was two-tailed. After the homogenous test, the normality test was conducted to find out students' normal distribution. All measure was using SPSS version 16.0.

At the post-test, the students were required to write a story within 45 minutes. Two same raters conducted the scoring at the pre-test. The rubric was also the same as the pre-test. One-way ANOVA with a level of significance 0.05 was conducted. Then continued with Sceffe post hoc to find which among the groups had the significant difference.

3.4.2. Observation

At the treatment stage, the classroom interaction was recorded digitally. The recording was using a mobile phone with a tripod. The researcher conducted the teaching-learning. While the researcher was teaching the content, the researcher observed and took notes accordingly.

3.4.3. Questionnaire

Students in the game-based groups had to fill in a questionnaire to examine the students' experience after game-based learning treatment. The number of items in the questionnaire was 18. The questionnaire was piloted and scored its validity and reliability. The validity of the questionnaire was conducted with Pearson Coefficient Correlation with the level of significance at 0.05, and the reliability of the questionnaire was conducted with Cronbach's Alpha with a level of significance at 0.07. Besides, the questionnaire was also validated by two experts. The first expert was to validate the language of the questionnaire, and the second expert was to validate the content and context.

The questionnaire for this research was aimed at examining students' experience toward game-based learning. The questionnaire was adapted from Flow theory by Chikzentmihalyi (1990). Schaffer, his student, developed the seven flow conditions based on the theory. The seven condition flow includes: knowing what to do, knowing how to do it, knowing how well you are doing,

knowing where to go, high perceived challenges, high perceived skills, freedom from distractions (Schaffer, 2013). He also published a measure, the Flow Condition Questionnaire (FCQ) on white paper.

The questionnaire was five-scaled Likert scale, ranged from strongly agree to strongly disagree. All of the items' direction was positive. The students were asked to answer based on their experience during the game-based learning process.

After the calculation of the questionnaire, its frequency table was built by using SPSS. The procedure was to see the students' perception of their learning experience within game-based learning.

Figure 3.4-1 Flow loop model with flow conditions

3.4.4. Interview

At the interview stage, students were interviewed to answer the question about the game-based learning process. The interview was structured, which means that the participants received the same question. Then, the answer was coded and reported according to the finding.

The interview was conducted on three participants in game-based learning groups with the six total participants. The participants were purposively selected

from the low and high scorer in the experiment groups. Thus, the interview was conducted after the post-test.

The data was recorded through mobile phone software. The result of the interview was summarized in short, and only a small clip of representation of the interview was displayed in the discussion.

The design of the interview was a structured interview, whereas a guideline had been set for all interviewees to answer. A guideline of the structured interview is presented in Appendix II. The interview's questions were developed around the learner's experience on game-based learning and how they perceived it. The questions are open-ended with a total of 24 questions, including the background of the interviewee.

The interview data display is presented in a sequence. Although the subjectivity of data analysis in the interview could not be separated from the researcher's interpretation, the researcher had reported the data as objectively as possible with its synthesis with other sources of data such as questionnaire and observation.

3.5. Research procedure

1. The researcher visited the institution which the research took place. The researcher interviewed some lecturer and staffs about teaching-learning in the institution especially to estimate the students' level proficiency. Ascertaining students' level proficiency was required to measure the level of difficulty of the game and instruction design adjustment for the treatment. Besides, the visit was to check the availability of the tools required for the research such as projector, policy for mobile phone in the classroom, and internet connection.

2. A pilot study was launched to test the questionnaire's validity and reliability. The pilot study was at the same institution with different subjects. The pilot study was a one-time teaching. In the end, the students were asked to fill in Flow questionnaire.

3. The pre-test required students to write a story about a riddle clip played. After watching the clip, the instruction of the writing was as follow:

> "Write a story from the riddle played. Use your language abilities to develop your story. Feel free to modify, cut, or add the content of your writing. Use your own language ability in narrating your story but be sure to write in the past tense. The length of the writing is expected more than 100 words written."

During writing, no tools and communication device were allowed. It was expected that students wrote independently for the production.

4. The result of the pre-test was tested for homogeneity and normality. The purpose was to group the students into GBL with inquiry strategies group, GBL with collaborative strategies group, and control group. The design allowed for the equal writing skill distributed evenly among the group.

5. Once the preparations were complete, the treatment was conducted within four meetings; the time allocated per meeting was 90 minutes. As mentioned in previous research design, group1 applied game-based learning with inquiry strategy, group2 applied game-based learning with collaborative strategy, and group3 applied task-based with lecturing as the control group.

   Group1 and group2 were the experiment groups. Meanwhile, group3 was the control group.

6. During the treatments, the data was recorded. Coding and memoing were conducted. While the treatments were on progress, the researcher explored the response toward game-based learning implementation from the teacher's point of view.

7. At the post-test, a test was conducted to measure students' improvement in writing through writing a story with the following instruction:

> "Write a story from the riddle played. Use your language abilities to develop your story. Feel free to modify, cut, or add the content of your writing. Use your own language ability in narrating your story but be sure to write in the past tense. The length of the writing is expected more than 100 words written"

The criteria were the same as implemented in the pre-test. The duration of writing was 45 minutes. During writing, students were expected to write individually in all groups. Also, no tools were allowed.

8. After post-test writing, students in the experimental groups were asked to fill in Flow Condition Questionnaire about game-based learning. The questionnaire was asking about their experience during the teaching-learning process. Flow Condition Questionnaire measured how the students perceive teaching-learning within GBL framework.

9. At last, six students were selected for an interview. The participants of the interview were the students in experimental groups. The interview asked how the students' perspective from the strategies implemented, especially around game-based learning. Three random samples were selected randomly from each of the group. The samples were selected to represent the result from the post-test. Representative students from the low and high scorer were selected to join the interview.

Pilot study → Pre-test → Treatment → Post-test → Questionnaire and Interview

Figure 3.5-1 Research timeline

3.6. Data analysis

The data was described through descriptive and inferential statistics. The pre-test and post-test were statistically measured through one-way ANOVA with the level of significance at 0.05. The software that was used was SPS version 16.0. For pre-test, a hypothesis from ANOVA homogenous test accepted H0. The hypothesis was:

H0 = there is no difference between variables' means

After the homogenous test was conducted, a normality test was conducted with Kolmogorov-Smirnov statistics by using SPSS. The normality test's purpose was to distribute the students evenly. Hence, the groups had equality students' distribution.

After the treatments, the post-test was conducted to compare the means from the experiment groups. The post-tests results were calculated with one-way ANOVA with the level of significance at 0.05. The procedure was to answer research question 1 which asked the difference in means comparison between the groups. Based on the calculation, the result was as follows:

H1 = there is a difference between variables' means

Because H0 was rejected at post-test, which indicated that there was a difference, another statistics measurement was conducted to find out the difference. The procedure was using Scheffe post hoc with the level of significance at 0.05. Post hoc procedure was to answer research question 2 which asked the difference between GBL with inquiry strategies and GBL with collaborative strategies. The results were interpreted to see different means among the groups. In this case, the process was to see which groups had the benefit from the treatment. The results told the difference of the treatment of all groups. In short, all the groups' means were compared statistically with ANOVA to compare the result between the control and experiment groups.

Another data was from observation. The observation was described based on the finding on the learning process. The observation is to answer research question 3 which asked how the process of teaching-learning with GBL. The finding was presented in descriptive analysis. The analysis was separated between GBL with inquiry strategies and GBL with collaborative strategies. No statistics measure in reporting observation.

Another data was questionnaire and interview. The questionnaire was analyzed by classifying the answer and put some answers in descriptive statistics. The validity and reliability of the questionnaire was measured by statistics. The validity used Pearson Coefficient Correlation and the reliability used Cronbach's Alpha. The result removed one item from the questionnaire because it did not reach the level of significance at 0.05. Then, frequency answer from the questionnaire was displayed on the table for each item in the question. The analysis was conducted in each indicator thematically. On the other hand, interview data were recorded. The data was coded analyzed descriptively. Since the interview was structured, the analysis was sequenced. The question was

organized orderly. Descriptive analytics was conducted to identify the students' perception of their experience when learning with the game-based method. Both instruments were to answer research question number 4.

3.6.1. Raters' profile and reliability

To obtain less subjective data interpretation, the researcher had decided to leave the pre- and post-test text scoring to other raters. The number of raters in this study was two. They were the lecturers at the institution where this research was conducted. The raters had the same background of teaching experience for more than five years as English lecturers.

The method to obtain a reliable measure for the raters was by training them. The training was in a small discussion on how to rate by using the rubric provided.

Also, the reliability of the two raters was tested in statistics with the Intraclass Correlation Coefficient (ICC). The measurement was using Pearson Correlation Coefficient to see the reliability between the two raters' score.

Correlations

		Rater1	Rater2
Rater1	Pearson Correlation	1	.666[*]
	Sig. (2-tailed)		.035
	N	10	10
Rater2	Pearson Correlation	.666[*]	1
	Sig. (2-tailed)	.035	
	N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

Table 3.6-1 Pearson correlation coefficient calculation

The result of Pearson test was 0.666. Based on the statistical analysis, both raters had a strong positive correlation in scoring writing product which means that if the rater1 scores high for the text, rater2 will likely score high as well and vice versa. From the analysis, it is concluded that the two raters had excellent reliability in scoring writing product.

0 - 0.3	Weak correlation
0.3 - 0.6	Medium correlation
0.5 – 1	Large correlation

Table 3.6-2 Pearson correlation table

3.6.2. Pre- and post-test clip reliability

In the pre- and post-test, students were required to watch a riddle clip before writing narrative text. In order to provide the reliability of the clip, the researcher had tested the content of the clip with an online vocabulary tool at lextutor.ca. The tool measured the vocabulary frequency in the content of the clip. To have good clip reliability, the clip was selected to have the same amount of vocabulary frequency for both pre- and post-test.

The researcher had to make the transcription from the clip. The clip was transcribed and uploaded to lextutor.ca. Pre-test clip has 106 words, and the post-test clip has 115 words. The length of the text is similar to each other, which means it has an equal length for implementation. After uploading the text, the researcher received the amount of K-2 in the text. K-2 is the most frequent words in English. Thus, by calculating the amount of K-2 in a text, the researcher could have some evidence about the difficulty of the text for students at to understand.

The result from measuring vocabulary in the text showed that the pre-test clip has 89.63% of K-2 words. Meanwhile, the result from measuring vocabulary in the text showed that the post-test clip has 92.17% of K-2 words.

Figure 3.6-1 Pre-test clip K-2 measurement

Figure 3.6-2 Post-test clip K-2 measurement

With the result, the researcher had assumed that the clip for pre-test had a little difference in the difficulty level as the clip for post-test. Likewise, there were not many gaps between the pre-test clip text and the post-test clip text for students to understand.

It was extremely difficult to find two clips with the exactly the same level of words frequency with a similar topic. Therefore, although there was a little different (2.54%) between the text in pre- and post-test, the researcher had decided to use both clips for writing a narrative.

3.6.3. Normality and validity of the questionnaire
3.6.3.1. Questionnaire's validity

The pilot survey was conducted with 17 students to measure the validity and reliability of the questionnaire. The validity of the questionnaire was conducted with Pearson Correlation Coefficient. The calculation with the statistics was using SPSS. The calculation was conducted at two-tailed with error probability at 0.05.

Based on the pilot survey, 17 out of 18 questions were valid at 0.05 level of significance. Question number eight was invalid. The question had the indicator to reveal where the student should go if they encounter difficulty during learning. The p-value of the question was at 0.363. Hence, the item was removed from further use ($p = 0.363 < 0.05$) leaving 17 questions for the real test.

3.6.3.2. Questionnaire's reliability

To measure internal consistency, Cronbach's Alpha was deployed. The reliability of the questionnaire was conducted with Cronbach's Alpha because the data type was ordinal. The calculation was automatically using SPSS version 16.0. The reliability of Cronbach's Alpha was set at 0.70 for social science.

Reliability Statistics

Cronbach's Alpha	N of Items
.897	18

The Alpha coefficient of the questions was at 0.897 which was higher than the level of significance ($p = 0.897 > 0.70$), suggesting that the items have high consistency. The result of the measurement had revealed that the question items were reliable.