CHAPTER III

METHODOLOGY

This chapter describes several points related to the research methodology. The points that are going to be explained are the research method and design, population and samples, research instrument, and data collection and analysis.

3.1 Research Method and Design

This study is a quantitative study to investigate the relationship between students' foreign language aptitude and their English achievement in the form of summative test scores. Therefore the study applies a descriptive statistic method with a correlational statistic as mentioned by Arikunto (1988) that:

> "Correlational research is a research that is intended to find out whether or not there is a correlation. If there is, how firm the correlation is and how significant it is."

The design of the study is *ex post facto* design. For that reason, this design is the most frequently used in studies (Hatch and Farhady, 1982:26). They also state that:

> "Ex post facto designs are often used when the researcher does not have control over the selection and manipulation of independent variable. This is why researcher looks at the type and/or the degree of relationship between the variables rather than at a cause-effect relationships."

The study would like to use students foreign language aptitude measured by foreign language aptitude test to be correlated with their English achievement represented in their English summative test scores. Therefore, the variables investigated in this research are:

1. The students' foreign language aptitude scores measured by the foreign language aptitude test as the independent variable.

2. The students' English achievement shown by their English summative test scores as the dependent variable.

3.2 Population and Samples

This research is conducted on the first grade students of SMUN 14 Bandung as the population. This grade is selected with the assumption that on this level, they are introduced to many new or developed materials in English learning comparing to what they have got in junior high school grade.

The first grade of SMUN 14 Bandung consists of eight classes with approximately 40 students in each class. Since the population was grouped into classes, then the study used cluster sampling in which the unit of sampling was not the individual but rather group of individuals (Borg and Call, 1979:187). Two classes, X-H and X-C, will be chosen as the sample. In this matter, the first class will be used for trying out the instrument and the other class as the real sample. Forty-two subjects in the last class are randomly drawn as the true sample and this is in line with McMillan and Schumacher (1984) who stated that for a correlational research, the samples must be at least 30 subjects (Ruseffendi, 1998:92).

3.3 Research Instrument

As the research instrument, a kind of test will be utilized in this research. Ruseffendi (1988:102) asserts that test and non-test type can be used to get the data needed in a research. Based on this description, the test type used is in the form of foreign language aptitude test.

3.3.1 The Foreign Language Aptitude Test

The MLAT was developed for measuring the native speakers of English (Broersma, 1989). Thus, a foreign language aptitude test for Indonesian is going to be utilized and used as the instrument in this research. This test consists of three parts i.e. paired associates, language analysis, and sound symbol association. Those three parts are modeled after the sample of the short version of Modern Language Aptitude Test. This short version was applied by Sasaki (1993) in which she adapted the form and also developed the questions as the new language aptitude battery for Japanese speaker (Sasaki, 1993:326). Those three parts are:

1. Paired associates, which focuses on the rote of memory aspect of learning foreign languages. In this part, Makasai language will be drawn on for the questions. This language is used with the assumption that it is new to the subject tested. This part consists of 10 items and the sample question of this part is:

   Pre-test activity: The instructor shows the students ten words in a new language and gives some time (40-50 seconds) to them to memorize the way those words are written and also the meaning of each words in Indonesian. (Done by showing a word list as the media), for example:

dia'	= cangkul	a'a-gua	= langit-langit
kado	= gergaji	fi-a'a	= mulut

   Instruction: Based on the words list shown before, choose the correct word that has the same meaning as the key word in each question! (Written and read by the instructor in Indonesian language)

   Key word : fi-a'a

   Options : a. cangkul c. langit-langit
   b. mulut d. gergaji

2. Language analysis, which tests students' recognition, analogy, and understanding of far greater range of syntactic structures. This part is going to assess students' ability in recognizing and analyzing the syntactic structure of sentences in Indonesian language. The sentences are adapted from the collection of Indonesian language test questions used for UMPTN (Ujian Masuk Perguruan Tinggi Negeri/Selection Test for Entering Public University) preparation book. This part consists of 100 items and the sample question is:

   Instruction: Choose a word in the sentence that has the same function as the underlined word in the key sentence! (Written and read by the instructor in Indonesian language)

   Key sentence : Merokok itu berbahaya bagi kesehatan.

   Question : Setelah makan, ayah merokok sambil membaca koran di beranda depan.
   A B C D E

3. Sound symbol association that requires students' ability to associate sound with symbols and depends somewhat on knowledge of Indonesian vocabulary. This part consists of 60 items and the sample question is:

   Instruction: Choose the appropriate word, which is the correct meaning or has close relationship with the word read by the instructor! (Written and read by the instructor in Indonesian language)

   Word : Bank read as /baŋ/

   Options :
   a. investor
   b. buku
   c. tabungan
   d. saham

3.4 Data Collection and Analysis

In collecting the data, a foreign language aptitude test will be administered as the first step. Secondly, students' English summative scores are going to be collected from their English teacher evaluation data.

Having collected the data, they will be analyzed through the following steps.

1. Determining the scores.
2. Classifying the scores.
3. Displaying the data by using frequency distribution.
4. Calculating descriptive statistics to describe how each student perform on both tests.
5. Computing the coefficient correlation between students' foreign language aptitude and their English summative test scores by the use of Pearson Product Moment coefficient correlation or Spearman Rank-Order correlation formula. The formula to obtain the coefficient correlation are:

   Pearson product moment coefficient of correlation (r)

$$r_{xy} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum y^2 - (\sum Y)^2)}}$$

X: Foreign language aptitude test score

Y: English summative test score

N: Sample

(Hatch & Farhady, 1982:201)

Spearman rank order correlation, ρ(rho):

$$\rho = 1 - \frac{6\left(\sum D^2\right)}{N\left(N^2 - 1\right)}$$

D= Difference of the rank

N= Number of subject

(Hatch & Farhady, 1982:206)

3.5 Trying out the Instrument

As stated by Harris (1969:13), a good test possesses several traits: reliability, validity, and practicality. This means that any test used must be proper in terms of dependable or stable, objectives, and applicable to particular situation. Furthermore, he also states that in order to be acquainted with the requirements of a good test, the objective test becomes a standard when it consists of pre-tested material. In other words, all items must have been firstly tried out.

3.5.1 Reliability

Test reliability is defined as "the extent to which the results can be considered consistent or stable" (Brown, 1988:98). To estimate test reliability, the writer used split-half method in which the steps are:

- Scoring the even-numbered and odd-numbered items on a test separately;
- Calculating the correlation between those two subjects;
- The result of the coefficient is calculated by using Spearman-Brown prophecy formula to determine the reliability of the full test. The formula is:

$$r_k = \frac{2r_1}{1+r_1}$$

r_k = reliability of the full test
r_1 = reliability of the half test

(Hatch and Farhady, 1982:247)

- Interpret the result of the computation based on the following criteria of reliability:

0.00—0.20 : Very low

0.20—0.40 : Low

0.40—0.70 : Moderate

0.70—0.90 : High

0.90—1.00 : Very high

(Guilford in Ruseffendi, 1994:120)

- Determining the Discriminating Power Index

In addition, a test is said to be reliable if the items are able to discriminate between high- and low- examinees. The following is the formula to estimate the discriminatory power of item:

$$D = \frac{RU}{TU} - \frac{RI}{TI}$$

D = Discriminating power index

Ru = The number of the upper group who answer the item correctly

RI = The number of the lower group who answer the item correctly

TU = The total number of upper group

TI = The total number of lower group

(Arikunto, 1986:201-205)

Afterward, the result is classified according to the following classification:

0.00—0.20 : Poor

0.20—0.40 : Satisfactory

0.40—0.70 : Good

0.70—1.00 : Excellent

(Arikunto, 2001:218)

3.5.2 Validity

Test validity is defined as " the degree to which a test measures what it claims to be measuring" (Brown, 1998:101). For that purpose, the instrument was carefully constructed based on the variable of the study investigated.

To find out the validity of the items, Pearson Product Moment formula is used:

$$r_{x_y} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum y^2 - (\sum Y)^2)}}$$

X= The score of the item

Y= Total score

Thus, the result is compared by using the T-Test formula to the t-table at the α level of .05 (Arikunto, 1986:69). The T-test formula is:

$$t_{observed} = \frac{r\left(\sqrt{n-2}\right)}{\sqrt{1-r^2}}$$

In addition to the test validity, a test is also said to be valid if the items neither too hard nor too easy for the population tested. Therefore the item difficulty or difficulty index must be determined by computing the percentage of the sample who answered each item correctly. The formula used to calculate the difficulty index is:

$$P = \frac{R}{T}$$

P= Difficulty index

R= Number of correct items

T= Number of responses

(Gronlund, 1968:102)

Subsequently, the result is categorized based on the following category:

$0.00 \leq P \leq 0.30$: Difficult

$0.30 \leq P \leq 0.70$: Moderate

$0.70 \leq P \leq 1.00$: Easy

(Arikunto, 2001:208)

3.5.3 Practicality

For the purpose of practicality, the test used in this study is an objective test in the form of multiple choice. This test is used since it is considered to be easily scored and interpreted.

3. 6 The Result of the Try Out

In order to fulfill the qualification of a standard objective test, the instrument was firstly tried out to class X-H, the subjects of the same kind as those for whom the tests were being designed. The try out was held on 8th June 2005. Besides allowing the writer to check the items in terms of difficulty level and discriminating power index, pretesting also provided an opportunity to check the test directions and to figure out the estimated time required for the students to work the items of the test.

The results revealed that the direction of the test was relatively clear to the subjects since they knew how to work the test. In addition to that, the time allowance required for the test was approximately 30 minutes.

3.6.1 Reliability and Discriminating Power Index

To check the reliability of the test, split-half method (Spearman-Brown Prophecy formula) is used by firstly applying it separately towards the two last parts (the sentence analysis and the spelling cues) of the test and excluding the first part (the paired associates). This is because Paired Associates is a short-term memory performance task and according to Boring and William (1961), as cited in Guilford (1954:147-148), it is categorized as a different test type. In this matter, following and adapting the standard of paired associates, which focuses on the time allotment, in the

original version of MLAT was preferable due to the assumption that MLAT is a standardized test.

The results showed that the reliability coefficient for those two parts were 0.95 and 0.82. These resulting values are considered very high and high, indicating that those two parts are adequately reliable to measure students' ability in analyzing sentences and associating sound and meaning. This also occurs due to the length of items used in the instrument, therefore, the selection of items used in the real test was easily done. See appendix for more detail analysis.

By the use of Gronlund's formula, the discriminating power index of part II and III share different categories. To make it clear, the detail scores and categorization of the discriminating power index is displayed in the following table.

Table 3.6.1.1
Discriminating Power Index of Part II "Sentence Analysis"

No	DP	Classification	No	DP	Classification	No	DP	Classification	No	DP	Classification
1	.10	Poor	26	.40	Good	51	.05	Poor	76	.10	Poor
2	.05	Poor	27	.15	Poor	52	.30	Satisfactory	77	.45	Good
3	.25	Satisfactory	28	.10	Poor	53	.15	Poor	78	.50	Good
4	.25	Satisfactory	29	.05	Poor	54	.35	Satisfactory	79	.20	Satisfactory
5	.35	Satisfactory	30	.20	Satisfactory	55	.45	Good	80	.10	Poor
6	.00	Poor	31	-.05	Poor	56	.45	Good	81	.40	Good
7	.10	Poor	32	.25	Satisfactory	57	.05	Poor	82	.15	Poor
8	.05	Poor	33	.20	Satisfactory	58	.35	Satisfactory	83	.30	Satisfactory
9	.30	Satisfactory	34	.20	Satisfactory	59	.35	Satisfactory	84	.45	Good
10	.25	Satisfactory	35	.35	Satisfactory	60	.10	Poor	85	.20	Satisfactory
11	.55	Good	36	.45	Good	61	.05	Poor	86	.35	Satisfactory
12	.30	Satisfactory	37	-.05	Poor	62	.30	Satisfactory	87	.25	Satisfactory
13	.50	Good	38	.25	Satisfactory	63	.30	Satisfactory	88	.25	Satisfactory
14	-.05	Poor	39	.15	Poor	64	.55	Good	89	.30	Satisfactory
15	.25	Satisfactory	40	.55	Good	65	.20	Satisfactory	90	.35	Satisfactory
16	.40	Satisfactory	41	.15	Poor	66	.35	Satisfactory	91	.10	Poor
17	.35	Satisfactory	42	.10	Poor	67	.15	Poor	92	.15	Poor
18	.30	Satisfactory	43	.20	Satisfactory	68	.20	Satisfactory	93	.20	Satisfactory
19	.25	Satisfactory	44	.25	Satisfactory	69	.00	Poor	94	.20	Satisfactory
20	.15	Poor	45	.20	Satisfactory	70	.15	Poor	95	.25	Satisfactory

21	.20	Satisfactory	46	.30	Satisfactory	71	.10	Poor	96	.00	Poor
22	.40	Good	47	.15	Poor	72	.70	Excellent	97	.40	Good
23	.55	Good	48	.40	Good	73	.00	Poor	98	-.15	Poor
24	.15	Poor	49	.55	Good	74	.30	Satisfactory	99	.25	Satisfactory
25	.25	Satisfactory	50	.40	Good	75	.15	Poor	100	.10	Poor

Table 3.6.1.2
Discriminating Power Index of Part III "Spelling Cues"

No	DP	Classification	No	DP	Classification
1	.20	Satisfactory	31	.20	Satisfactory
2	.00	Poor	32	.25	Satisfactory
3	.00	Poor	33	.15	Poor
4	.15	Poor	34	.25	Satisfactory
5	.10	Poor	35	.15	Poor
6	.35	Satisfactory	36	.05	Poor
7	.20	Satisfactory	37	-.05	Poor
8	.00	Poor	38	.30	Satisfactory
9	.15	Poor	39	.20	Satisfactory
10	-.05	Poor	40	.00	Poor
11	.10	Poor	41	.05	Poor
12	.25	Satisfactory	42	.25	Satisfactory
13	.15	Poor	43	.00	Poor
14	.55	Good	44	.05	Poor
15	.30	Satisfactory	45	.20	Satisfactory
16	.15	Poor	46	.20	Satisfactory
17	-.05	Poor	47	.20	Satisfactory
18	-.25	Satisfactory	48	.20	Satisfactory
19	.15	Poor	49	.50	Good
20	.45	Good	50	.10	Poor
21	.10	Poor	51	.20	Satisfactory
22	.15	Poor	52	.00	Poor
23	.05	Poor	53	.05	Poor
24	.10	Poor	54	.05	Poor
25	.10	Poor	55	-.05	Poor
26	.45	Good	56	.25	Satisfactory
27	.15	Poor	57	.10	Poor
28	-.20	Poor	58	.00	Poor
29	.45	Good	59	.05	Poor
30	.05	Poor	60	.30	Satisfactory

3.6.2 Validity and Difficulty Level

As mentioned earlier, Pearson Product Moment formula was applied in calculating the validity of the test items, and yet excluded the "Paired Associates" part. Having found out the result, it was then compared to the t-table by means of the T-test formula. In this matter, the t-observed is compared to the t-table of 40 samples at the α level of .05 (1.686). The result of validity calculation including the classification of difficulty index of each items are presented in the following tables. For more detail analysis, see appendix.

Table 3.6.2.1
Validity and Difficulty Index of Part II "Sentence Analysis"

No	t-obs	DI	Category	No	t-obs	DI	Category	No	t-obs	DI	Category	No	t-obs	DI	Category
1	1.237	.95	I/E	26	2.606	.35	V/D	51	1.867	.83	V/E	76	1.945	.85	V/E
2	1.427	.83	I/E	27	2.271	.83	V/E	52	4.525	.65	V/M	77	2.799	.63	V/M
3	1.405	.28	I/D	28	3.749	.90	V/E	53	2.735	.78	V/E	78	3.367	.50	V/M
4	2.598	.33	V/D	29	0.119	.33	I/D	54	3.179	.43	V/M	79	1.569	.25	I/D
5	3.072	.73	V/E	30	2.682	.10	V/D	55	2.983	.58	V/M	80	2.276	.95	V/E
6	1.245	.35	I/D	31	0.227	.93	I/E	56	4.527	.63	V/M	81	5.493	.80	V/E
7	-0.002	.25	I/D	32	2.895	.78	V/E	57	2.239	.88	V/E	82	4.200	.93	V/E
8	1.802	.78	V/E	33	3.139	.75	V/E	58	4.359	.83	V/E	83	4.072	.80	V/E
9	1.775	.25	V/D	34	1.268	.25	I/D	59	3.330	.73	V/E	84	4.318	.43	V/M
10	0.833	.68	I/M	35	1.962	.43	V/M	60	1.653	.70	I/E	85	0.599	.30	I/D
11	3.383	.53	V/M	36	2.993	.78	V/E	61	2.368	.83	V/E	86	3.093	.53	V/M
12	2.974	.65	V/M	37	0.548	.68	I/M	62	3.435	.70	V/E	87	3.391	.68	V/M
13	4.723	.75	V/E	38	2.059	.58	V/M	63	2.425	.55	V/M	88	1.970	.68	V/M
14	0.128	.53	I/M	39	3.629	.93	V/E	64	4.189	.68	V/M	89	2.596	.80	V/E
15	1.620	.58	I/M	40	3.504	.53	V/M	65	2.946	.75	V/E	90	2.456	.53	V/M
16	2.127	.50	V/M	41	3.811	.93	V/E	66	3.537	.63	V/M	91	1.313	.40	I/M
17	3.104	.73	V/E	42	1.881	.85	V/E	67	1.385	.23	I/D	92	3.014	.93	V/E
18	2.207	.75	V/E	43	1.639	.30	I/D	68	0.595	.25	I/D	93	2.628	.80	V/E
19	2.806	.83	V/E	44	1.467	.63	I/M	69	0.644	.30	I/D	94	5.198	.90	V/E
20	1.939	.73	V/E	45	2.103	.80	V/E	70	2.979	.88	V/E	95	2.568	.68	V/M
21	2.697	.45	V/M	46	2.533	.70	V/E	71	0.935	.20	I/D	96	-0.234	.40	I/M
22	2.892	.40	V/M	47	1.861	.73	V/E	72	6.689	.65	V/M	97	2.756	.55	V/M
23	3.073	.48	V/M	48	2.968	.35	V/D	73	1.598	.60	I/M	98	-1.244	.08	I/D

24	2.276	.88	V/E	49	4.397	.58	V/M	74	3.944	.80	V/E	99	3.170	.13	V/D
25	2.435	.33	V/D	50	3.501	.70	V/E	75	1.983	.23	V/D	100	1.723	.75	V/E

Table 3.6.2.2
Validity and Difficulty Index of Part III "Spelling Cues"

No	t-obs	DI	Category	No	t-obs	DI	Category
1	1.968	.65	V/M	31	0.814	.60	I/M
2	-	1.00	V/E	32	1.652	.48	I/M
3	1.075	.75	I/E	33	2.515	.93	V/E
4	2.331	.58	V/M	34	3.195	.83	V/E
5	3.075	.80	V/E	35	2.122	.83	V/E
6	3.782	.58	V/M	36	1.472	.98	I/E
7	1.571	.75	I/E	37	0.077	.28	I/D
8	0.888	.90	I/E	38	4.138	.85	V/E
9	0.371	.93	I/E	39	1.571	.75	I/E
10	-0.966	.98	I/E	40	-0.555	.55	I/M
11	2.334	.95	V/E	41	1.472	.98	I/E
12	2.212	.83	V/E	42	2.868	.88	V/E
13	0.701	.93	I/E	43	-	1.00	V/E
14	3.598	.58	V/M	44	0.681	.48	I/M
15	2.783	.80	V/E	45	4.715	.90	V/E
16	1.620	.93	I/E	46	2.597	.80	V/E
17	-0.561	.83	I/E	47	3.699	.90	V/E
18	-1.320	.78	I/E	48	2.995	.90	V/E
19	1.028	.73	I/E	49	3.570	.60	V/M
20	2.605	.48	V/M	50	1.574	.95	I/E
21	1.823	.80	V/E	51	4.178	.90	V/E
22	2.039	.78	V/E	52	0.023	.55	I/M
23	1.472	.98	I/E	53	0.516	.98	I/E
24	2.021	.95	V/E	54	0.890	.98	I/E
25	0.330	.95	I/E	55	0.434	.88	I/E
26	4.569	.63	V/M	56	2.303	.83	V/E
27	1.146	.88	I/E	57	0.793	.85	I/E
28	-0.765	.80	I/E	58	-	1.00	V/E
29	3.440	.73	V/E	59	-0.037	.98	I/E
30	1.368	.73	I/E	60	1.506	.55	I/M

t-obs : t-observed
DI : Difficulty Index
I/V : Invalid/Valid

E : Easy
M : Moderate
D : Difficult

As asserted by Hatch and Larazaton (1991: 449), the poor items, including those, which were found to be easy or difficult, should be revised or discarded. Thus, for the real test, several items from those two parts, which were considered valid and had the category easy or difficult are selected and modified (either stems and options) to be re-used. As the result, the real test consists of 50 items comprising 10 items in "Paired Associates", 25 items in "Sentence Analysis", and 15 items in "Spelling Cues".