Designing and Evaluating
Quantitative Research
in Education

(3rd Revision)

M. Holandyah, M.Pd

Penerbit dan Percetakan

NoerFikri Offset

Designing and Evaluating Quantitative Research in Education (3nd Revision)

M. Holandyah, M.Pd

Desain Cover : Sigit Dwi Sucipto
Setting dan tata letak : Toni Wicaksono

Perpustakaan Nasional Katalog dalam Terbitan (KDT)
Anggota IKAPI (No. 012/SMS/13)

Dicetak oleh Noer Fikri Offset, Palembang

Noer Fikri Offset

Jl. Mayor Mahidin No. 142
Palembang – Indonesia ✉ 30126
Telepon : 0711 366625
Faks : 0711 366625
Email : noerfikri@gmail.com

Cetakan ke 3, Maret 2021

ISBN : 978-602-1307-05-2

FOREWORD

Many thanks to Allah SWT, the God of the universe for giving His great blessing and almighty in completing this book. This book entitles "**Designing and Evaluating Quantitative Research in Education (3rd Revision)**". This book is intended to help students of English education study program in designing and evaluating their research study since it is arranged based on a research outline of an experimental study. Each topic in this book will guide them in determining an experimental research title, formulating a research problem, choosing their research design, selecting their research sample, and testing their research hypothesis.

It is expected that this book can be as guidance for anyone who has problem in analyzing and testing their research hypothesis with the assistance of SPSS application program. Furthermore, it is also intended that this book can give more knowledge and information for those who are designing and evaluating their experimental research study.

Finally, the writer realizes that this book is far from perfectness. Criticism and suggestions are eagerly expected to make this book better in the future.

Palembang, Maret 2021
MH

CONTENTS

>><<

LIST OF APPENDICES

>><<

LIST OF TABLES

LIST OF FIGURES

>><<

INTRODUCTION

A quantitative research refers to a systematic empirical investigation of social phenomena via statistical, mathematical or computational techniques. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena.

There are several types of quantitative research. They are: survey research, correlation research, causal-comparative research, and experimental research. Each type has its own typical characteristics. In experimental study, it provides a method of hypothesis testing. After experiments define a problem, they propose a tentative answer, or hypothesis.

After the treatment has been administered for an appropriate length of time, researchers observe or measure the groups receiving different treatments to see if they differ. If the average scores of the two groups on the posttest do differ and researchers cannot find any sensible alternative explanations for this difference, so they can conclude that the treatment did have any effects and is likely the cause of the difference.

At the end, a hypothesis testing is formulated based on a research problem. In testing the hypothesis comparing two means scores of two independent groups, “t-test” is used to compare two means scores of two independent groups. They are: **independent sample t-test**; is used to measure *means significant difference or influence* comparing two means scores of two independent groups. While, **paired sample t-test**; is used to measure *means significant improvement, means paired differences*, and *means significant difference* comparing means scores in each group after having matched the samples. However, Analysis of Variance (ANOVA) is used to compare two means scores of more than two groups. They are: **One-way ANOVA**; is used to measure *significant difference* comparing more than two variables in one group, Analysis of Covariance (**ANCOVA)**; is used to measure *significant different effects* comparing more than two variables in two groups, **Two-way ANOVA**; is used to measure *significant interaction effects* which has one moderator variable and two independent groups, and **Two-way MANOVA**; is used to measure *significant main effects* which has two moderator variables and two independent groups. Then, there are two more research instruments used in testing the hypothesis. They

are: **Regression Analysis**; is used to measure the *effects, impacts, influence*, or *significant linear relationship* comparing two or more independent groups, and **Correlation Analysis**; is used to compare the *relationship* or the *correlation* between two independent groups.

A. Research Title

In determining a research title in an experimental study, there are at least three conditions to be considered. They are supposed to be supported as following: (1) practical teaching strategy to the students; (2) applicable teaching procedures supported with some relevant references; and (3) some previous related studies on the investigated teaching strategy. The following are some examples in determining a research title in experimental research study.

1. **Teaching** reading comprehension using **Collaborative Strategic Reading (C-S-R) strategy**…
2. **Using Numbered Heads Together (N-H-T) strategy** in teaching reading comprehension…
3. The **application** of **Personal Vocabulary Notes (P-V-N) strategy** in teaching vocabulary…
4. **Applying T-Chart strategy** in teaching narrative writing
5. The **implementation** of **Draw, Label, Caption (D-L-C) strategy** in teaching descriptive writing…
6. **Implementing Verbal, Visual, Words,** and **Association (V-V-W-A) strategy** in teaching English…
7. Improving Students' Reading Comprehension Achievement Using **PreP (PreReading Plan) Strategy** ….
8. The **effectiveness** of **Role, Audience, Format,** and **Topic (R-A-F-T) strategy** in teaching descriptive writing …
9. The **effects** of **motivation** on students' reading comprehension average scores ….
10. The **impacts of emotional intelligence** on students' reading comprehension average scores …
11. The **influence** of reading assignments on students' reading final scores …
12. The **correlation** between students' vocabulary and grammar scores on students' writing average scores ….
13. The **effectiveness** of students' learning styles on students' reading comprehension average scores taught using **5E-learning cycle strategy** ….

14. The **influence** of students' learning styles on students' speaking skill taught using **Talking Chips technique** …
15. The **significance** of students' learning styles and students' emotional intelligence on students' reading comprehension achievement taught using **Two Column Method Note-Taking** ….

B. Research Background

In research background, it presents how to identify problems by students' teachers in classroom activities. It describes a brief overview of factual information, phenomenon about some important facts. To support the paragraph on the investigated language skill, it should be based on some preliminary studies. It is illustrated based on the writer's direct observation (provided with check lists) in teaching and learning process before doing an actual teaching for his or her research study. Moreover, the paragraph can be supported on students' or teachers' perceptions and problems which are commonly encountered during the process of teaching and learning activities in the classroom with the investigated skill. Furthermore, the factual information is described based on a direct interview or a given questionnaire to identify the problems investigated in the study.

C. Research Problem

Research problem becomes core of the study. It aims to find the key of the problem in the study. Fraenkel, Wallen and Hyun (2012, p.27) state that a research problem is exactly that—a problem that someone would like to research. A problem can be anything that a person finds unsatisfactory or unsettling, a difficulty of some sort, a state of affairs that needs to be changed, anything that is not working as well as it might. Problems involve areas of concern to researchers, conditions they want to improve, difficulties they want to eliminate, questions for which they seek answers. Furthermore, Fels (2009, p.1) states that a research question is a clear, focused, concise, complex and arguable question around which the writers center to their research. Research questions help writers focus on their research study by providing a path through the research and writing process. The followings are examples of constructing a research question related to the investigated research title in the form of short-answer question:

[2]

1. Is there any ***significant difference*** on students' reading comprehension achievement taught using **Think-Aloud strategy** and teacher's method …
2. Does **Project-Based Learning strategy *significantly influence*** the students' descriptive writing achievement …
3. Is there any ***significant difference*** of students' learning style and students' reading comprehension score taught using **Shared Reading strategy** and teacher's method …
4. Is **Generating Interaction between Schemata and Text (G-I-S-T) strategy *effective*** in teaching reading comprehension …
5. Is there any ***significant improvement*** on students' reading comprehension score taught using **Question Answer Relationship (Q-A-R) strategy** …
6. Does i**nteractive writing strategy** give ***significant improvement*** on students' descriptive writing achievement …
7. Are ***means paired differences*** of students' reading comprehension average score taught using **Peer-Assisted Writing strategy** higher than conventional strategy …
8. Are there any ***significant different effects*** of learning styles on students' reading comprehension average scores taught using **Image, Elaborate, Predict,** and **Confirm (I-E-P-C)** and conventional strategies …
9. Do learning styles ***influence*** students' reading comprehension average score taught using **LGL (List-Group-Label)** and conventional strategies …
10. Do learning styles give ***significant different effects*** on students' reading comprehension average score taught using **OK4R** and conventional strategies …
11. Is there any ***effects/impact*** of motivation on students' reading comprehension average scores …
12. Is there any ***significant linear relationship*** between motivation and students' reading comprehension average scores …
13. Do students' reading assignments and students' reading final scores have ***significant linear relationship*** …

[3]

14. Is students' reading assignments ***related to*** students' reading comprehension average scores …
15. Is there any ***correlation*** between students' vocabulary and grammar scores on students' writing average scores …
16. Do students' vocabulary and grammar scores have any ***correlation*** on students' writing average scores …
17. Are there any ***significant interaction effects*** of learning styles in (visual, auditory, kinesthetic) categories on students' speaking skill taught using **Talking Stick** and conventional strategies …
18. Do learning styles in (visual, auditory, kinesthetic) categories have ***significant interaction effects*** on students' reading comprehension average score taught using **Story Face** and conventional strategies …
19. Are there any ***significant main effects*** of learning styles in (visual, auditory, kinesthetic) categories and motivation in (high, average, low) categories on students' reading comprehension scores taught using **Semantic Feature Analysis** and conventional strategies …
20. Do learning styles in (visual, auditory, kinesthetic) categories and motivation in (high, average, low) categories have ***significant main effects*** on students' speaking skill taught using **Word, Sentence, Question,** and **Answer (W-S-Q-A)** and conventional strategies …

D. Research Hypotheses

Research hypotheses are known as tentative answers for the study. Fraenkel, et.al. (2012, p.83) state that a hypothesis is, simply put, a prediction of the possible outcomes of a study. A hypothesis forces researchers to think more deeply and specifically about the possible outcomes of a study. From the statement, it can be assumed that research hypothesis is necessary in doing a research study since it is a possible outcome or a tentative answer of a research study. Research hypothesis offers two choices of outcomes or answers. They are rejected and accepted hypotheses. **Null hypothesis** (H*o*) is a negative statement of the research hypotheses. It is written in a negative sentence which implies that the hypothesis is rejected. The null hypothesis (H*o*) is rejected whenever the p-output is higher than 0.05. Then, **Alternative hypothesis** (H*a*) is a positive

statement of the research hypothesis. It is written in a positive sentence which implies that the hypothesis is accepted. The alternative hypothesis (H*a*) is accepted whenever the p-output is equal or lower than 0.05. The research hypothesis is formulated as follows.

1. **H*o*:** There is ***no significant difference*** on students' vocabulary scores taught using **Think, Talk,** and **Write (T-T-W)** and teacher's method …

H*a*: There is a ***significant difference*** on students' vocabulary scores taught using **Think, Talk,** and **Write (T-T-W)** and teacher's method ….

2 **H*o*:** **Mood, Understand, Recall, Detect, Elaborate,** and **Review (MURDER) strategy *does not give significant influence*** on students' reading comprehension average scores ….

H*a*: **Mood, Understand, Recall, Detect, Elaborate,** and **Review (MURDER) strategy *gives significant influence*** on students' reading comprehension average scores ….

3. **H*o*:** There is ***no significant difference*** of learning styles on students' reading comprehension achievement taught using **Survey, Question, Predict, Read, Respond,** and **Summarize (SQP2RS) strategy** …

H*a*: There is a ***significant difference*** of learning styles on students' reading comprehension achievement taught using **Survey, Question, Predict, Read, Respond,** and **Summarize (SQP2RS) strategy** …

4. **H*o*:** **Anticipation Guide strategy** is ***not effective*** in teaching reading comprehension …

H*a*: **Anticipation Guide strategy** is ***effective*** in teaching reading comprehension ….

5. **H*o*:** **Preview, Predict, Prior Knowledge,** and **Purpose (4-P) strategy *does not give*** *means* ***significant improvement*** on students' reading comprehension average scores …

H*a*: **Preview, Predict, Prior Knowledge,** and **Purpose (4-P) strategy *gives significant improvement*** on students' reading comprehension average scores ….

[5]

6. **H*o*:** ***Means paired differences*** on students' reading comprehension average scores taught using **rainbow dot strategy** ***are not higher*** than teacher's method …

   H*a*: ***Means paired differences*** on students' reading comprehension average score taught using **rainbow dot strategy** ***are higher*** than teacher's method ….

7. **H*o*:** There are no ***significant different effects*** on students' learning styles on students' reading comprehension average scores taught using **Somebody-Wanted-But-So (S-W-B-S)** and teacher's method ….

   H*a*: There are ***significant different effects*** on students' learning styles on students' reading comprehension average scores taught using **Somebody-Wanted-But-So (S-W-B-S)** and teacher's method ….

8. **H*o*:** Students' learning styles ***do not give significant different effects*** on students' reading comprehension average scores taught using story grammar and teacher's method ….

   H*a*: Students' learning styles ***give significant different effects*** on students' reading comprehension averages scores taught using story grammar and teacher's method ….

9. **H*o*:** There is ***no significant linear relationship*** between motivation and students' reading comprehension average scores ….

   H*a*: There is a ***significant linear relationship*** between motivation and students' students' reading comprehension average scores ….

10. **H*o*:** There is no ***significant linear relationship*** between students' reading assignments and students' reading final scores …

    H*a*: There is a ***significant linear relationship*** between students' reading assignments and students' reading final scores …..

11. **H*o*:** There is ***no correlation*** between students' vocabulary and grammar scores on students' writing average scores ….

    H*a*: There is a ***correlation*** between students' vocabulary and grammar scores on students' writing average scores ….

12. **H*o*:** There are ***no significant interaction effects*** of students' learning styles on students' reading comprehension average

scores taught using SQ3R and teacher's method …..

H*a*: There are ***significant interaction effects*** of students' learning styles on reading comprehension achievement taught using SQ3R and teacher's method …..

13. **H*o*:** Learning styles ***have no significant interaction effects*** on students' reading comprehension average scores taught using SQ3R and teacher's method …..

    H*a*: Learning styles ***have significant interaction effects*** on students' reading comprehension average scores taught using SQ3R and teacher's method …..

14. **H*o*:** There are ***no significant main effects*** of students' learning styles and students' emotional intelligence on students' reading comprehension average scores taught using Think-Pair-Share and teacher's method …..

    H*a*: There are ***significant main effects*** of students' learning styles and students' emotional intelligence on students' reading comprehension average scores taught using Think-Pair-Share and teacher's method …..

15. **H*o*:** Learning styles and motivation ***have no significant main effects*** on students' reading comprehension average scores taught using Think-Pair-Share and teacher's method ….

    H*a*: Learning styles and motivation ***have significant main effects*** on students' reading comprehension average scores taught using Think-Pair-Share and teacher's method ….

E. Hypothesis Testing

To prove the formulated research problem, the instruments of research hypotheses testing is required. The null hypothesis (H*o*) is accepted whenever the p-output is higher than 0.05. It means that the alternative hypothesis (H*a*) is rejected. On the other hand, the alternative hypothesis is accepted (H*a*) whenever the p-output is equal or lower than 0.05. It means that the null hypothesis (H*o*) is rejected. The followings are examples of research hypothesis testings and research questions to prove whether the hypothesis is accepted or rejected.

1. **Independent Sample T-Test;** is used to analyze the following research questions:
   - Is there any ***significant difference*** on students' reading comprehension average scores taught using **Collaborative Strategic Reading (C-S-R) strategy** and teacher's method …
   - Does **Collaborative Strategic Reading (C-S-R) strategy** give ***significant influence*** on students' reading comprehension average scores …
   - Does **Collaborative Strategic Reading (C-S-R) strategy *influence*** students' reading comprehension average scores …
   - Is **Collaborative Strategic Reading (C-S-R) strategy *effective*** in teaching reading comprehension …

2. **Paired Sample T-Test:** is used to analyze the following research questions:
   - Are ***means paired differences*** on students' reading comprehension average scores taught using **Somebody-Wanted-But-So (S-W-B-S)** higher than teacher's method …
   - Does **Somebody-Wanted-But-So (S-W-B-S)** strategy give ***means significant improvement*** on students' reading comprehension average scores ...
   - Does **Somebody-Wanted-But-So (S-W-B-S)** strategy ***improve*** students' reading comprehension average scores …
   - Is there any ***significant improvement*** on students' reading comprehension average scores taught using **Somebody-Wanted-But-So (S-W-B-S)** strategy …
3. **Regression Analysis;** is used to analyze the following research questions:
   - Is there any ***effect/impact/influence*** of motivation on students' reading comprehension average scores …
   - Does motivation give any ***impact/effect/influence*** on students' reading comprehension average scores …
   - Do motivation and students' reading comprehension average scores have ***significant linear relationship*** …

- Is there any ***significant linear relationship*** between students' reading assignments and students' reading final scores …
- Are students' reading assignments ***related to*** students' reading final scores …

4. **Correlation Analysis;** is used to analyze the following research questions:
   - Is there any ***correlation*** between students' vocabulary and grammar scores on students' writing average scores …
   - Do students' vocabulary and grammar scores have any ***correlation*** on students' writing average scores …

5. **One-way ANOVA;** is used to analyze the following research questions:
   - Is there any ***signficant difference*** on students' reading comprehension average scores taught using **5E-learning cycle**, **semantic mapping**, and **teacher's method** …
   - Is there any ***significant difference*** of motivation in (high, average, low) categories on students' reading comprehension average scores taught using **5E-learning cycle strategy** …
   - Do emotional intelligence in (high, middle, low) categories give ***significant difference*** on students' reading comprehension average scores taught using **5E-learning cycle strategy** …
   - Do learning styles in (visual, auditory, kinesthetics) categories ***influence*** students' reading comprehension average scores taught using **5E-learning cycle** strategy …

6. **Analysis of Covariance (ANCOVA);** is used to analyze the following research questions:
   - Are there any ***significant different effects*** of learning styles in (visual, auditory, kinesthetic) categories on students' reading comprehension average scores taught using **Survey, Question, Predict, Read, Respond,** and **Summarize (SQP2RS) and conventional strategies** …
   - Do emotional intelligence in (high, average, low) categories give ***significant different effects*** on students' reading comprehension

average scores taught using **Survey, Question, Predict, Read, Respond,** and **Summarize (SQP2RS) and conventional strategies** …

7. **Two-way ANOVA;** is used to analyze the following research questions:
   - Are there any ***significant interaction effects*** of learning styles in (visual, auditory, kinesthetic) category on students' reading comprehension average scores taught using **PQRST** and conventional strategies …
   - Do emotional intelligence in (high, middle, low) categories have ***significant interaction effects*** on students' reading comprehension average scores taught using **PQRST** and conventional strategies …

8. **Two-way MANOVA;** is used to analyze the following research questions:
   - Are there any ***significant main effects*** of learning styles in (visual, auditory, kinesthetic) categories and emotional intelligence in (high, middle, low) categories on students' reading comprehension average scores taught using **Image, Elaborate, Predict,** and **Confirm (I-E-P-C)** and conventional strategies …
   - Do learning styles in (visual, auditory, kinesthetic) categories and emotional intelligence in (high, middle, low) categories have ***significant main effects*** on students' reading comprehension average scores taught using **Image, Elaborate, Predict,** and **Confirm (I-E-P-C)** and conventional strategies …

F. Criteria of Hypothesis Testing

In criteria of testing of the hypothesis, the students' posttest scores in control and experimental groups are analyzed using research instrument testing. The result analysis depends on the problem investigated. The research hypothesis is determined based on the following criteria.

1. **Measuring significant difference/influence**

a. **Using Independent Sample T-Test**; in measuring significant difference/influence of two groups, an independent sample t-test is used. It is administered to measure two independent variables. The criteria of testing the hypothesis is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.05, the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.
   - If the p-output (Sig.2-tailed) is higher than 0.05, the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

b. **Using One-way ANOVA**; is used to measure significant difference/influence which has two or more variables in one group. The criteria of testing the hypothesis is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.05, the alternative hypothesis (H*a*) is accepted, and the null hypothesis (H*o*) is rejected.
   - If the p-output (Sig.2-tailed) is higher than 0.05, the alternative hypothesis (H*a*) is rejected, and the null hypothesis (H*o*) is accepted.

2. **Measuring means paired differences**; in measuring mean paired differences, paired sample t-test is used. The criteria of testing the research hypothesis is as follows:
   - If means paired differences on students' reading comprehension average score taught using **Image, Elaborate, Predict,** and **Confirm (I-E-P-C)** strategy is higher than conventional strategy, the null hypothesis (H*o*) is rejected and the alternative hypothesis (H*a*) is accepted.
   - If means paired differences on students' reading comprehension average score taught using **Image, Elaborate, Predict,** and **Confirm (I-E-P-C)** strategy are not higher than conventional strategy, the null hypothesis (H*o*) is accepted and the alternative hypothesis (H*a*) is rejected.

3. **Measuring a significant improvement;** in measuring a significant improvement, paired sample t-test is used. The criteria of testing the hypothesis is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.025, the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.
   - If the p-output (Sig.2-tailed) is higher than 0.025, the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

4. **Measuring significant difference effects**; Analysis of Covariance (ANCOVA) is used to measure significant different effects which have more than two variables in both groups. The criteria of testing the hypothesis is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.05, the alternative hypothesis (H*a*) is accepted, and the null hypothesis (H*o*) is rejected.
   - If the p-output (Sig.2-tailed) is higher than 0.05, the alternative hypothesis (H*a*) is rejected, and the null hypothesis (H*o*) is accepted.

5. **Measuring effectiveness**; In measuring the effectiveness between two strategies, independent sample t-test is used.The criteria of testing the hypothesis is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.05, the alternative hypothesis (H*a*) is accepted, and the null hypothesis (H*o*) is rejected.
   - If the p-output (Sig.2-tailed) is higher than 0.05, alternative hypothesis (H*a*) is rejected, and the null hypotheis (H*o*) is accepted.

6. **Measuring effects, impact, influence, or significant linear relationship**; in measuring effects, impact, influence, or significant linear relationship; regression analysis is used. The criteria of testing the hypothesis is as follows:

- If the p-output (Sig.2-tailed) is lower than 0.05, the alternative hypothesis (H*a*) is accepted, and the null hypothesis (H*o*) is rejected.
- If the p-output (Sig.2-tailed) is higher than 0.05, the alternative hypothesis (H*a*) is rejected, and the null hypothesis (H*o*) is accepted.

7. **Measuring correlation**; in measuring correlation, Pearson Correlation Coefficient is used. Sugiyono (2012, p.231) mentions the criteria of testing the hypothesis in measuring correlation as follows:
   - If the p-output (Sig.2-tailed) is higher than 0.199, the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.
   - If the p-output (Sig.2-tailed) is lower than 0.199, the null hypothesis (H*o*) is accepted, and Ha is rejected.

8. **Measuring significant interaction effects**; in measuring significant interation effects, two-ways ANOVA are used. According to Wuensch (2010, p.4), the criteria of hypothesis testing in finding significant interaction effects using two-ways ANOVA is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.024, the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.
   - If the p-output (Sig.2-tailed) is higher than 0.024, the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

9. **Measuring significant main effects**; in measuring significant main effects, two-way MANOVA is used. Wuensch (2010, p.4) states that the criteria of testing the hypothesis in finding significant main effects using two-ways MANOVA is as follows:
   - If the p-output (Sig.2-tailed) is lower than 0.05, the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.
   - If the p-output (Sig.2-tailed) is higher than 0.05, the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

G. Criteria Levels in Assessing (Written) Writing and (Oral) Speaking Contents.

The scoring systems in writing and speaking skills are really different from scoring students' reading comprehension. It needs criteria levels of two or more raters (judges) to assess the two language skills called **inter-rater reliability**. It is relevant to what Brown (2004:20) says that inter-rater reliability occurs when two or more scores yields inconsistent scores of the same test, possibly for lack of attention to scoring criteria, experience, inattention, or even preconceived biases. From the statement, it can be stated that two or more raters are required to avoid the inconsistency or the biases from scoring of the two language skills (speaking and writing). The followings are criteria contents in assessing students' writing and speaking achievements.

a. Assessing students' writing achievements

Weigle (2002, p.116) states that there are at least five criteria components in assessing students' writing achievements. They are: content (13-30), organization (7-20), vocabulary (7-20), language use (5-25), and mechanics (2-5). It needs more than one raters or judges to assess students' writing achievements. The following table is criteria levels in assessing students' writing achievements with some criteria contents.

Table 1

Assessing Writing Achievement with Five Criterial Contents

SCORE	LEVEL	CRITERIA
Content (13-30)	30-27	**Excellent to Very Good**: knowledge, substantive, through developing thesis, relevent to topic.
	26-22	**Good to Average**: some knowledge of subject, adequate range, limited development of thesis, mostly relevant to topic, but lacks detail.
	21-17	**Fair to Poor**: limited knowledge of subject, little substance, inadequate development of topic.
	16-13	**Very Poor**: does not show knowledge of subject,

		non-substantive, not pertinent, or not enough to evaluate.
Organization (7-20)	20-18	**Excellent to Very Good:** fluent expression, ideas clearly stated/supported, succinct, well-organized, logical sequencing, cohesive.
	17-14	**Good to Average:** somewhat choppy, loosely organized but main ideas stand out, limited support, logical but incomplete sequencing.
	13-10	**Fair to Poor:** non-fluent, ideas confused or disconnected, lacks of logical sequencing and development.
	9-7	**Very Poor:** does not communicate, no organization, OR not enough to evaluate.
Vocabulary (7-20)	20-18	**Excellent to Very Good:** sophisticated range, effective word/idiom choice and usage, word form mastery, appropriate register.
	17-14	**Good to Average:** adequate range, occasional errors of word/idiom form, choice, usage but meaning not obscured.
	13-10	**Fair to Poor:** limited range, frequent errors of word/idiom form, choice, usage, meaning confused or obscured.
	9-7	**Very Poor:** essentially translation, little knowledge of English vocabulary, idioms, word form, OR not enough to evaluate.
Language Use (5-25)	25-22	**Excellent to Very Good:** Effective complex construction, few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions.
	21-18	**Good to Average:** effective but simple construction, minor problems in complex construction, several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning seldom obscured.
	17-11	**Fair to Poor:** major problem in simple/complex construction, frequent errors of negation, agreement, tense, number, word order/function,

		articles, pronouns, prepositions and/or fragments, run-ons deletions, meaning confused or obscured.
	10-5	**Very Poor:** virtually no mastery of sentence construction rules, dominated by errors, does not communicate, OR not enough to evaluate.
Mechanics (2-5)	5	**Excellent to Very Good:** demonstrates mastery of conventions, few errors of spelling, punctuation, capitalization, paragraphing.
	4	**Good to Average:** Occasional errors of spelling, punctuation, capitalization, paragraphing but meaning not obscured.
	3	**Fair to Poor:** frequent errors of spelling, punctuation, capitalization, paragraphing, poor handwriting, meaning confused or obscured.
	2	**Very Poor:** no mastery of conventions, dominated by errors of spelling, punctuation, capitalization, paragraphing, handwriting illegible, or not enough to evaluate.
T o t a l :	100	

Source: Weigle. (2002, p.116). *Assessing Writing*. London: Cambridge University Press.

Then, the five criteria levels of (written) writing contents are described in the following table as example of assessing students' writing achievements with a single rater.

No	**Students' Name**	**RATER A: Lenny Marzulina, M.Pd**					**TOT**
		CNT	**ORG**	**VOC**	**LANGU**	**MCH**	
		(13-30)	**(7-20)**	**(7-20)**	**(5-25)**	**(2-5)**	
1	Damayanti						
2	Kusmayati						
3	Devi Susanti						
4	Fauzaniaty						
5	Rizky Andi						
6	Farihah						
7	Mellyza						

8	Irina M						
9	Darmawan						
10	Mukhlis						
11	Ayu Cecilia						
12	Alit Wigati						
13	Joni Iman						
14	Nur Aprianti						
15	Iskandar R						
16	Maghfiroh						
17	Sri Astuti						
18	Kholil Abld						
19	Agus Arif						
20	Sukmawati						

Where: CNT= Content; **ORG**= Organization; **VOC** = Vocabulary; **LANGU**= Language Use; and **MCH** = Mechanics.

b. Assessing Oral (speaking) Content

Hughes (1989, p.111) states that there are five criteria in assessing oral (speaking) content. They are: pronunciation (2-15), grammar (5-25), vocabulary (3-20), fluency (3-20), and comprehension (3-20). And It needs more than one raters to assess oral speaking skill. The following is the table of criteria levels in assessing oral (speaking) content.

Table 2

Assessing Oral (Speaking) with Five Criteria Contents

SCORE	LEVEL	CONTENT
Pronunciation (2-15)	2	Pronunciation problem so severe to make speech virtually unintelligible.
	3	Very hard to understand because of problems. Must frequently repeat in order to make him/herself understood.
	5	Pronunciation problems necessitate concentration on the part of the listener and occasionally lead to misunderstanding.

	10	Always intelligible, though one is conscious of a definite accent and occasional inappropriate intonation patterns.
	15	Pronunciation and intonation approximate that of a native speaker.
Grammar (5-25)	5	Errors in grammar and word order so severe so to make speech virtually unintelligible.
	10	Grammar and word order errors make comprehension difficult.
	15	Make frequent errors of grammar and word order that occasionally obscured meaning.
	20	Occasionally makes grammatical and/or word errors that do not obscure meaning.
	25	Grammatical usage and word order approximate that of a native speaker.
Vocabulary (3-20)	3	Vocabulary limitations so extreme as to make conversation virtually impossible.
	5	Misuses words and very limited vocabulary, comprehensive quite difficult.
	10	Students frequently uses the wrong words; conversation somewhat limited because of inadequate vocabulary.
	15	Students occasionally use inappropriate terms and/or most rephrase ideas because of lexical inadequacies.
	20	Use of vocabulary and idioms approximate that of a native speaker.
Fluency (3-20)	3	Speech is so halting and fragmentary as to make conversation virtually impossible.
	5	Usually hesitant: Often forced into silence by language limitations.
	10	Speech is everyday conversation and classroom discussion frequently disrupted by the students' search for the correct manner of expression.
	15	Speech in everyday conversation and classroom discussions generally fluent, with occasionally lapses while the student searches for frequently

		manner of expression.
	20	Speech in everyday conversation and classroom discussion fluent and effortless, approximating that of a native speaker.
Comprehension (3-20)	3	Cannot be said to understand simple conversation.
	5	Has great difficulty following what is said. Can comprehend only "social conversation" spoken slowly and with frequent repetitions.
	10	Understands most of what is said at slower-than normal speed with repetitions.
	15	Understand nearly everything at normal speech, although occasional repetition may be necessary.
	20	Understand everyday conversation and normal classroom discussion without difficulty.
T o t a l:	100	

Source: Hughes. (1989:p.111). *Testing for Language Teachers*. UK: Cambridge University Press.

Then, the five criteria levels of oral speaking contents are described in the following table as example of assessing students' oral speaking with a single rater.

No	**Students' Name**	**RATER B: Dian Erlina, S.Pd., M.Hum**					**TOT**
		PRNC (2-15)	**GRM (3-25)**	**VOC (3-20)**	**FLUCY (3-20)**	**CMPRH (3-20)**	
1	Damayanti						
2	Kusmayati						
3	Devi Susanti						
4	Fauzaniaty						
5	Khadafi						
6	Darmawan						
7	Yunica						
8	Irina						
9	Darmawan						
10	Mukhlis						

11	Ayu Cecilia						
12	Alit Wigati						
13	Joni Iman						
14	Nur Aprianti						
15	Iskandar R						
16	Maghfiroh						
17	Sri Astuti						
18	Kholil Abld						
19	Agus Arif						
20	Sukmawati						

Where: PRNC = Pronunciation; **GRM** = Grammar, **VOC** = Vocabulary; **FLUCY** = Fluency; **CMPRH** = Comprehension

H. Research Instrument Analysis

Before implementing research treatments in experiential and control groups, a tryout on research instrument should be administered to estimate the **validity** and **reliability** of research instrument for students' pretest and posttest activities. The followings are steps to analyze the validity and reliability test of the obtained scores based on the result of a tryout analysis. They are as follows.

1. Validity Test

The quality of the instruments used in research is very important, for the conclusions that researchers draw are based on the information they obtain using these instruments. Faenkel and Wallen (1990, p.138) state that the term "validity" refers to the appropriateness, meaningfulness, and usefulness of any inferences of researcher draws based on the data obtained through the use of an instrument. It means that validity test is used to measure whether the obtained data of an instrument is valid or not. In this part, there are two kinds of validity test to be administered for research instrument. They are validity test of each question item and content validity.

a. Construct Validity

Hughes (1989, p.26) states that a test, part of a test, or a testing technique is said to have construct validity if it can

be demonstrated that it measures just the ability which it is supposed to measure. Furthermore, Sugiyono (2010, p.177) states that **expert judgments** is required to estimate the construct validity. After constructing the instruments related to some aspects measured, then it is consulted to achieve some expert judgement from at least three validators to evaluate whether the components of the instrument are valid or not to be applied in research activities. In this part, the construct validity of the research instruments involves two types. They are question items for pretest and posttest activities, and lesson plans for control and experimental groups. In estimating those two research instruments, the forms of validator instuments are prepared (Example of validator form for test instruments and lesson plans can be seen in Appendix).

b. Validity of Each Question Item

Validity test of each question item is used to indicate whether the test item of the instruments in each question is valid or not. To know whether it is valid or not, the score of significance (r-output) should be compared with the score of "r-table" product moment. A question item is considered valid if "r-output" is higher than "r-table" (Basrowi and Soenyono, 2007, p.24). The following is example of a research study in analyzing a questionnaire on students' emotional intelligence. There are 10 question items with N (sample) is 14 respondent (students) with 4 choices in answering the questionnaire. They are: 0 for *always*, 1 for *sometimes*, 2 for *seldom*, and 3 for *never*. After the questionnaire is distributed to the research sample to estimate the validity of the instrument, each question item on respondence's answer from the questionnaire is then analyzed using **Pearson correlation coefficients**. Before analyzing the obtained scores, table analysis should be made. The following is table analysis on respondence's answers

from emotional intelligence questionnaire with 10 question items and 14 samples (N).

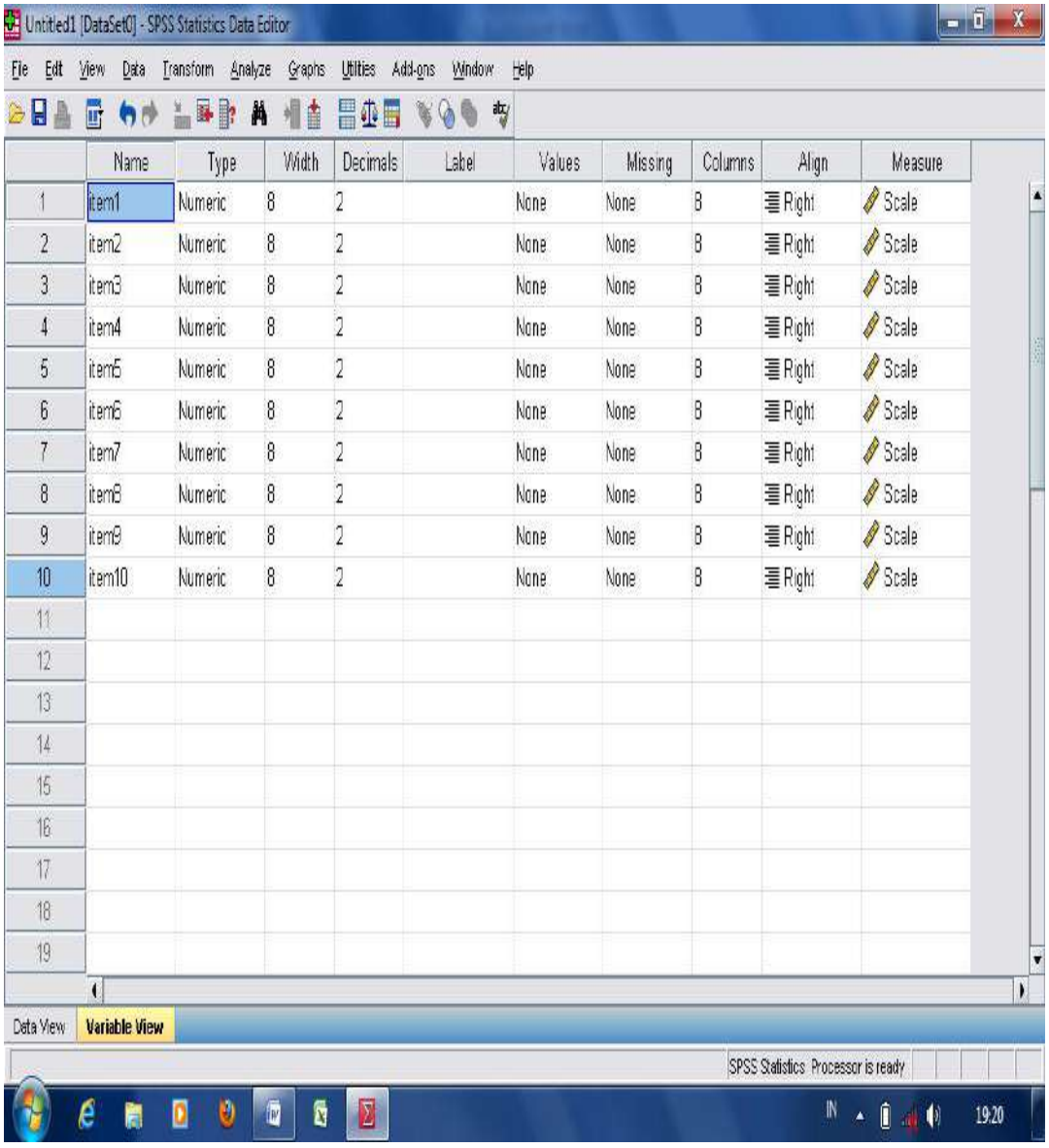
Table 3

Analysis of Each Question Item on Emotional Intelligence Questionnaire

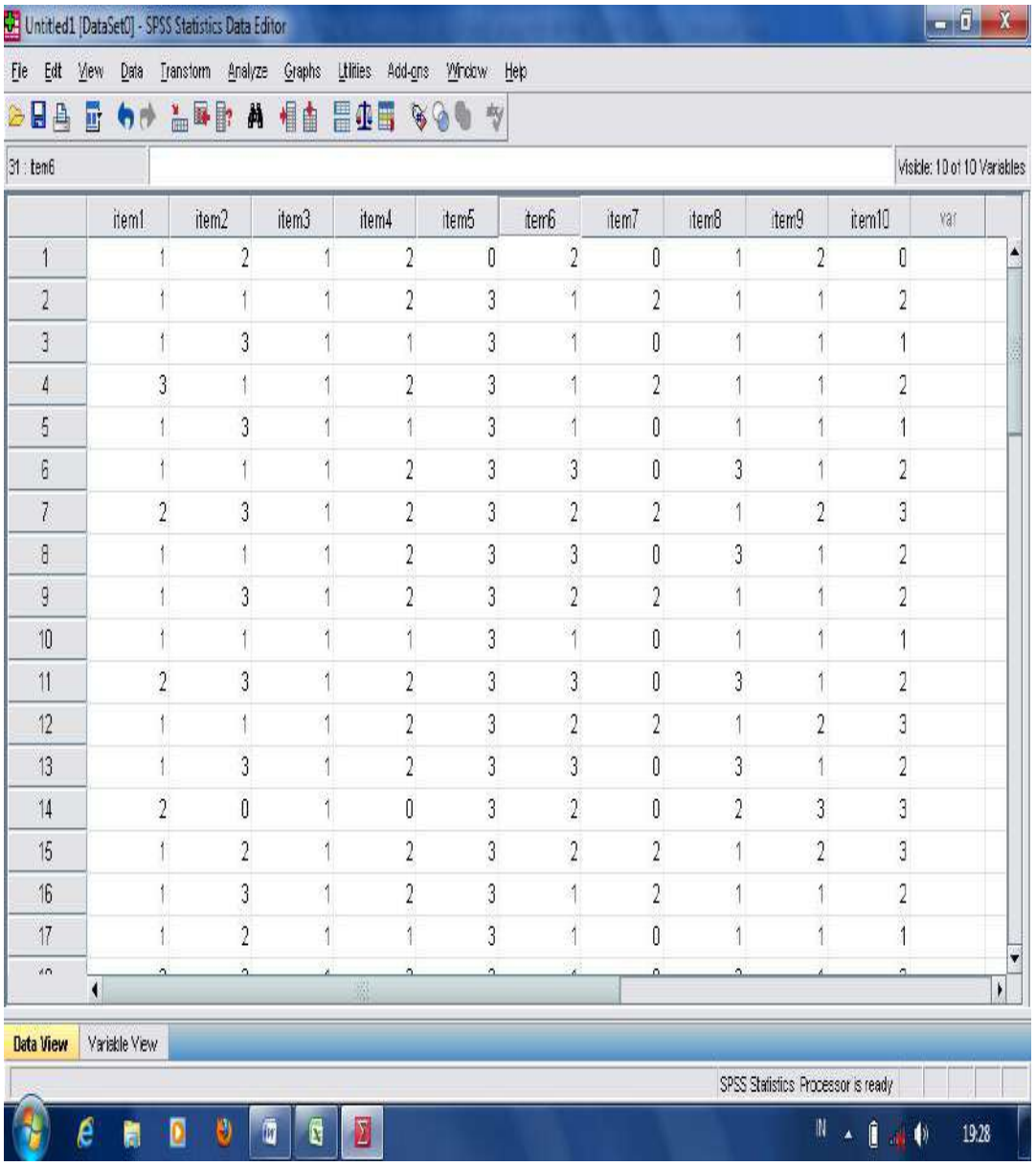
N	Emotional Intelligence Questionnaire Item Number:										Score
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
1	0	2	2	1	2	2	1	2	0	2	14
2	1	0	1	2	0	1	0	1	0	1	7
3	3	1	0	0	2	2	3	0	2	0	12
4	0	3	2	2	1	0	2	2	1	1	14
5	2	0	2	1	3	1	0	0	0	2	11
6	1	3	3	0	2	3	1	1	3	1	18
7	3	2	0	2	1	0	3	0	2	0	12
8	0	0	2	1	2	1	0	3	0	2	11
9	3	1	1	2	0	2	1	0	1	3	14
10	1	3	2	0	3	0	2	2	3	0	16
11	0	0	3	2	1	2	0	0	1	0	9
12	1	2	1	1	2	1	1	2	2	2	14
13	2	0	2	0	1	0	2	0	2	0	9
14	0	0	2	2	2	2	1	0	1	2	12

Then, the result analysis found in “r-output” table shows that the question item is categorized valid whenever the significance (2-tailed) of “r-output” is higher than “r-table” with N (sample) = 14 is 0.532 (the score of significance with a certain number of sample can be seen in r-table Product Moment). The following are steps to run the analysis of validity test in each question item using SPSS:

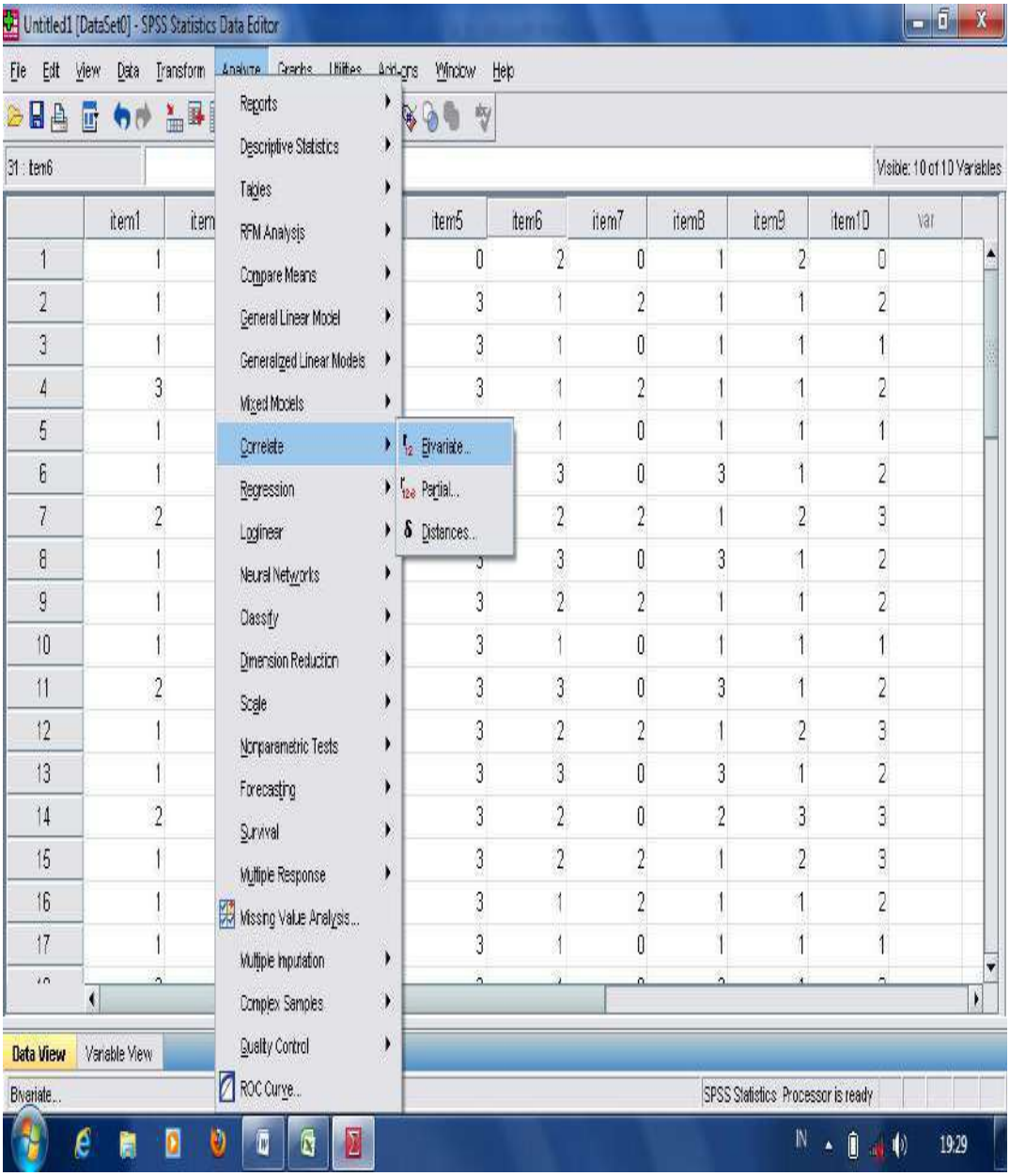
- Open the worksheet of the application in SPSS program;
- Then rename the “variable view” into (e.g. Item1, Item2, and...)



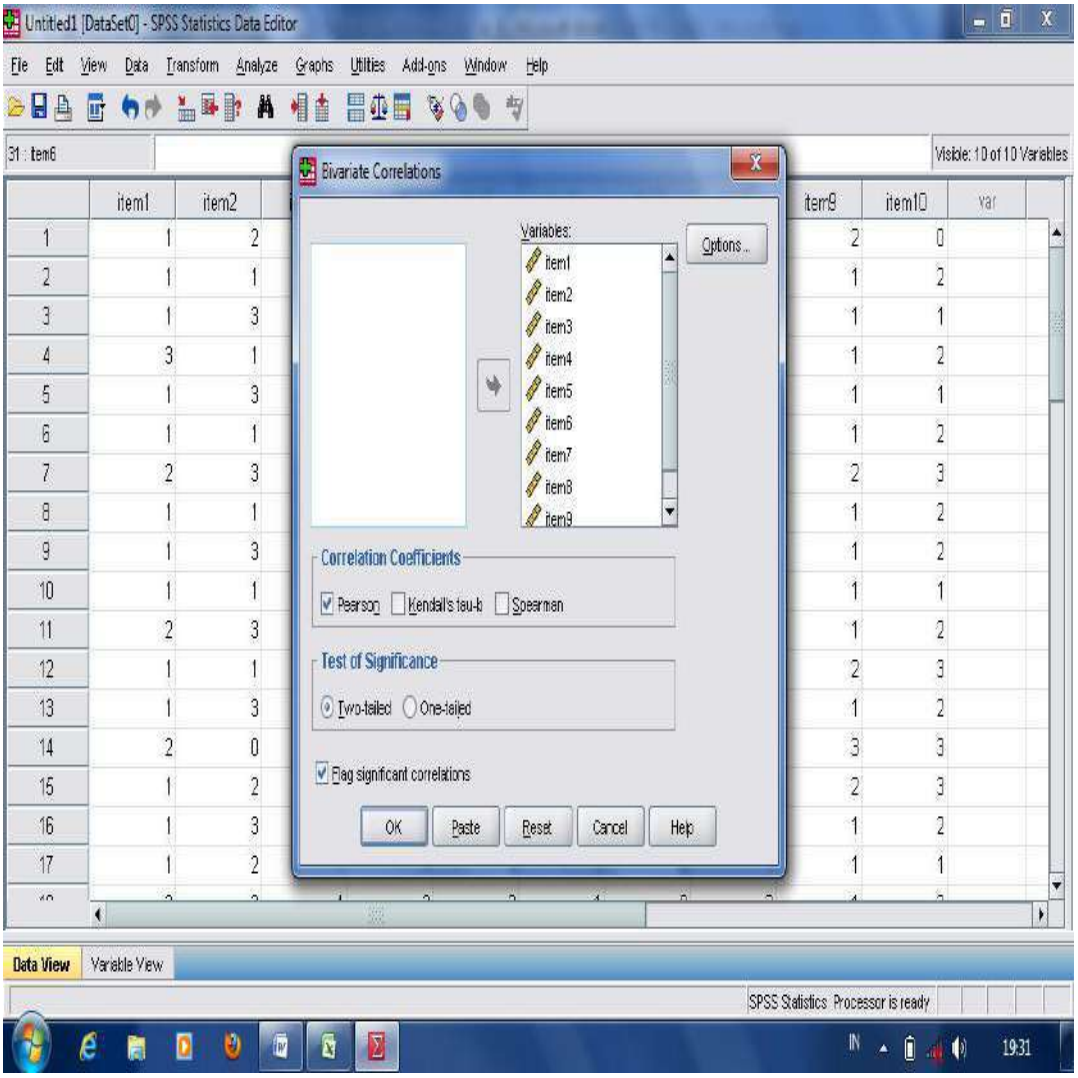
- Move into “data view” and input all the scores in each category”;



- Then, follow the next step to analyze the question items: Click "**Menu"**, "**Analyze"**, "**Correlate"**, and "**Bivariate"**…



- Move all items into "variables column" and check "**Pearson**"; Choose (2-tailed) in Test of Significance; then click OK.



- Check the validity of each question item into r-table with a certain number of sample whether it is valid or not. The following table shows the result analysis of each question item on the students' emotional intelligence questionnaire analyzed using Pearson correlation coefficient.

Correlations

		item 1	item 2	item 3	item 4	item 5	item 6	item 7	item 8	item 9	item10
item1	Pearson Correlation	1	-.033	-.692**	-.182	-.187	-.109	.515	-.591*	.311	-.153
	Sig. (2-tailed)		.910	.006	.534	.521	.710	.060	.026	.278	.601
	N	14	14	14	14	14	14	14	14	14	14
item2	Pearson Correlation	-.033	1	.005	-.244	.215	-.104	.489	.414	.583*	-.145
	Sig. (2-tailed)	.910		.987	.401	.460	.725	.076	.141	.029	.620
	N	14	14	14	14	14	14	14	14	14	14
item3	Pearson Correlation	-.692**	.005	1	-.123	.252	.261	-.546*	.204	-.044	.058
	Sig. (2-tailed)	.006	.987		.675	.384	.368	.043	.484	.881	.845
	N	14	14	14	14	14	14	14	14	14	14
item4	Pearson Correlation	-.182	-.244	-.123	1	-.583*	-.039	-.290	-.154	-.547*	.322
	Sig. (2-tailed)	.534	.401	.675		.029	.894	.314	.599	.043	.262
	N	14	14	14	14	14	14	14	14	14	14
item5	Pearson Correlation	-.187	.215	.252	-.583*	1	.024	.022	.273	.208	-.011

	Sig. (2-tailed)	.521	.460	.384	.029		.935	.940	.344	.475	.969
	N	14	14	14	14	14	14	14	14	14	14
item6	Pearson Correlation	-.109	-.104	.261	-.039	.024	1	-.349	-.205	-.063	.351
	Sig. (2-tailed)	.710	.725	.368	.894	.935		.222	.482	.830	.218
	N	14	14	14	14	14	14	14	14	14	14
item7	Pearson Correlation	.515	.489	-.546*	-.290	.022	-.349	1	-.190	.626*	-.529
	Sig. (2-tailed)	.060	.076	.043	.314	.940	.222		.515	.017	.052
	N	14	14	14	14	14	14	14	14	14	14
item8	Pearson Correlation	-.591*	.414	.204	-.154	.273	-.205	-.190	1	-.115	.220
	Sig. (2-tailed)	.026	.141	.484	.599	.344	.482	.515		.695	.451
	N	14	14	14	14	14	14	14	14	14	14
item9	Pearson Correlation	.311	.583*	-.044	-.547*	.208	-.063	.626*	-.115	1	-.530
	Sig. (2-tailed)	.278	.029	.881	.043	.475	.830	.017	.695		.051
	N	14	14	14	14	14	14	14	14	14	14
item10	Pearson Correlation	-.153	-.145	.058	.322	-.011	.351	-.529	.220	-.530	1

Sig. (2-tailed)	.601	.620	.845	.262	.969	.218	.052	.451	.051	
N	14	14	14	14	14	14	14	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

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

Then, the result of significant score of Pearson correlation (r-output) is compared with "r-table" product moment. The r-table for a certain number of samples (N) is shown as in the following table.

Table 4
The r-table of Product Moment
for a Certain Number of Samples (N)

N	Significant Level		N	Significant Level		N	Significant Level	
	5%	1%		5%	1%		5%	1%
3	0.997	0.999	**27**	0.381	0.487	**55**	0.266	0.345
4	0.950	0.990	**28**	0.374	0.478	**60**	0.254	0.330
5	0.878	0.959	**29**	0.367	0.470	**65**	0.244	0.317
6	0.811	0.917	**30**	0.361	0.463	**70**	0.235	0.306
7	0.754	0.874	**31**	0.355	0.456	**75**	0.227	0.296
8	0.707	0.834	**32**	0.349	0.449	**80**	0.220	0.286
9	0.666	0.798	**33**	0.344	0.442	**85**	0.213	0.278
10	0.632	0.765	**34**	0.339	0.436	**90**	0.207	0.270
11	0.602	0.735	**35**	0.334	0.430	**95**	0.202	0.263
12	0.576	0.708	**36**	0.329	0.424	**100**	0.195	0.256
13	0.553	0.684	**37**	0.325	0.418	**125**	0.176	0.230
14	0.532	0.661	**38**	0.320	0.413	**150**	0.159	0.210
15	0.514	0.641	**39**	0.316	0.408	**175**	0.148	0.194
16	0.497	0.623	**40**	0.312	0.403	**200**	0.138	0.181
17	0.482	0.606	**41**	0.308	0.398	**300**	0.113	0.181

18	0.468	0.590	**42**	0.304	0.393	**400**	0.098	0.128
19	0.456	0.575	**43**	0.301	0.389	**500**	0.088	0.115
20	0.444	0.561	**44**	0.297	0.384	**600**	0.080	0.105
21	0.433	0.549	**45**	0.294	0.380	**700**	0.074	0.097
22	0.423	0.537	**46**	0.291	0.380	**800**	0.070	0.091
23	0.413	0.526	**47**	0.288	0.372	**900**	0.065	0.086
24	0.404	0.515	**48**	0.284	0.368	**1000**	0.062	0.081
25	0.396	0.505	**49**	0.281	0.364			
26	0.388	0.496	**50**	0.279	0.361			

Source: Sugiyono. (2012, p.373). *Statistika Untuk Penelitian*. Bandung: Alfa Beta.

From the result analysis in validity test of each question item above, it is found that there are 5 questions considered invalid since the scores of significance are lower than 0.532. They are question item1, question item3, question item7, question item8, and question item9. Then, it is found that 5 questions items are considered valid since the scores of significance are higher than 0.532. They are question item2, question item4, question item5, question item6, and question item10. Then, the result analysis of validity test in each question item is displayed in the following table.

No	Validity Test of Each Question Item	Sig.(2-tailed) of Pearson Correlation (r-output)	r-table score	Result
1	Item 1	0	0.532	Invalid
2	Item 2	0.910	0.532	**Valid**
3	Item 3	0.006	0.532	Invalid
4	Item 4	0.534	0.532	**Valid**
5	Item 5	0.521	0.532	**Valid**
6	Item 6	0.710	0.532	**Valid**
7	Item 7	0.060	0.532	Invalid
8	Item 8	0.026	0.532	Invalid
9	Item 9	0.278	0.532	Invalid
10	Item 10	0.601	0.532	**Valid**

Another example is given when a researcher is going to analyze the validity of each question item with multiple choice questions. There are 20 multiple choice questions, and the N-sample is 30 students. Then, each question item is analyzed for its validity. From students' answer on multiple choice questions, the correct answer is labeled 1, and the wrong answer is labeled 0. The analysis of validity test of each question item in the form of multiple choice questions is displayed in the following table.

Table 5

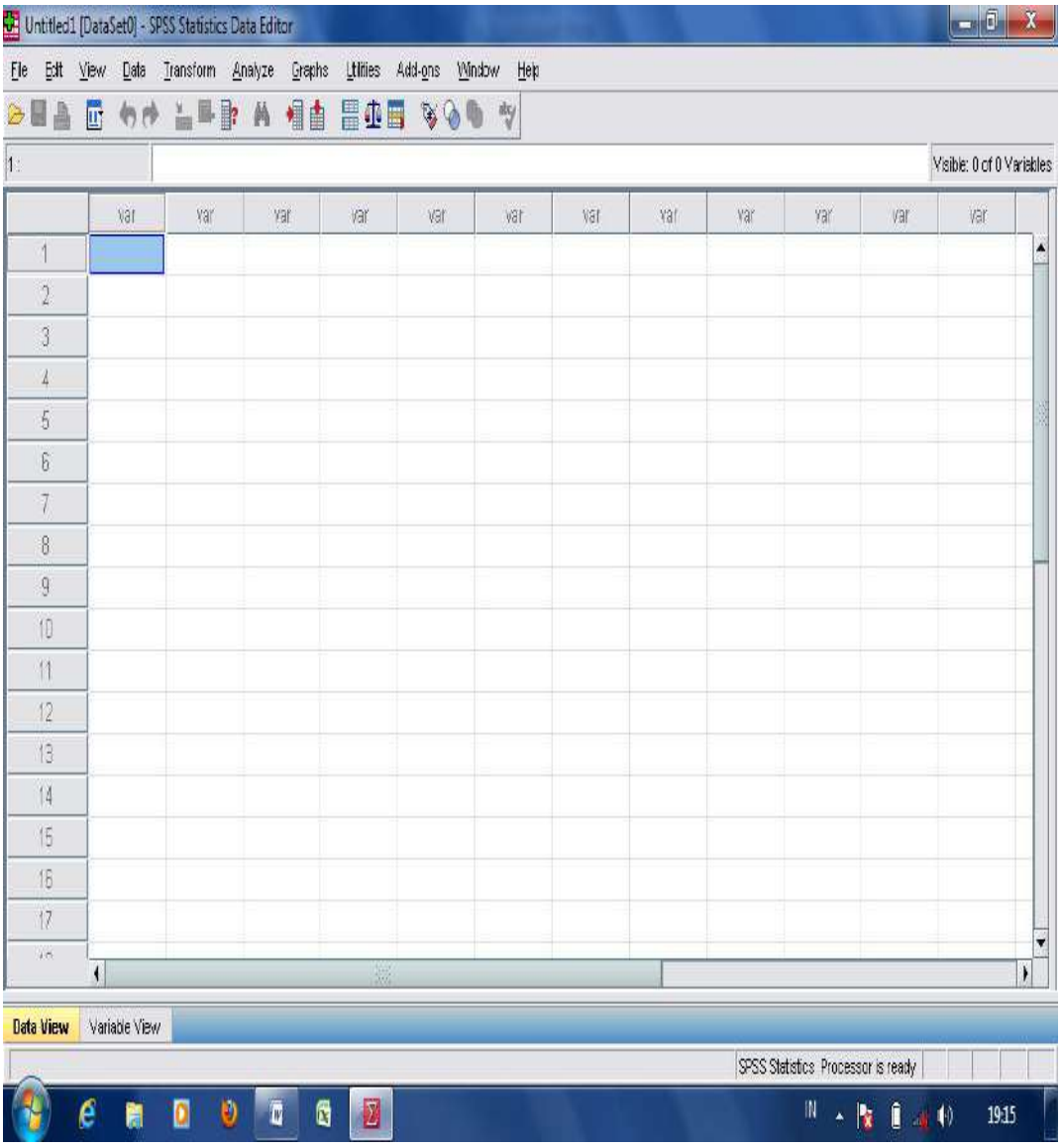
Analysis of Each Question Item on Reading Comprehension Test with Multiple Choice Questions

N	Reading Comprehension Question Item Number:																				Score
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1	0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	1	1	13
2	0	1	1	0	1	0	1	1	0	1	1	0	1	1	0	0	1	1	1	0	12
3	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	1	1	1	0	12
4	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	12
5	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	16
6	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	15
7	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	16
8	1	1	1	0	1	1	0	0	1	1	1	1	0	0	1	1	1	0	0	1	13
9	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	16
10	1	1	1	0	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	15
11	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	17
12	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	15
13	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	14
14	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	1	0	13
15	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	16
16	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	15
17	0	0	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	13
18	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	0	0	1	1	15
19	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0	1	1	15

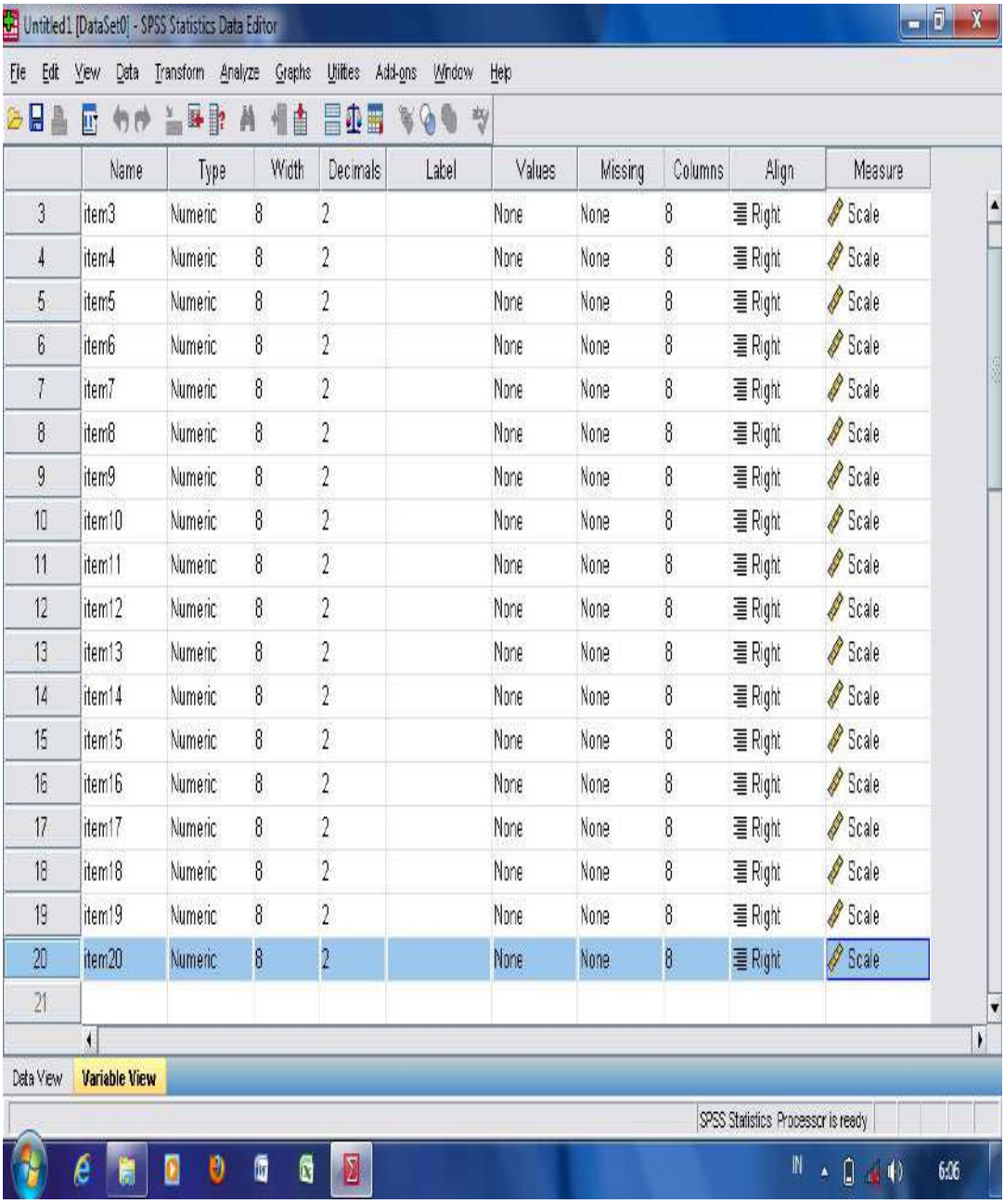
20	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	0	17
21	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	18
22	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	1	1	1	13
23	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	1	1	1	0	1	14
24	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	0	1	1	1	1	15
25	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	14
26	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	14
27	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	16
28	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1	16
29	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0	16
30	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	16

From the result of the students' answer in testing the validity of each question item, it is categorized valid whenever the significance (2-tailed) of the r-output is higher than the r-table with N-sample 30 is 0.349 (the score of significance with a certain number of samples can be seen in the r-Table). Then, the analysis of validity test is done as in the following steps:

- Open the worksheet of the application in SPSS program;

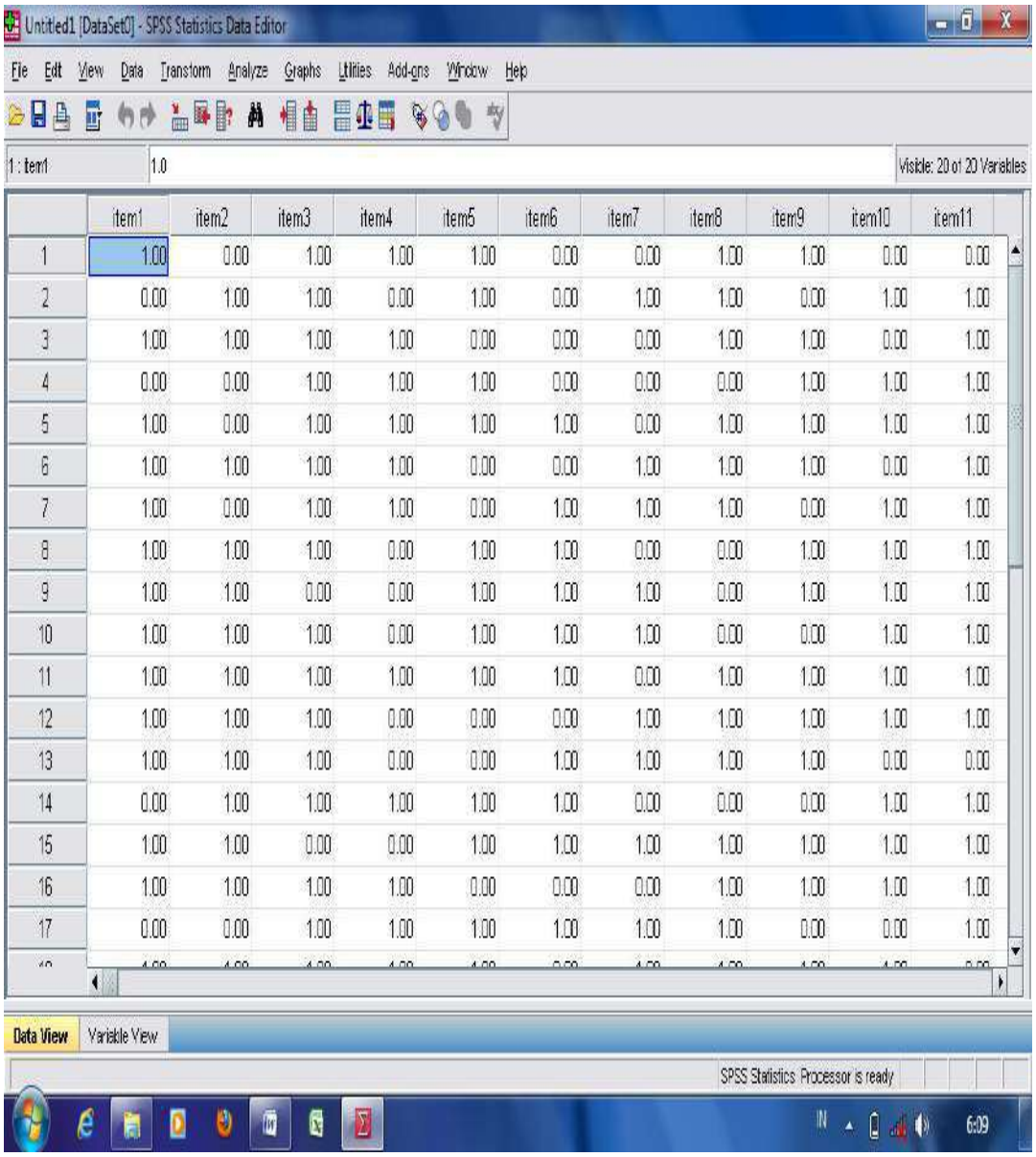


- Rename the “variabel view” into (e.g. Item1, Item2, and...)

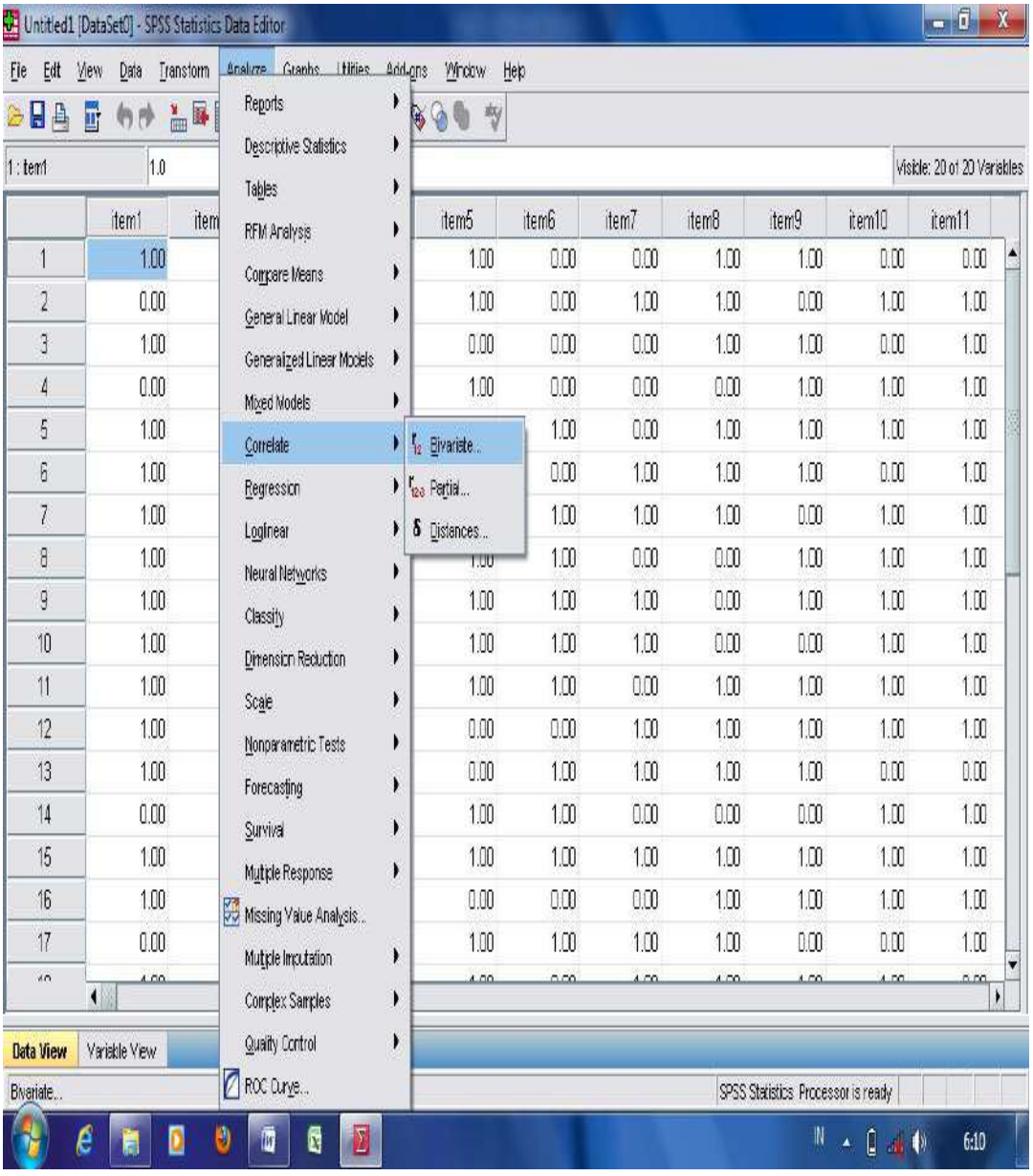


- Move into “data view” and input the score in each category”;“

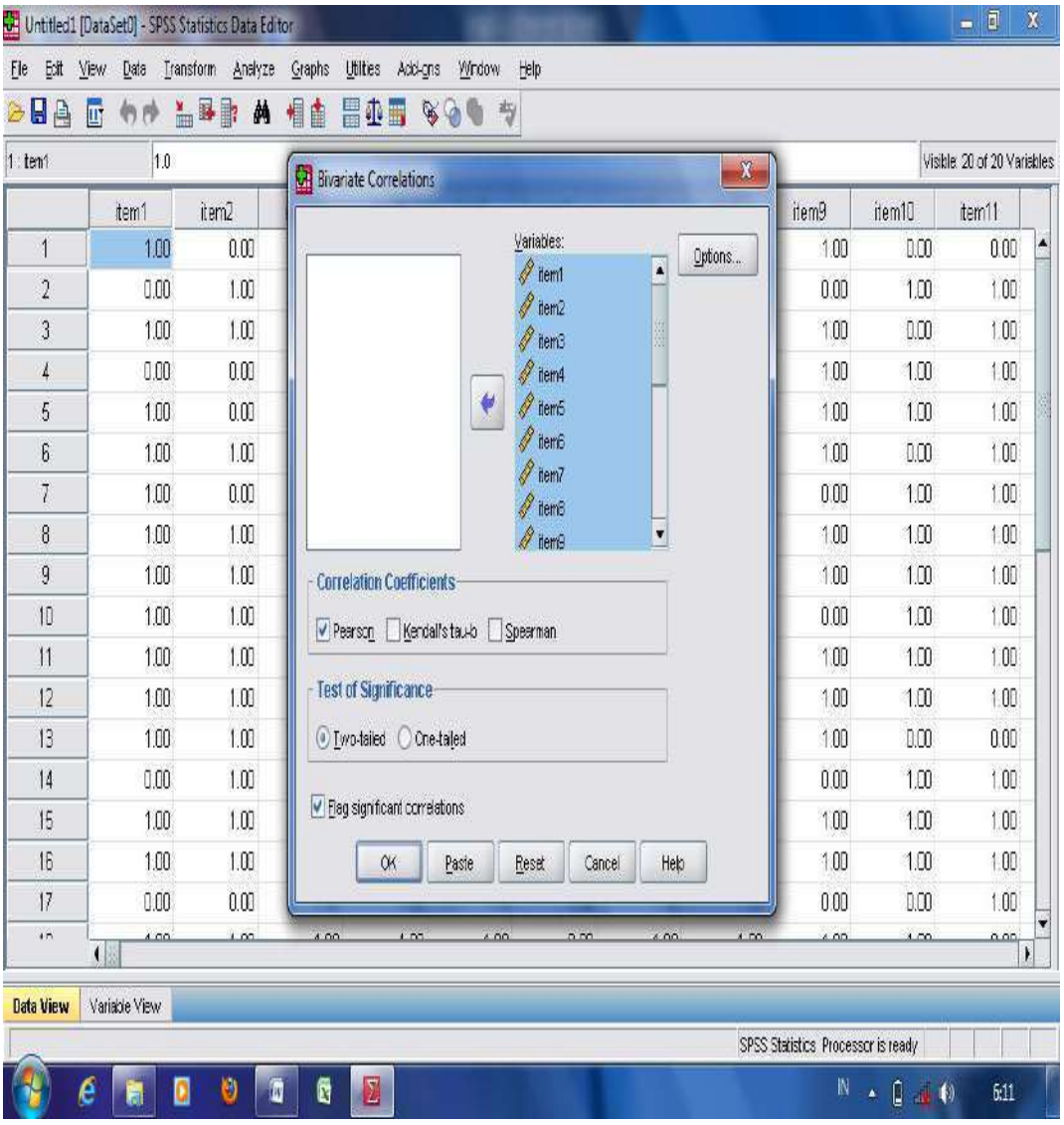
[34]



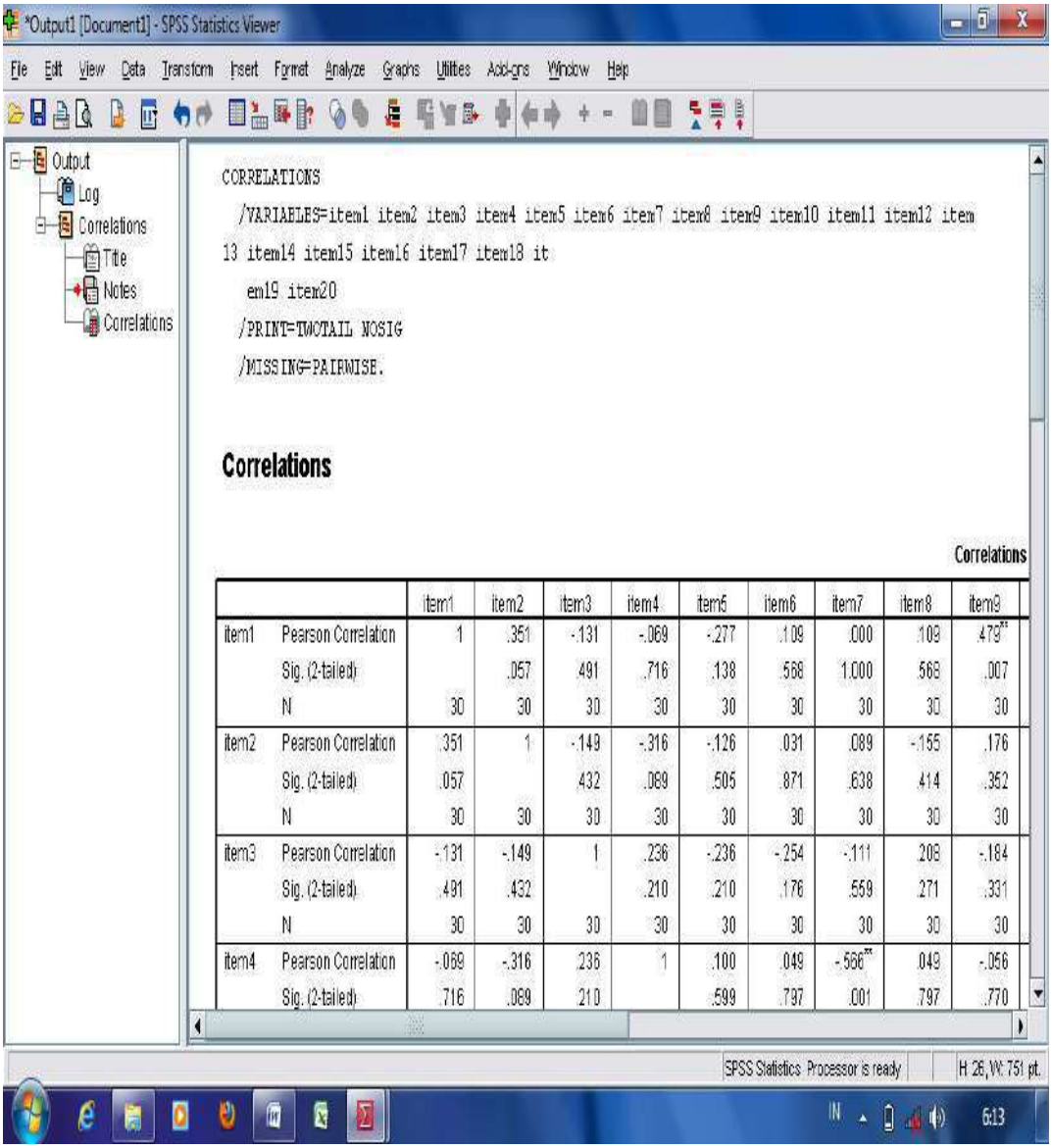
- Then, follow the next step to analyze the question items: Click "**Menu**", "**Analyze**", "**Correlate"**, and "**Bivariate**"…



- Move all items into "variables column" and check "**Pearson"**, Choose (2-tailed) in Test of Significance; then click OK.



- Check the validity of each question item into r-table with a certain number of sample (N=30) whether it is valid or not.
- The following table shows the result analysis of each question item on the students' emotional intelligence questionnaire analyzed using Pearson correlation coefficient.



From the result analysis of validity of each question item in the above table, it is found that there are 10 questions considered invalid. They are question item no1, no2, no5, no9, no11, no12, no16, no18, no19, and no20 since the scores of significance are lower than 0.361. Then, 10 questions items are considered valid. They are question item no3, no4, no6, no7, no8, no10, no13, no14, no15, and no17 since the scores of

significance are higher than 0.361. The result analysis of each question item is displayed in the following table.

No	Validity Test	Sig.(2-tailed) of Pearson Correlation	r-table score	Result
1	Item no 1	0	0.361	Invalid
2	Item no 2	0.057	0.361	Invalid
3	Item no 3	0.491	0.361	**Valid**
4	Item no 4	0.716	0.361	**Valid**
5	Item no 5	0.138	0.361	Invalid
6	Item no 6	0.568	0.361	**Valid**
7	Item no 7	1.000	0.361	**Valid**
8	Item no 8	0.568	0.361	**Valid**
9	Item no 9	0.007	0.361	Invalid
10	Item no 10	0.747	0.361	**Valid**
11	Item no 11	0.299	0.361	Invalid
12	Item no 12	0.115	0.361	Invalid
13	Item n0 13	0.797	0.361	**Valid**
14	Item n0 14	0.827	0.361	**Valid**
15	Item no 15	0.366	0.361	**Valid**
16	Item no 16	0.134	0.361	Invalid
17	Item no 17	0.366	0.361	**Valid**
18	Item no 18	0.299	0.361	Invalid
19	Item no 19	0.354	0.361	Invalid
20	Item no 20	0.272	0.361	Invalid

c. Content Validity

Hughes (1989:22) states a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc., with which it is meant to be concerned. A content validity is very important since it is an accurate measure of what it is supposed to measure. In order to judge whether or not a test has content validity, a

specification of the skills or structures should be made based on the curriculum and syllabus. Then, the result analysis in constructing the content validity is presented in the test of specification table including: objectives of the test, text's title, test indicators, number of test items, total of the questions, type of test, and answer keys. The following is example of constructing the test of specification table as it is shown in Table 6.

Table 6

Test of Specification Table

Objectives	Test Materials	Indicators	Number of items	Tot	Types of Test	Answer Key
The students are able to respond the written meanig of reading text	The lost caterpillar	The students are able: - to identify main idea - to find the detail and factual informati on - to find inferenced word - to find a concludin g sentence	1, 5, 6 2, 3, 4 7, 8 9, 10	3 3 2 2	Multiple Choices	a, c, b a, d a, c c, d

2. **Reliability Test**

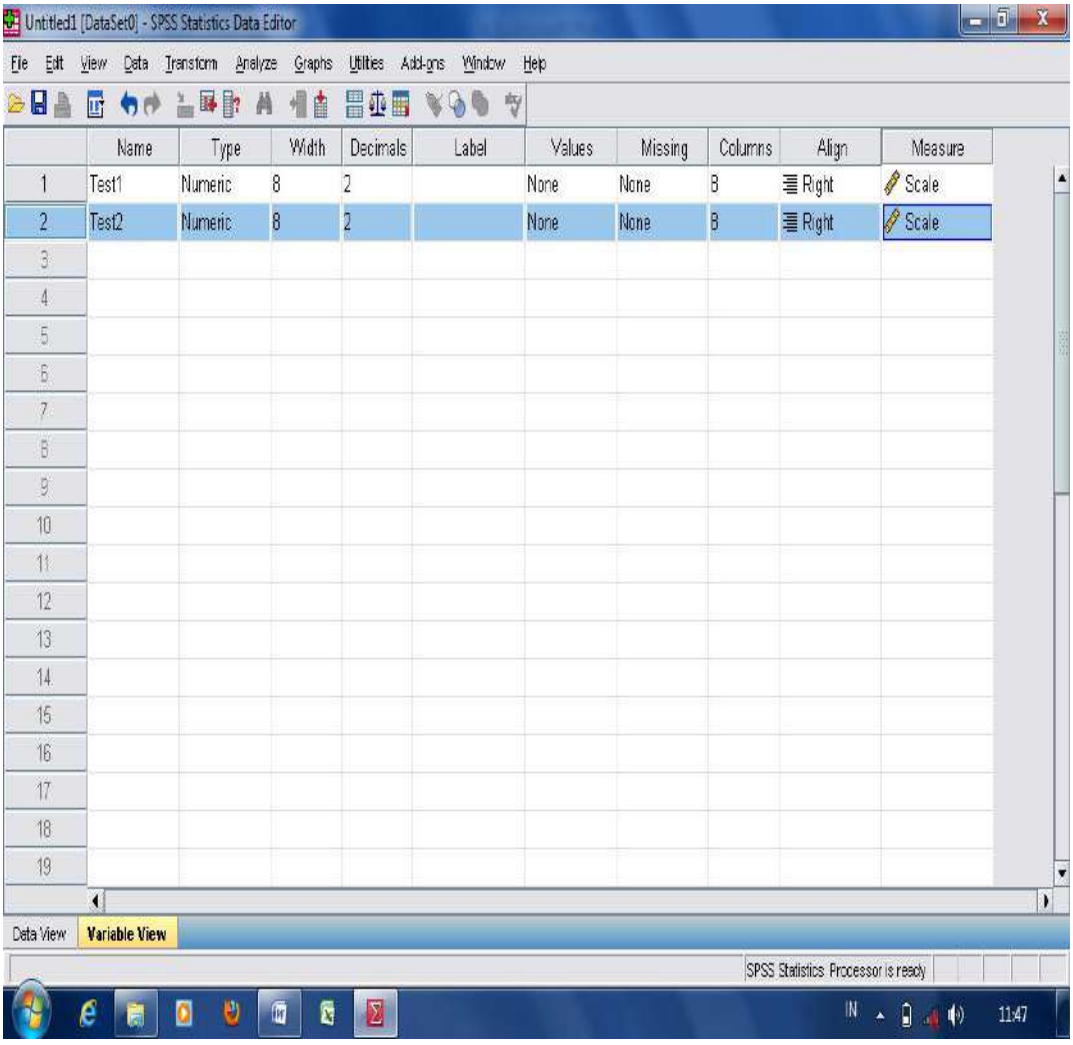
Fraenkel and Wallen (1990, p.133) state that reliability test is used to measure the consistency of the two scores obtained for each individual from one administration of an instrument to another and from one set of items to another. Or it can be stated that a student who receives a high score the first time he takes the test to receive a high score the next time he takes the test. The score would probably not be identical, but they should be close. Then, Fraenkel and Wallen (1990, p.134) mention that a reliability coefficient expresses a relationship between scores of the same individual on the same instrument at two different times. Furthermore, Fraenkel and Wallen (1990, p.136) state that the score is considered reliable if the score of significance is at least or preferably higher than 0.70. There are three best known ways to obtain a reliability coefficient. They are as follows.

a. **Test-Retest Method**

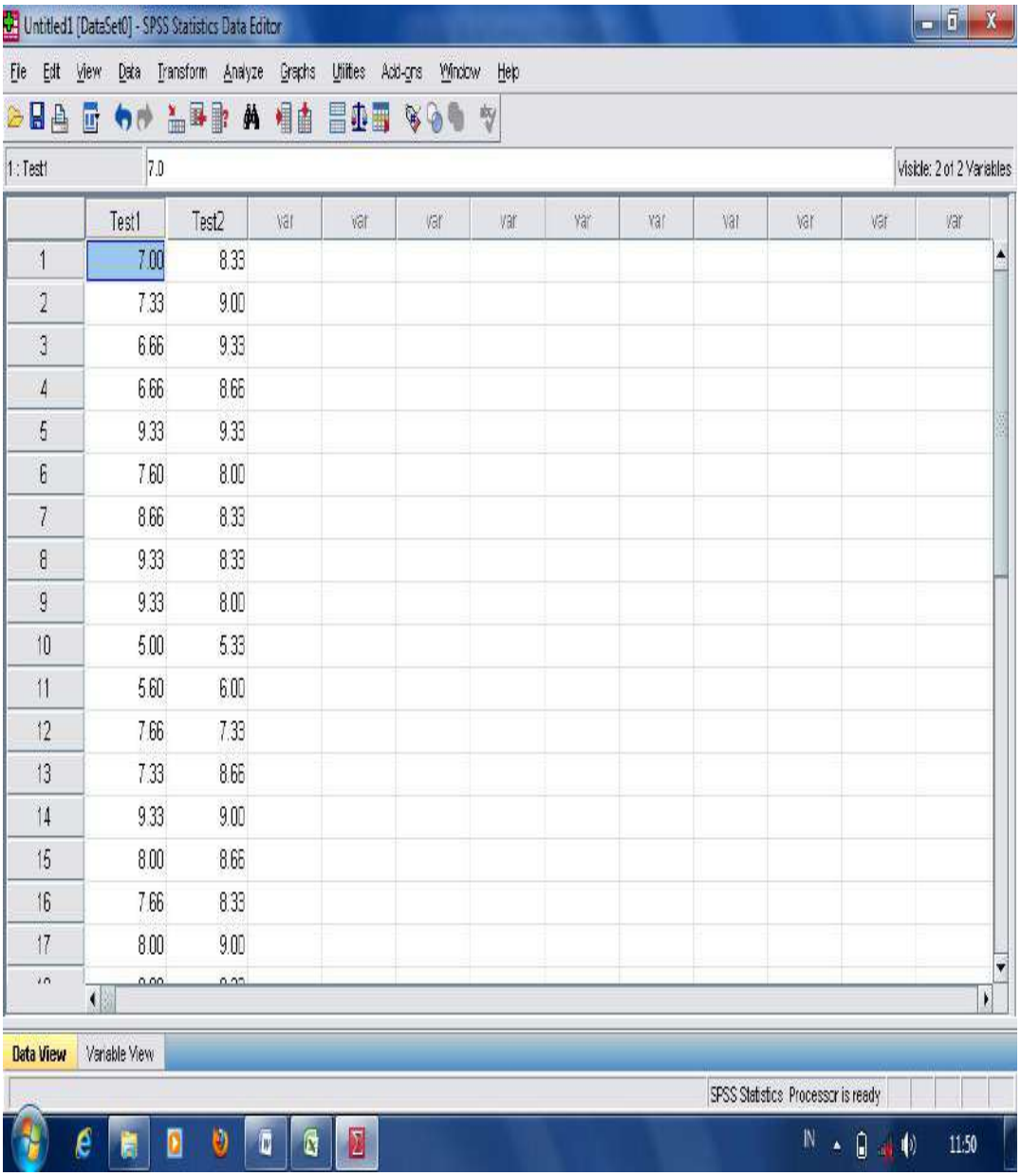
In checking for evidence of test-retest reliability, an appropriate time interval should be selected. This interval should be that during which individuals would be assumed to retain their relative position in a meaningful group. Test-retest method measures the stability of test scores over time which involves administering the same instrument twice to the same group of individiuals after a certain time interval has elapsed (Fraenkel and Wallen, 1990, p.139). A reliability coefficient is then calculated to indicate the relationship between the two sets of scores obtained. Reliability coefficients will be affected but the length of time that elapses between the two administrations of the test. The longer the time interval, the lower the reliability coefficient is likely to be, since there is a greater likelihood of changes in the individuals taking the test. To measure the test-retest method, **Pearson correlation coefficient** found in SPSS is used. The followings are steps to start the analysis:

- Open the worksheet in the application of SPSS program; and Move the cursor into “variable view”; then, type the column

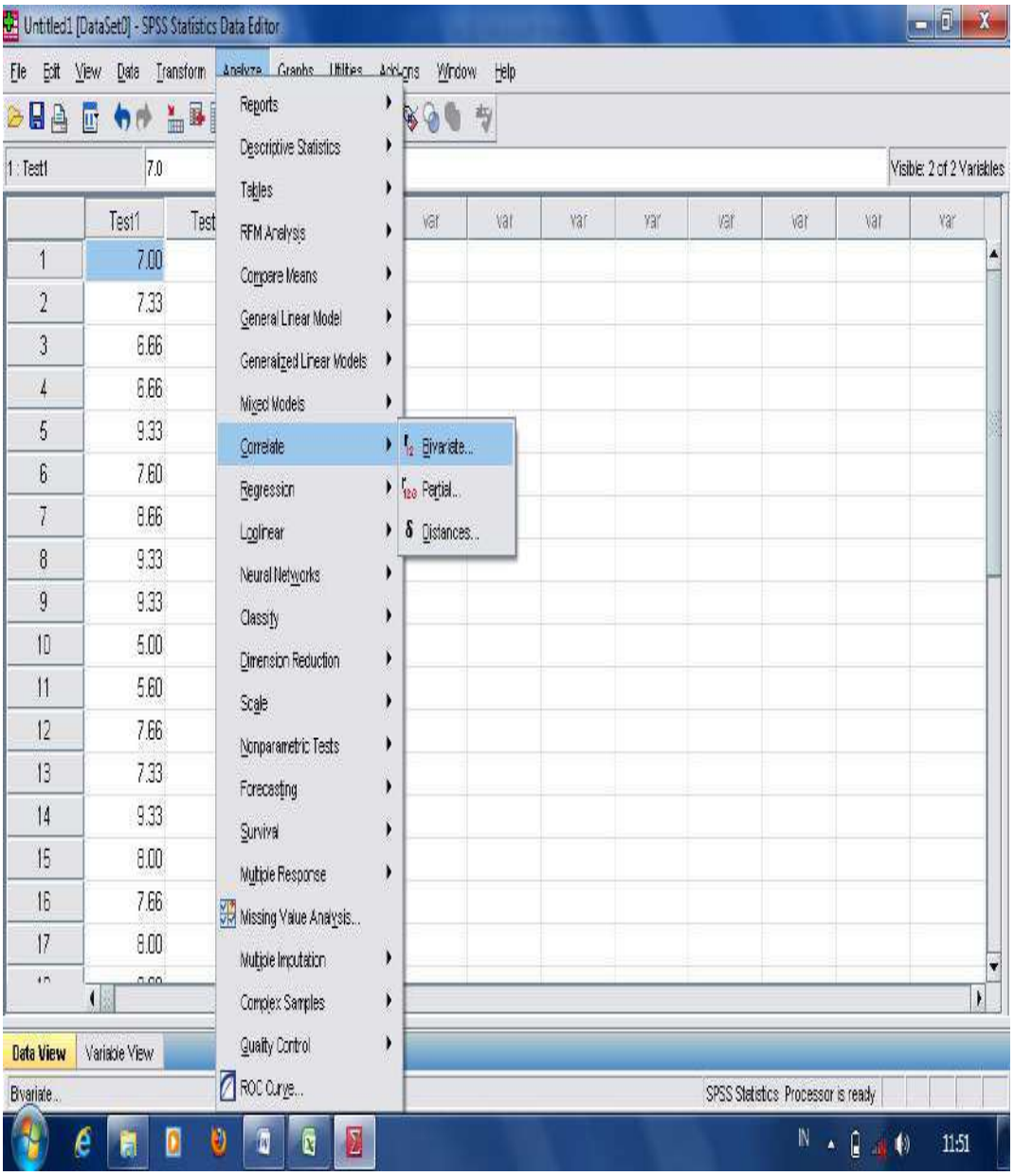
into the category which is going to be compared namely: Test1, and Test2;



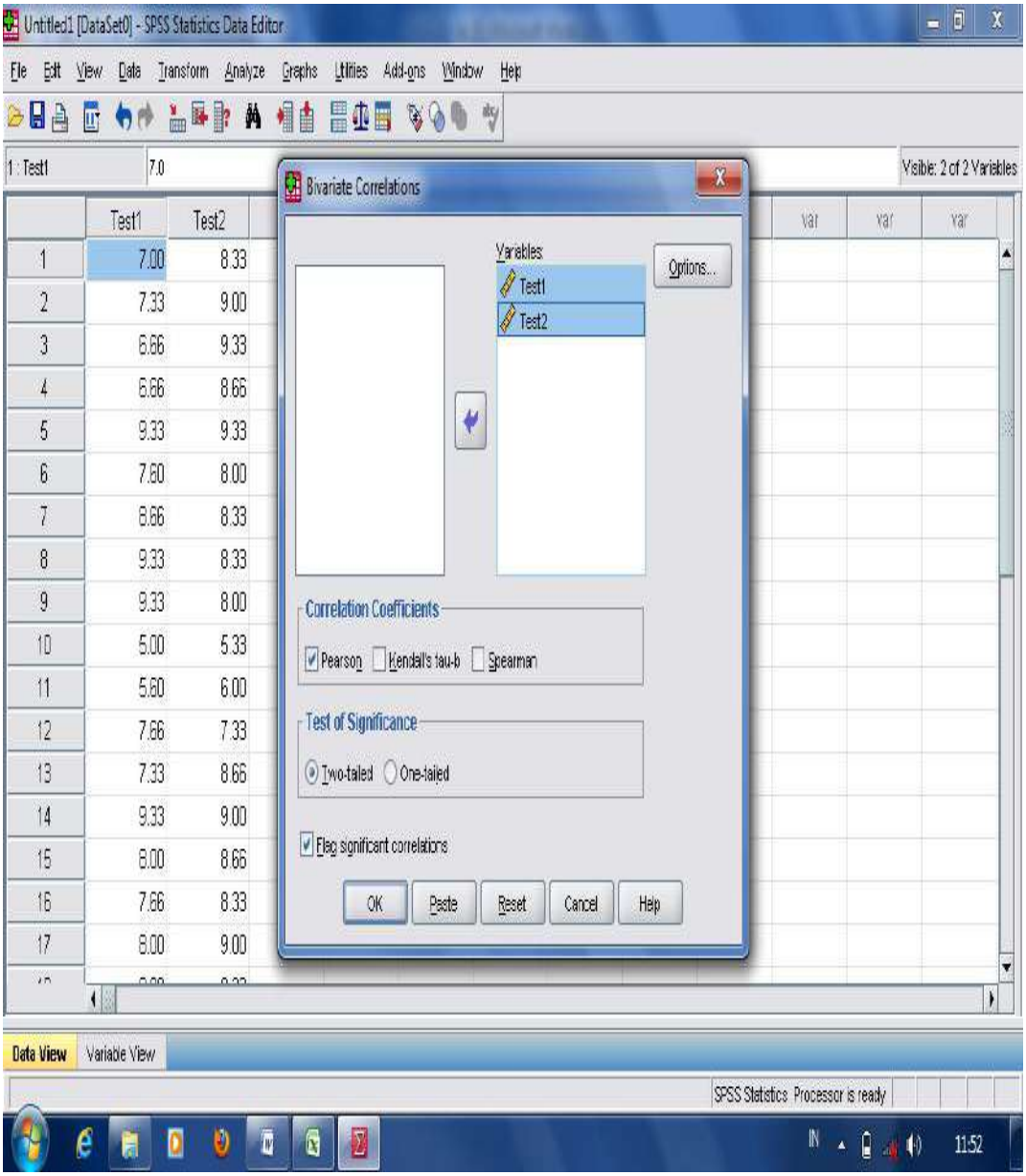
- Move the cursor into “data view” and then, input the score of tryout analysis in each category;



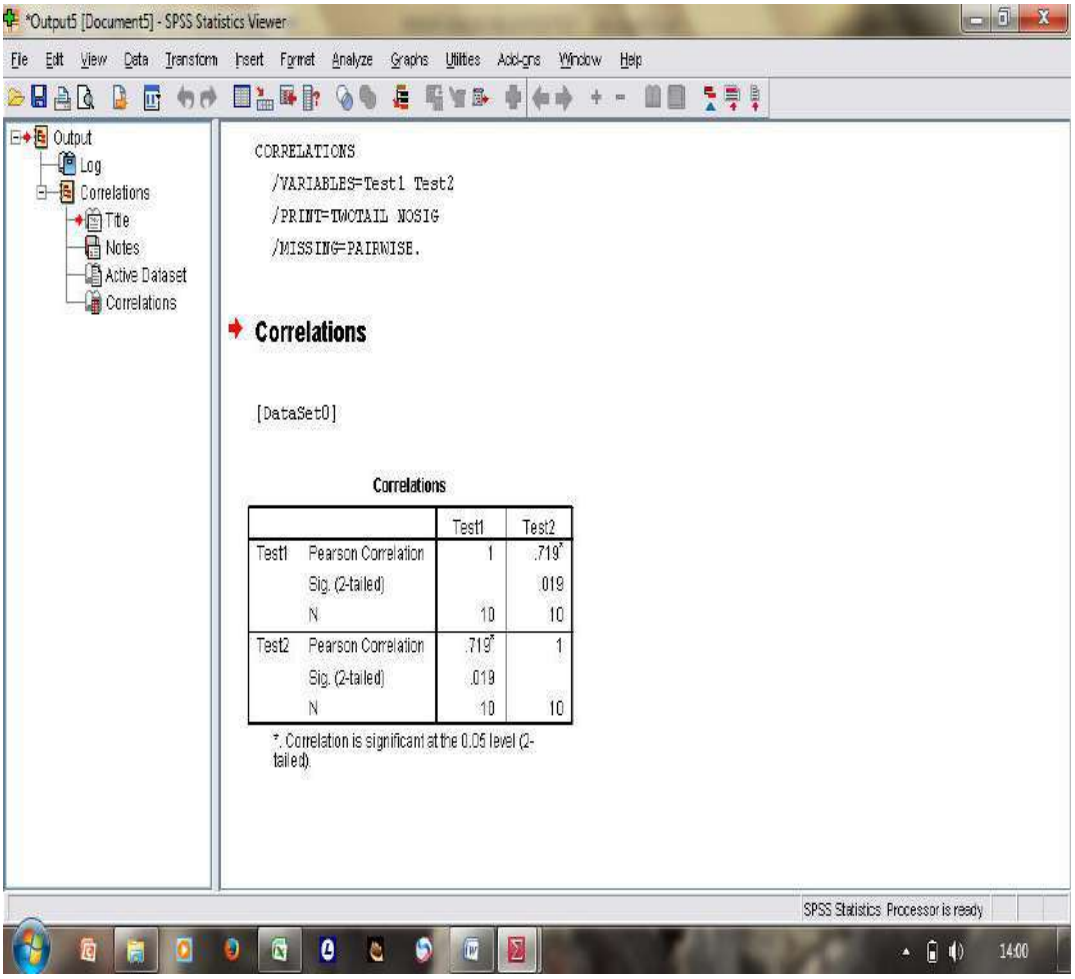
- After that, click "**Analyze**" then choose "**Correlate**". Move the cursor into right side and select "**Bivariate**";



- Then, move all the variables in the right side into the column of variable; then press OK.



- The result analysis in measuring test-retest method is displayed as follows.



The result analysis of reliability test shows that the score of **Pearson Correlation** is 0.719. From the p-output, it can be stated that the test instrument is reliable since it is higher than 0.70.

b. Equivalent-Forms Method and Equivalent-Forms *plus* Test-Retest Method

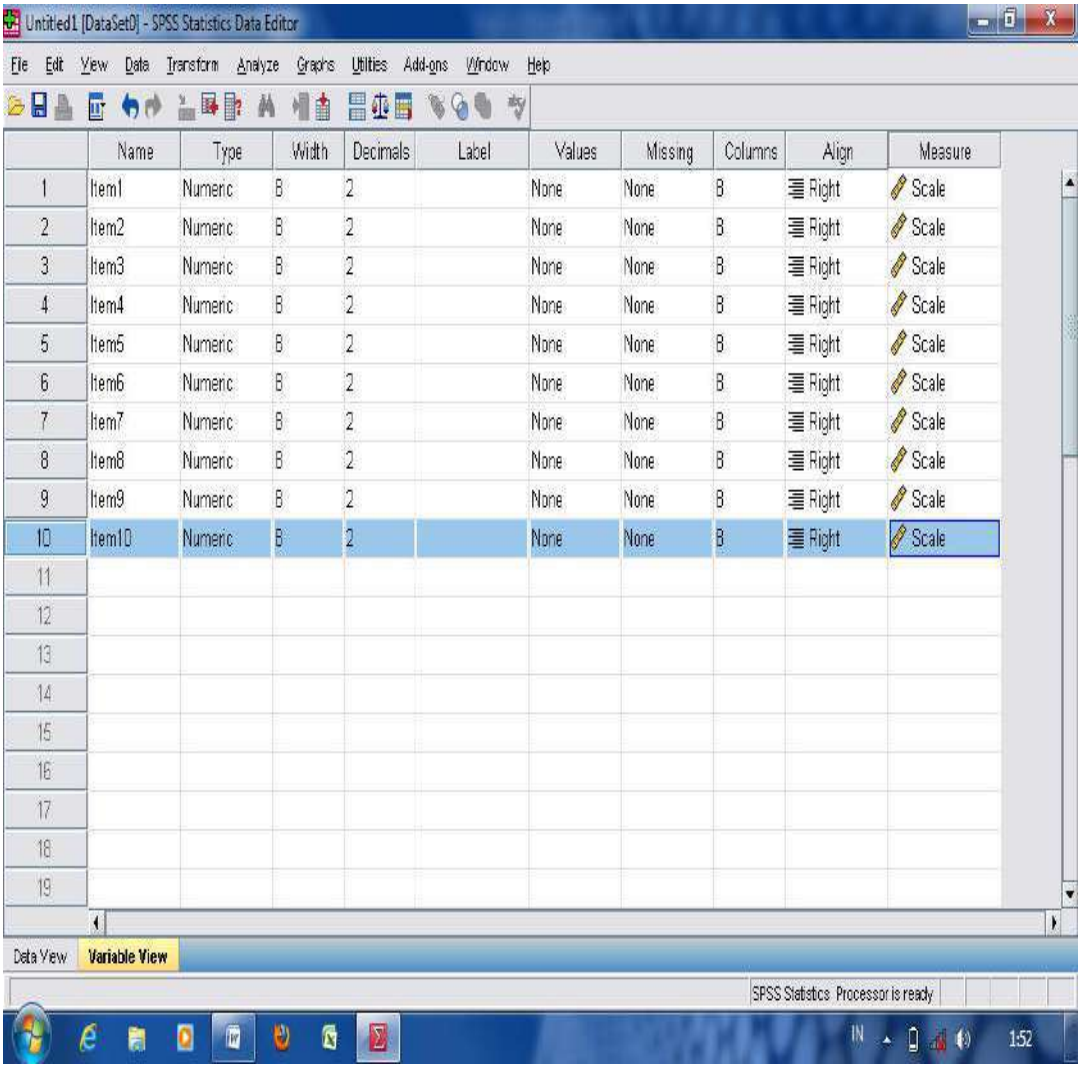
The equivalent-forms method measures the consistency of test scores over two different, but equivalent, forms of an instrument to the same group of individuals at the same time (Fraenkel and Wallen, 1990, p.139). While, the equivalent-forms plus test-retest method is used to measure the consistency of scores over two different forms and a time

interval. Although the questions are different, they should sample the same content and they should be constructed separately from each other. A reliability coefficient is then calculated between the two sets of scores obtained. A high coefficient would indicate strong evidence of reliability that the two forms are measuring the same thing.

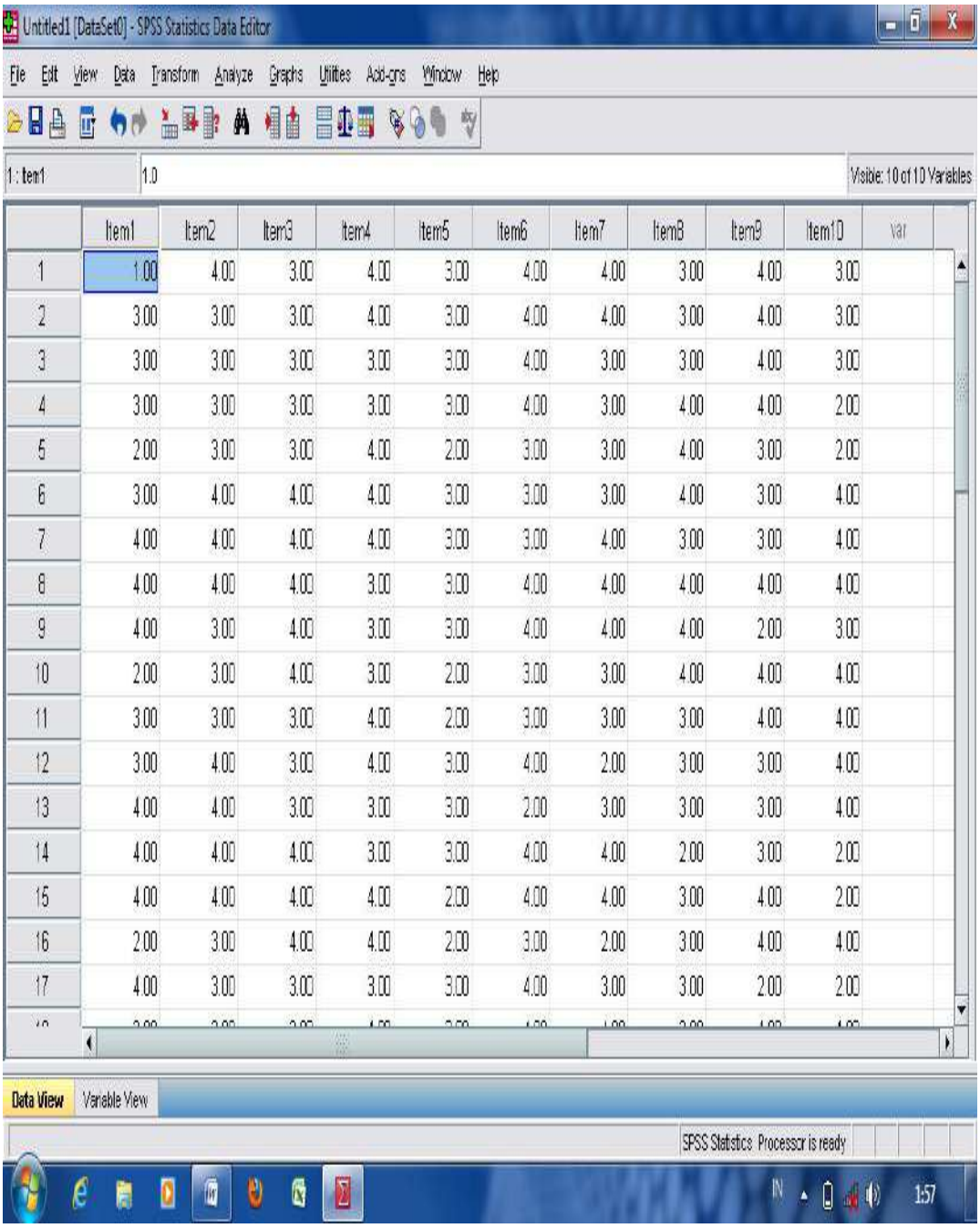
c. Internal Consistency Method

Internal consistency method is used to estimate the reliability which involves comparing different sets of items that are part of an instrument (Fraenkel and Wallen, 1990, p.139). There are several types of internal consistency methods in estimating reliability that require only a single administration of an instrument. They are: split-half procedure, Kuder-Richardson Approaches (KR21), and Alpha Coefficient.

1. **Split-half Procedure**; involes scoring two halves (usually odd items versus even items) of a test separately for each peson and then calculating a correlation for the two sets of scores. The coefficient indicates the degree to which the two halves of the test provide the same results, and hence describes the internal consistency of the test. The reliability coefficient is calculating using what is known as the **Spearman-Brown prophecy formula**. The follwing is example of statistical analysis in measuring the reliability test using SPSS: A research study in analyzing students' learning styles on students' reading comprehension subject. There are 10 question on students' learning style' questionnaire with sample (N) is 20 students. The followings are steps to start the reliability analysis using split half procedure found in SPSS application program:
   - Open the worksheet in the application of SPSS program;
   - Move the cursor into "variable view"; Then, type the coloum into the category that is going to be compared namely: Item1, Item2, and....

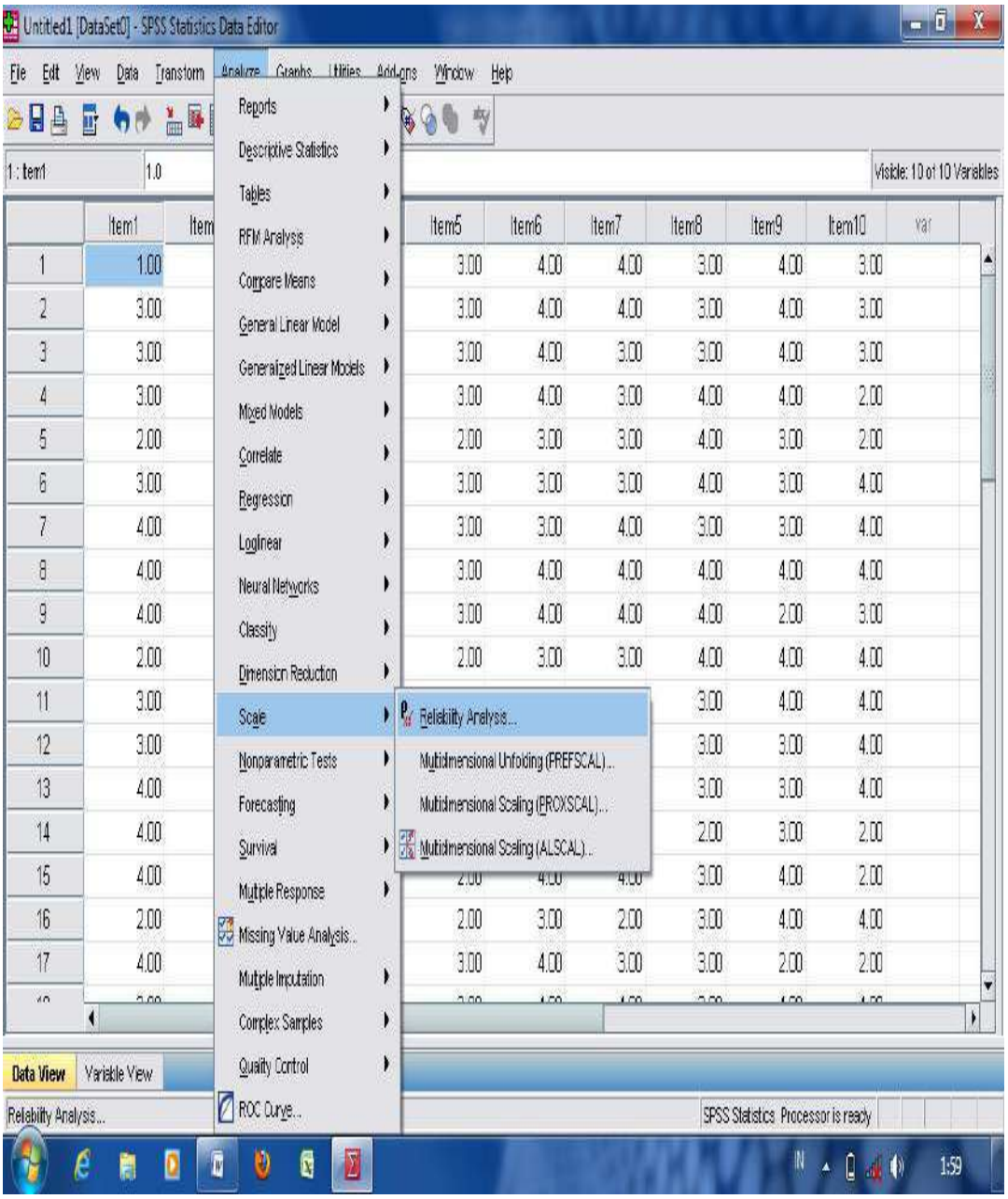


- Move your cursor into “data view”, and then, input the score of tryout analysis in each category;

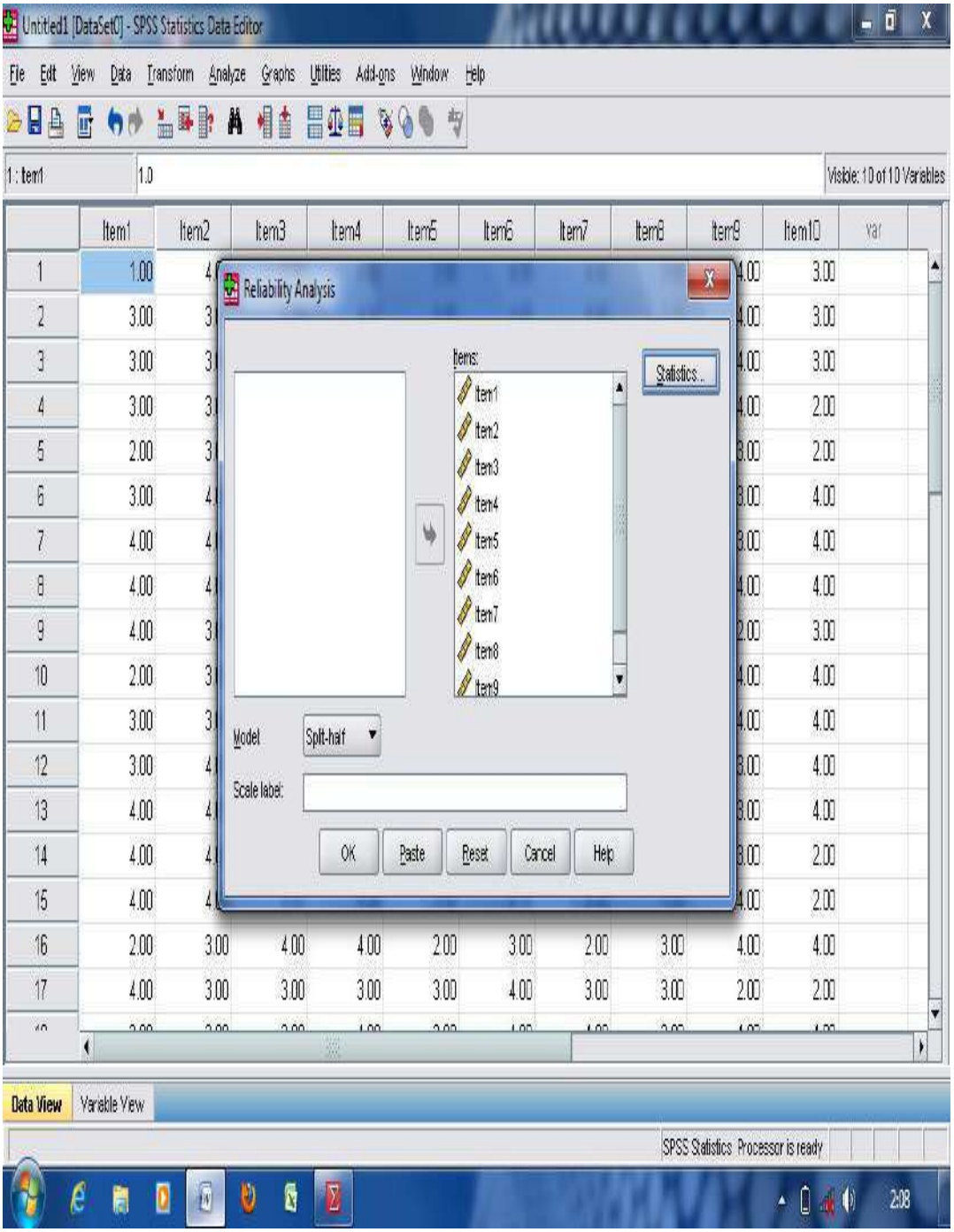


	item1	item2	item3	item4	item5	item6	item7	item8	item9	item10	var
1	1.00	4.00	3.00	4.00	3.00	4.00	4.00	3.00	4.00	3.00	
2	3.00	3.00	3.00	4.00	3.00	4.00	4.00	3.00	4.00	3.00	
3	3.00	3.00	3.00	3.00	3.00	4.00	3.00	3.00	4.00	3.00	
4	3.00	3.00	3.00	3.00	3.00	4.00	3.00	4.00	4.00	2.00	
5	2.00	3.00	3.00	4.00	2.00	3.00	3.00	4.00	3.00	2.00	
6	3.00	4.00	4.00	4.00	3.00	3.00	3.00	4.00	3.00	4.00	
7	4.00	4.00	4.00	4.00	3.00	3.00	4.00	3.00	3.00	4.00	
8	4.00	4.00	4.00	3.00	3.00	4.00	4.00	4.00	4.00	4.00	
9	4.00	3.00	4.00	3.00	3.00	4.00	4.00	4.00	2.00	3.00	
10	2.00	3.00	4.00	3.00	2.00	3.00	3.00	4.00	4.00	4.00	
11	3.00	3.00	3.00	4.00	2.00	3.00	3.00	3.00	4.00	4.00	
12	3.00	4.00	3.00	4.00	3.00	4.00	2.00	3.00	3.00	4.00	
13	4.00	4.00	3.00	3.00	3.00	2.00	3.00	3.00	3.00	4.00	
14	4.00	4.00	4.00	3.00	3.00	4.00	4.00	2.00	3.00	2.00	
15	4.00	4.00	4.00	4.00	2.00	4.00	4.00	3.00	4.00	2.00	
16	2.00	3.00	4.00	4.00	2.00	3.00	2.00	3.00	4.00	4.00	
17	4.00	3.00	3.00	3.00	3.00	4.00	3.00	3.00	2.00	2.00	

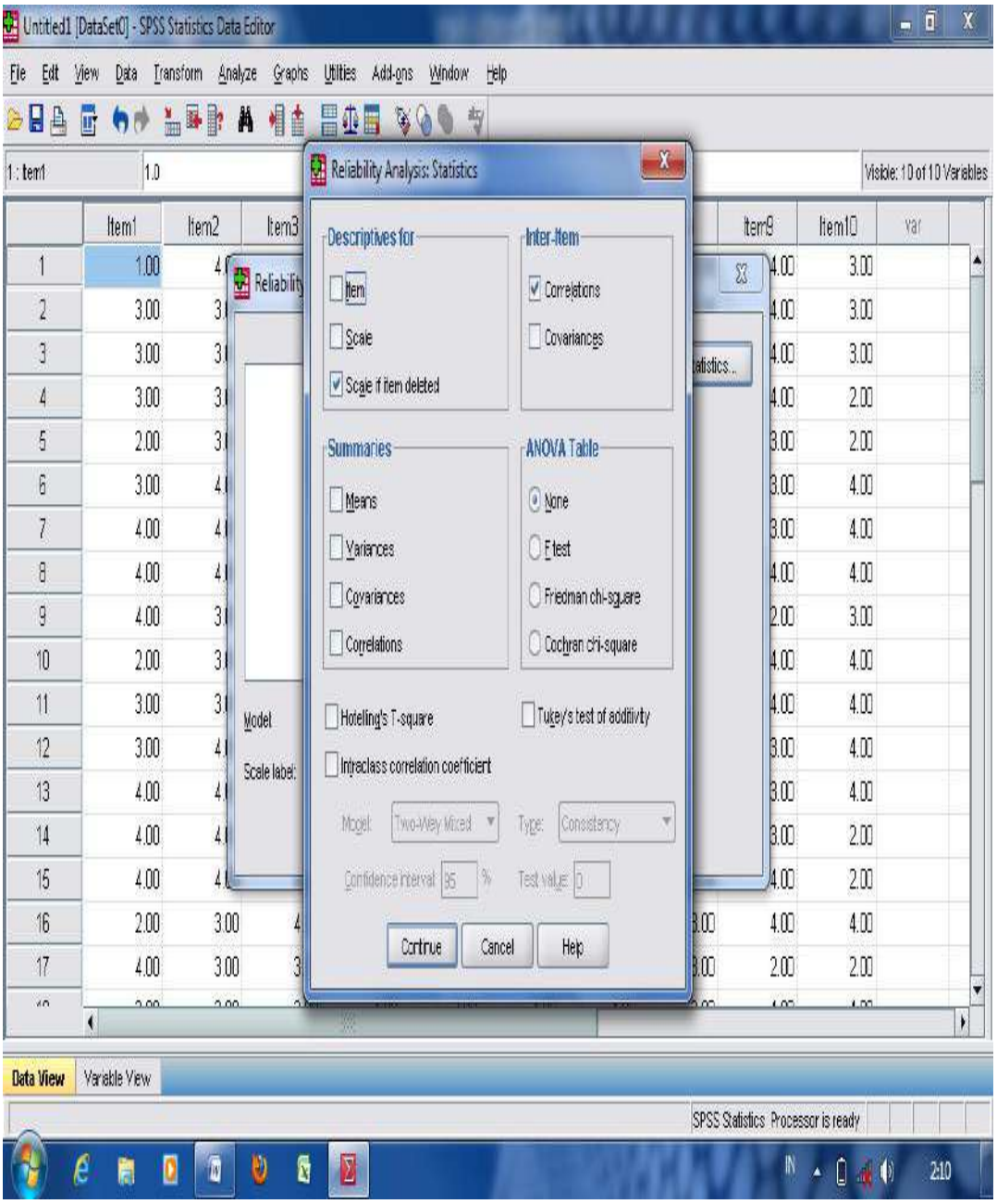
- Then click "**Analyze"**, "**Scale"**, and "**Reliability Analysis**";



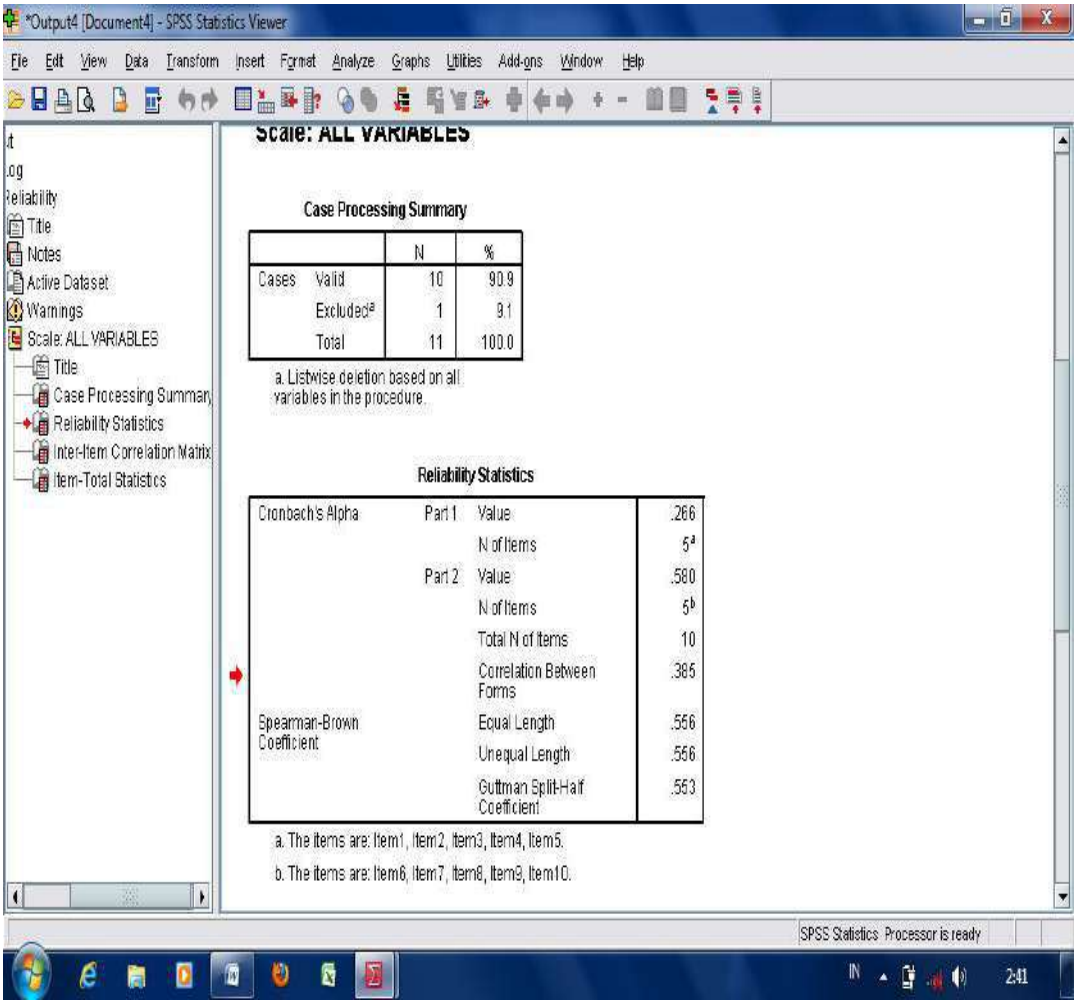
- Move all items into the right box provided, then change the model box into "**Split-Half**";



- Then, press the statistic button; check in the "**scale if item deleted**" box, and the "**correlations**" box; and Click Continue and OK.



- The ouput of the result analysis in reliability test using **split-half method** from students' learning styles on students' reading comprehension average score as follows.



From the result of measuring reliability test using split-half method, it was found that the p-output of **Gutmann Split-Half Coefficient** is 0.533. From the score it can be stated that the reliability of learning styles on students' reading comprehension average score is reliable since the p-output is higher than r-table (0.444) with sample (N) is 20 students.

2. **KR21;** the formula of KR21 require only three pieces of information—the number of items in the test, the mean, and the standard deviation. The formula of KR21 can be used only if it can be assumed that the items are of equal difficulty. The

formula of KR-21 reliability coefficiet is suggested by Fraenkel, et.al (2012:156) as follows.

$$\text{KR-21} = \frac{K}{K-1}\left[1-\frac{M(K-M)}{K(SD^2)}\right]$$

Where:

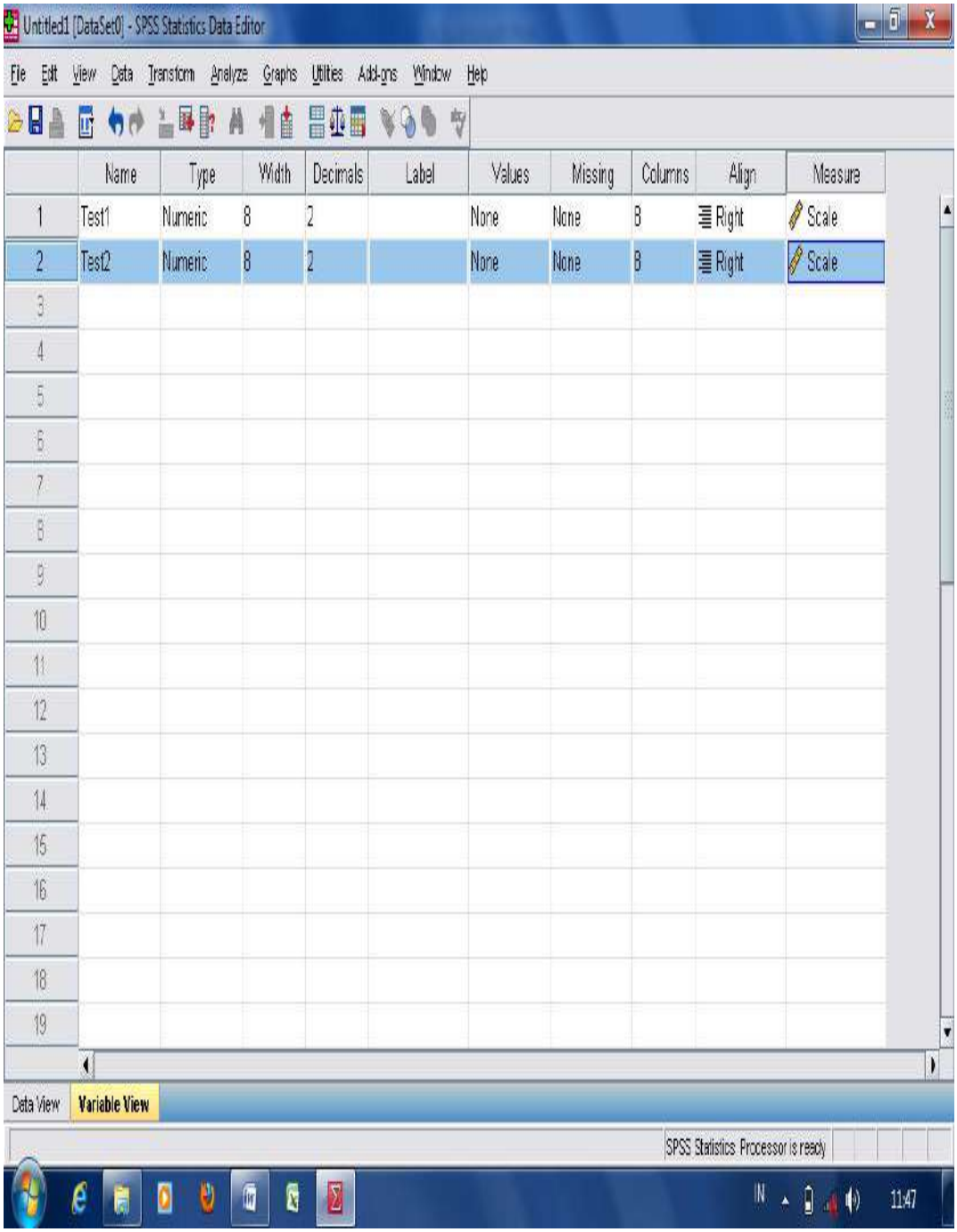
K = number of items on the test
M = mean of the set of test scores
SD = standard deviation of the set of test scores.

$$\text{SD} = \sqrt{\sum\frac{(x-\bar{x})^2}{N}}$$

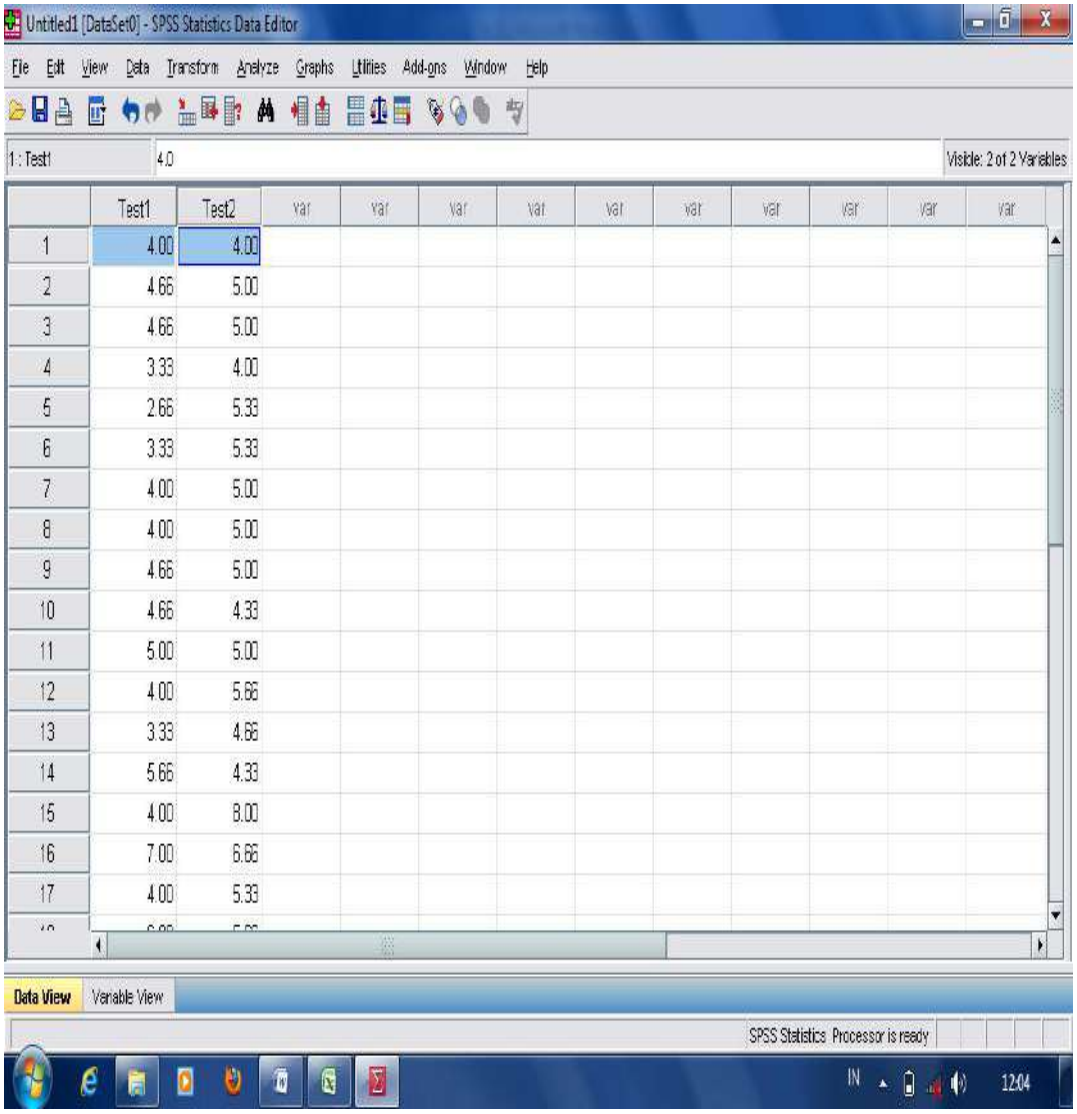
Where:

SD = standard deviation of the set of test scores
x = number of correct answer
$\bar{x}$ = students' means scores
N = number of students

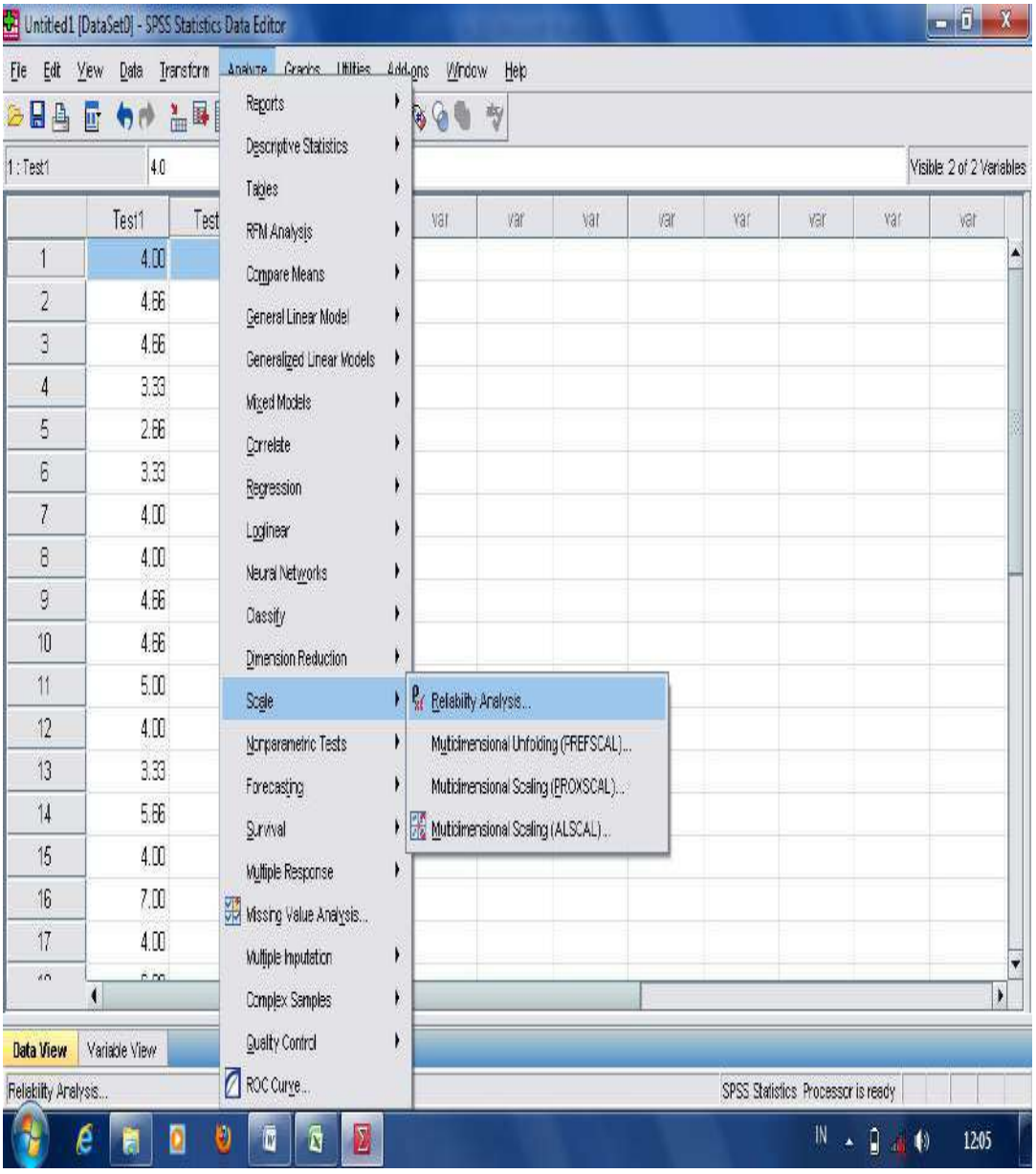
3. **Alpha Coefficient**; is also called **Cronbach Alpha** which is used to measure consistency of test scores over different parts of an instrument**.** Cronbach Alpha is a general form of the KR20 formula to be used in calculating the reliability items that are not scored right versus wrong, as in some "***essay tests"*** where more than one answer is possible. The followings are steps to start the reliability analysis using Cronbach Alpha found in SPSS application program:
   - Open the worksheet in the application of SPSS program; Move the cursor into "Variable View"; Then, type the coloum into the category that is going to be compared namely: Test1, and Test2;



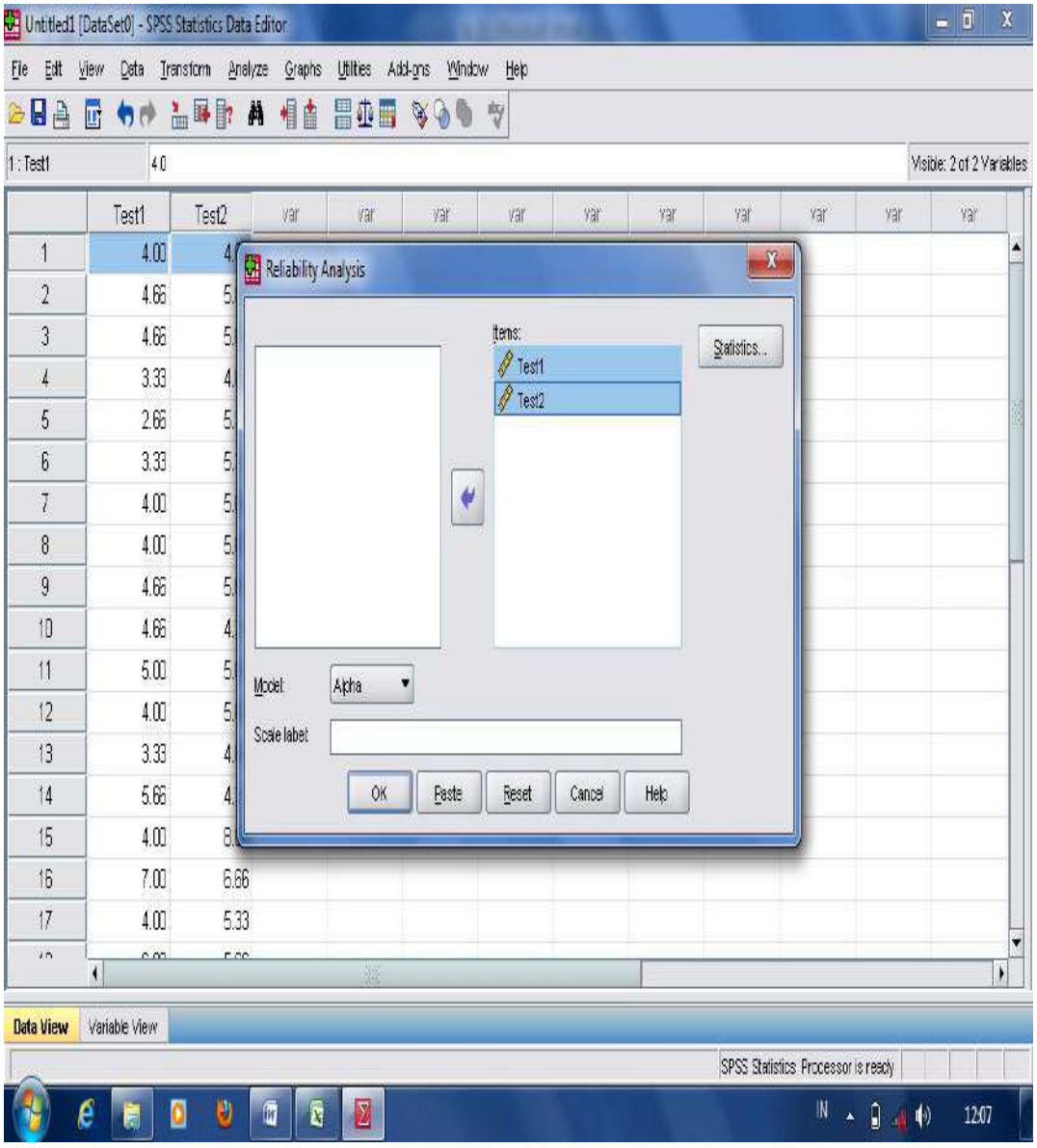
- Move your cursor into “data view”, and then, input the score of tryout analysis in each category;



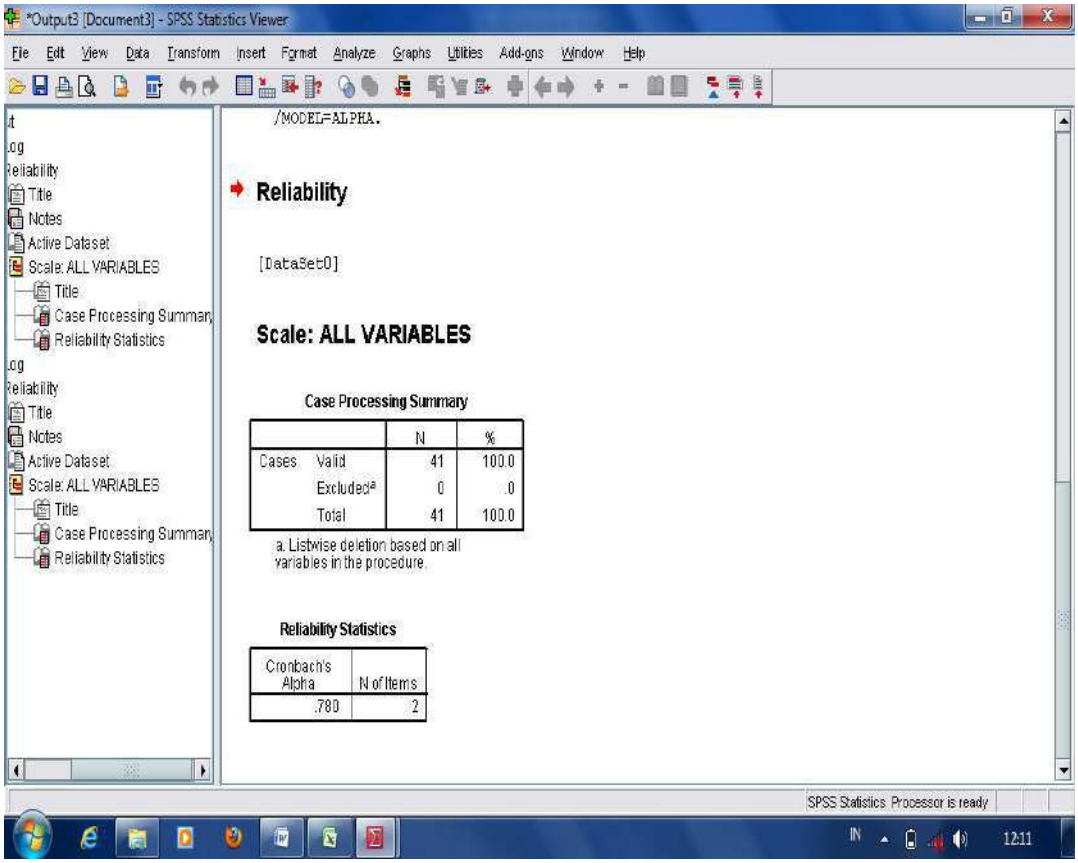
- Then, click "**Analyze**", "**scale"**, and "**reliability analysis**";



- Move all the scores into the box of items in the right side; then, press OK.



- The following is the result analysis in measuring reliability test using Cronbach's Alpha.



From the result analysis above, it is found that the p-output was 0.780. From the score, it can be stated that the students' test is reliable since the p-output is higher than 0.70.

I. Data Distributions

In analyzing the data distributions, there are two analyses to be done, they are described as followed.

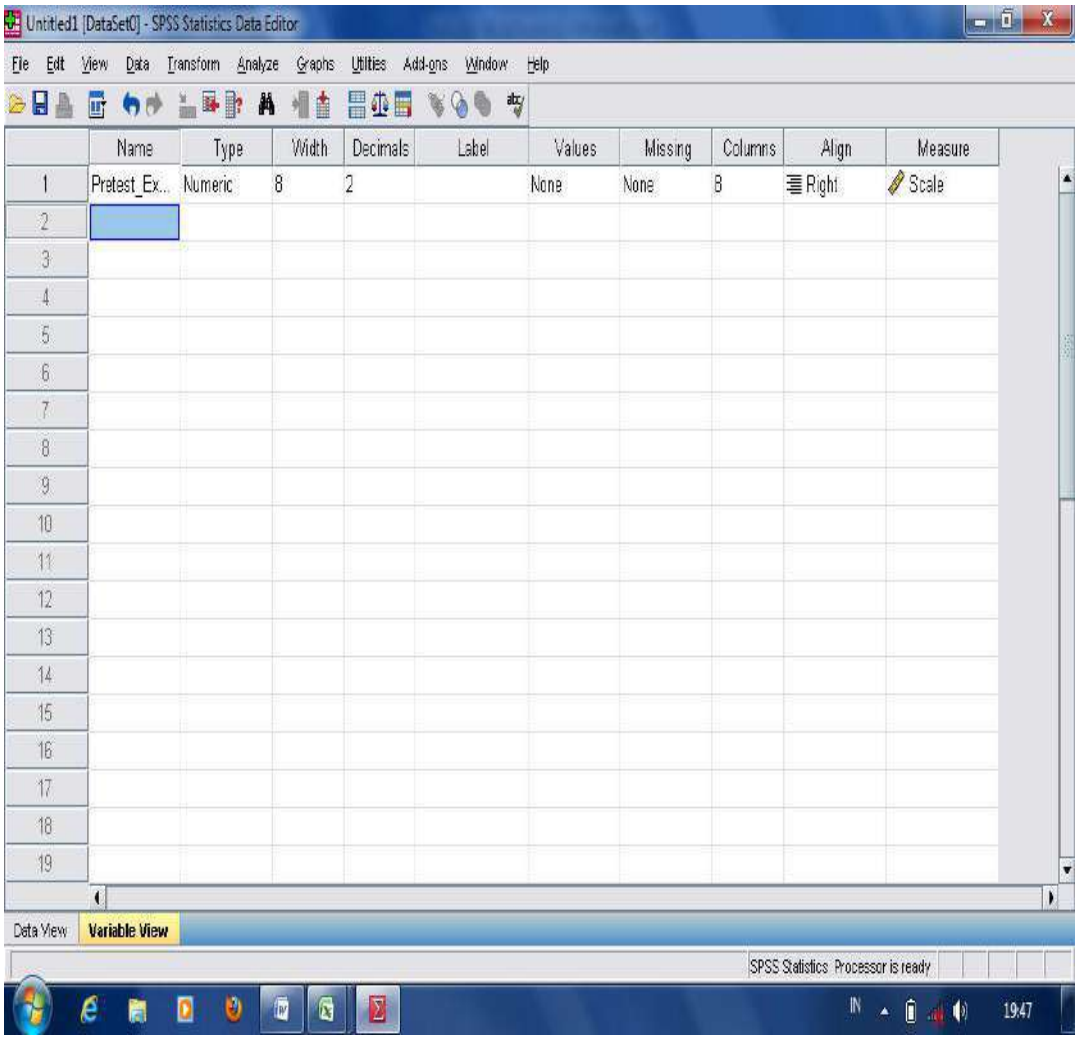
1. Data Discriptions

In data discriptions, there are two analyses to be done. They are: (a) distributions of frequency table, and (b) descriptive statistics. The scores are obtained from students' pretest and posttest in both groups (control and experiment).

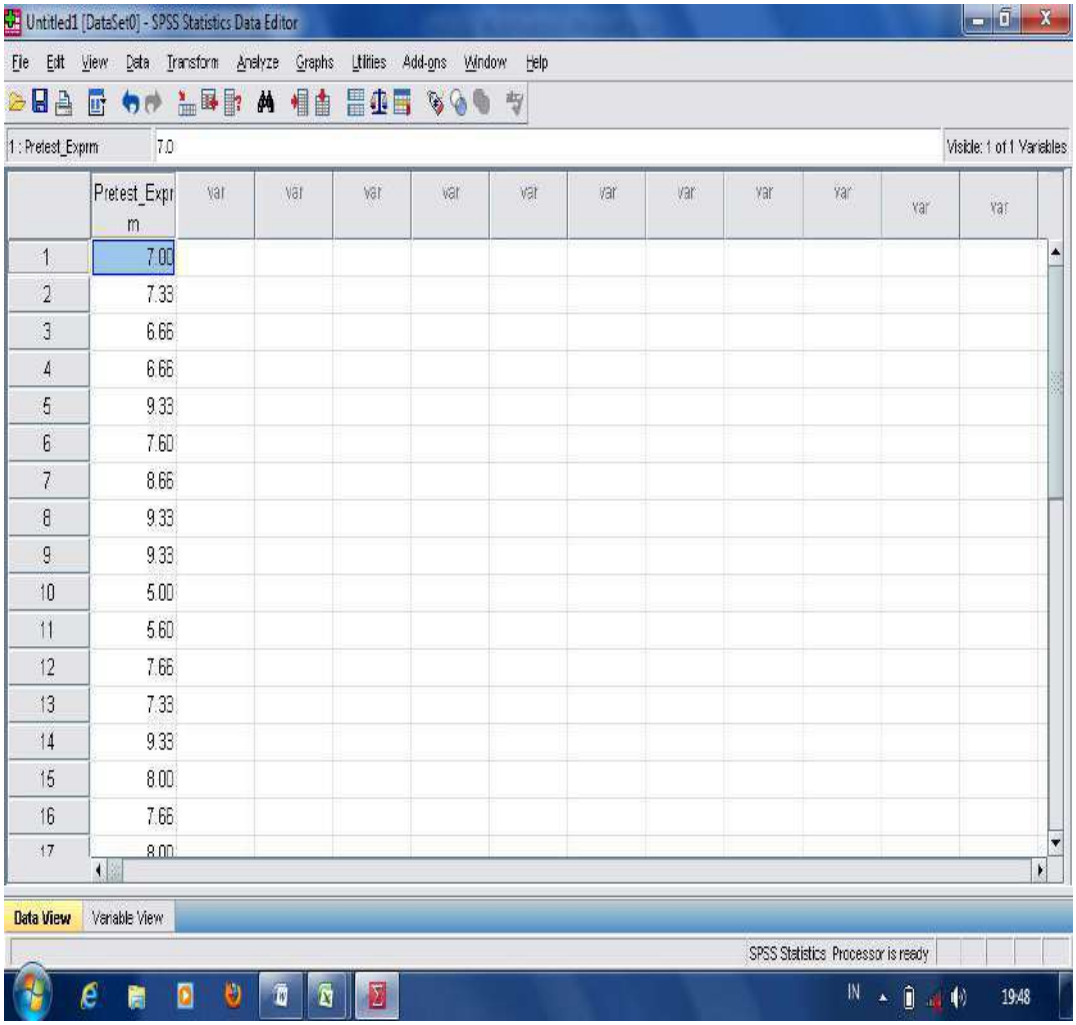
a. **Distributions of frequency data**; in this part, the students' scores are described by presenting a number of student who got

a certain score, and its score's percentage. To start the analysis in frequency table, it is done as in the following steps:

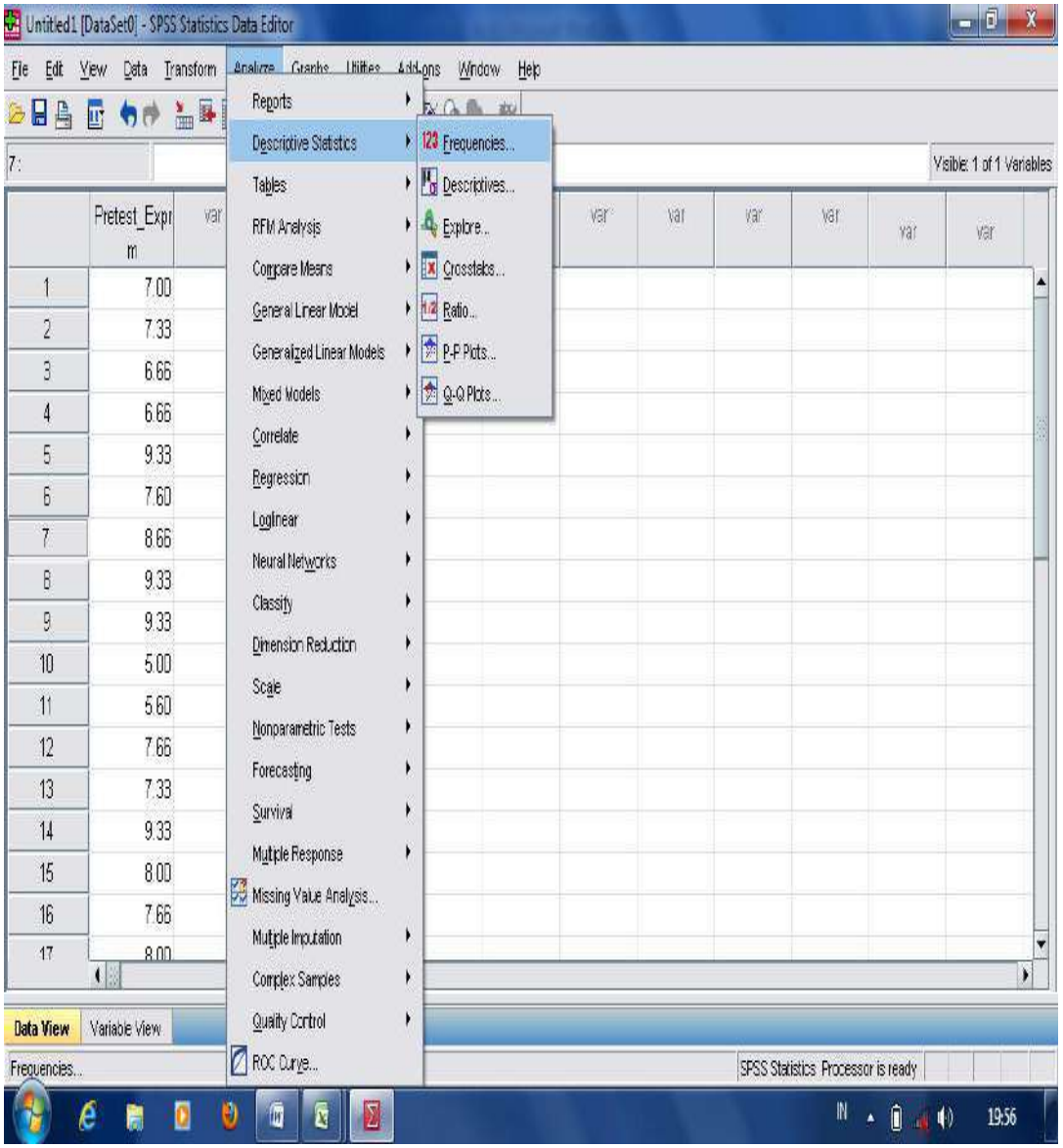
- Open the worksheet in the application of SPSS program; Move into a "variable view"; type the coloum into the category that is going to be compared for example, the variable is named as "Pretest experiment";



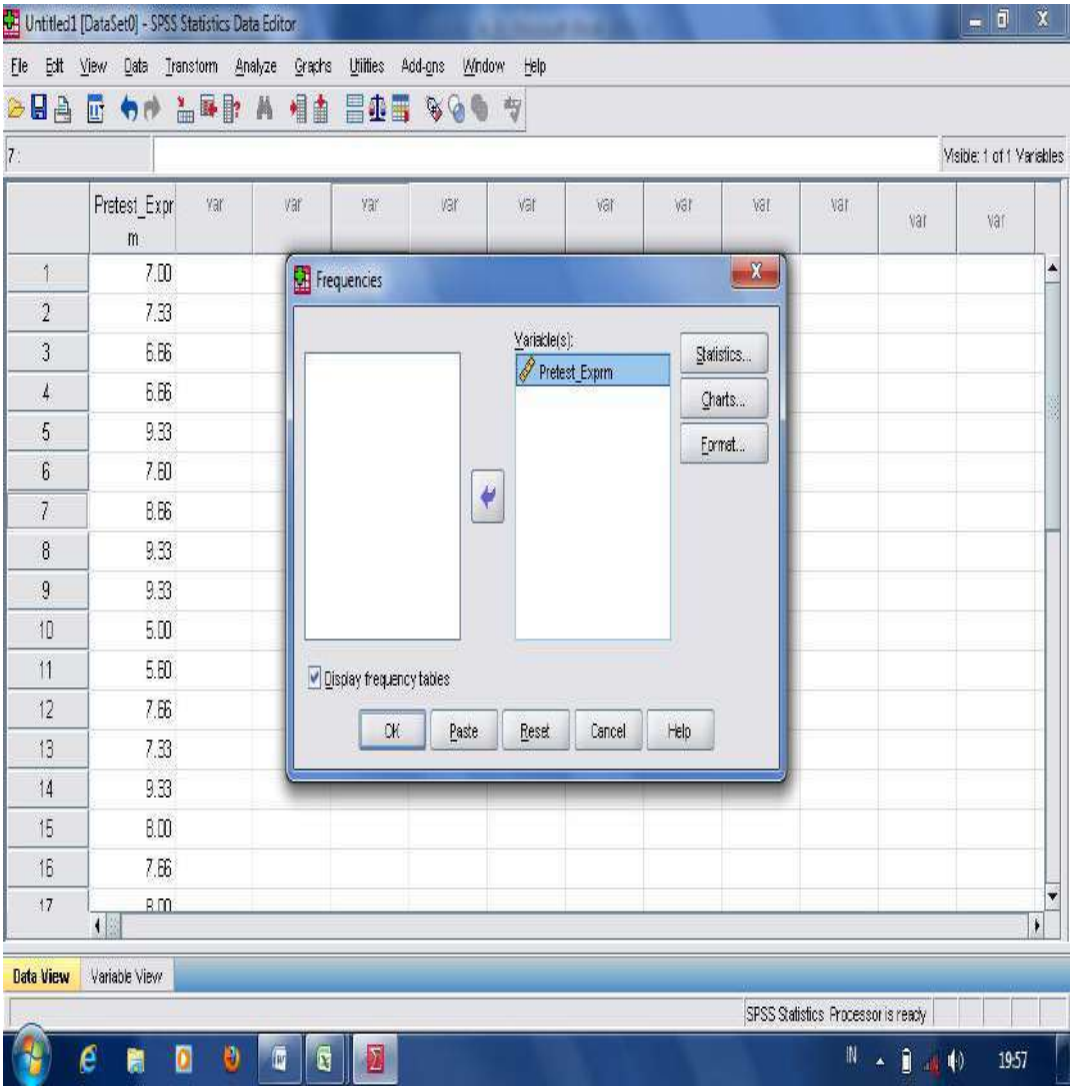
- And then input the scores of the category in data view column;



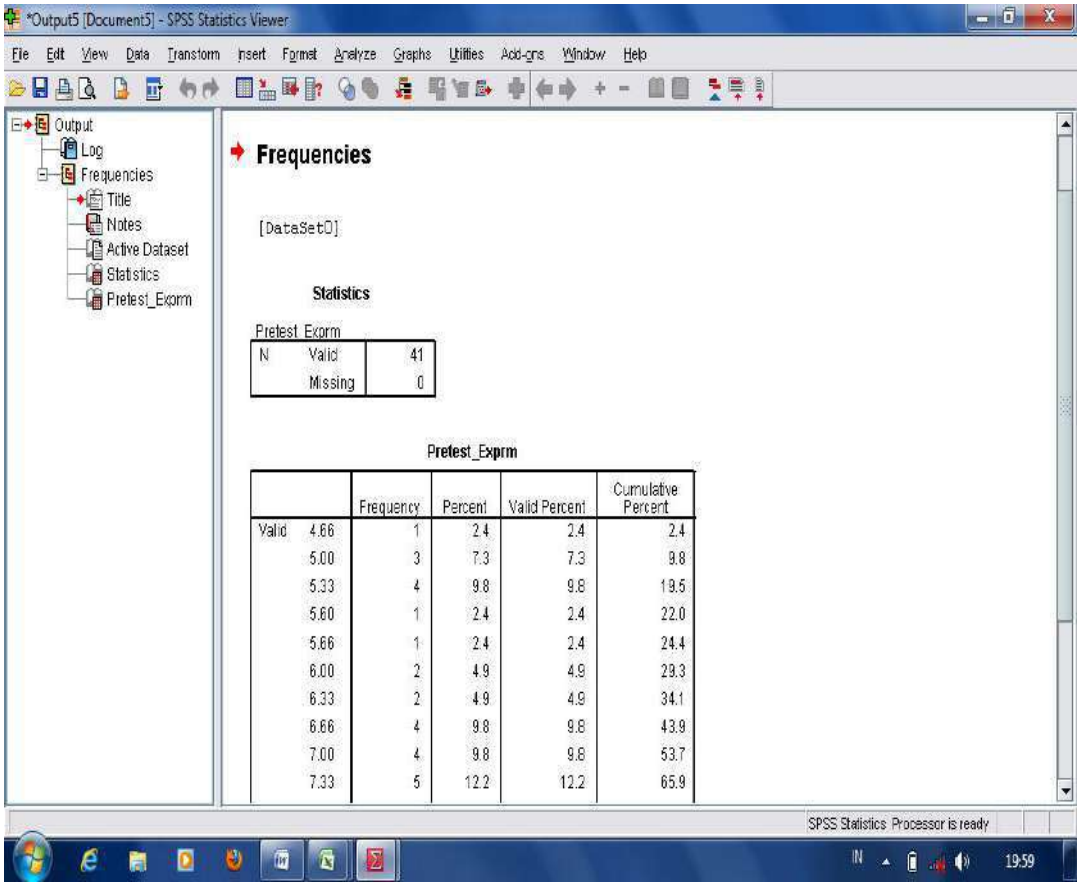
- Then, clik "**analyze"**, "**descriptive statistics"**, and "**frequency"**; press OK.



- Move the category of "**Pretest_Exprm**" into the variable column and then, press OK.



- The following is the result analysis in measuring frequency of the data obtained from students' pretest in experimental group.



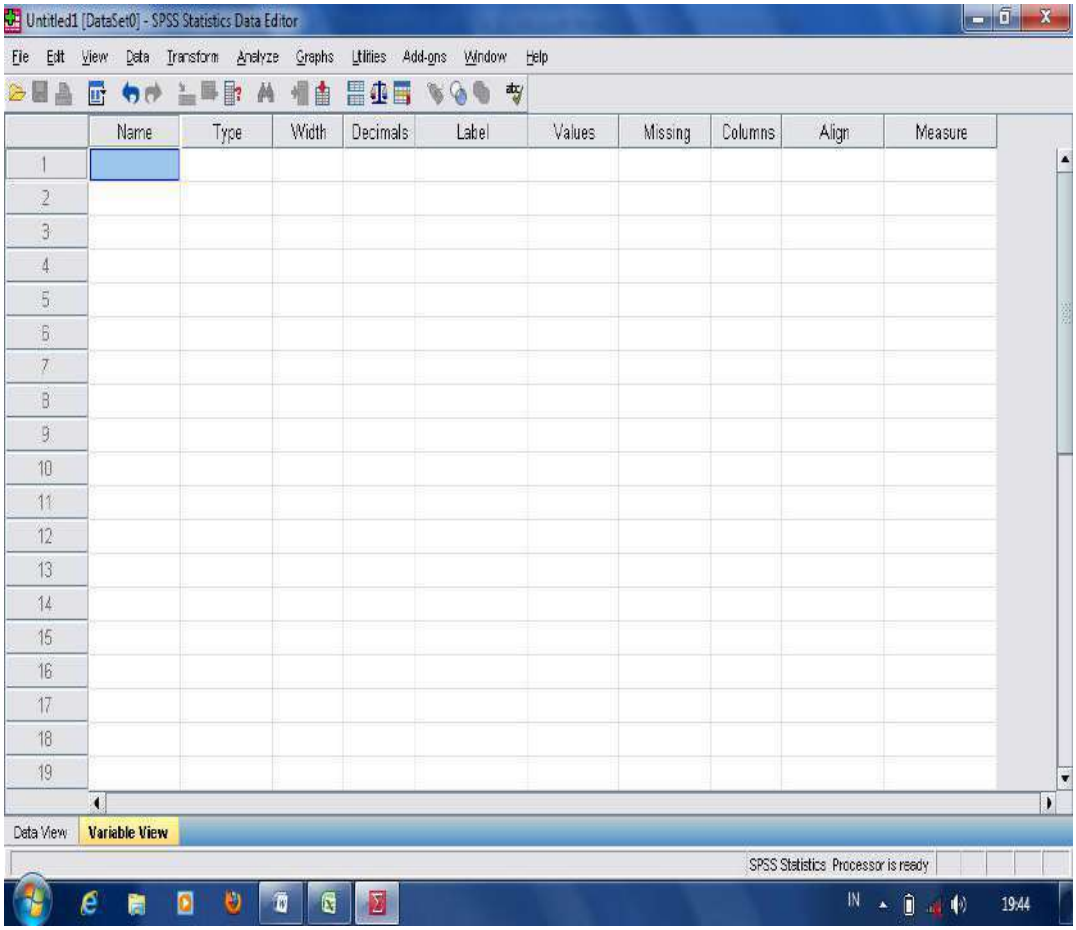
The illustration of the result analysis presented in the table is described as follows:

From the result analysis of frequency data, it was found that there are four students who got the score 4.00 (10%), one student obtained the score 4.33 (2.5%), four student achieved the score 4.66 (10%), seven students got the score 5.00 (17.5%), four students got the score 5.33 (10%), one student got the score 5.66 (2.5%), four student got the score 6.00 (10%), one student got the score 7.66 (2.5%), and two students got the score 8.00 (5%).

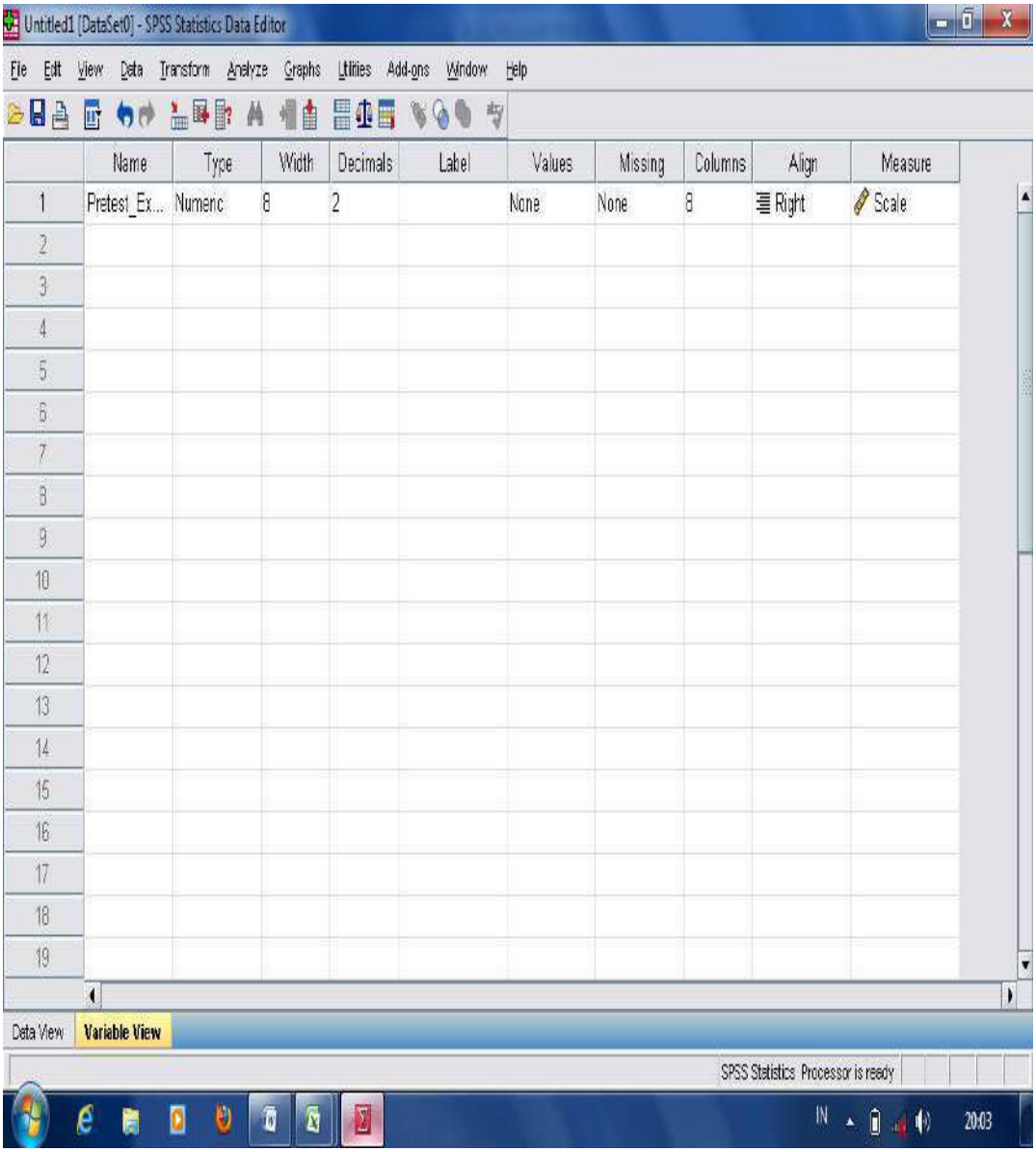
b. **Descriptive Statistics;** in this part, the data is obtained to get the lowest score (minimum), the highest score (maximum), mean score and the score of standard deviation. For example: the scores of students' pretest in experimental and control groups are analyzed to get its distribution in descriptive

statistics, so the analysis of the scores is done as in the following steps:

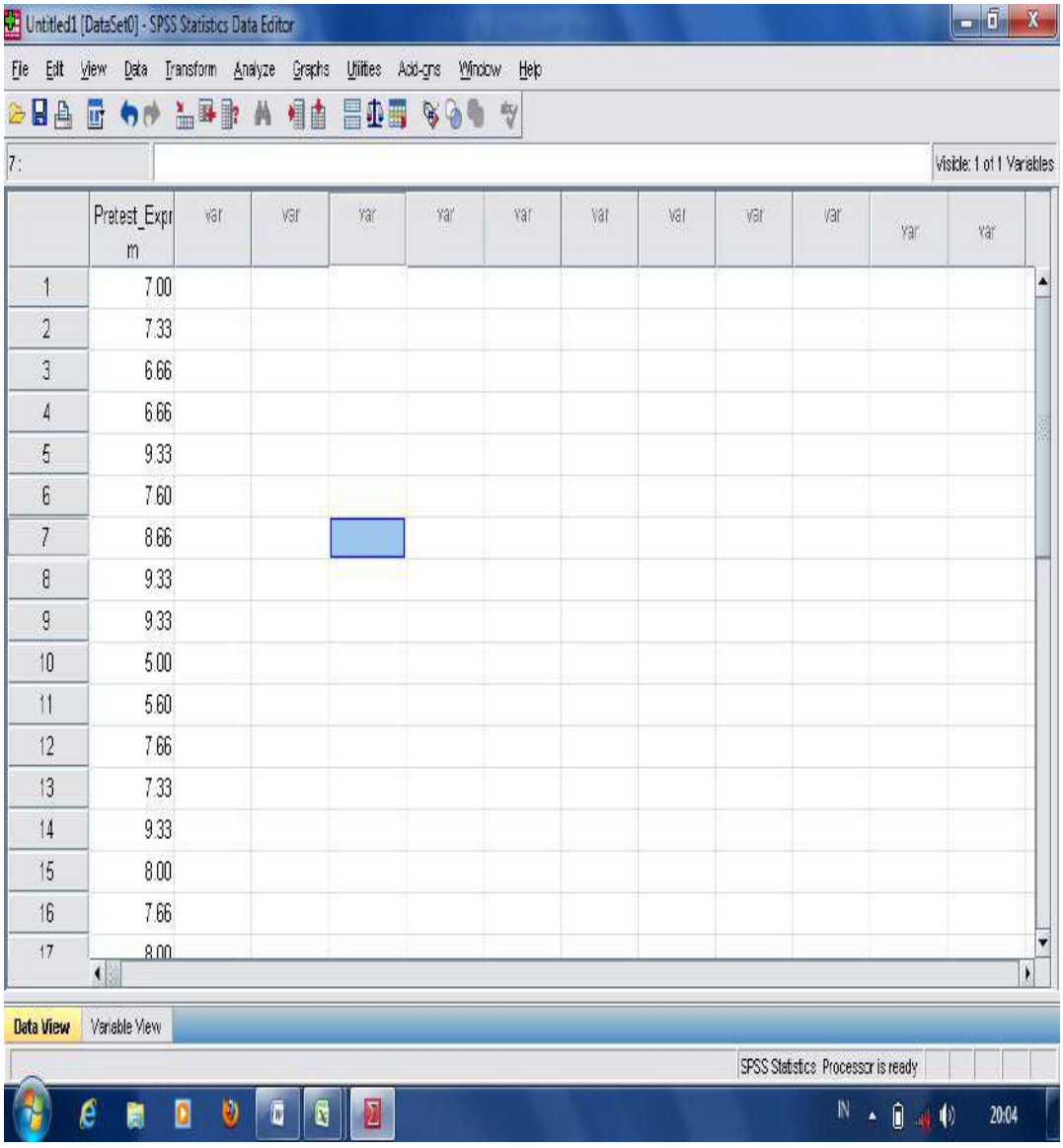
- Open the worksheet in the application of SPSS program;



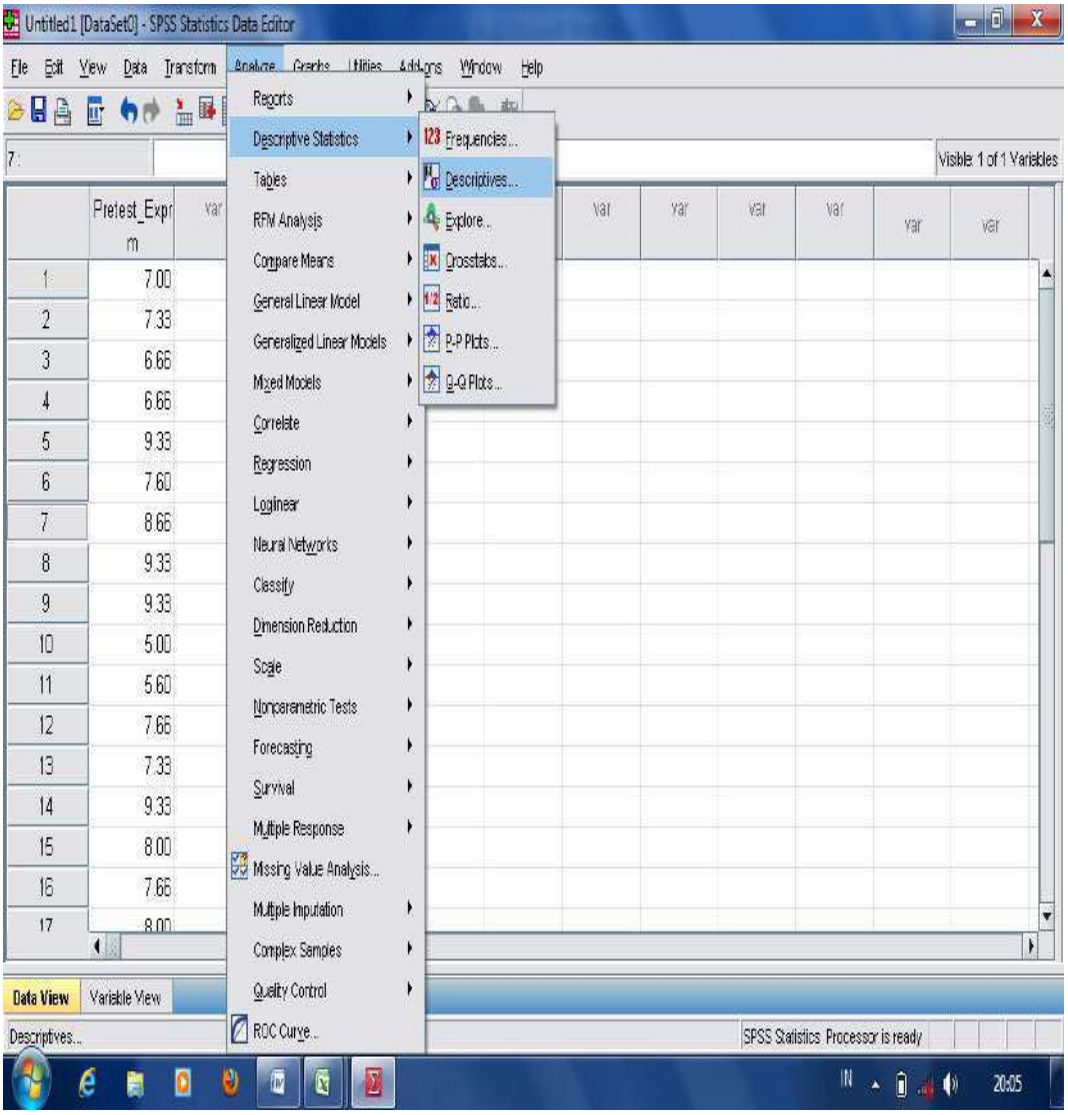
- Move the cursor into “variable view”; type the column into the category that is going to be compared for example, the variable named “**Pretest experiment**”;



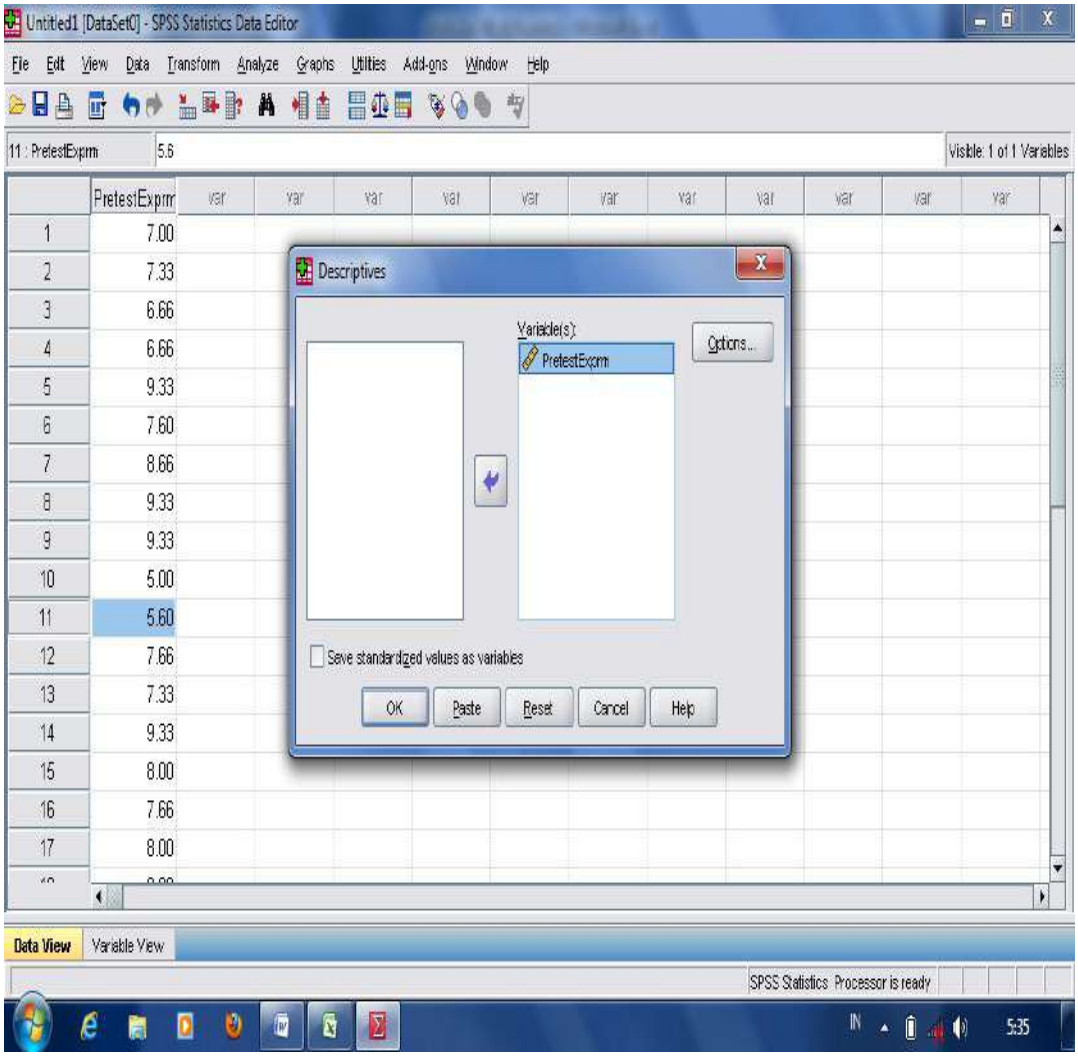
- And then input the scores of the category in data view column;



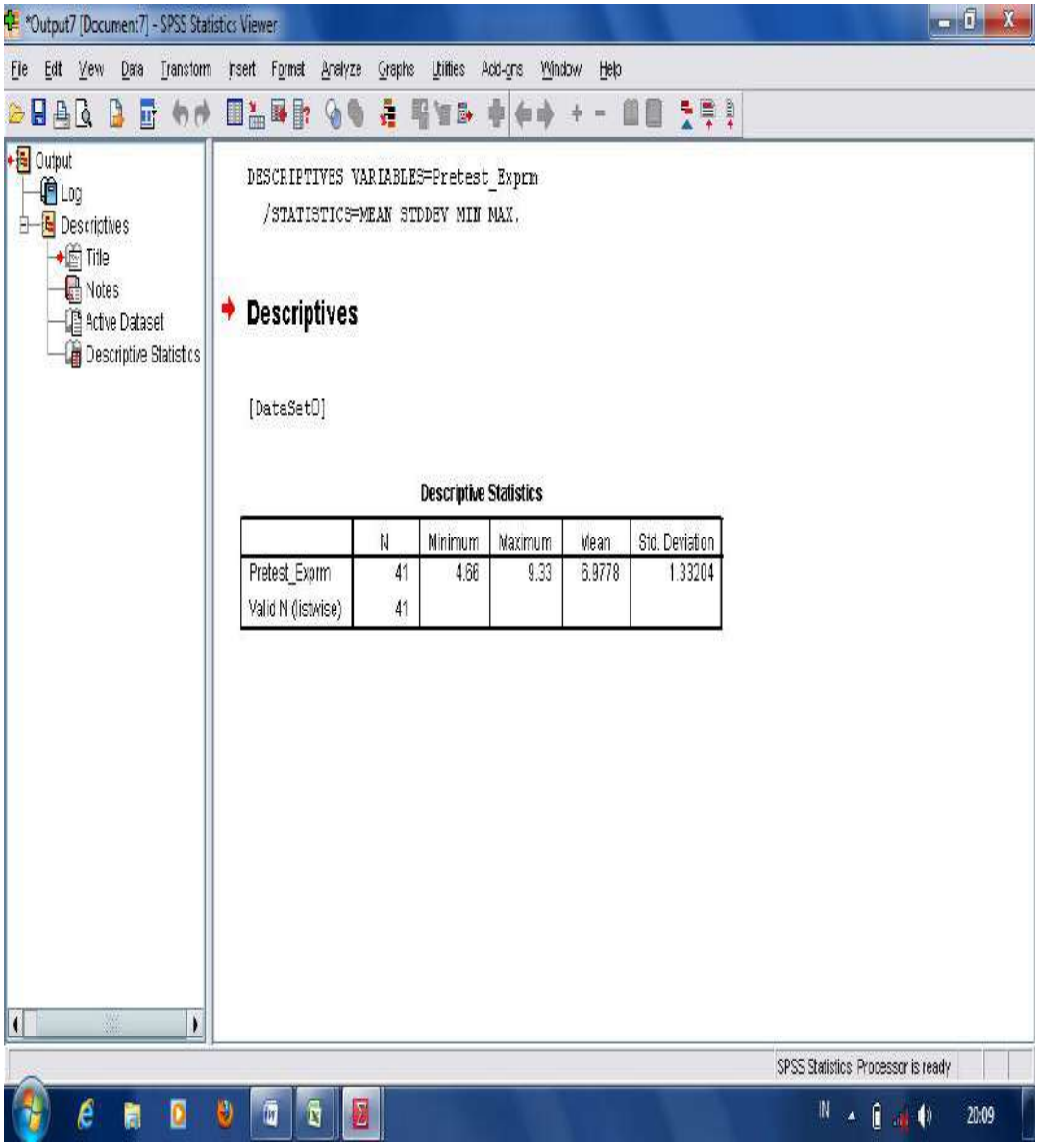
- Then, click "**analyze**", "**descriptive statistics**", and "**descriptive"**;



- Move the score data into “variable” column; then press OK.



- The result analysis of descriptive statistic is shown in the following table.



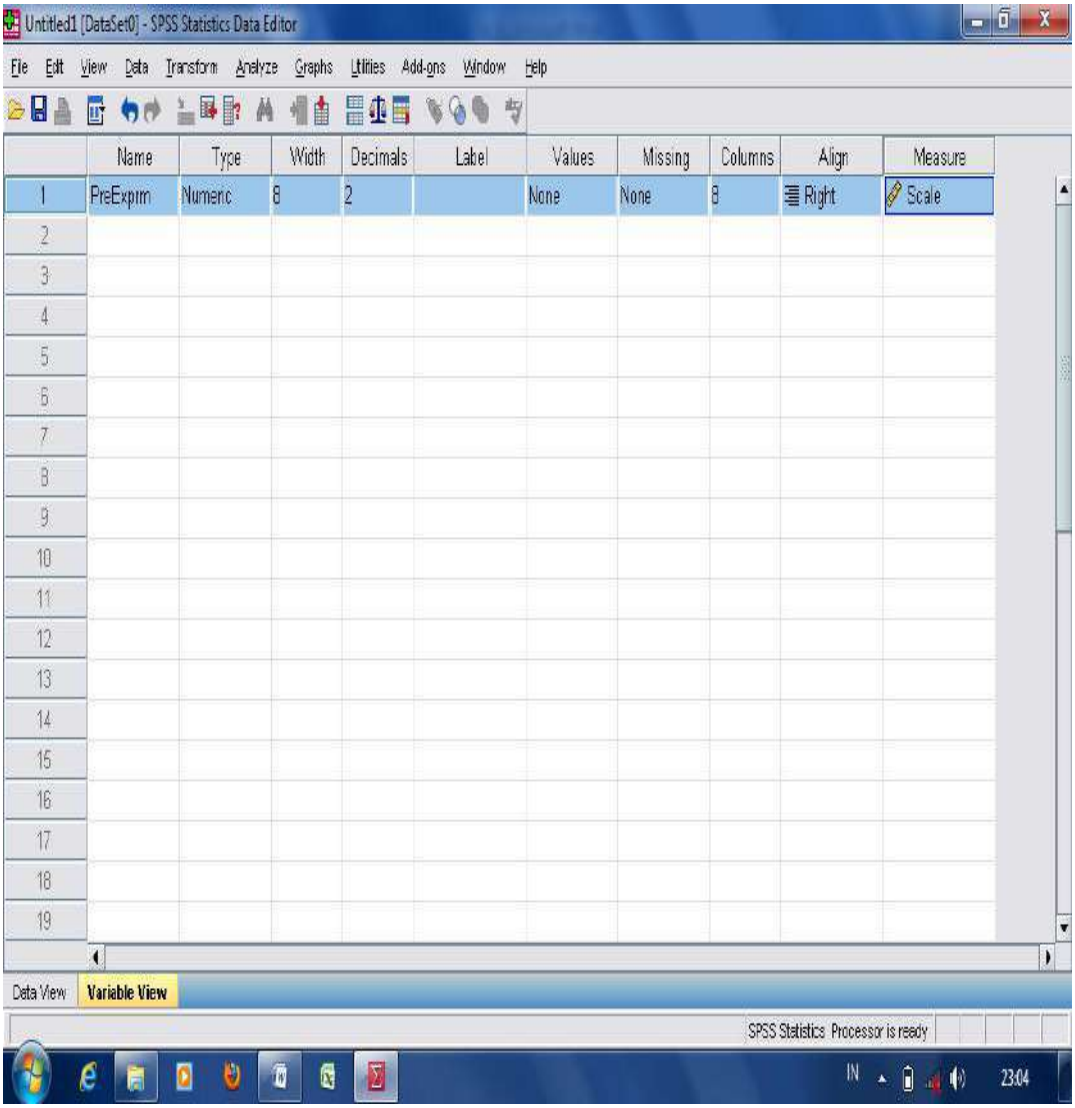
Then, the above table is illustrated as follows: The result analysis of descriptive statistics from students pretest scores in experimental group found that there are 20 students who are in the group of pretest experiment. The lowest score is 25, the highest score is 85, mean score is 58.25, and standard deviation is 16.87999.

2. **Prerequisite Analysis**

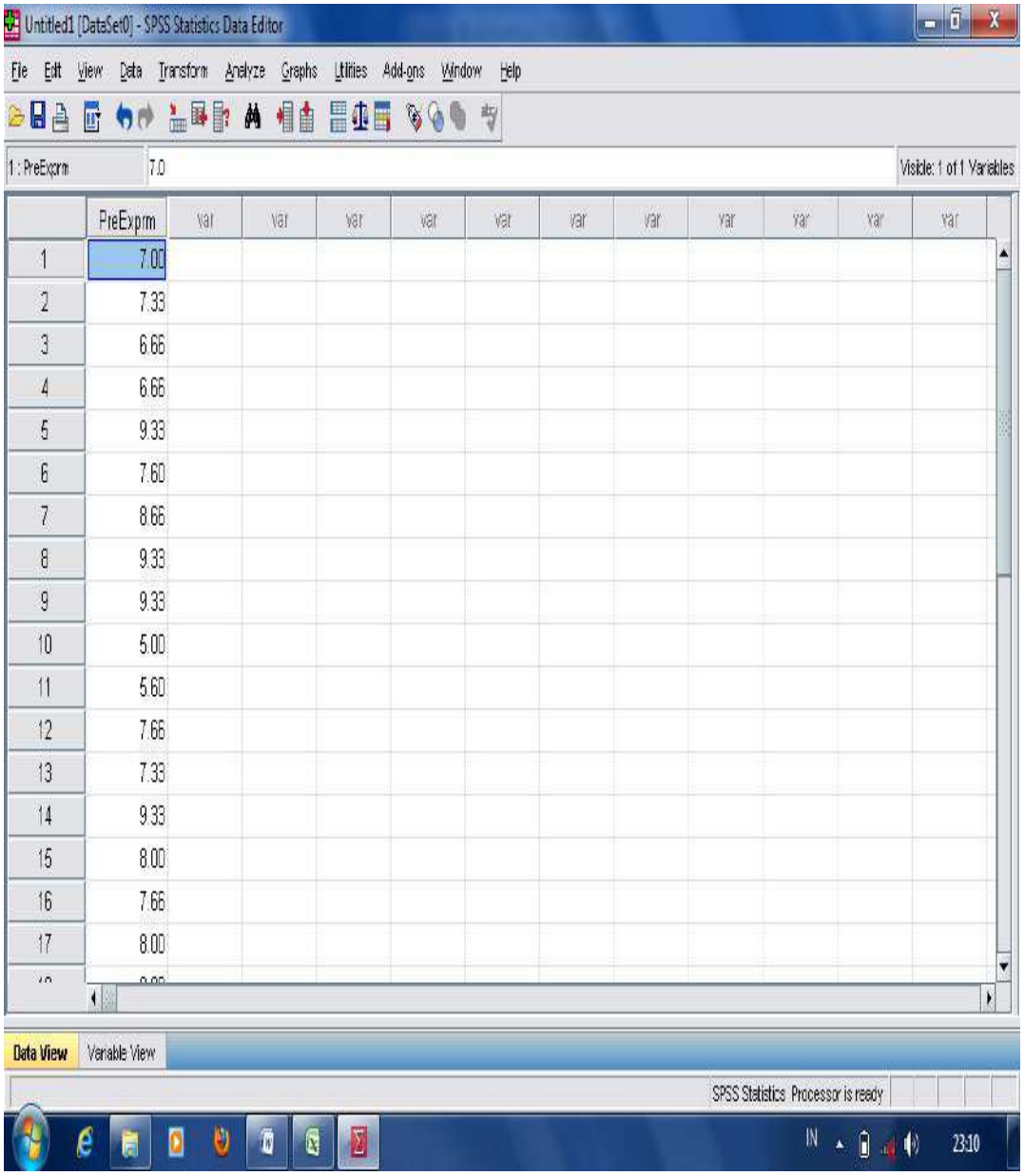
Prerequisite analysis is an analysis done before testing the research hypothesis. It estimates whether or not the obtained data from students' pretest and posttest scores in both groups (experiment and control) are distributed normal and homogen. To estimate the normality of the obtained data, **1-Sample Kolmogorov-Smrinov** or **Kolmogorov-Smrinov and Shapiro-Wilk** is used. Meanwhile, the homogeneity test is used to estimate whether the obtained data are distributed homogen or not, **Levene Statistics** is used. The data is distributed normal and homogen when the result analysis (p-output) of the normality and homogeneity test is higher than 0.05. However, when the obtained data is not distributed normal and homogen, the analysis of testing the research hypothesis can not be done. The following is prerequisite analysis in terms of analyzing normality and homogeneity test:

a. Normality Test; is used to measure the obtained data whether it is normal or not. The data is obtained from students' pretest and posttest in control and experimental groups. The test is considered normal whenever it is higher than 0.05. In analyzing the normality test, **1-Sample Kolmogorov-Smirnov Test** is used. The followings are the steps to start the analysis of the normality test.

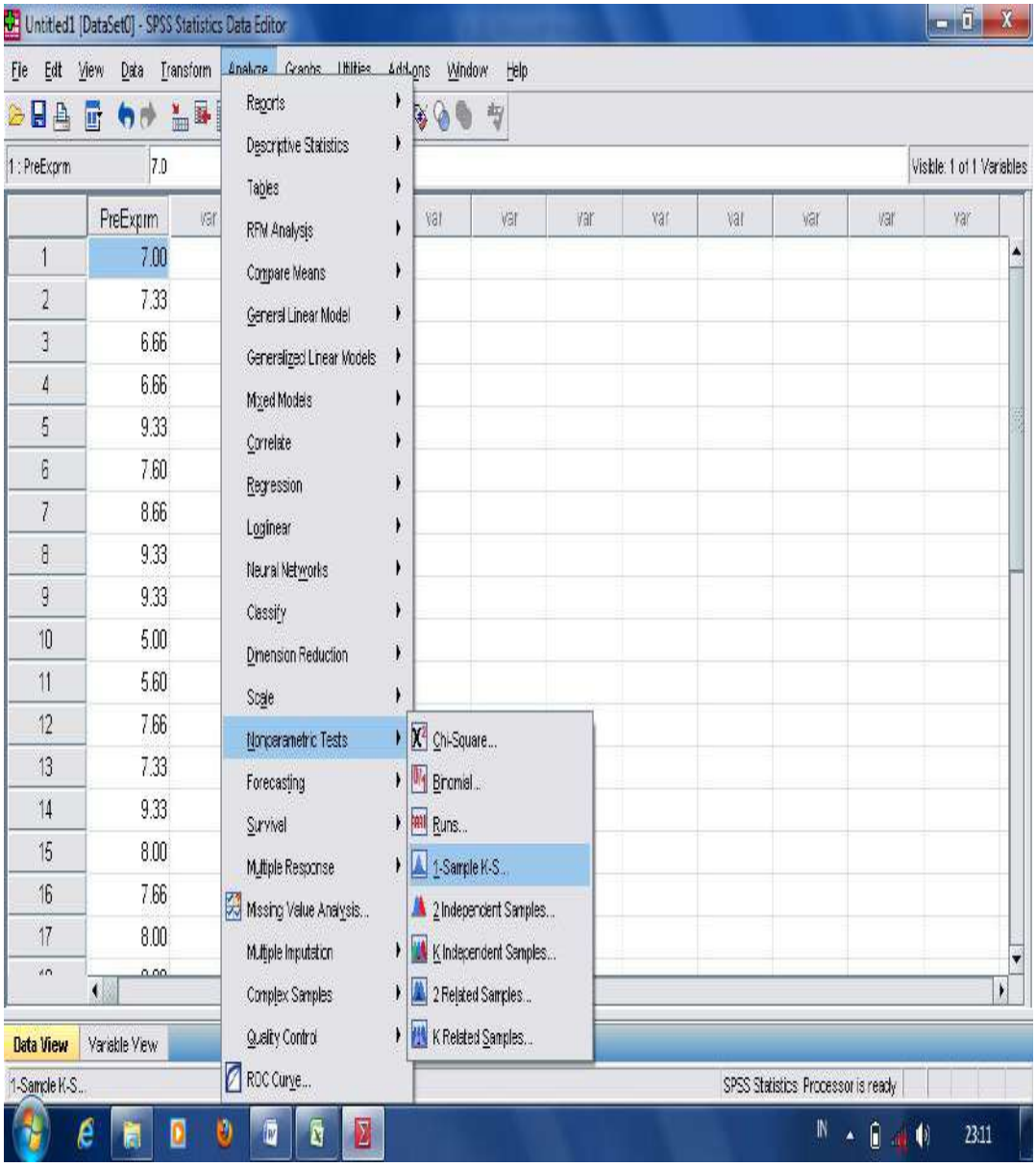
- Open the worksheet in the application of SPSS program; Then, move into a "Variable View"; type the column into the category of the variable named "**Pretest experiment**";



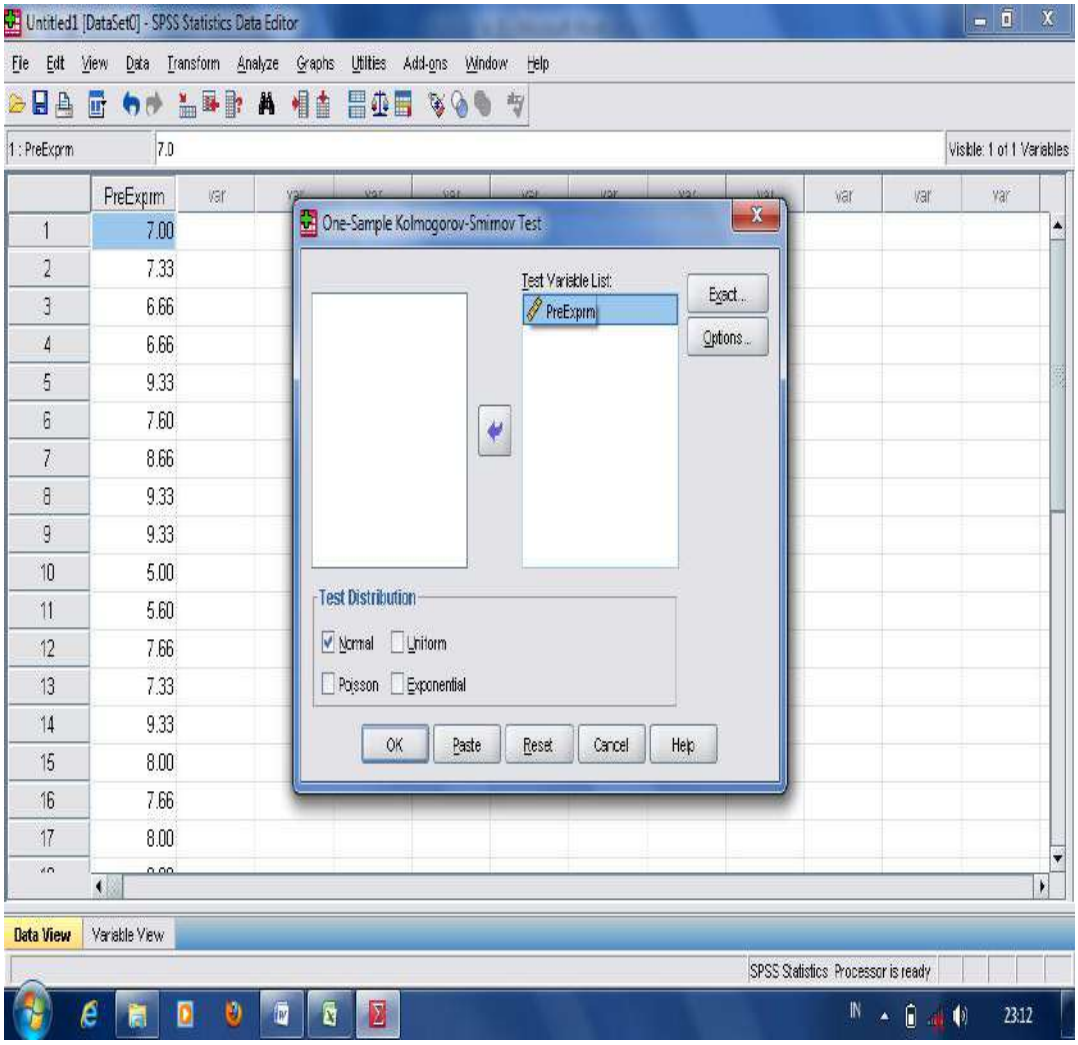
- And then input the scores of the category in data view column;



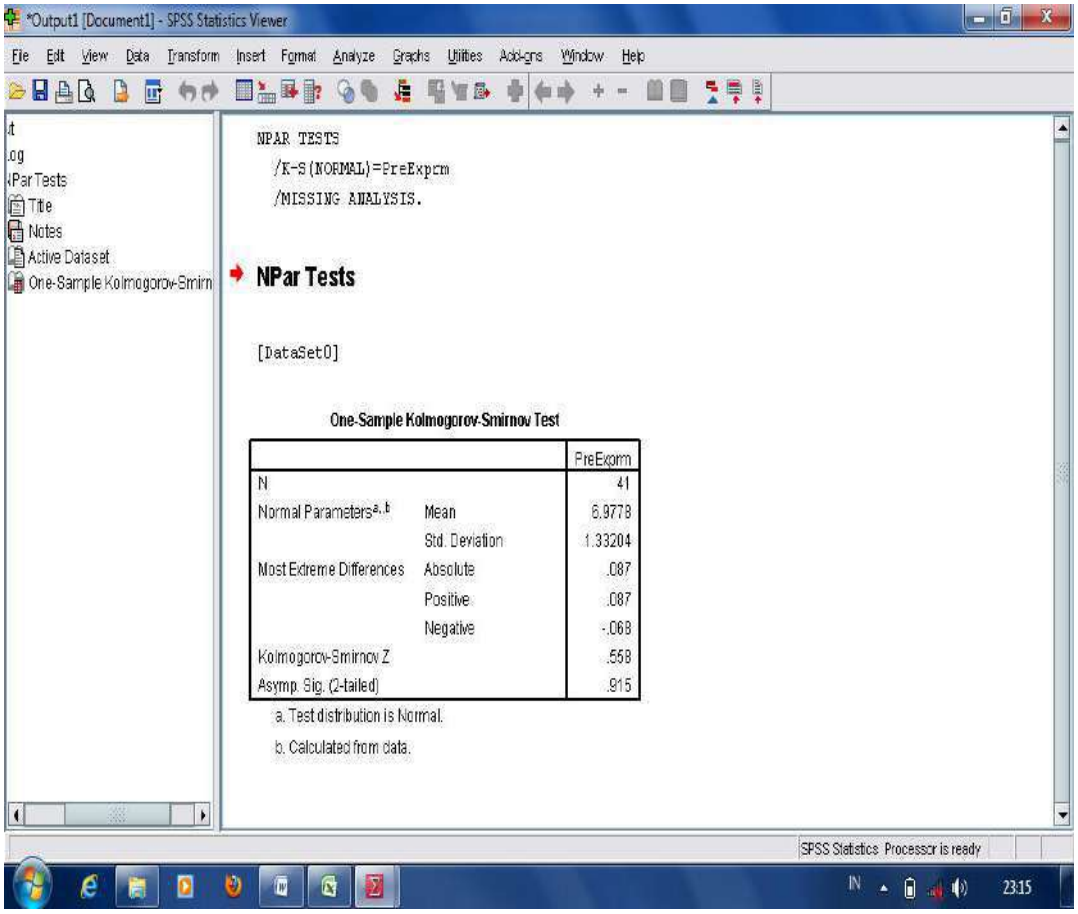
- Then, click "**analyze**", "**non-parametric test**", and **"1 sample K-S"**; Then, move all data into "**test variable list**"; and press OK;



- Then, move the score into the box of variable list and press OK.



- The result analysis of normality test which is obtained from students' pretest scores in experimental group is seen as follows.

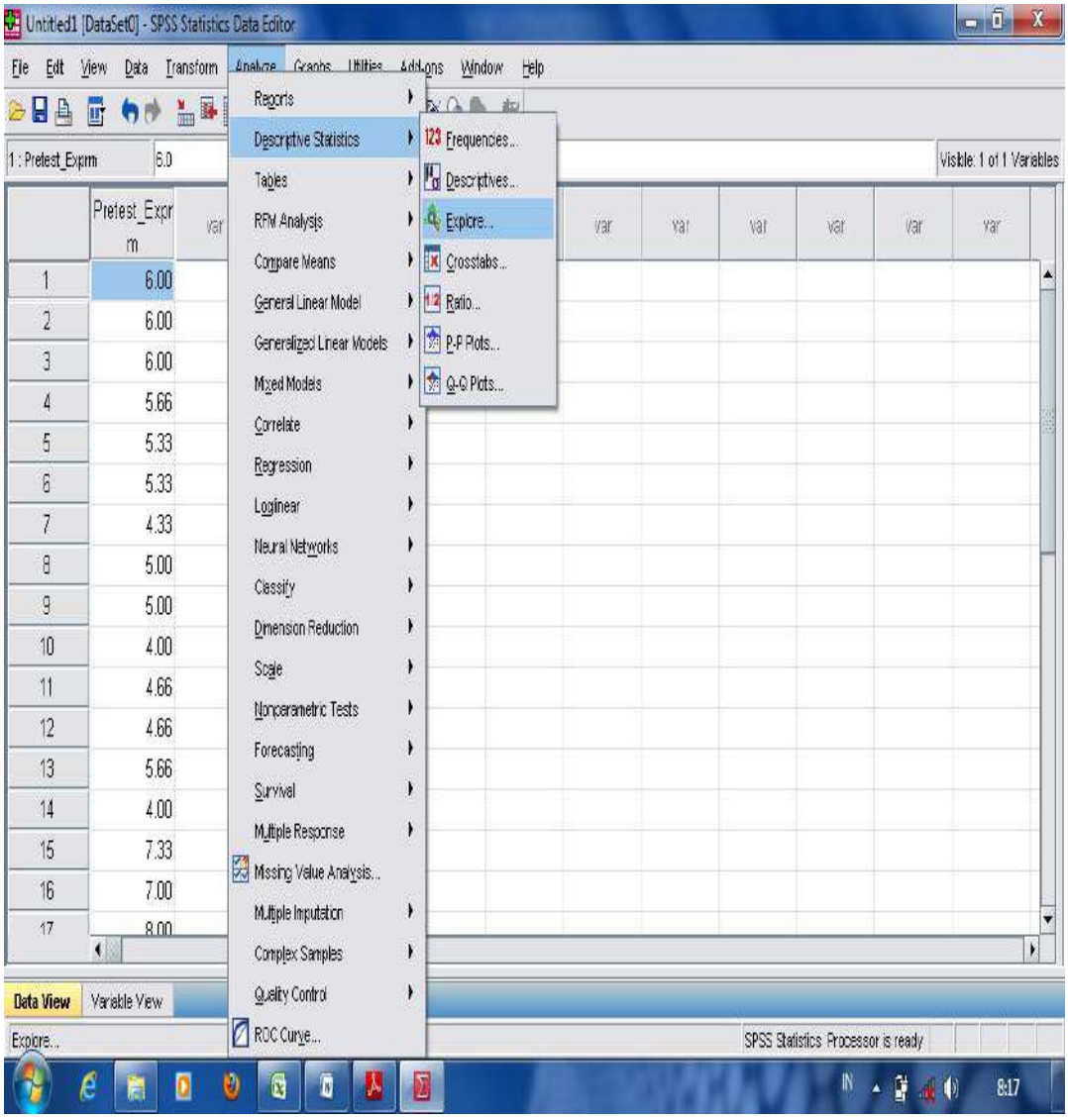


From the table of normality test above, it is found that the significance of normality test from students' pretest scores in experimental group is 0.558. From the score, it can be stated that the obtained data is categorized normal since it is higher than 0.05.

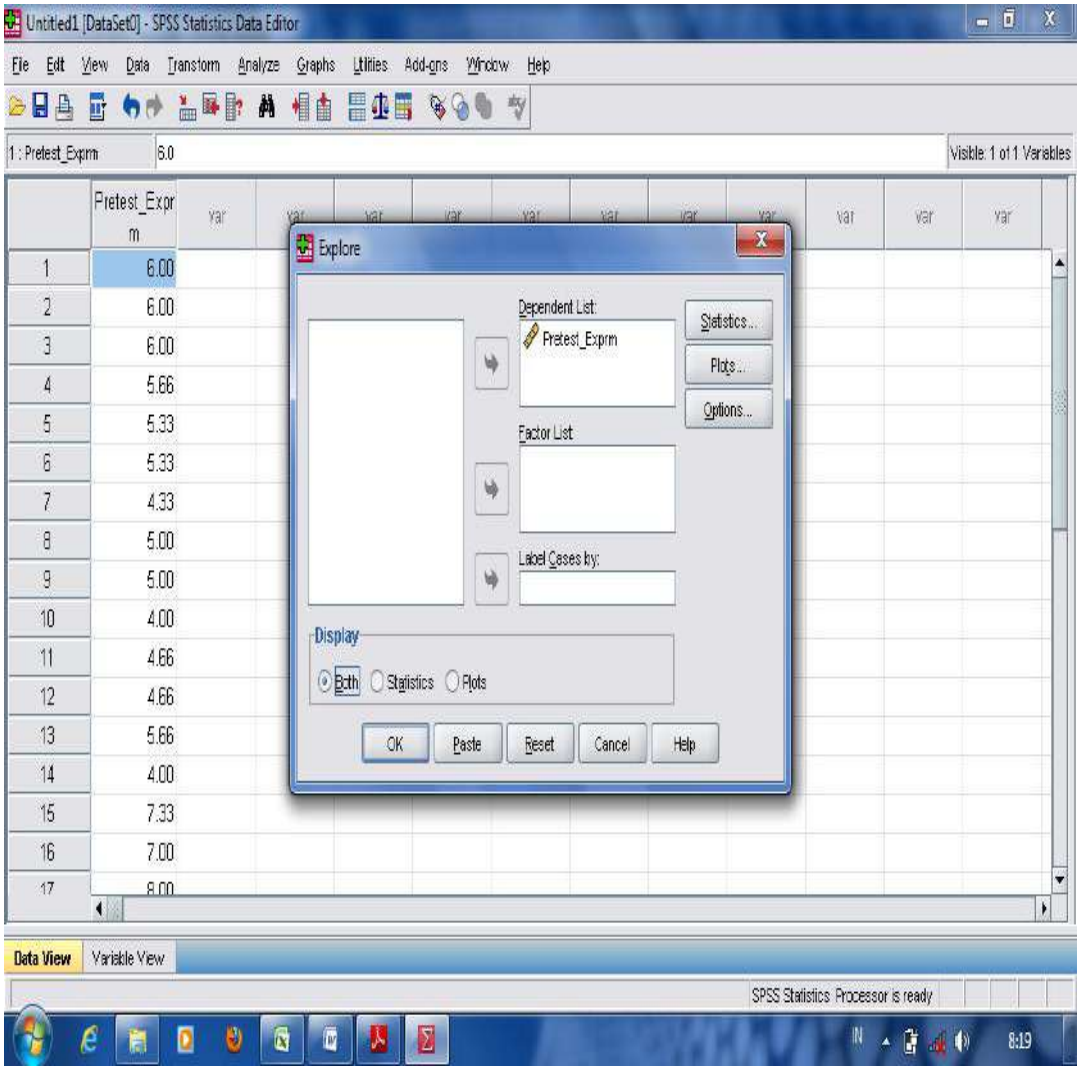
Besides that, there is another way to analyse normality test using "**Kolmogorov-Smrinov and Shapiro-Wilk**". The following are steps to run the analysis.

- Open the worksheet in the SPSS application program; and then, move into a "Variable View"; type the column into the category of the variable named "**Pretest experiment**";

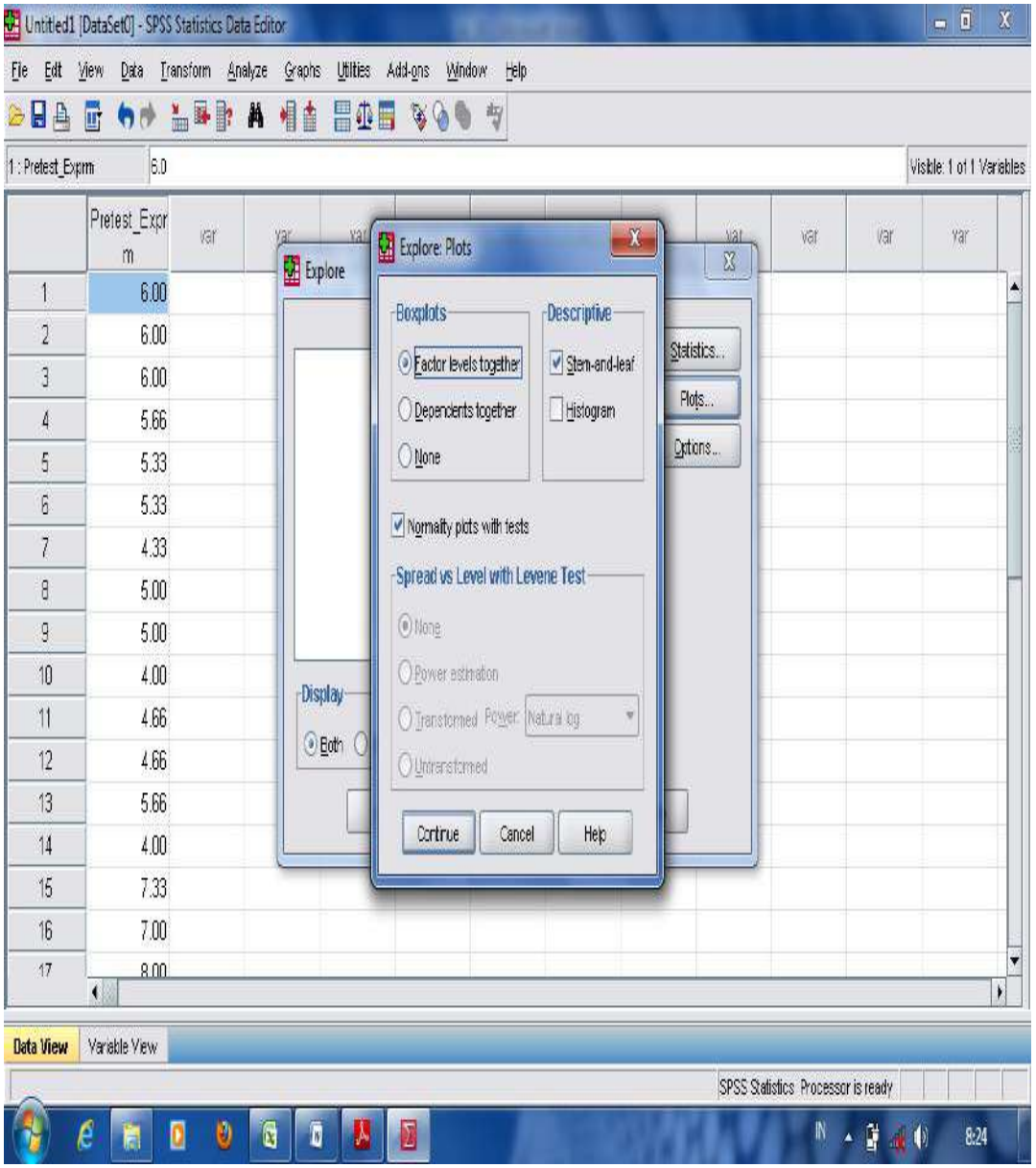
- Input all the scores of the category in data view clolum; Then, click Analyze, Descriptive Statistics, and Explore...



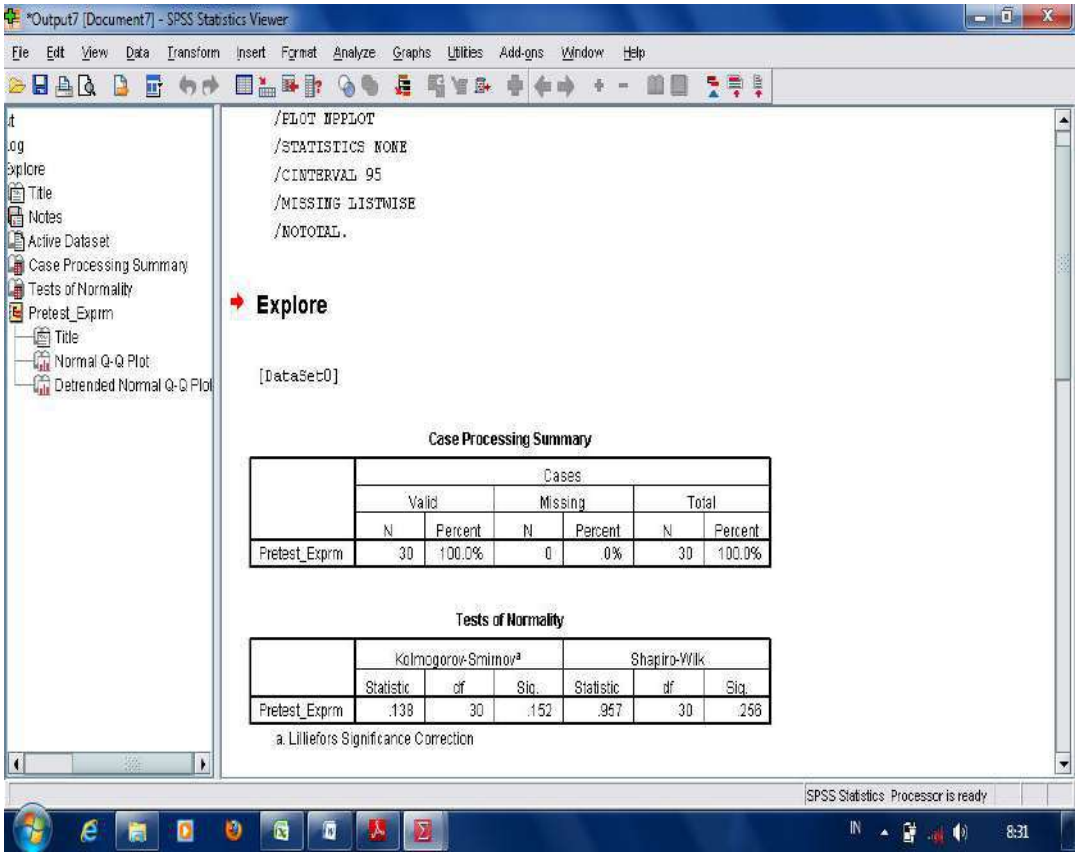
- Then, move the existing variable into the column of “Dependent List”;



- Then, press the provided box “plot”, and give a check mark in the column of “ Normality plots with test”. Press continue and OK;



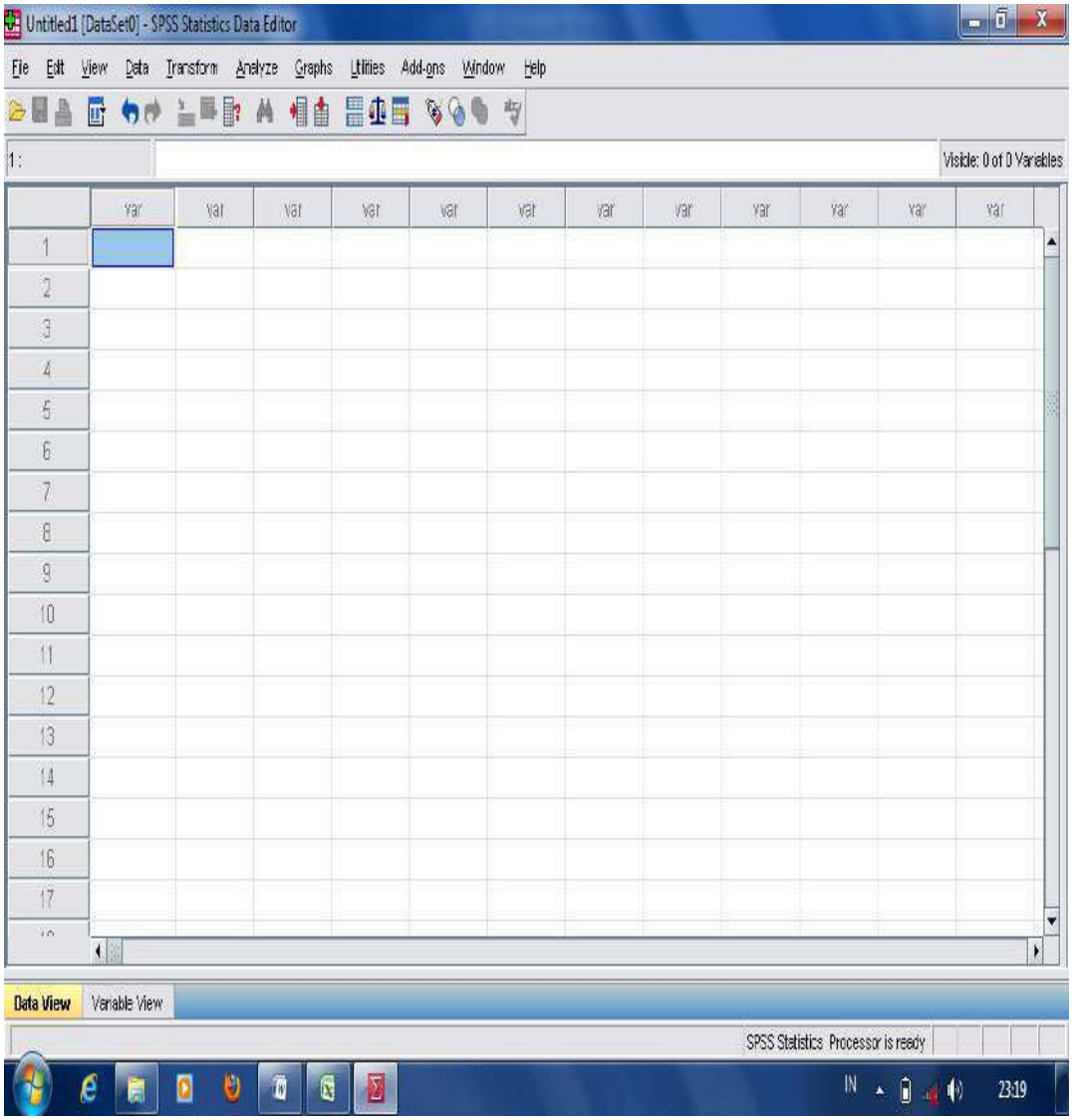
- Then result of normality test using "**Kolmogorov-Smrinov and Shapiro-Wilk**" is shown as in the following table of p-ouput:



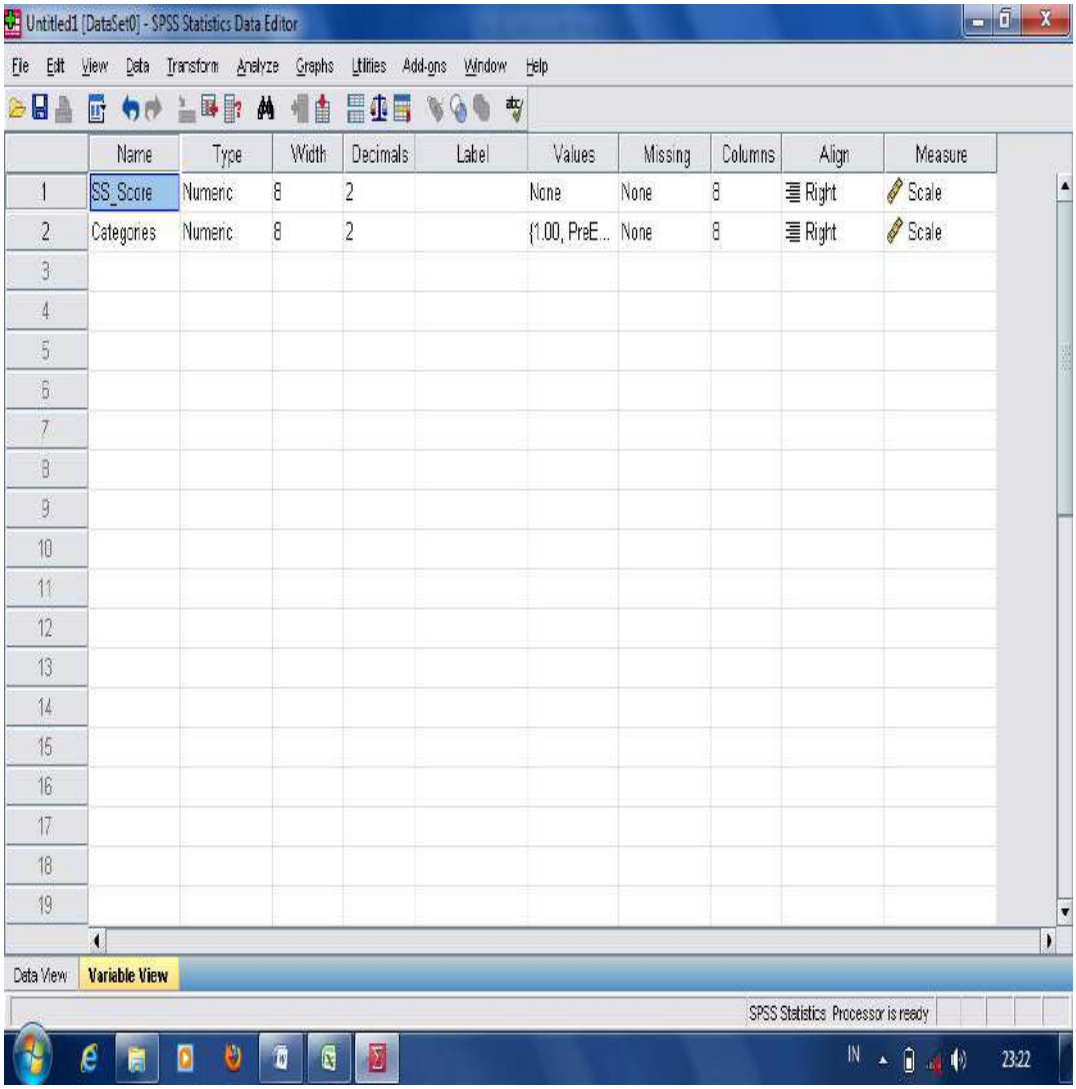
From the result analysis of normality test above, it is found that the significance of normality test from students' pretest scores in experimental group using Kolmogorov-Smrinov is 0.152, while using Shapiro-Wilk is 0.256. From the two scores, it can be stated that the obtained data is categorized normal since it is higher than 0.05.

b. Homogeneity Test; it is used to measure the obtained data whether it is homogen or not. The data can be categorized homogen whenever it is higher than 0.05. The obtained data are achieved from students' pretest and posttest scores in experimental and control groups. In analyzing the homogeneity test, **Levene Statitsics** in SPSS is used. The followings are steps to run the homogeneity test.

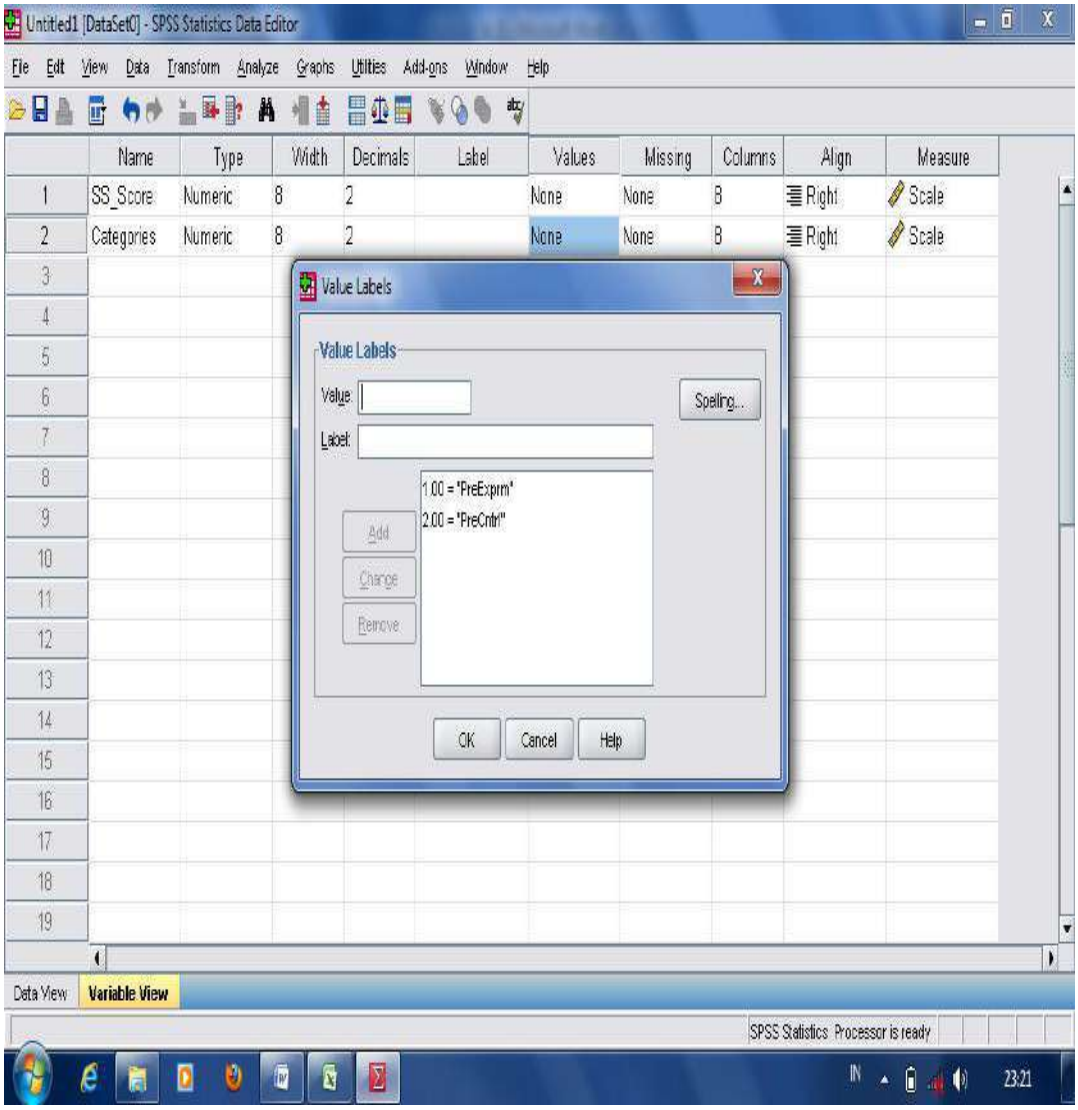
- Open the worksheet in the application of SPSS program;



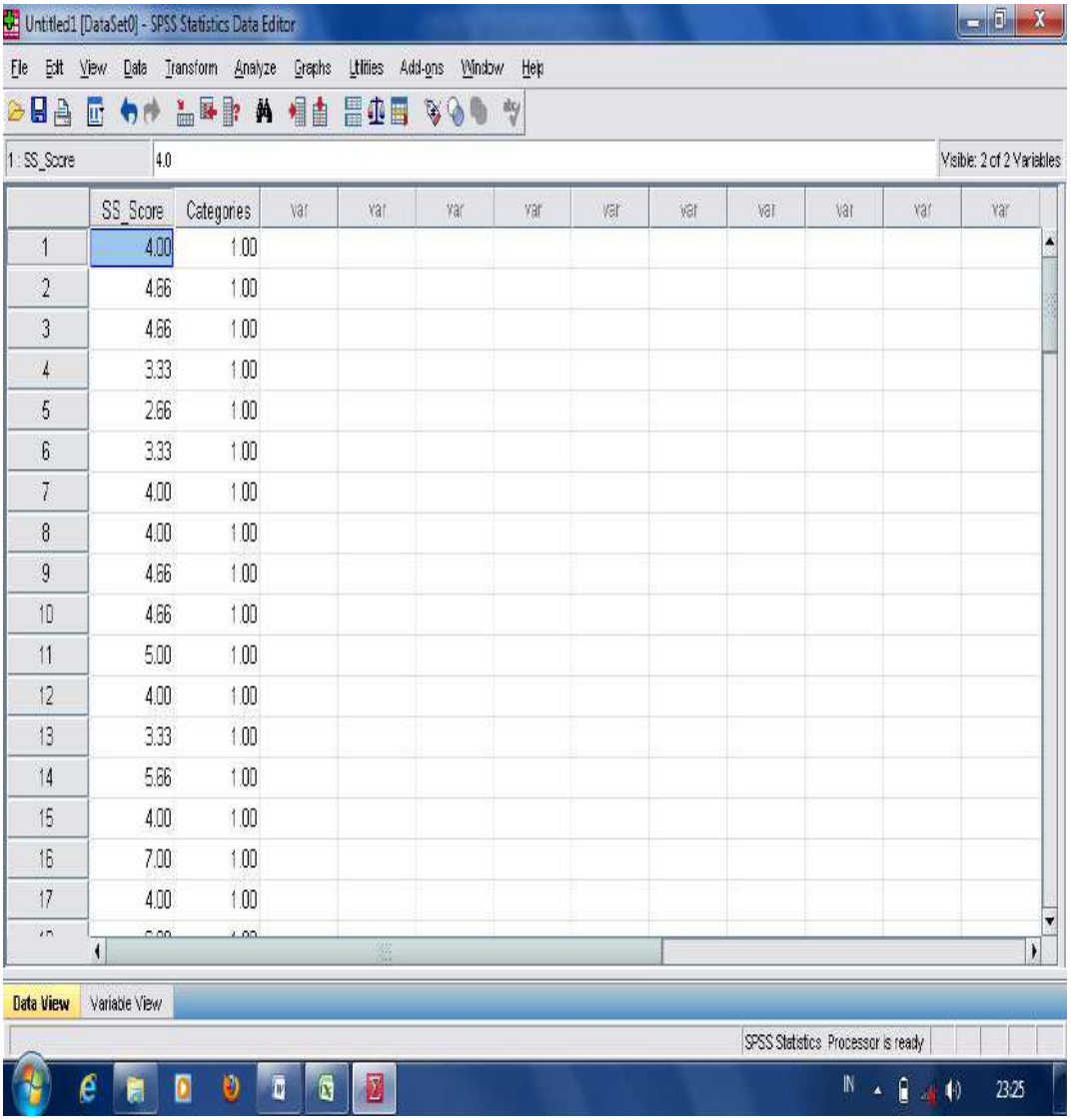
- Move into “variable view”, then type “**Ss_score**” in the first column, and “**category**” in the second column;



- In the variable view, go to "value" column, and then move into "value label",
- Then, type 1 into the value column and label column for "**pretest control**" and type 2 into the value column and label for "**pretest experiment**". Then, click "add" and press OK.

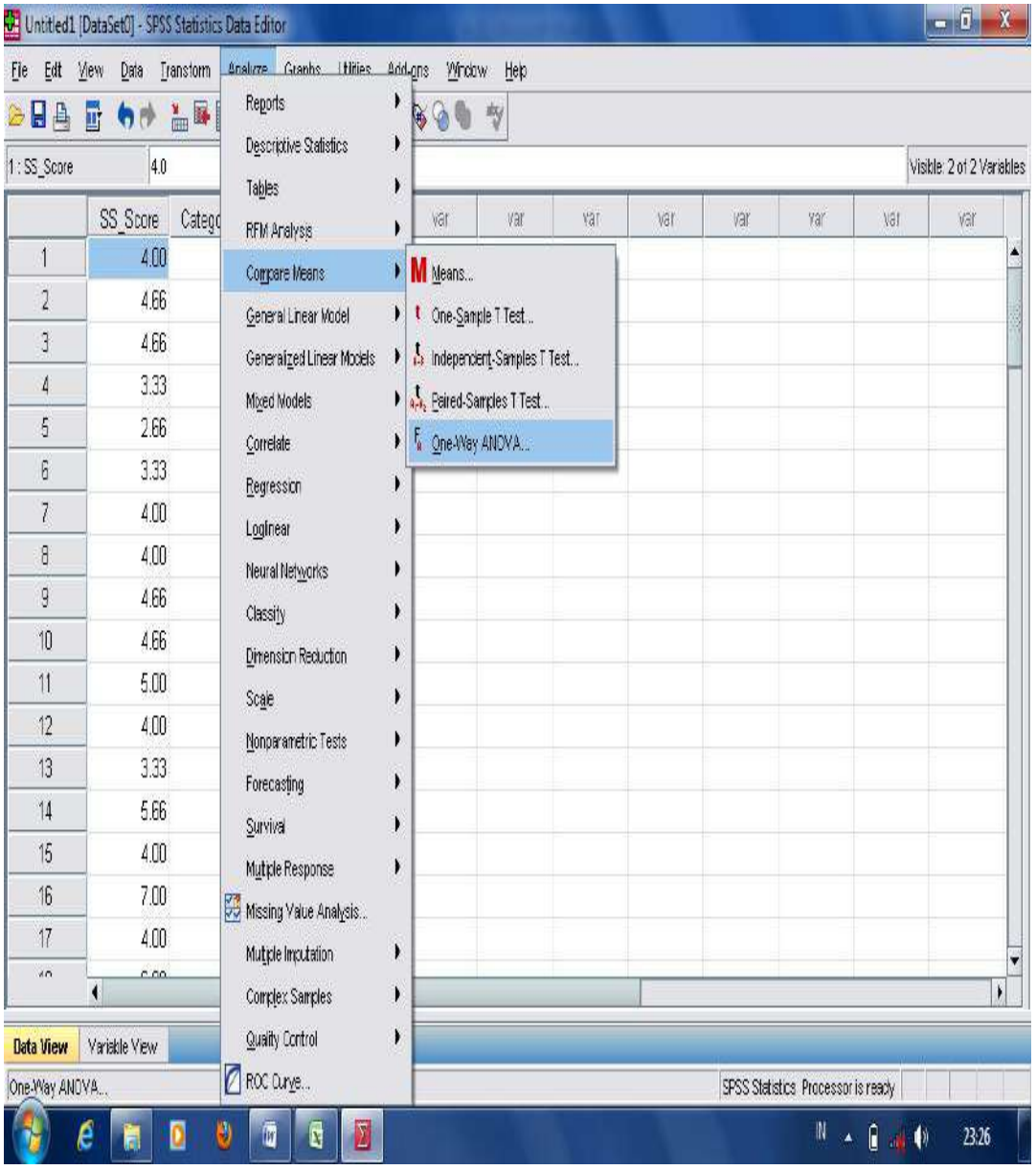


- Move into “data view” column, and input the obtained score into the “**Ss_score**” coloumn, and press 1 for pretest control and 2 for pretest experiment in the “**category**” coloumn;

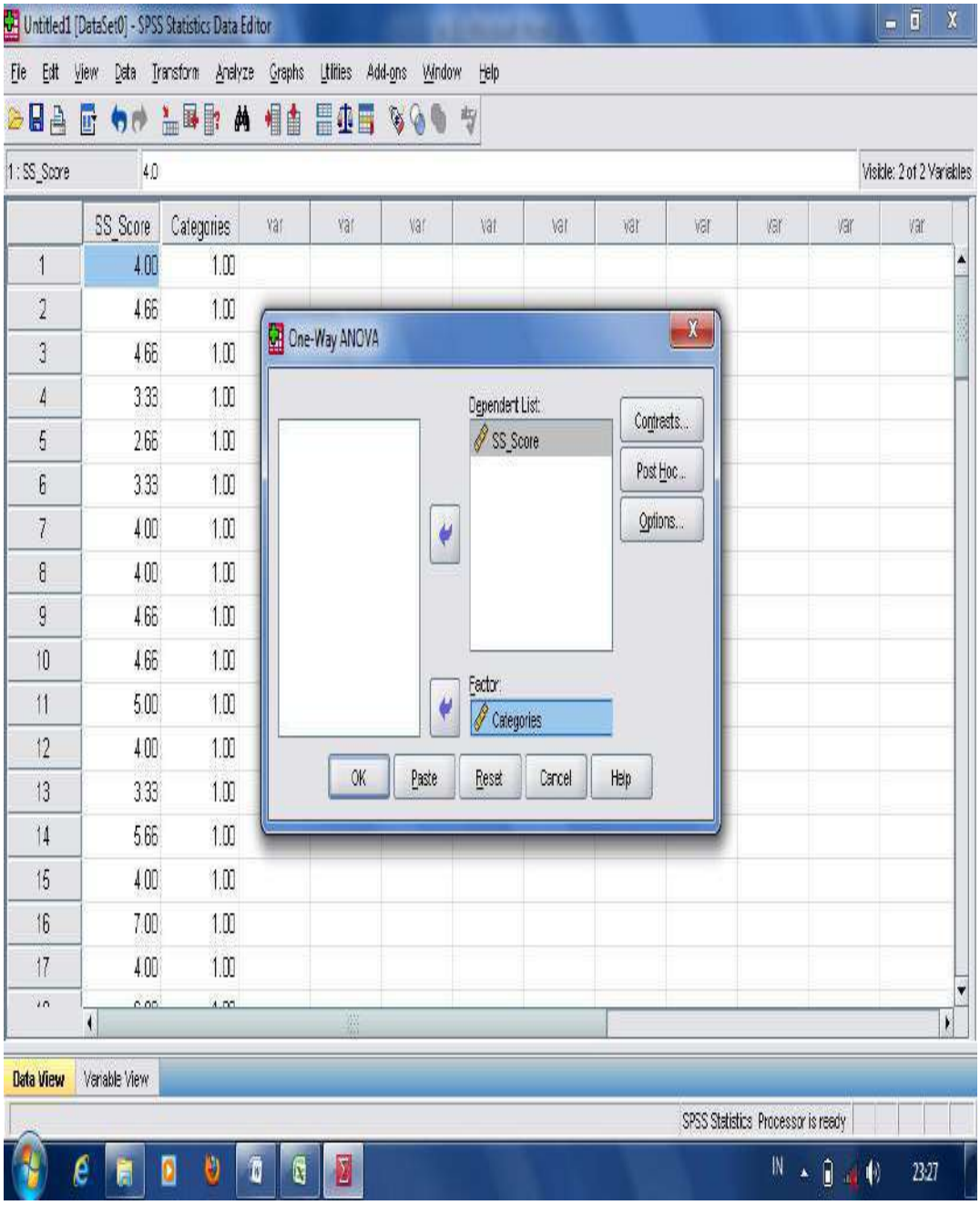


	SS_Score	Categories
1	4.00	1.00
2	4.66	1.00
3	4.66	1.00
4	3.33	1.00
5	2.66	1.00
6	3.33	1.00
7	4.00	1.00
8	4.00	1.00
9	4.66	1.00
10	4.66	1.00
11	5.00	1.00
12	4.00	1.00
13	3.33	1.00
14	5.66	1.00
15	4.00	1.00
16	7.00	1.00
17	4.00	1.00

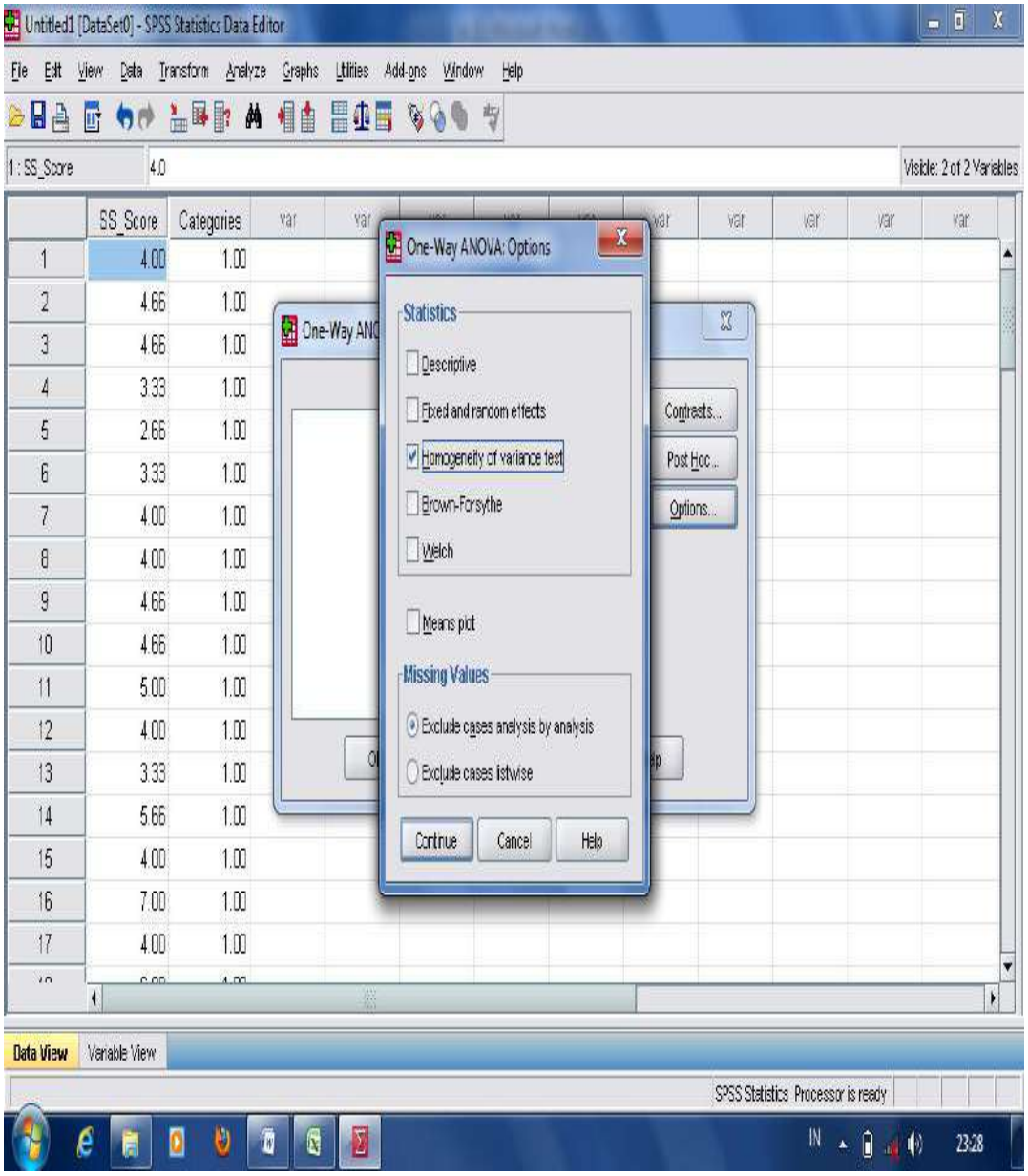
- Then, click "**Analyze"**, "**Compare Means"**, and choose "**One-way ANOVA"**;



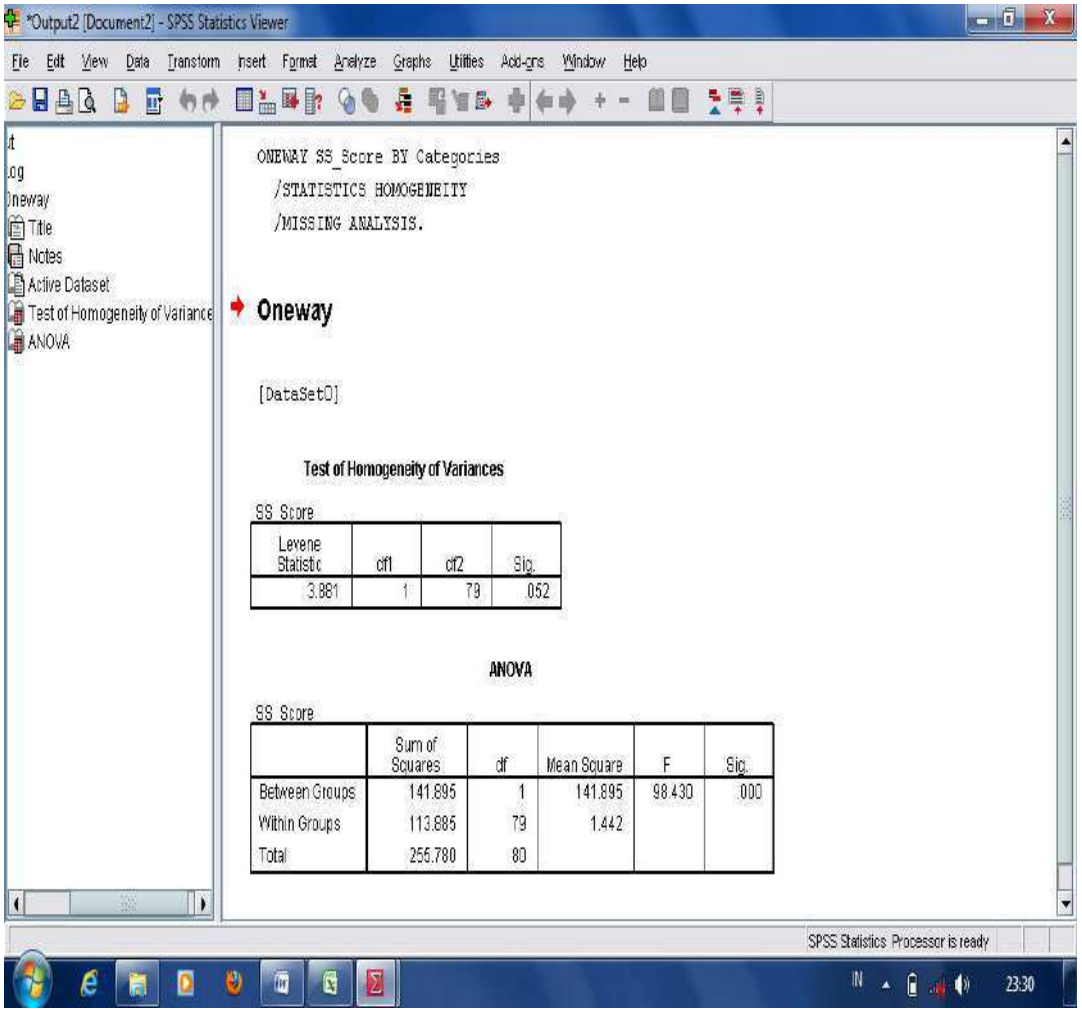
- Move the "**Ss_Scores**" into dependent column, and "**category**" into factor column;



- Go to "**option**" box, check on "**homogeneity test variance**" and press continues; Press OK.



- The following table shows the result analysis of homogeneity test using "**Levene Statistics"**.



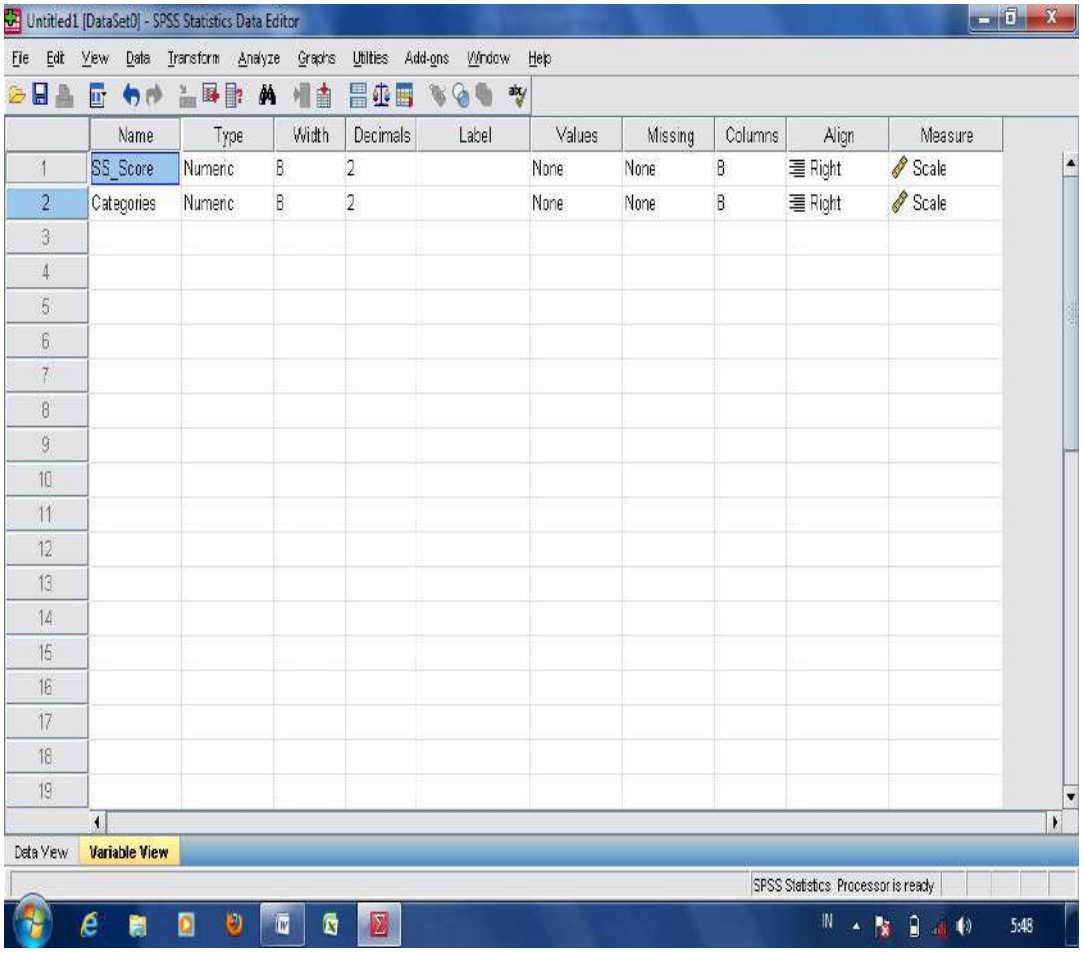
From the result analysis in the above table, it is found that the significant score is 0.052. From the p-output, it can be stated that the obtained score from students' pretest scores in control and experimental groups is homogen since the p-output is higher than 0.05.

J. Istruments for HypothesisTesting

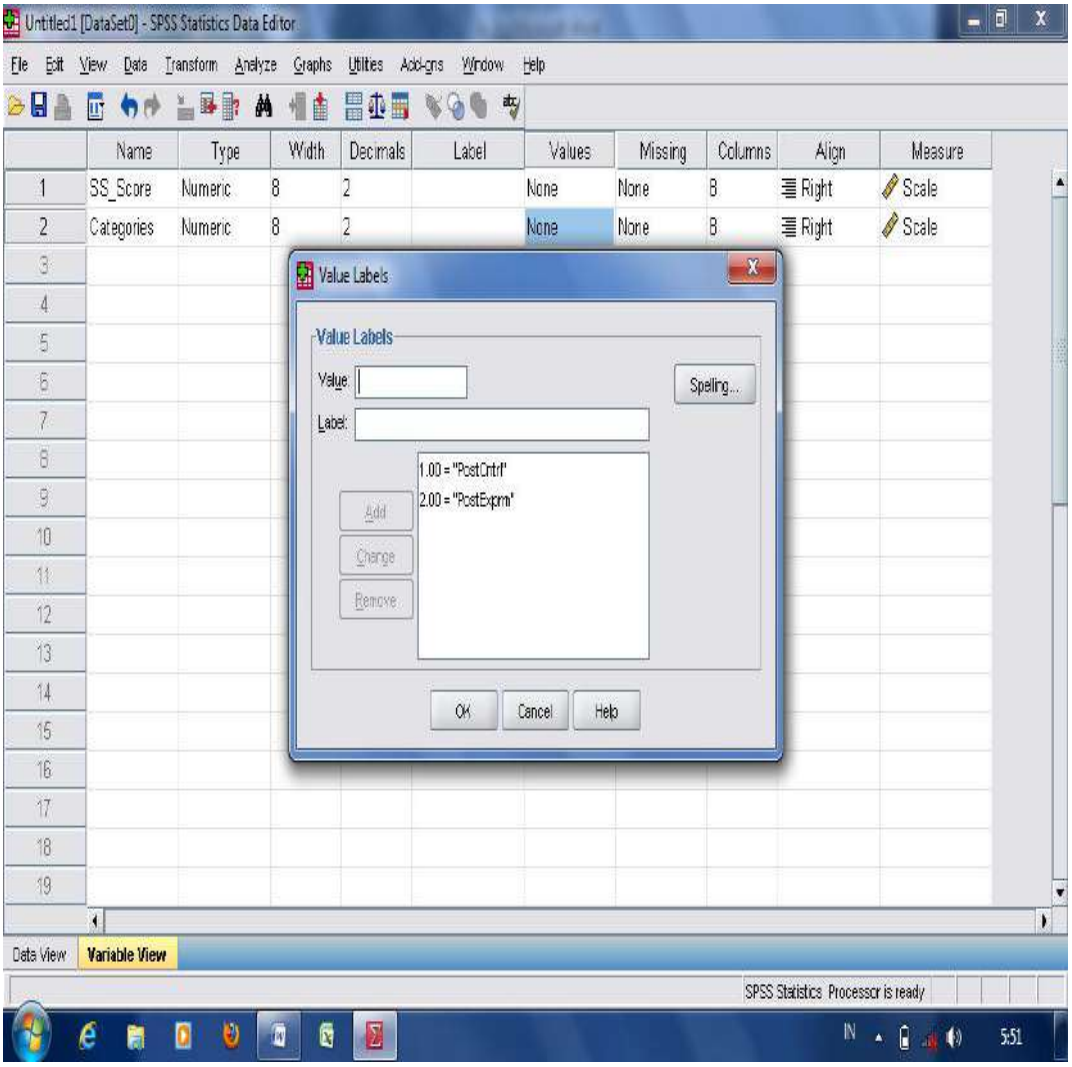
In testing the research hypothesis, the following testing instruments are used: independent sample t-test, paired sample t-test, regression analysis, correlation analysis, one-way ANOVA, analysis of covariance (ANCOVA), two-way ANOVA, and two-way MANOVA. They are further illustrated as follows.

1. **T-test**; is analytical procedures comparing two means only in two groups (Nunan, 1992, p.35). There are two kinds of t-test. They are independent sample t-test and paired sample t-test.
   a. **Independent Sample T-Test;** is used to compare the means scores of two independent groups. These types of t-tests are also used to compare groups of participants that are not related in any way. The groups are independent from one another. So, participants in one group have no relationship to participants in the second group. This is sometimes called a **between subjects design** which measures **means significant difference or significant influence** between two independent samples. The scores are obtained from students' posttest in both groups analyzed using "independent sample t-test". For example: a research study finding significant difference on students' reading comprehension average scores taught using KWLH and conventional strategies. The scores are obtained from students' posttest scores in both groups. Means significant difference or significant influence is found whenever the p-output is lower than mean significant difference at 0.05 levels. To start analysis using "Independent Sample T-Test" in SPSS, the steps are illustrated as follows:

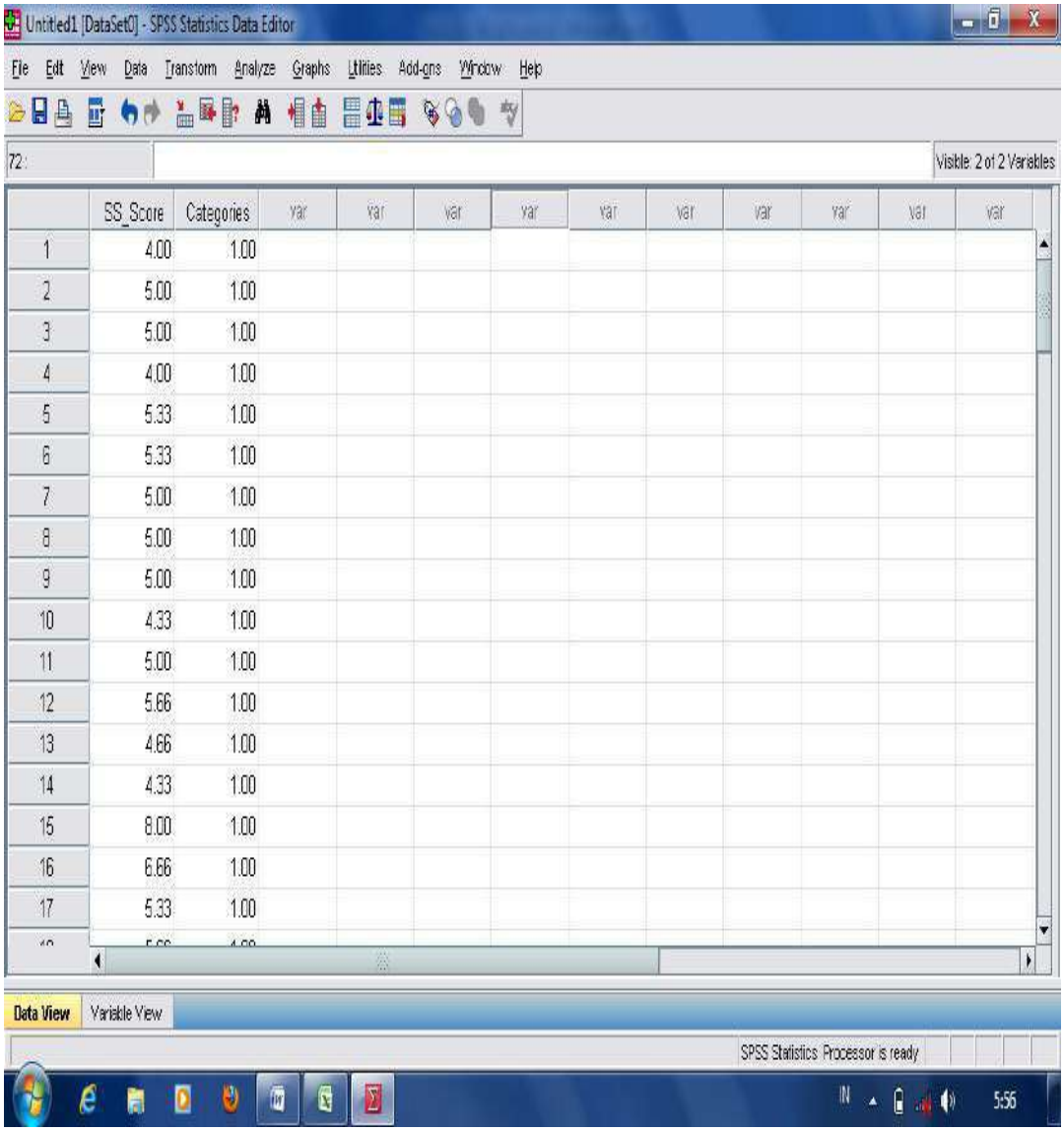
- Open the worksheet in the application of SPSS program;
- Move into Variable View, and type "**Ss_scores**" in the first column and "**categories**" in second column;



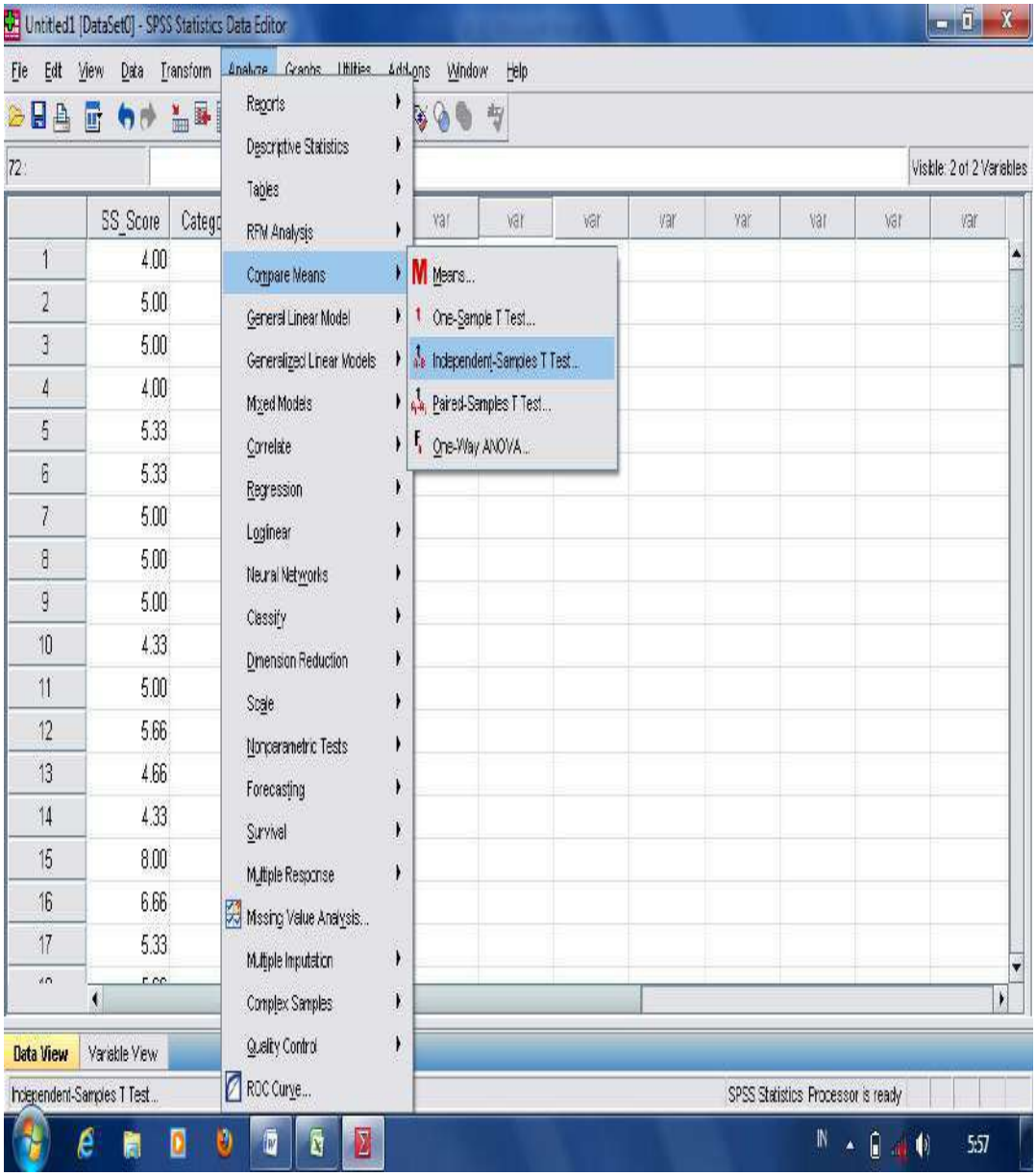
- Input the data into “values”; type 1 into the “Value” box, and type 1 for “**Posttest_Cntrl**” into the “Label” box, and 2 for “**posttest_Exprm**”. Then, press “add” to the column and OK.



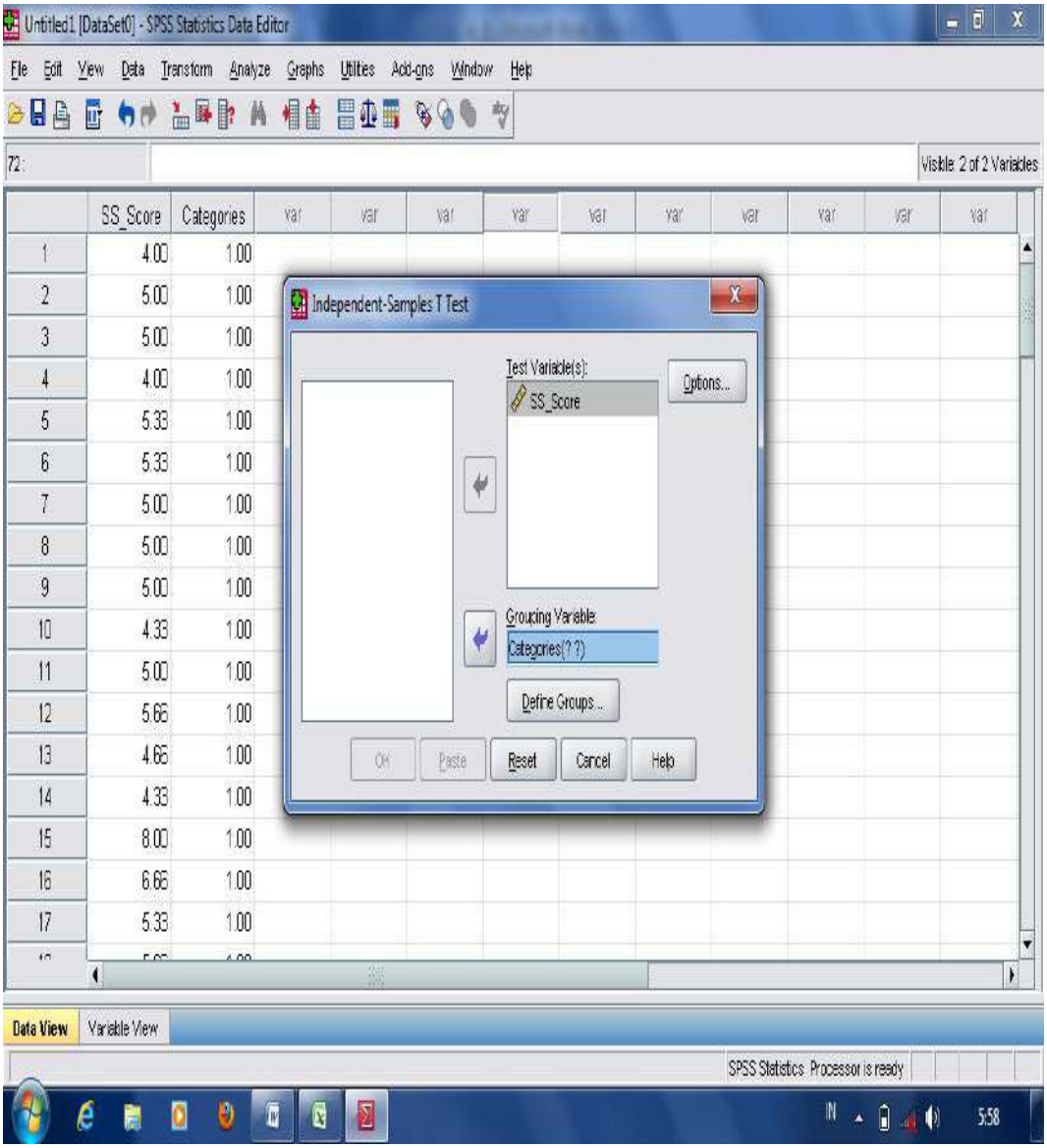
- Input the data score in the first column ("**Ss_scores**) for the score in the first category, then input "1" in the first category column and "2" for the second category;



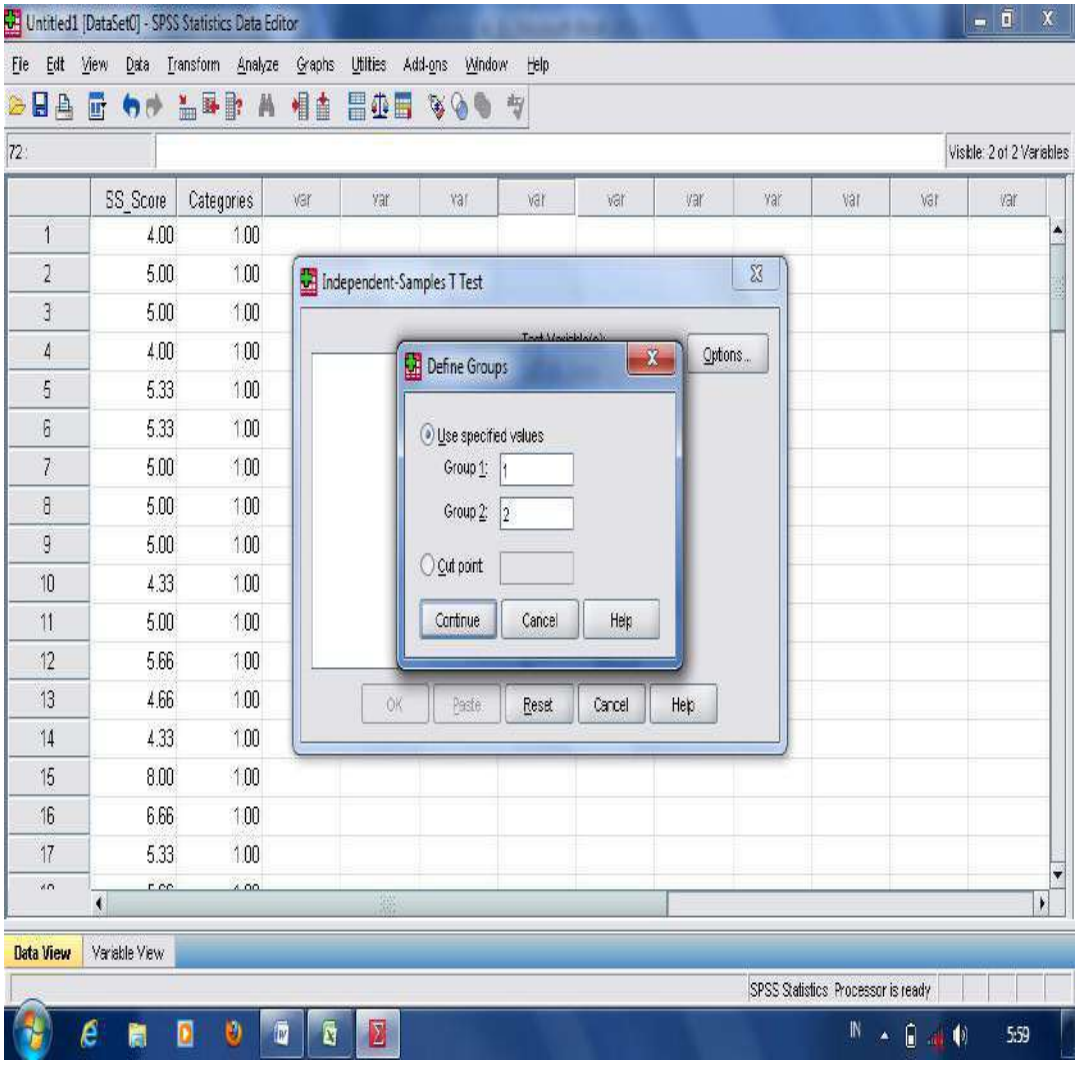
- Open "**Analyze**", choose "**Compare Means",** and click "**Independent Sample T-Test"**;



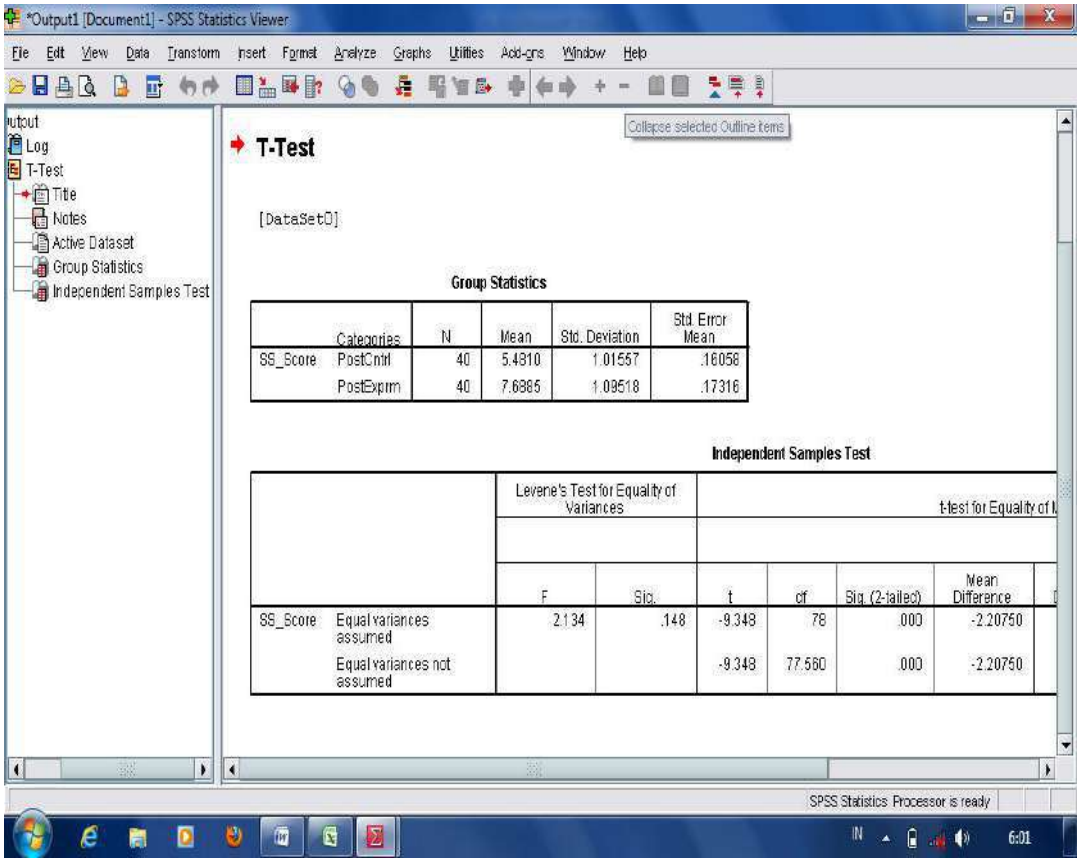
- Move "**Ss_scores**" in Test Variable Column and "**Categories**" into Grouping Variable;



- Click "**Define Groups**" – Type 1 for Group 1 and 2 for group 2; Press Continue.



- The result analysis (p-output) in measuring means significant difference or influence using independent sample t-test is shown as follows.

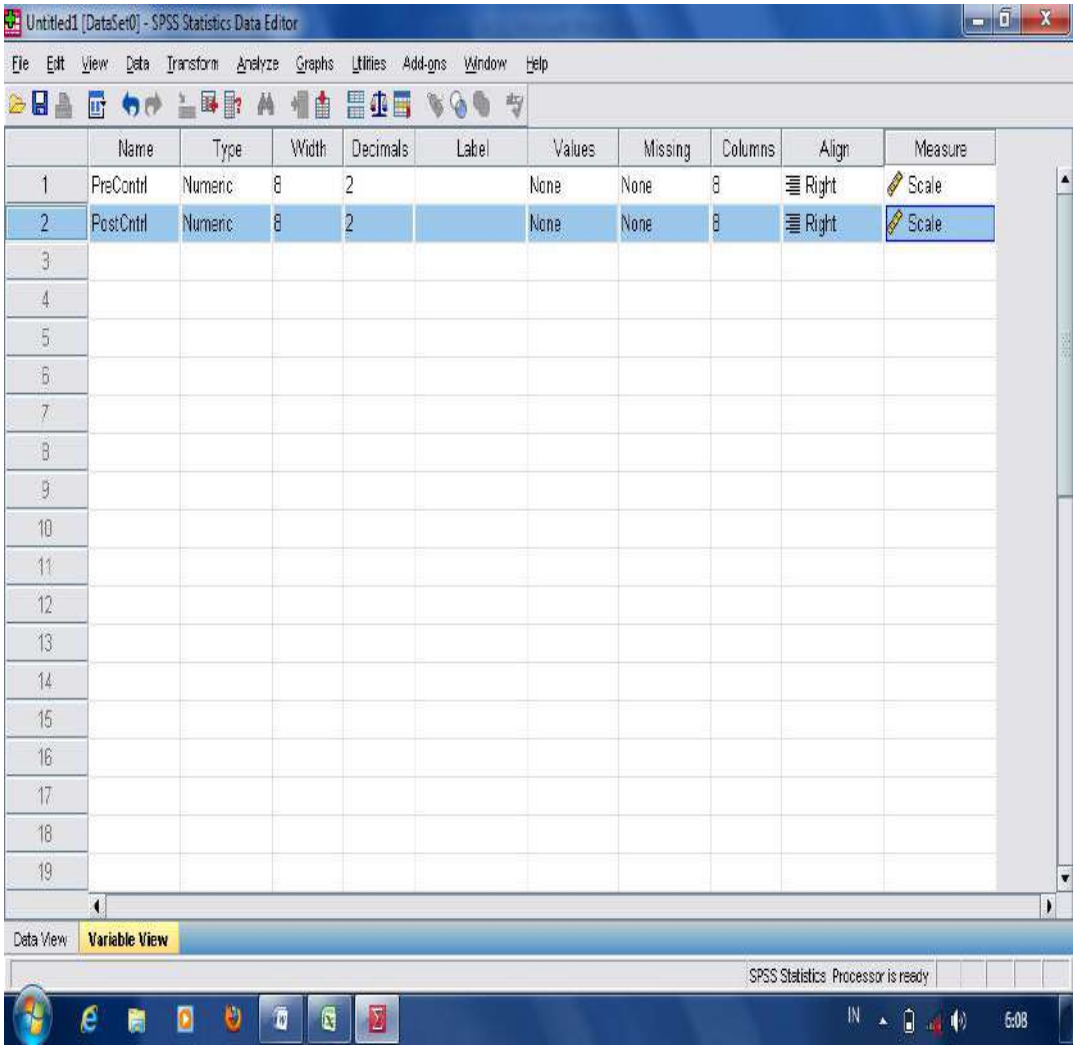


From the result analysis table in measuring means significant difference using independent sample T-Test, it is found that the p-output 0.000. It means that there is significant difference since the p-ouput is lower than mean significant difference at 0.05 levels.

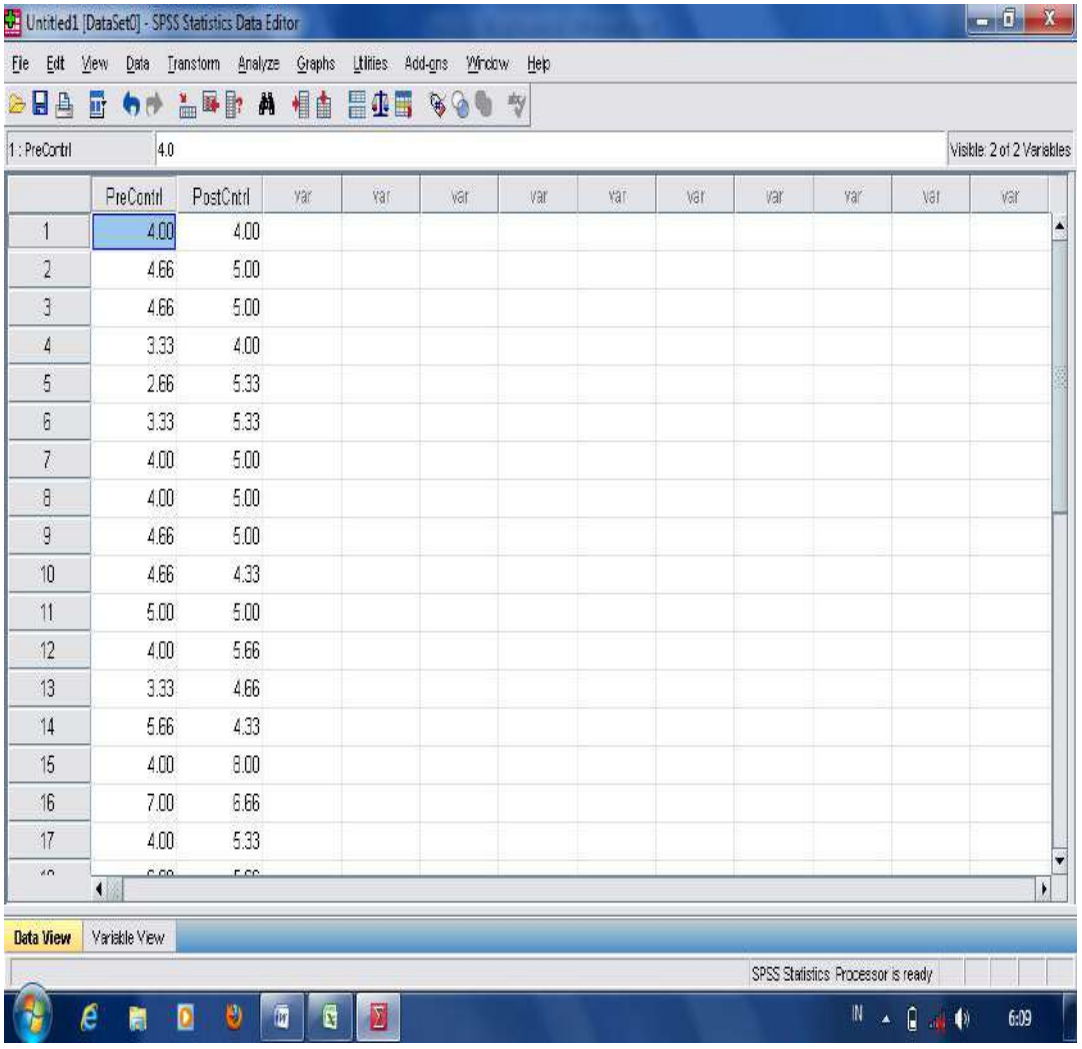
b. Paired Sample T-Test; these types of tests are used to compare groups that are related in some way. There are so many ways that participants in two groups can be related. One way is that participants in the first group are the same as participants in the second group. This is sometimes called a **repeated measures design**. A second way is that participants in the first group are genetically related to participants in the second group. For example, a pair of twins could be divided up so one twin participated with the first group and the other twin participated with the second group. A third way is if participants in one group

are matched with participants in a second group by some attribute.It is related to Hatch and Farhady (1982, p.122) that paired sample t-test is used when the two scores are matched from each student or score. Or in other words, it can be claimed that **paired sample t-test** is used to compare the results of two matched samples measuring: means significant improvement from students' pretest to posttest in each group, and means paired differences. However, (Hatch and Farhady, 1982, p.115) mention that the procedure for matched t-test is similar to the t-test for independent samples. From the preceeding statement, it can be assumed that after having matched the samples, paired sampe t-test can also be used to measure a **means significant difference** as it is used for independent sample t-test.

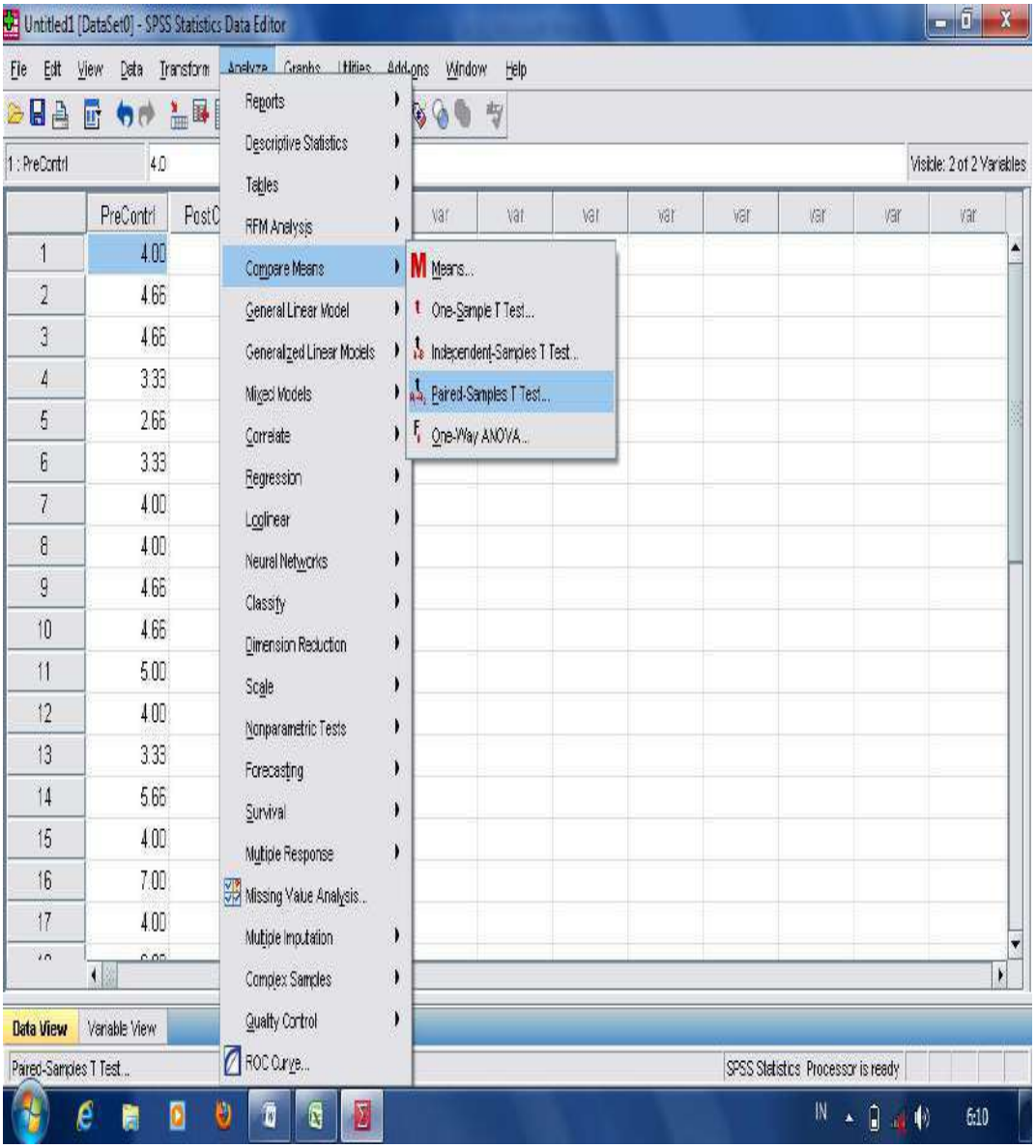
- **Means Paired Differences;** is used to compare the progress/improvement of the two means scores from pretest to posttest in each group (experiment and control). The following is example of a research study in finding means paired differences from students pretest to posttest scores on students' reading comprehension average scores taught using KWLH and conventional strategies. The analyses of the obtained scores are achieved from students' pretest to posttest scores in both groups. To start analysis using paired sample t-test in SPSS, it is shown as in the following steps:
  - Open the worksheet in the appliecation of SPSS program; Then, label the first column into "**Pretest_Cntrl**" and the second column into "**Posttest_Cntrl**";



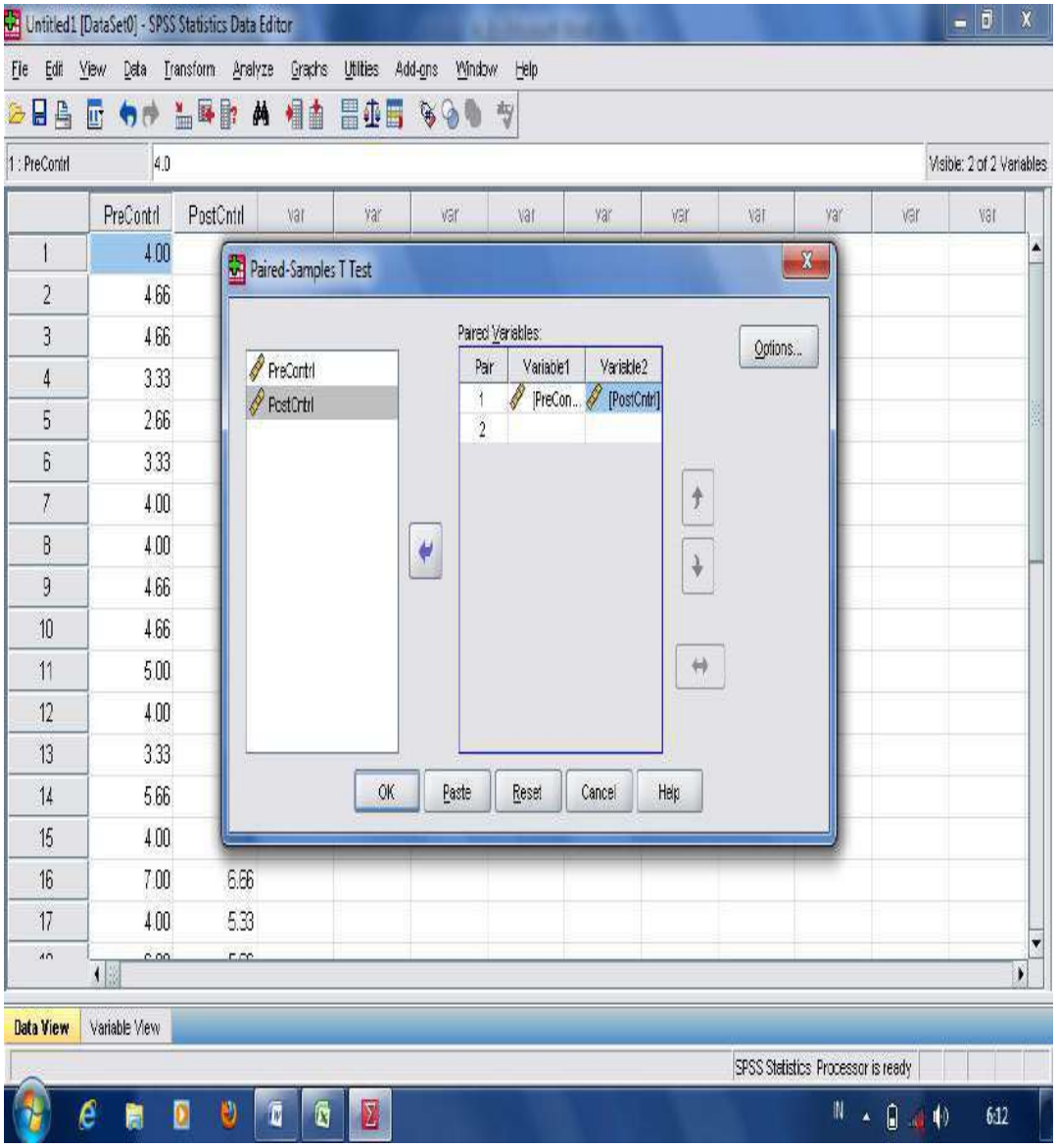
- Move the cursor into “data view” and input the data scores obtained in both columns;



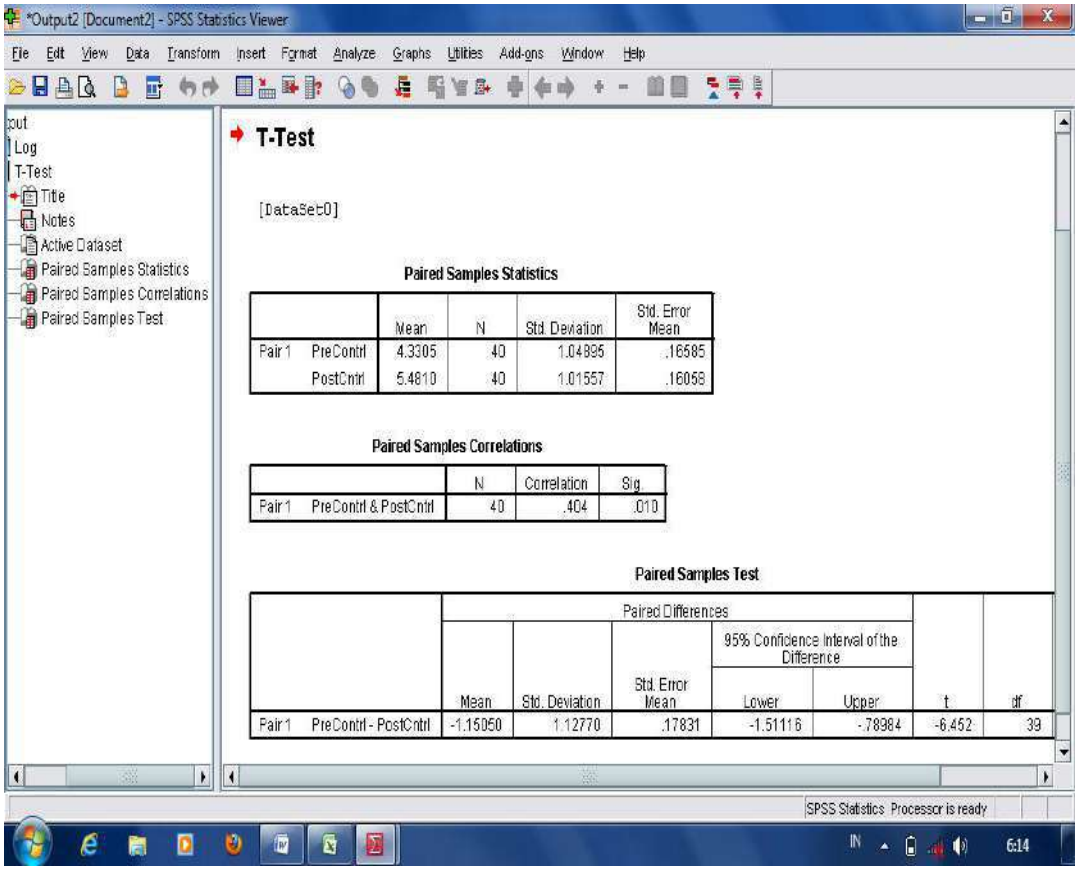
- Click "**Analyze"**, then move the cursor into the column "**Compare Means**". After that, choose "**Paired Sample T-Test**";



- A checked box of "**Paired Sample T-Test**" will be seen. Then, move the label "**pretest_cntrl**" into the first variable and the label "**posttest_Ctrl**" into the second variable; then, click OK.

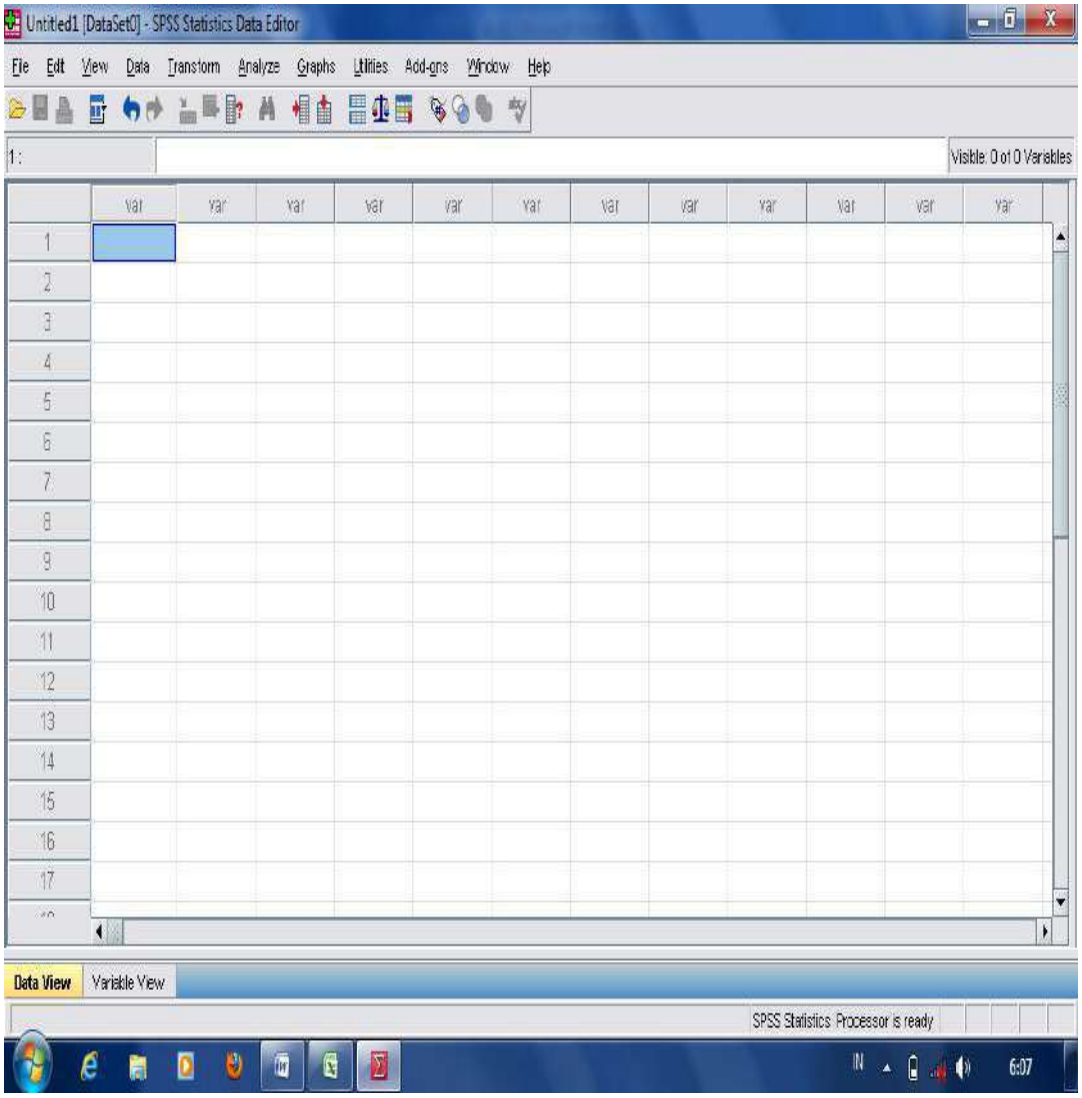


- The result of the output analysis in measuring means paired differences using Paired Sample T-Test is shown as follows.

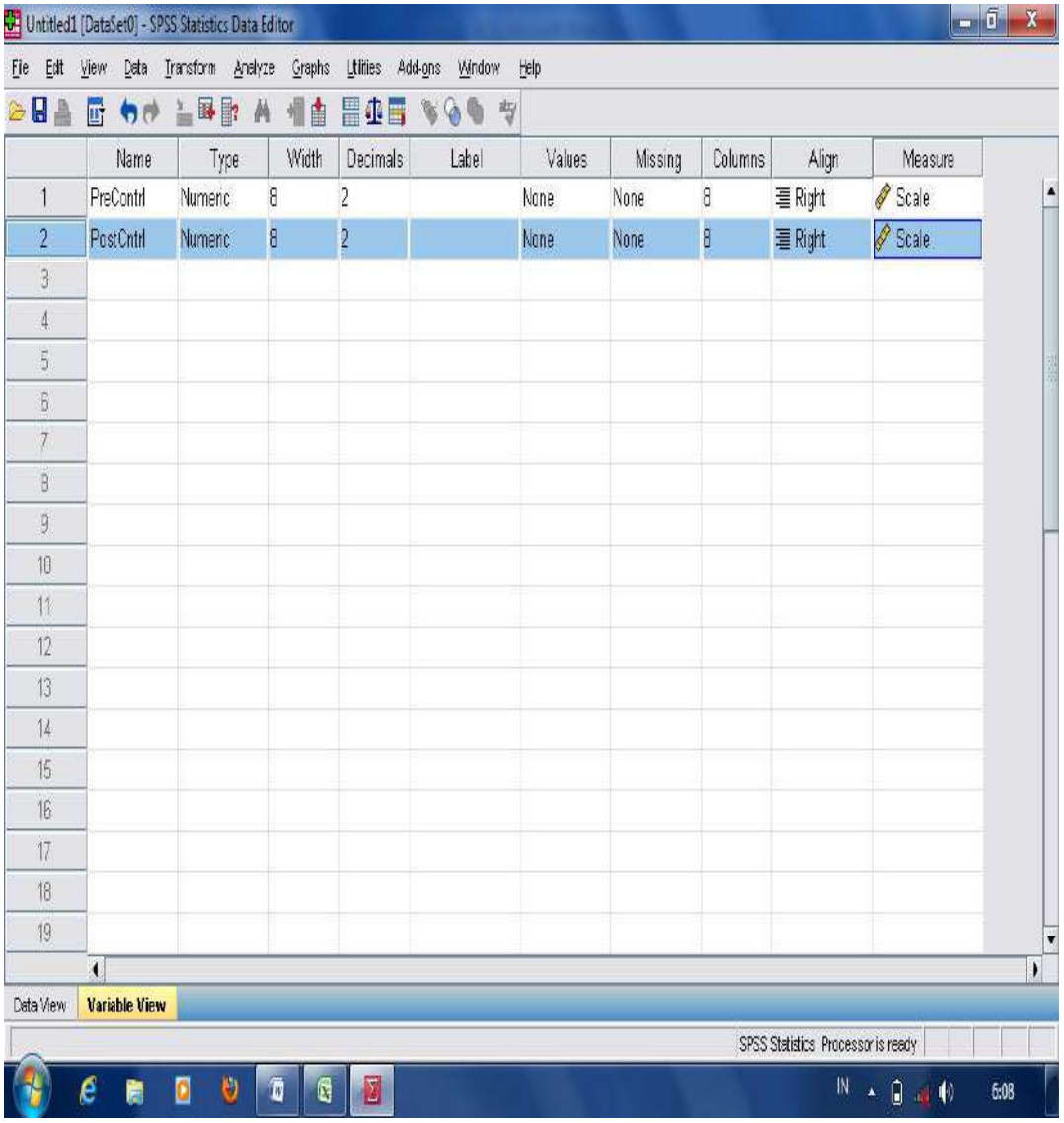


From the analysis table in measuring mean paired differences using "Paired Sample T-Test", it is found that means paired differences from students' pretest to posttest scores in control group is -1.15050.

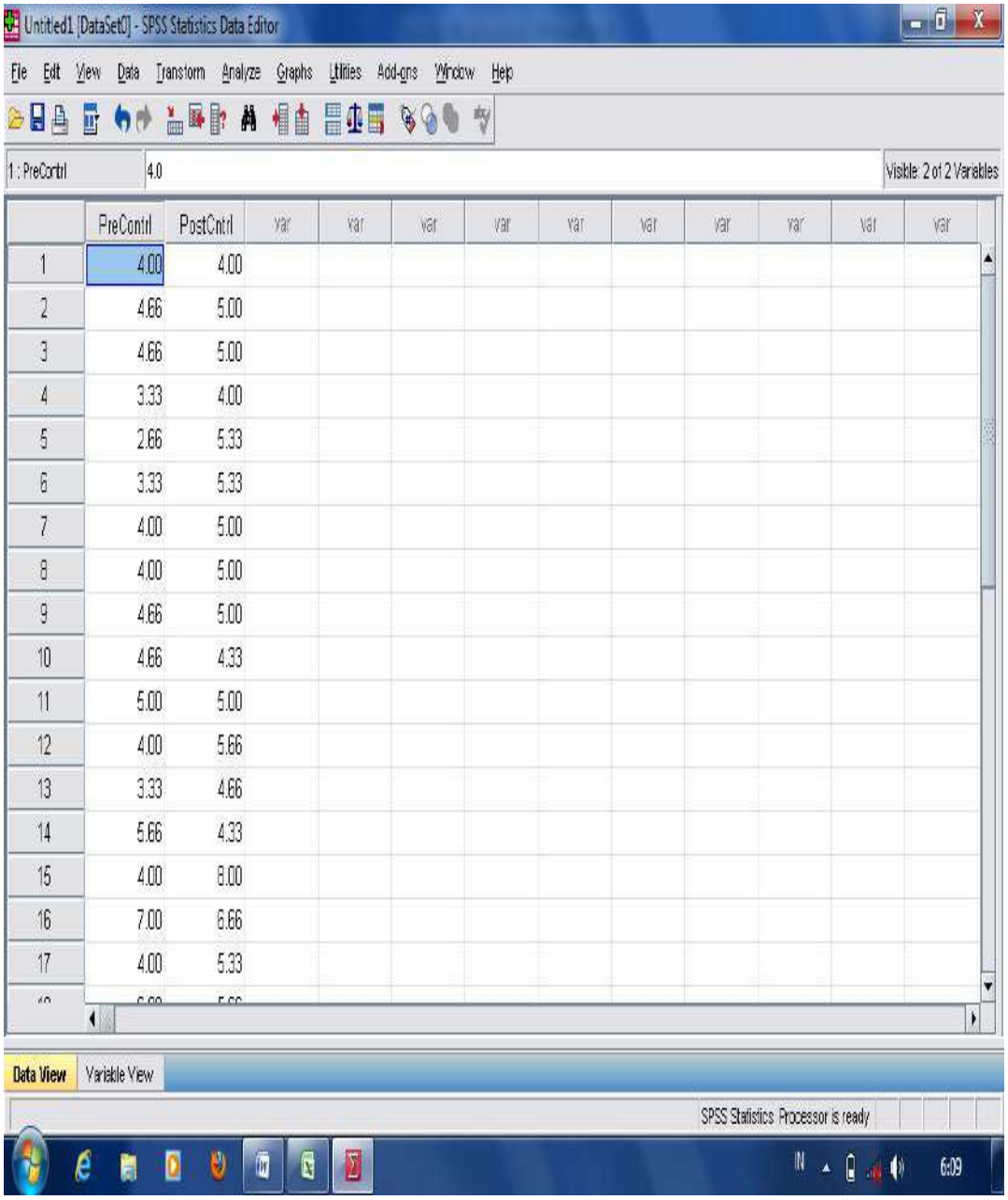
- **Means Significant Improvement;** is used to find the improvement of means scores in only one group. For example: a research study finding means significant improvement from students' pretest to posttest scores on students' reading comprehension average scores taught using conventional strategy. The scores and analysis are obtained from students' pretest to posttest scores taught using conventional strategy. To start analysis of means significant improvement using "Paired Sample T-Test", it is illustrated in the following steps:
  - Install and run the application of SPSS program;



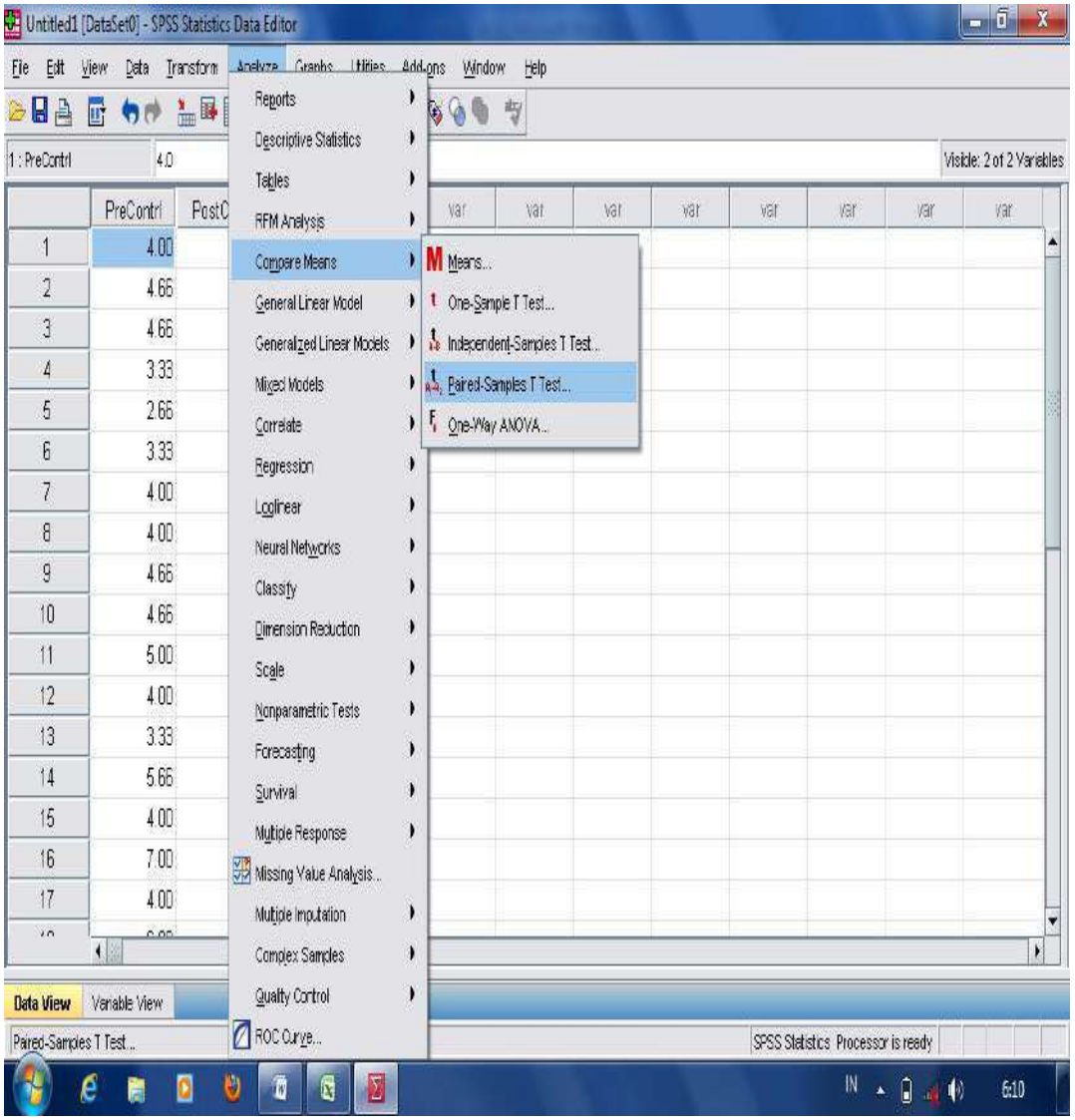
- Label the first column into "**Pretest_Cntrl**" and the second column into "**Posttest_Cntrl**";



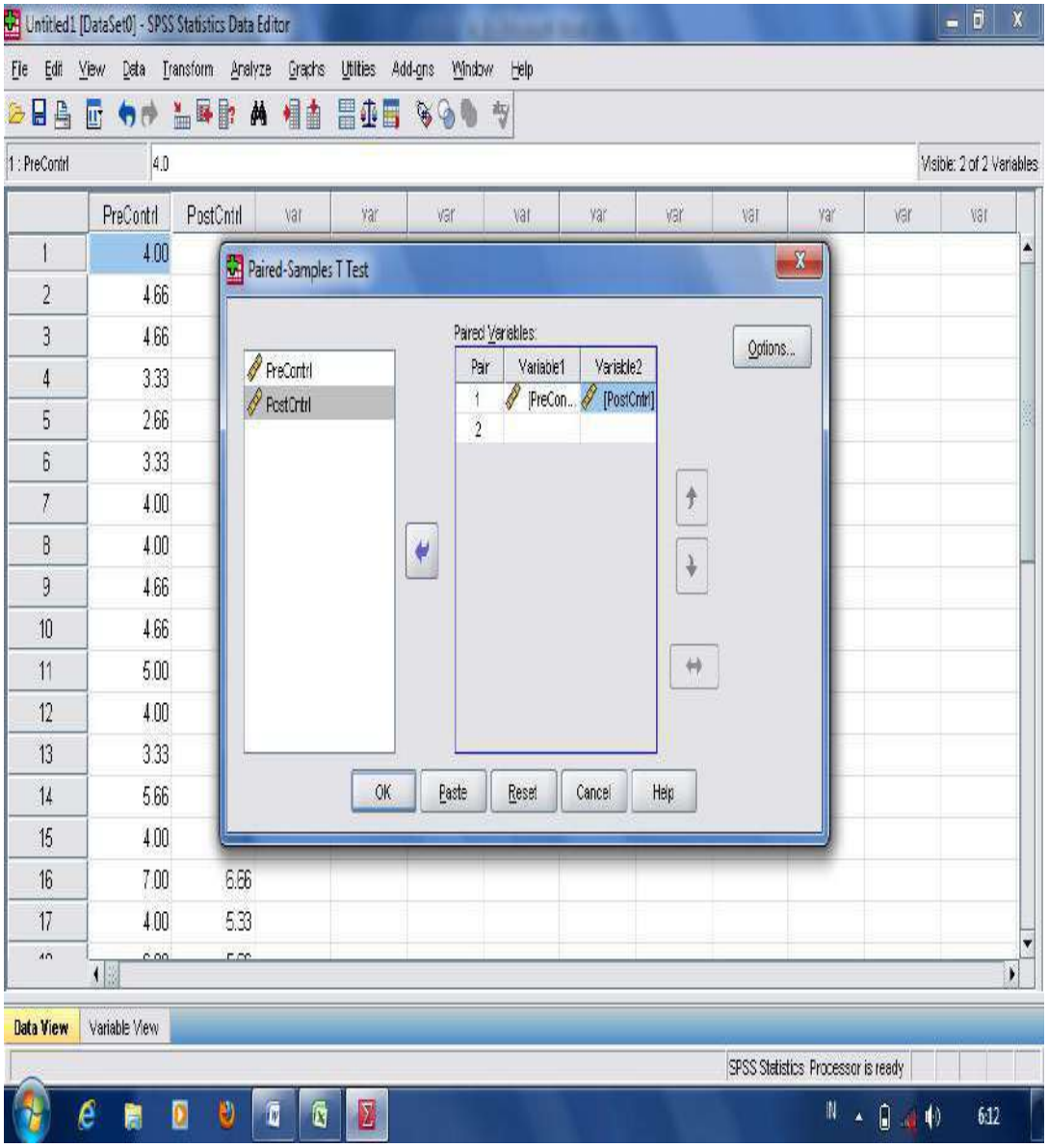
- Move the cursor into "data view" and input the data scores obtained in both columns;



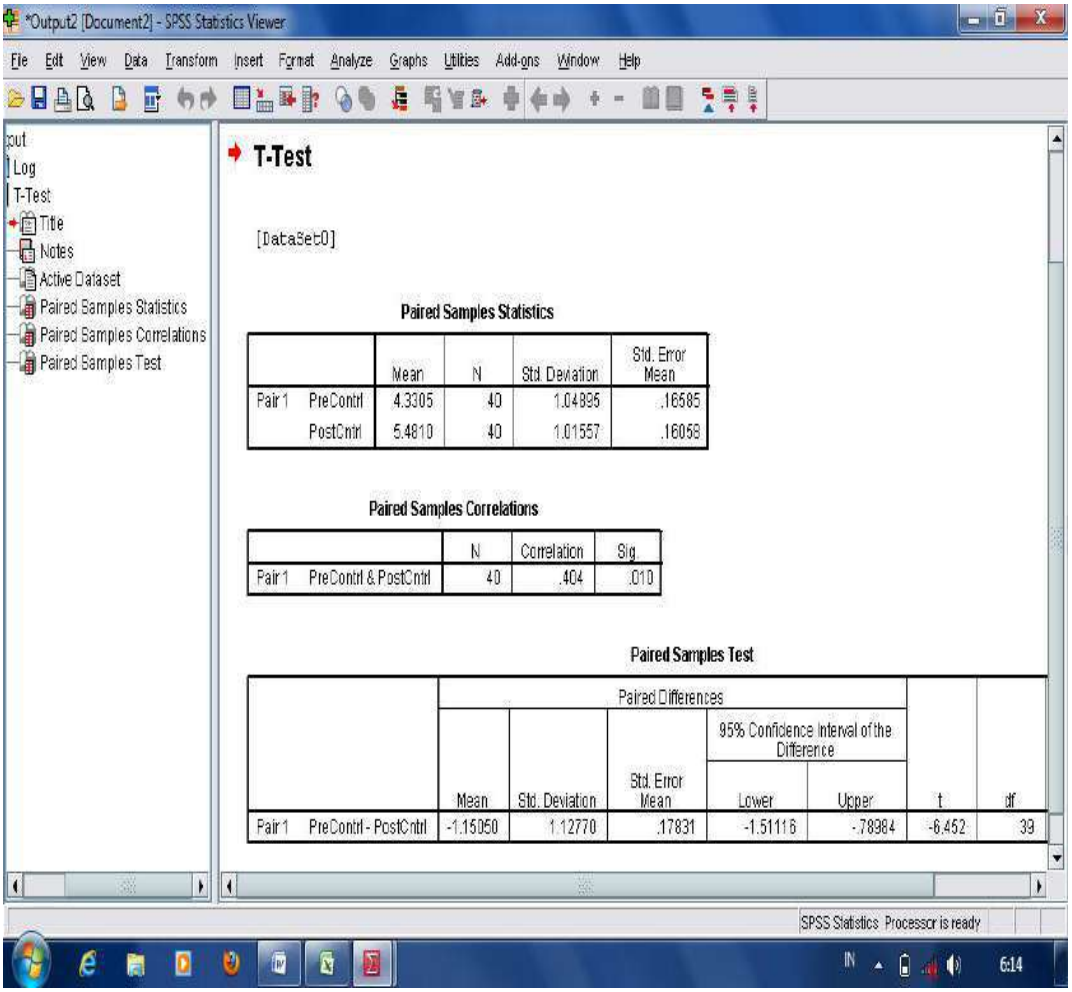
- Click "**Analyze"**, and then move the cursor into the column "**Compare Means**". After that, it continues to choose "**Paired Sample T-Test**";



- A checked box of "**Paired Sample T-Test**" will be seen. Then, move the label "**pretest_cntrl**" into the first variable and the label "**posttest_Ctrl**" into the second variable; then, click OK.



- The result of the output analysis in measuring means paired differences using paired sample t-test is shown as follows.

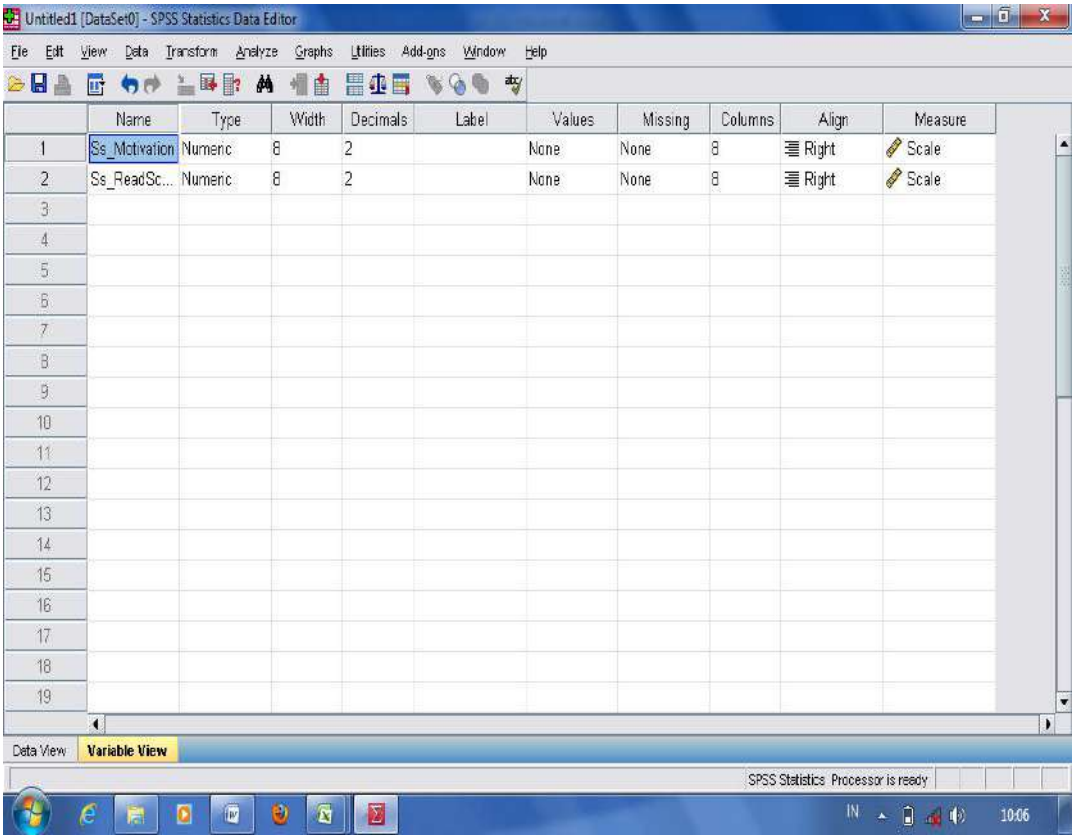


From the above table measuring means significant improvement using "Paired Sample T-Test", it was found the p-output 0.000. It means that there is means significant improvement on students' reading comprehension from pretest to posttest scores in control group since the p-output is lower than mean significant difference at 0.025 levels.

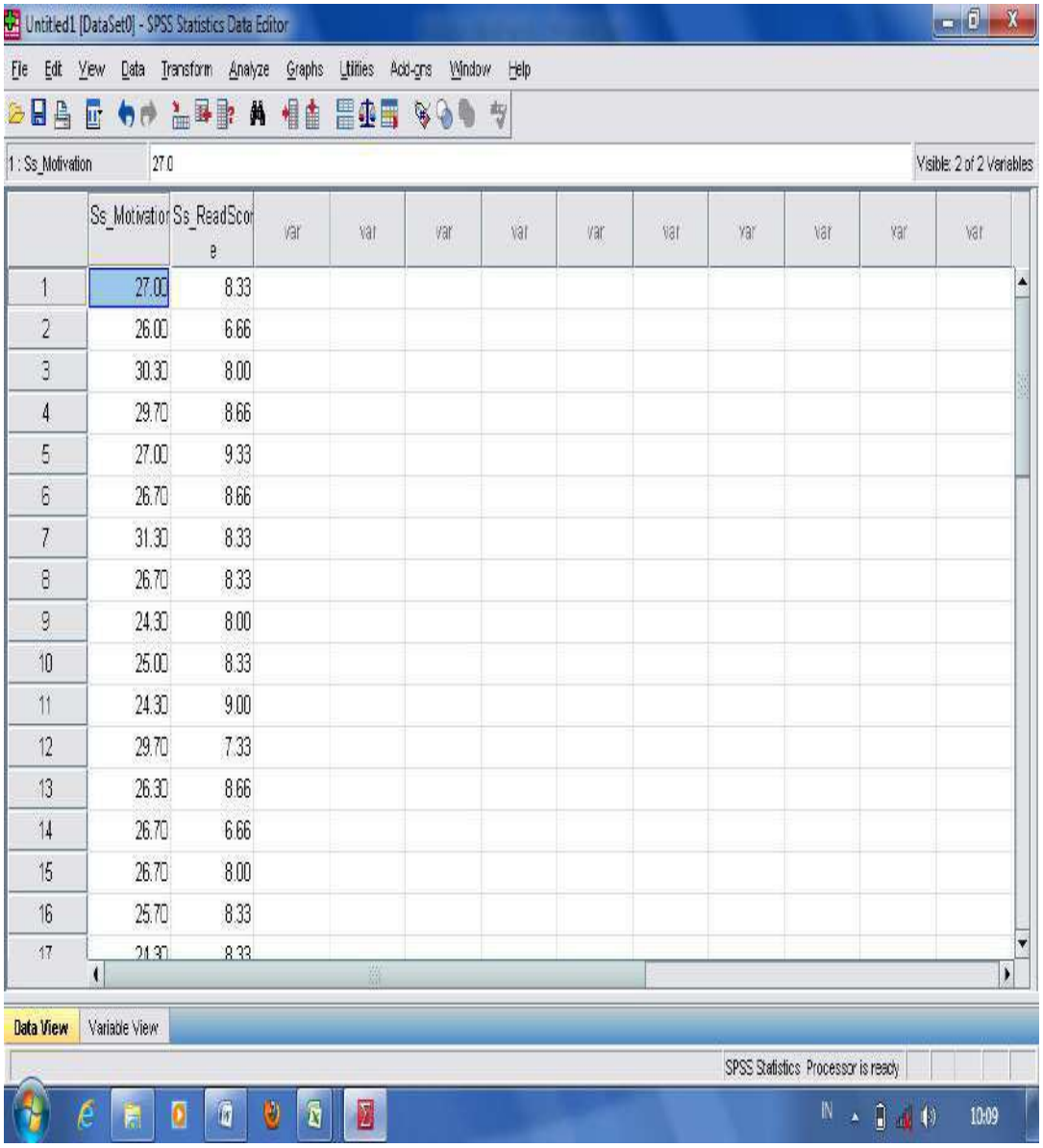
2. **Regression Analysis;** is used to measure or to predict the *effects/impacts/influence* of two or more variables whether or not there is strength of relationship between two variables (Basrowi and Soenyono, 2007, p.159). Research question in analyzing regression analysis in regression analysis is to find a **significant linear relationship or the effects/impact/influence** of one or more

predictor variables with one independent variable identifying the relationship between two variables. Significant linear relationship is found whenever the p-output (Sig.2-tailed) is lower than 0.05. The following is example of a research study in finding the **effects/impacts/influence of motivation** on students' **reading comprehension average scores**. To start "**Simple Linear Regression**" in measuring significant linear relationship between two variables (predictor and dependent), the following are steps to start:

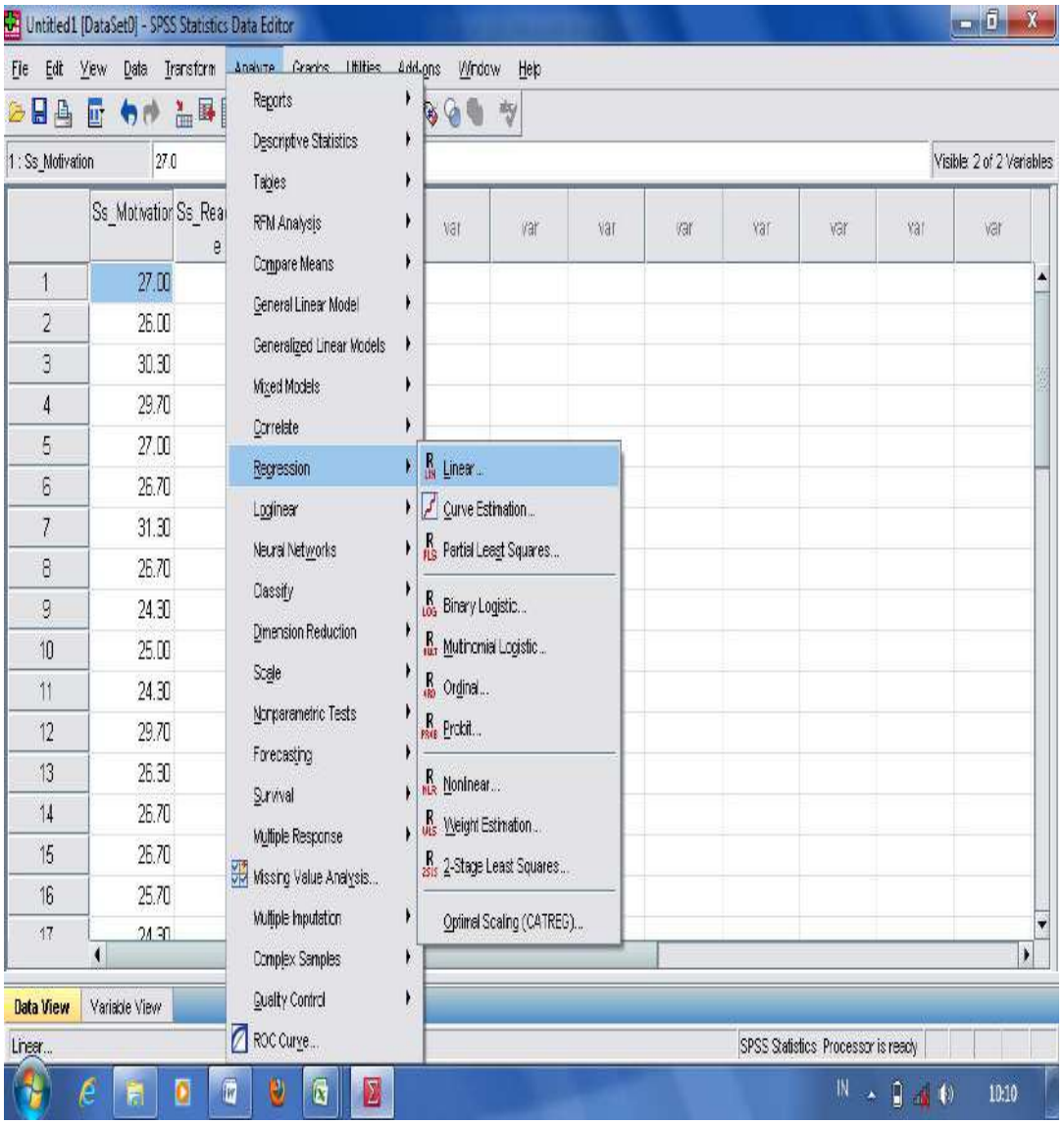
- Open the worksheet in the application of SPSS program;
- Move into a "variable view"; type the column into the category that is going to be compared namely: "**motivation**", and "**Ss_ReadScore**".



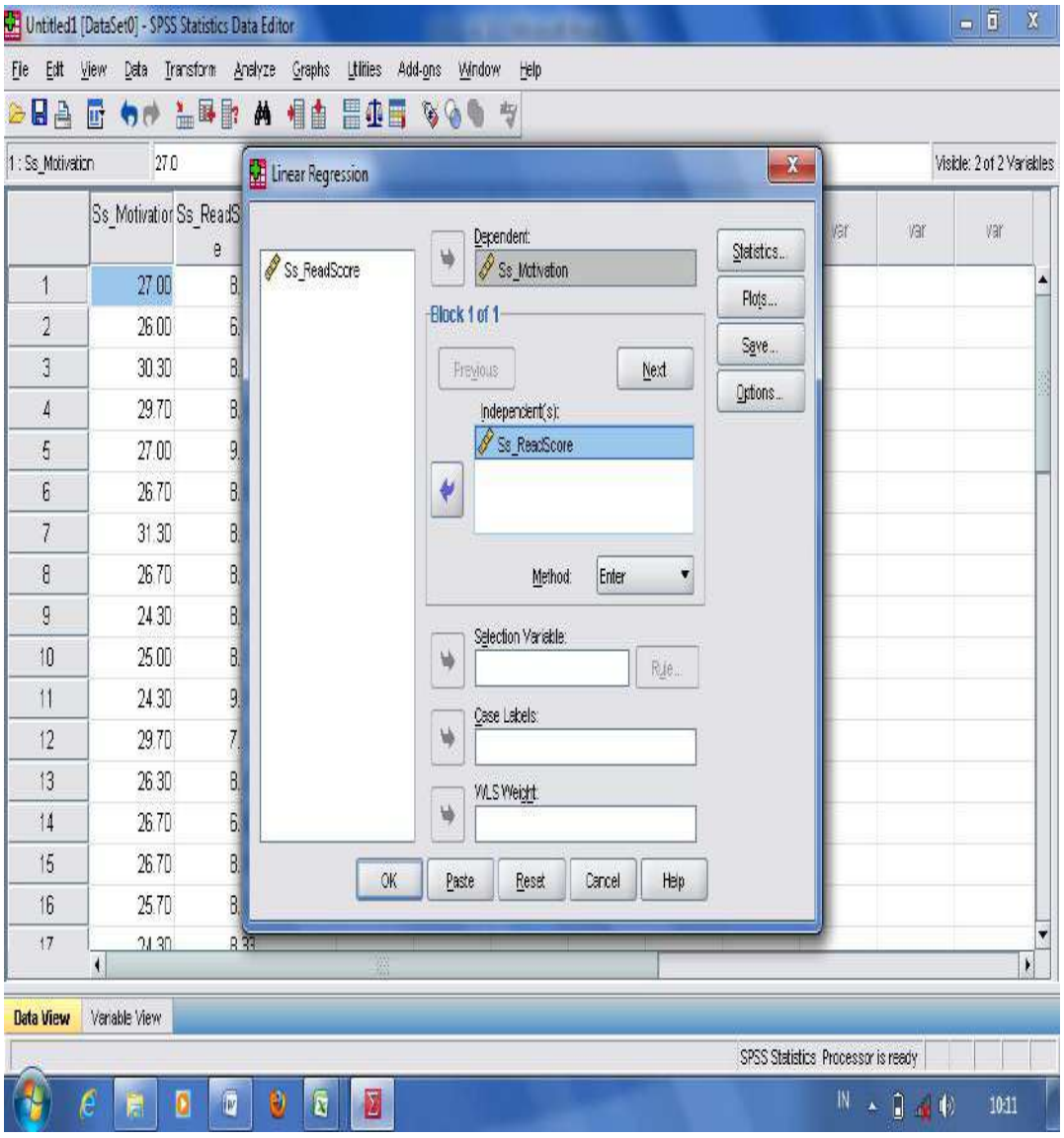
- Move into “data view”, input the score obtained in each category;



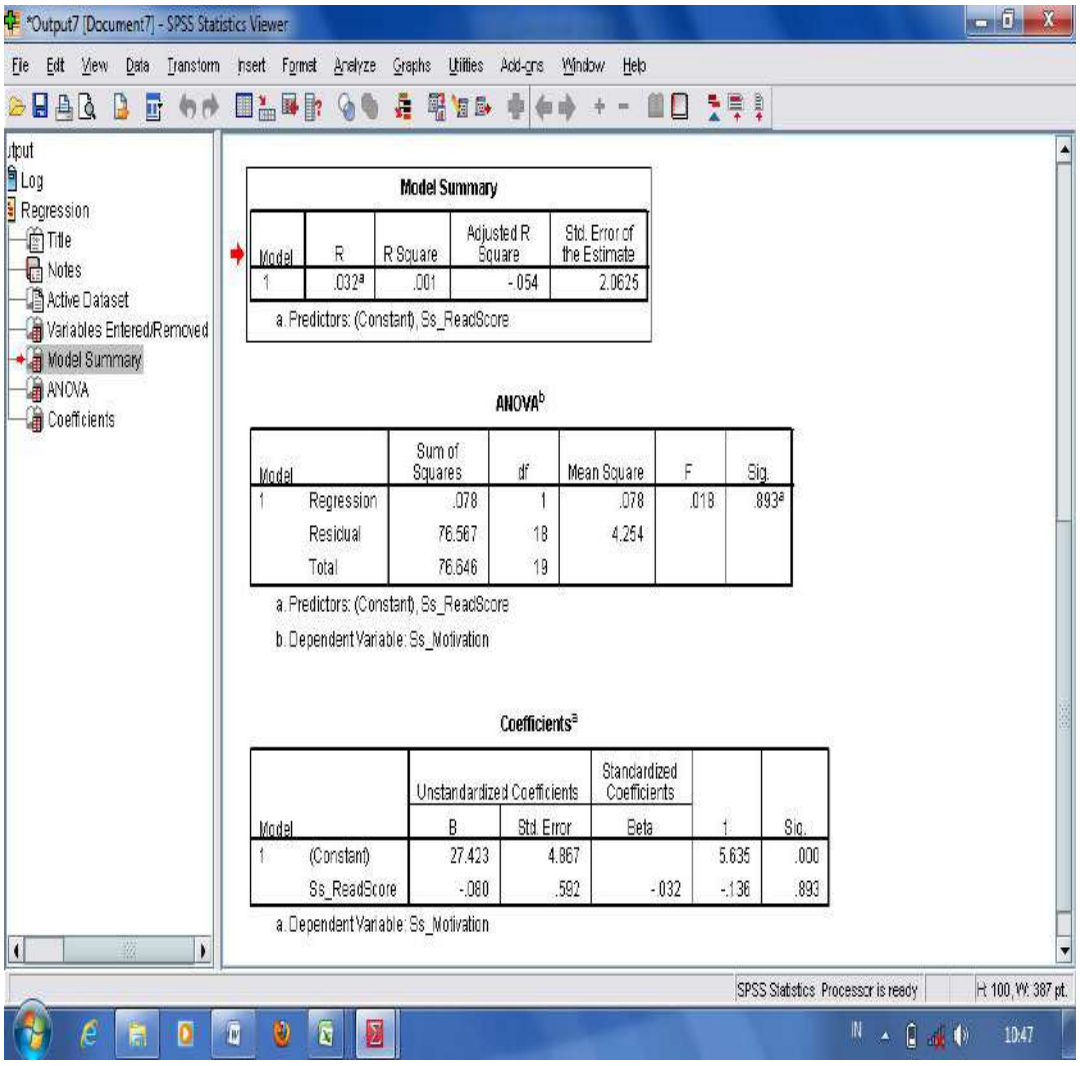
- After that, click “**Analyze**”, then choose “**Regression**” and move the cursor into the right column and choose “**Linear**”;



- Move the variable "**Ss_Scores**" into the dependent box provided in the right side, then "**Motivation**" and **Ss_ReadCmpr**" into independent box in the right side below the dependent box; then, click OK.



- The result analysis in measuring the regression analysis is displayed in the following table.



From the above table in measuring regression analysis, it is found that the significant score is 0.032. From the p-output, it can be stated that there is significant linear relationship between motivation and students' reading comprehension scores since the significant score of the p-output is lower than 0.05.

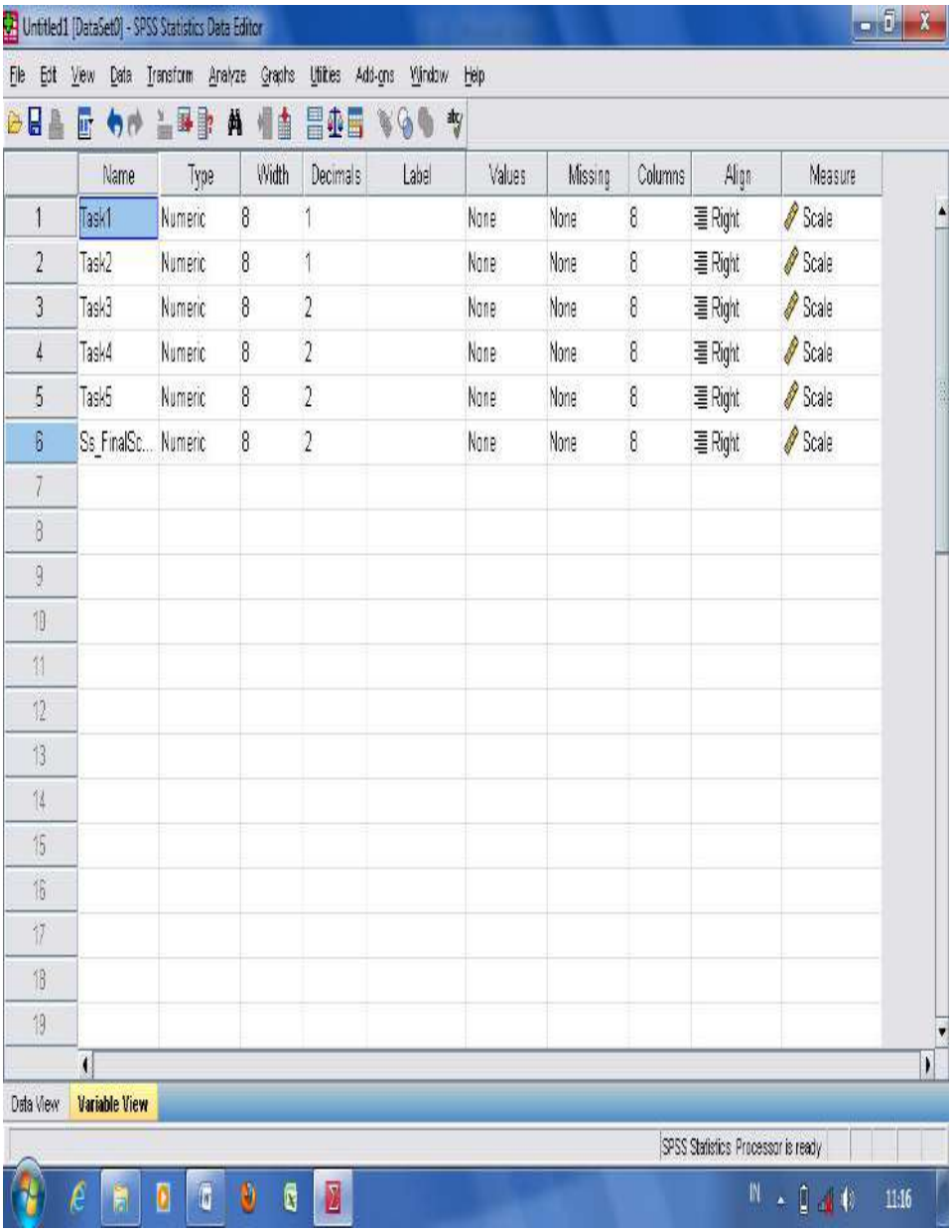
Then, the regression analysis is done to the variables which have one dependent variable, and two or more independent variables. A research study finding the **effects/impacts/influence** of the students'

reading assignments (considered as independent variable) on students' reading final scores (considered as dependent variable). The students' reading assignments and students' reading final scores are described in the following table.

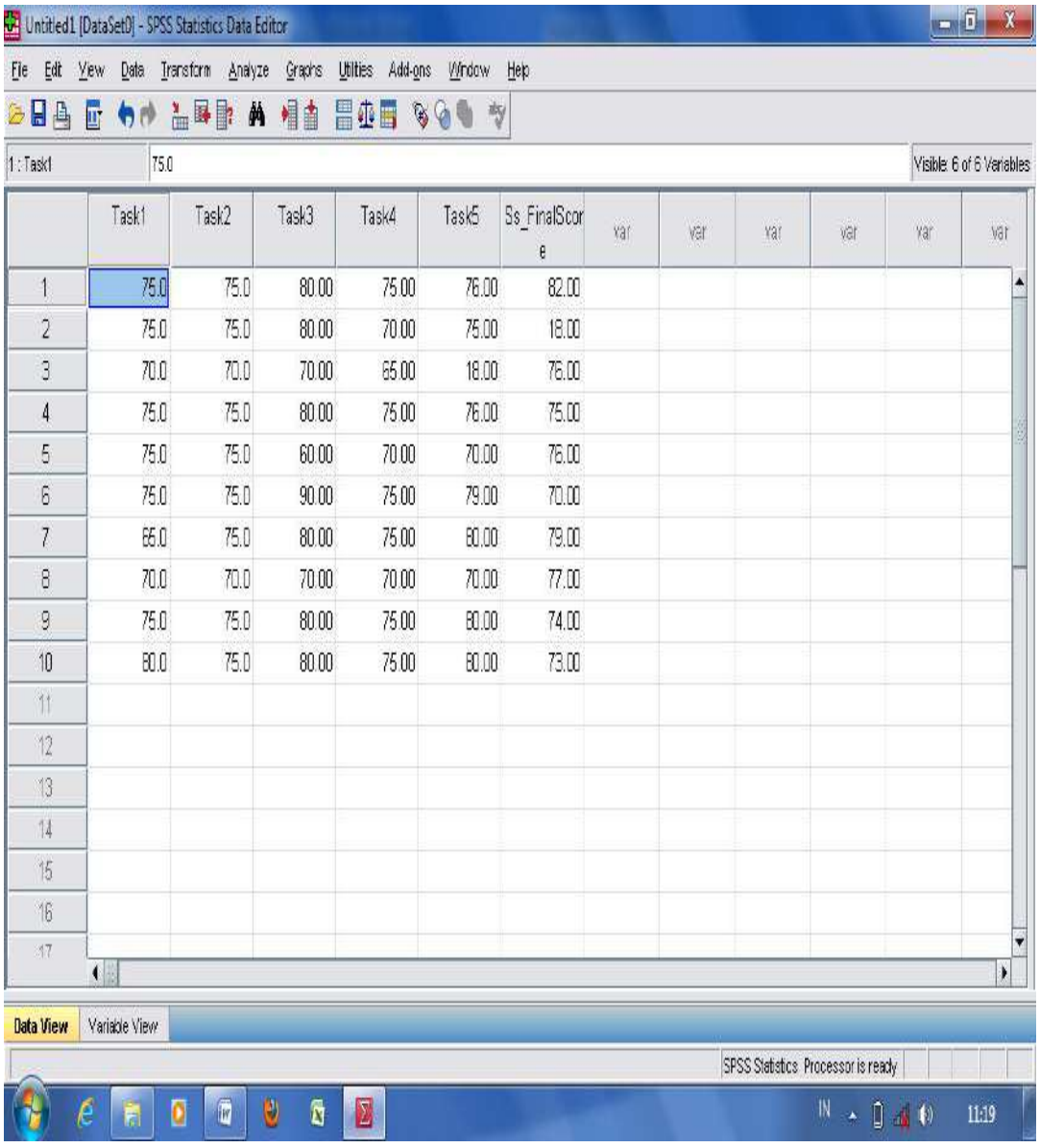
N	Ss_reading assignments' scores					Ss_reading final scores
	Task1	Task2	Task3	Task4	Task5	
1	75	75	80	75	76	82
2	75	75	80	70	75	18
3	70	70	70	65	18	76
4	75	75	80	75	76	75
5	75	75	60	70	70	76
6	75	75	90	75	79	70
7	65	75	80	75	80	79
8	70	70	70	70	70	77
9	75	75	80	75	80	74
10	80	75	80	75	80	73

Then, the above scores are analysis to find significant linear relationship between students' reading assignments' scores and students' reading final scores using SPSS application program as follows.

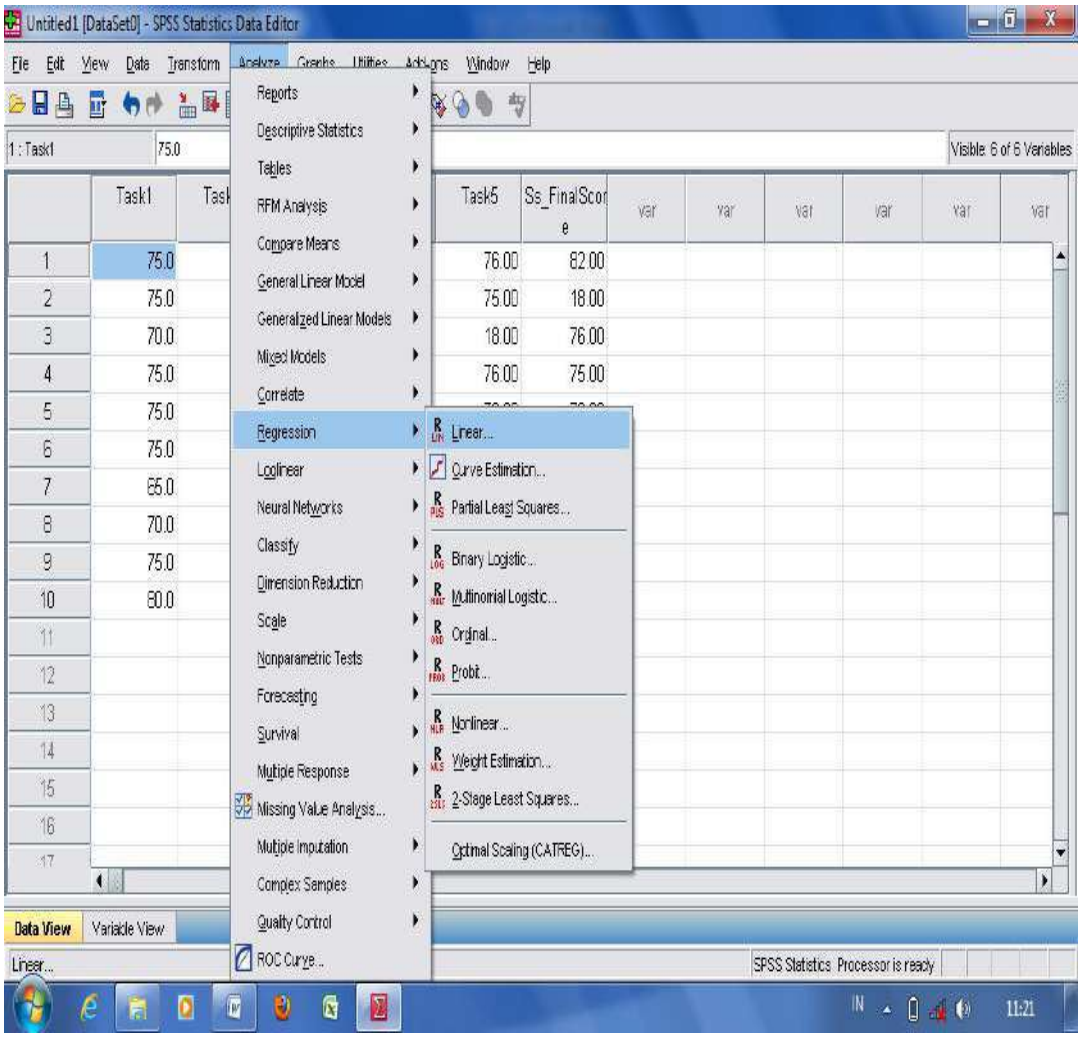
- Open the worksheet in the application of SPSS program;
- Move into a "Variable View"; type the column into the category that is going to be compared namely: "**Task1, Task2, Task3, Task4, Task5**", and "**Ss_FinalScore**".



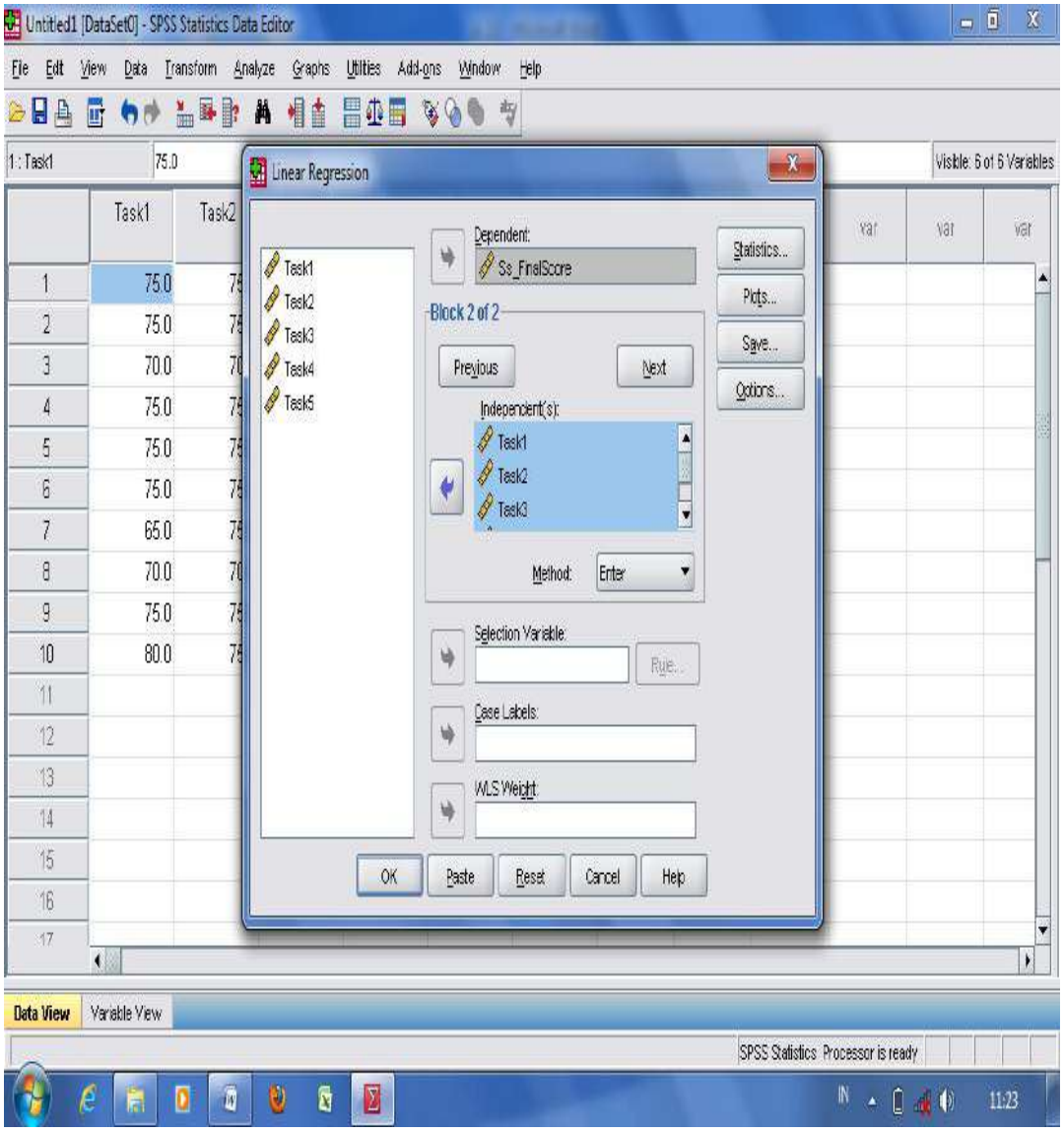
- Move into “data view”, input the score obtained in each category;



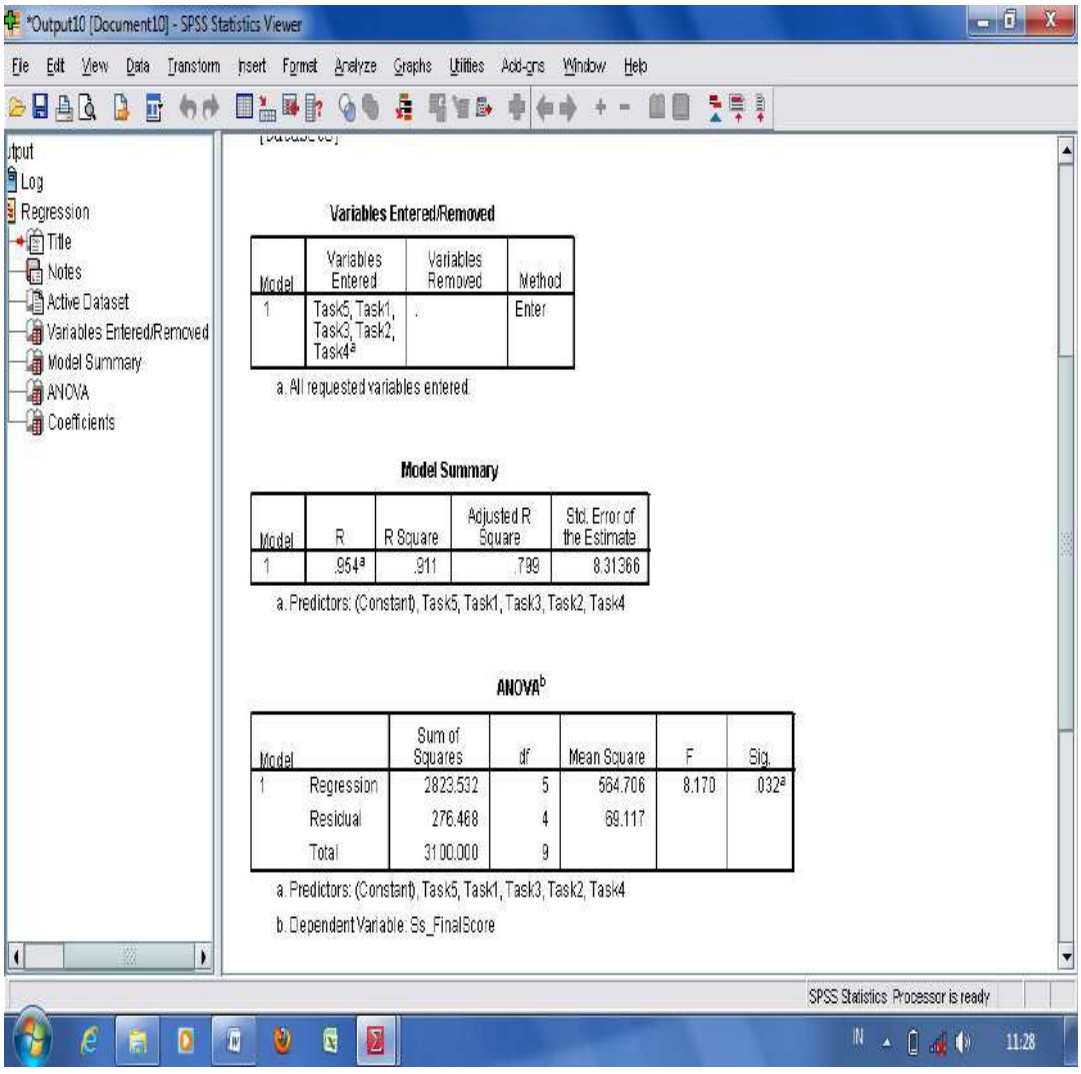
- After that, click "**Analyze**", then choose "**Regression**" and move the cursor into the right column and choose "**Linear**";



- Move the variable "**Ss_Scores**" into the dependent box provided in the right side, then "**Motivation**" and **Ss_ReadCmpr**" into independent box in the right side below the dependent box; then, clik OK.



- The result analysis in measuring the regression analysis is displayed in the following table of p-output.



From the table of the p-output in measuring significant linear relationship using regression analysis, it is found that the significant score is 0.954. From the p-output, it can be stated that there is no significant linear relationship between students' reading assignments and students' reading final scores since the significant score of the p-output is higher than 0.05.

3. **Correlation Analysis;** is used to measure the correlation between two variables whether the values of two variables are associated or

not (Achelis, 2013). To find the correlation between two variables, **Bivariate Correlation** is used. (Basrowi and Soenyono, 2007, p.109) state that in bivariate correlation, **Pearson Product-Moment Coefficient** is administered where the obtained data should distribute into normal. Furthermore, Sugiyono (2012, p.231) states that to interpret the significant score of the correlation coefficient (r-output), it should be referred to the interval coefficient for interpreting the correlation coefficient as it is displayed in the following table.

Table 7

Interval coefficient for interpreting
the correlation coefficient

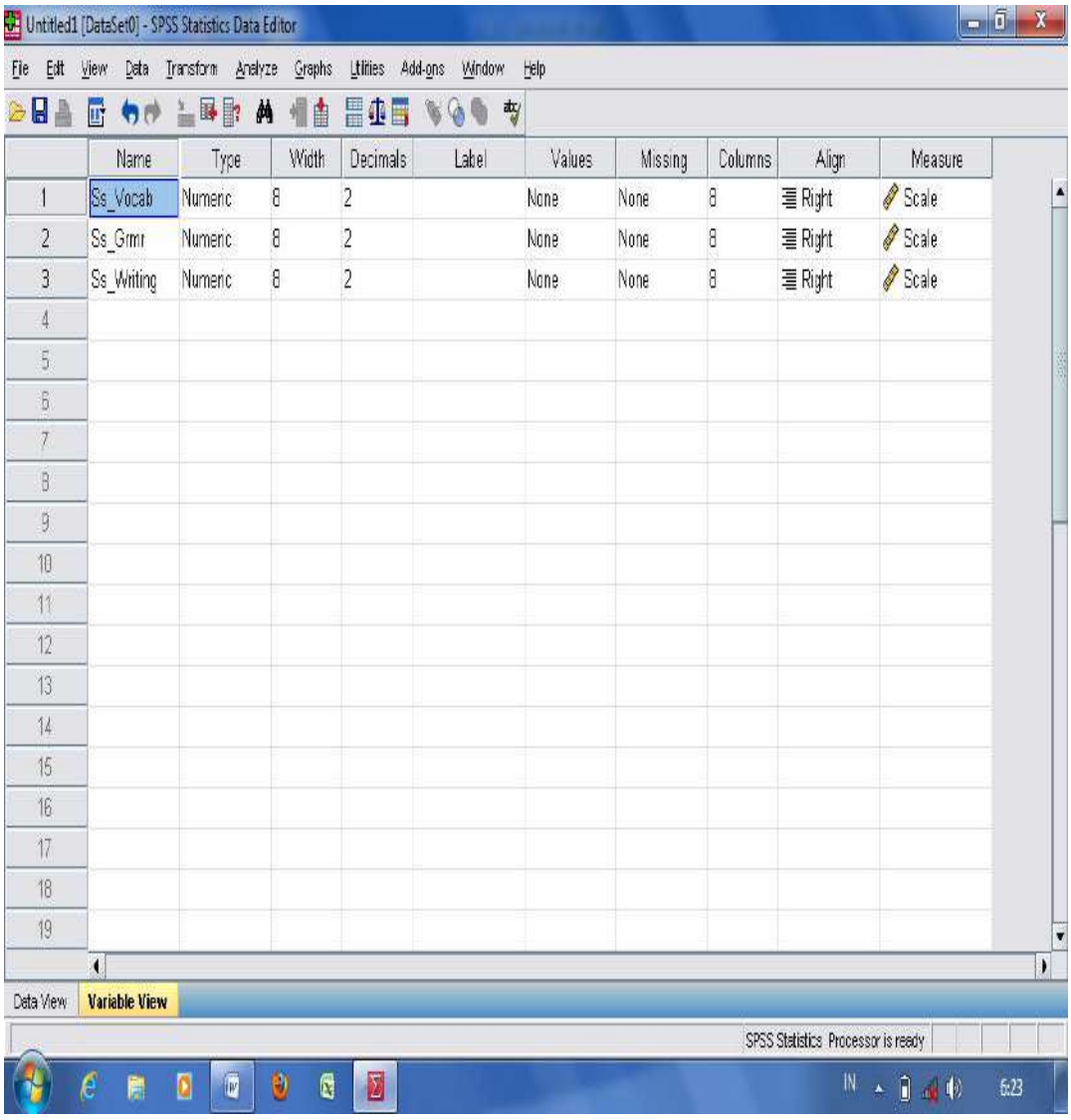
Interval coefficient	Correlation Level
0.00 – 0.199	very low (no correlation)
0.20 – 0.399	Low
0.40 – 0.599	middle
0.60 – 0.799	High
0.80 – 1.000	very high

Source: Sugiyono. (2012:p.231). *Statistika Untuk Penelitian*. Bandung: Alfa Beta.

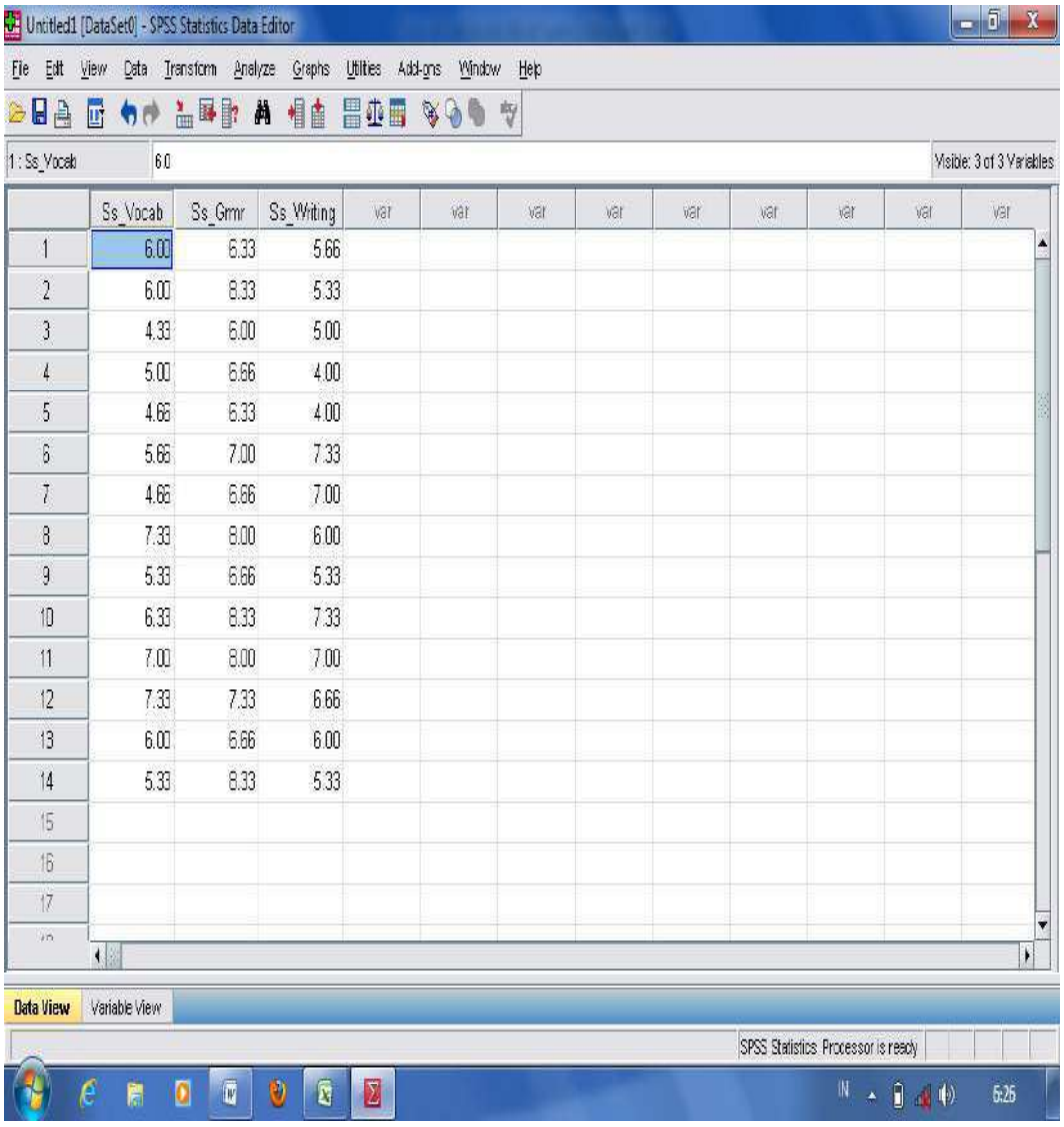
The following is example of a research study in finding the **correlation** between **students' vocabulary scores** and **students' grammar scores** on students' writing average scores. To start the correlation analysis, the **Pearson coefficient** is used in the following steps:

- Open the worksheet in the application of SPSS program;

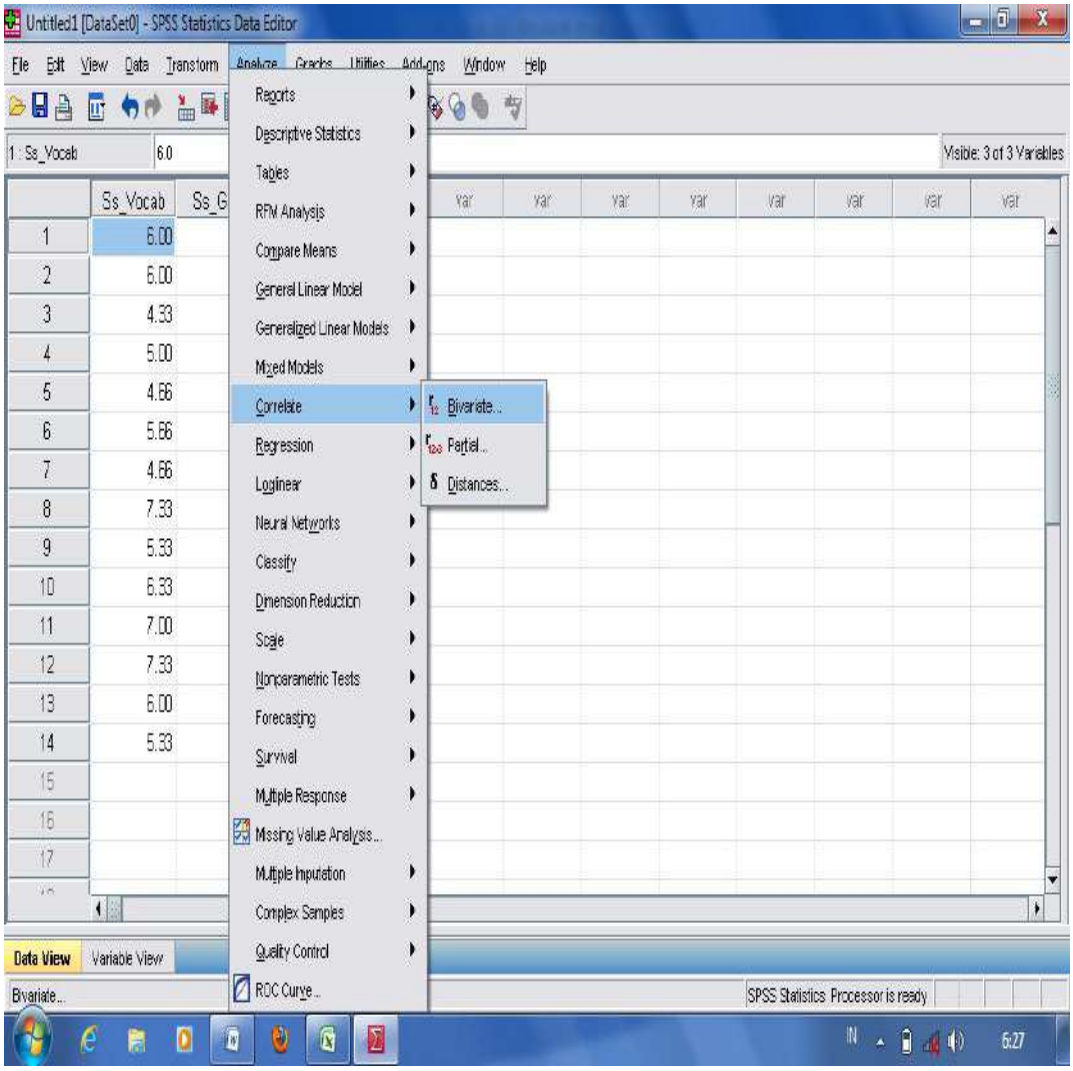
- Move into a "variable view"; type the column into the category that is going to be compared namely ("**Ss_Vocab"**, "**Ss_Grmr"**, and "**Ss_Writing**").



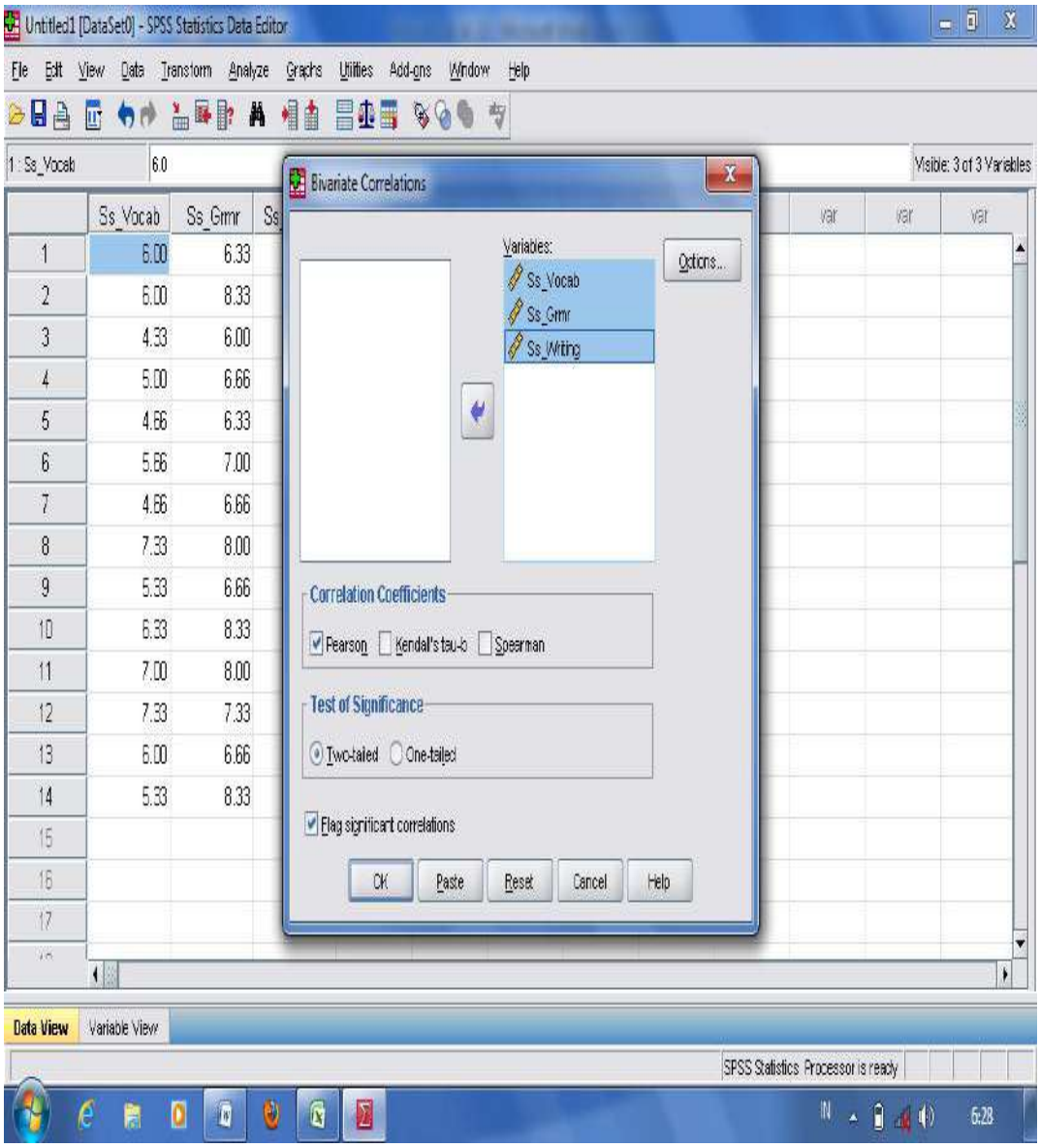
- Move the cursor into “data view” and then, input the obtained scores in each category;



- After that, click "**Analyze**", then choose "**Correlate**". Move the cursor into the right side and select "**Bivariate**";



- Then, move all the variable boxes in the left side into the column of variable; and press OK.



- The result of correlation analysis using "Pearson product moment correlation coefficient" is shown in the following table.

Correlations

		Ss_Vocab	Ss_Grmr	Ss_Writing
Ss_Vocab	Pearson Correlation	1	.604*	.229
	Sig. (2-tailed)		.022	.431
	N	14	14	14
Ss_Grmr	Pearson Correlation	.604*	1	-.055
	Sig. (2-tailed)	.022		.852
	N	14	14	14
Ss_Writing	Pearson Correlation	.229	-.055	1
	Sig. (2-tailed)	.431	.852	
	N	14	14	14

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

From the result analysis in measuring the correlation between students' vocabulary and students grammar score on students' writing average scores, it is found that the correlation is as follows.

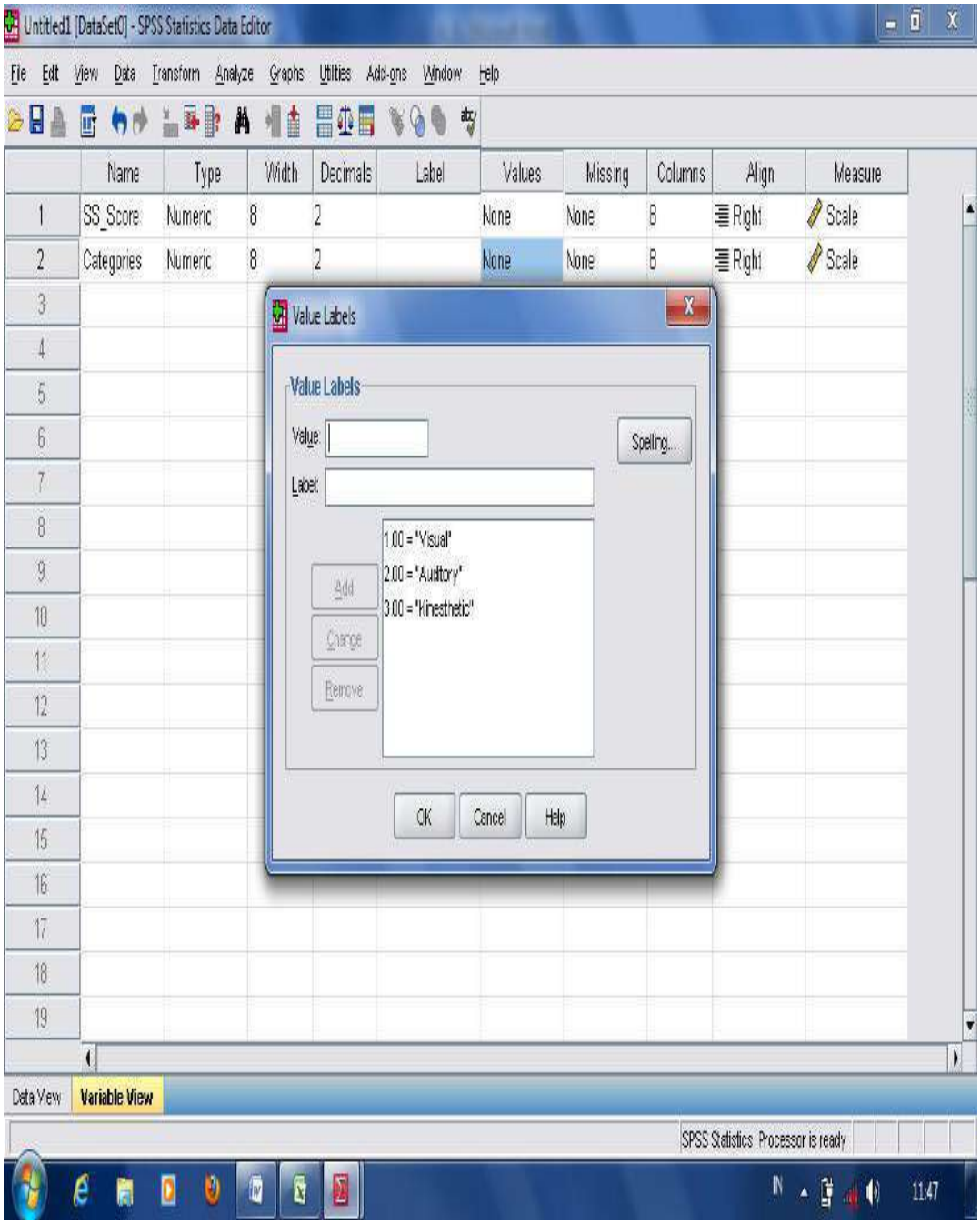
1. The correlation between students' vocabulary scores and students' writing average scores is 0.229 with significant score (2-tailed) is 0.431.
2. The correlation between students' grammar scores and students' writing average score is -0.55 with significant score (2-tailed) is 0.852.

Then, the interpretation of the significant score is concluded as follows.

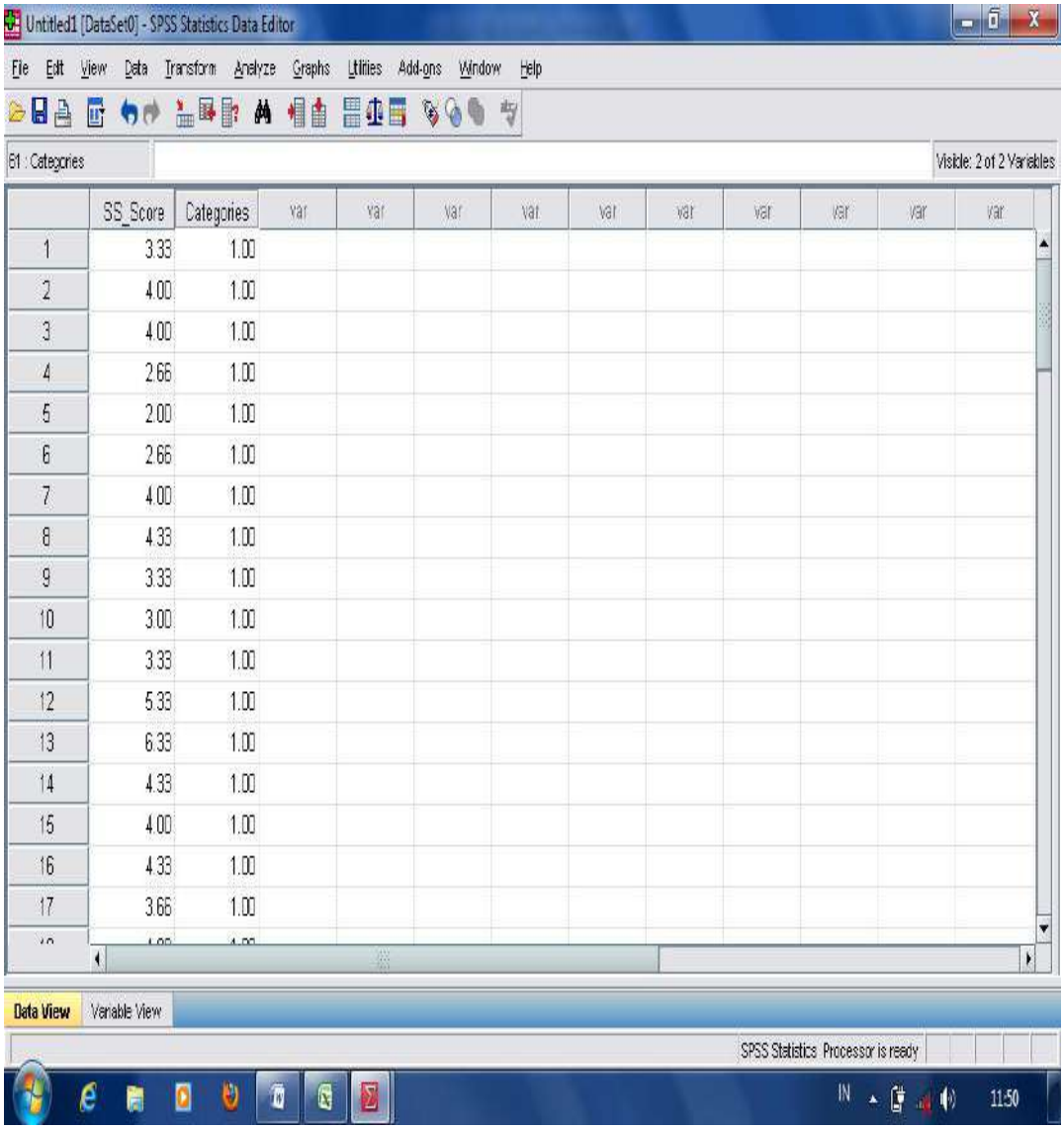
1. There is a ***middle*** correlation between students' vocabulary scores and students' writing average scores since the score of significance (0.431) is between 0.40 and 0.599.
2. There is a ***very high*** correlation between students' grammar scores and students' writing average scores since the score of significance (0.852) is between 0.80 and 1.000.

4. **Analysis of Variance (ANOVA);** is used to compare more than two means or more than two groups (Nunan, 1992, p.35). The following Analysis of Variance (ANOVA) is as ollows.
   a. **One-way ANOVA;** Hatch and Farhady (1982, p.128) state that one-way ANOVA enables to compare the means of more than two groups on one variable. From the statement, it can be assumed that one-way ANOVA is used to measure **significant difference or influence** to compare the means of more than two variables in one group of independent variable. The following is example of a research study in finding significant difference of learning styles on students' reading comprehension average score taught using KWLH strategy. There are three categories of students' learning style (visual, auditory, and kinesthetic). So, the analysis is to find significant difference of learning styles in three categories taught using KWLH reading strategy. A mean signficant difference is found whenever the p-output was lower than mean significant difference at 0.05 levels. The following are steps in analyzing significant difference using one-way ANOVA:

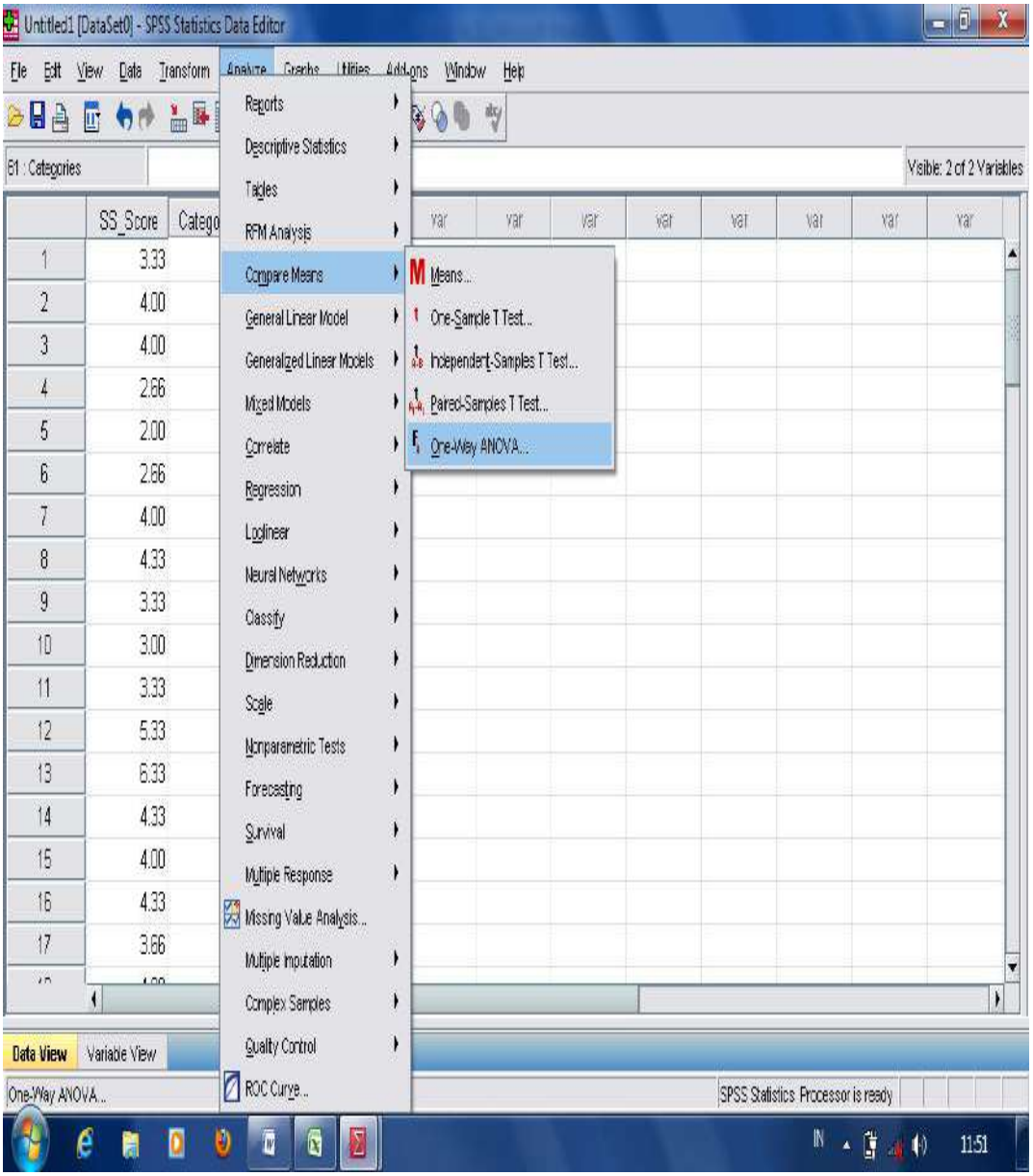
- Open the worksheet in the application of SPSS program;
- Type "**Ss_Scores**" in the first column and "**Categories**" in second column. Into the "value" box, and type 1 for "visual" in the "label" box, 2 for Auditory, and 3 for Kinesthetic, then press "add" to the column and OK.



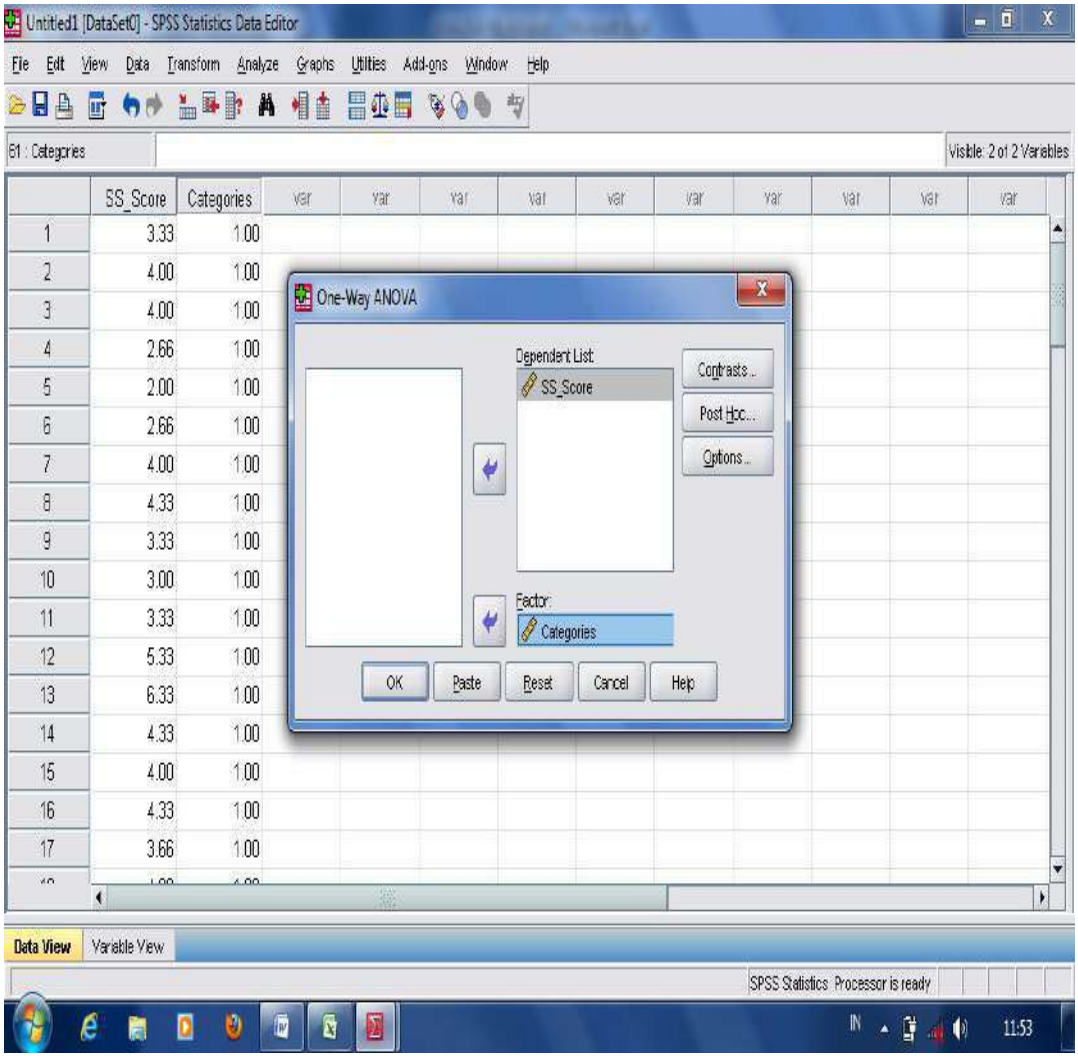
- Two columns will be seen, they are: the column of "**Ss-scores**" and "**category**". Then, input the data scores in each category;



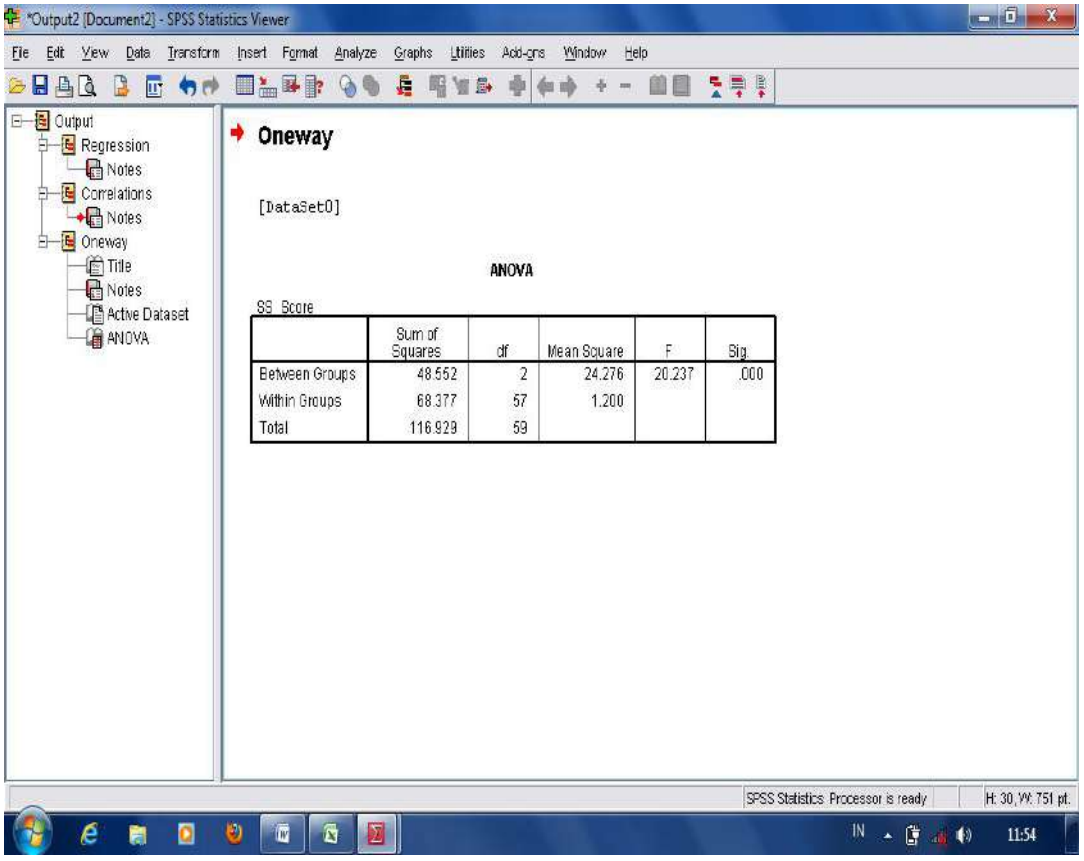
- Move the cursor into data view; Press "**Analyze"**, and "**Compare Means"**, and then, "**One-way ANOVA"**;



- Move "**Ss_score**" into dependent box, and "**category**" into factor box; then, press OK.



- The result analysis in measuring significant difference using one-way ANOVA is displayed in the following table.



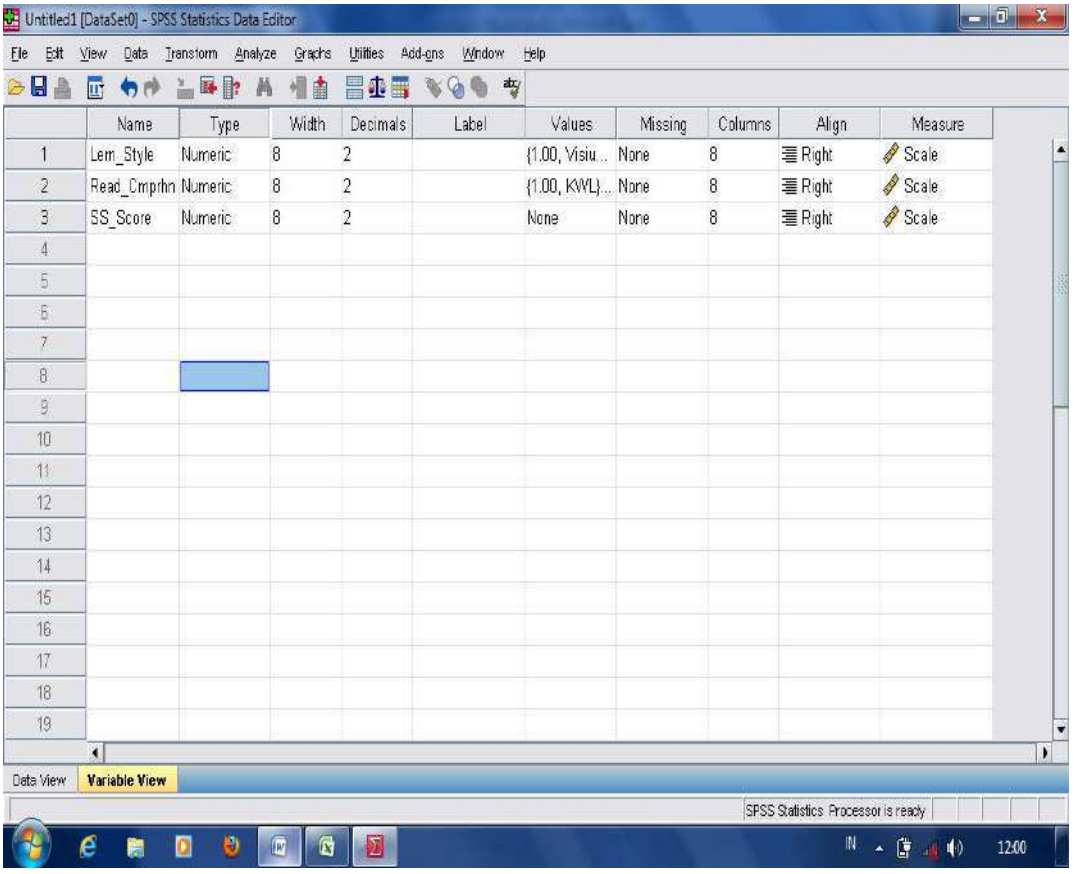
The result analysis in measuring significant difference using One-way ANOVA is found the p-output 0.000. From the result, it can be stated that there is significant difference since the p-output was lower than mean significant difference at 0.05 levels.

b. Analysis of Covariance (ANCOVA)

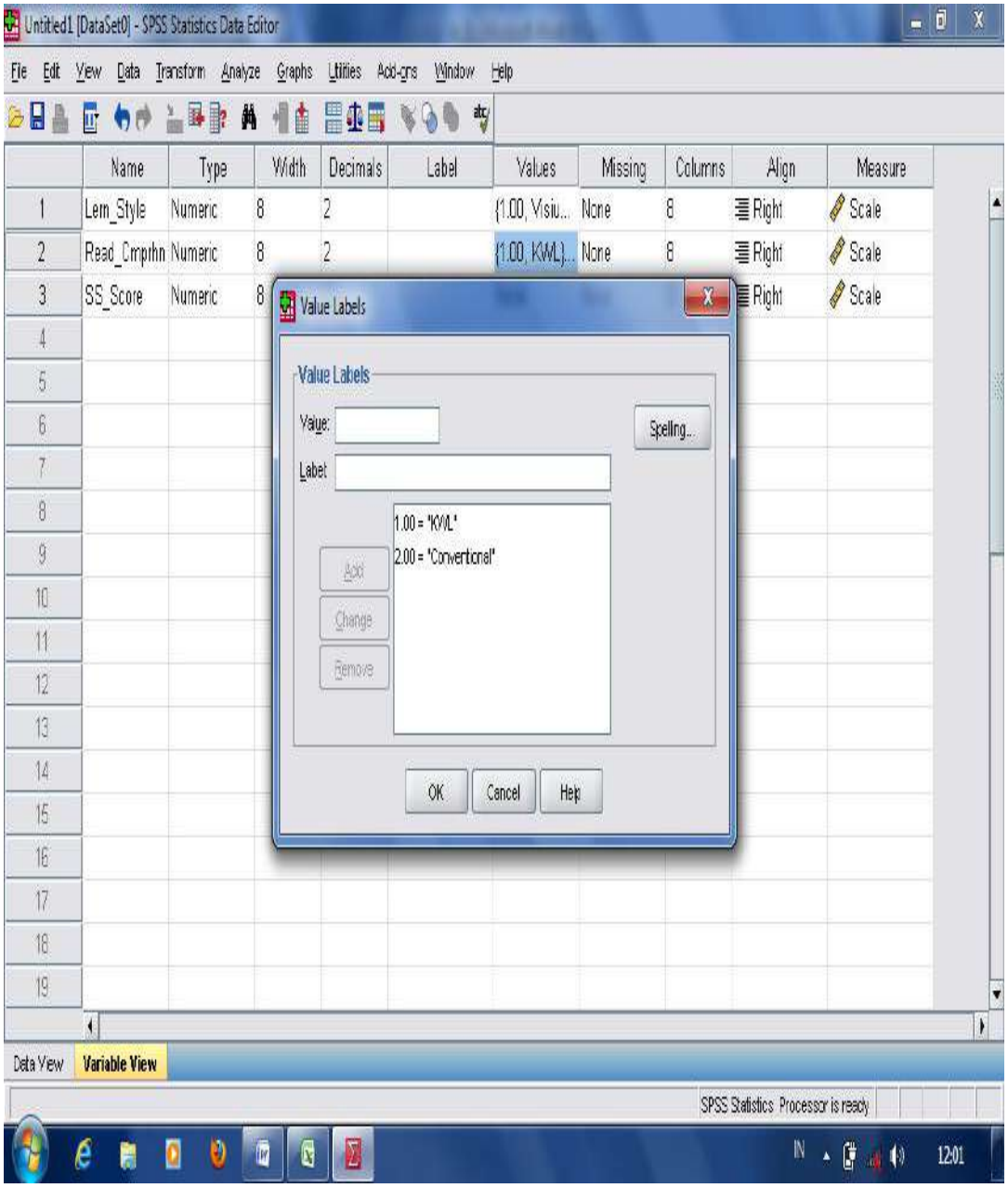
ANCOVA is used to compare the means of two or more independent variables in two independent group variables measuring ***significant different effects*.** The following is example of a reserach study which have two independent groups: **reading comprehension scores taught using two strategies** (KWLH and conventional) and **students' learning styles** (visual, auditory, kinesthetic). Then, ANCOVA compares between students' learning style in (visual, auditory, kinesthetic) categories and reading reading comprehension scores taught using two strategies (KWLH and conventional). Significant different effects are found whenever the p-output was

lower than mean significant difference at 0.05 levels. To run the statistical analysis using Analysis of Covarience (ANCOVA), the following steps can be done:

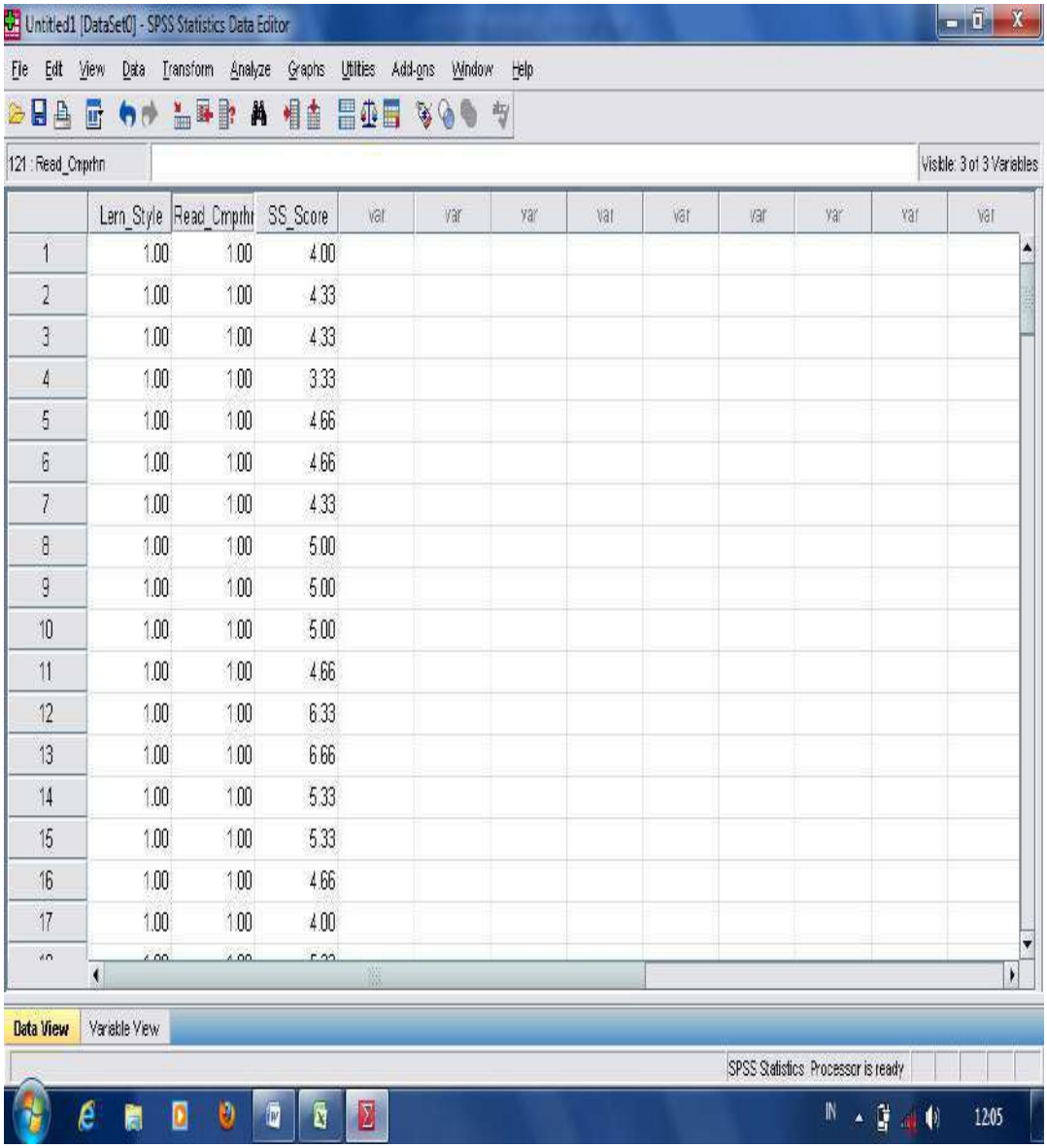
- Open the worksheet in the application of SPSS program;
- Move into "Variable View", and type Reading Strategy in the first column, Learning Styles in the second column, and "**Ss_Scores**" in the third column;



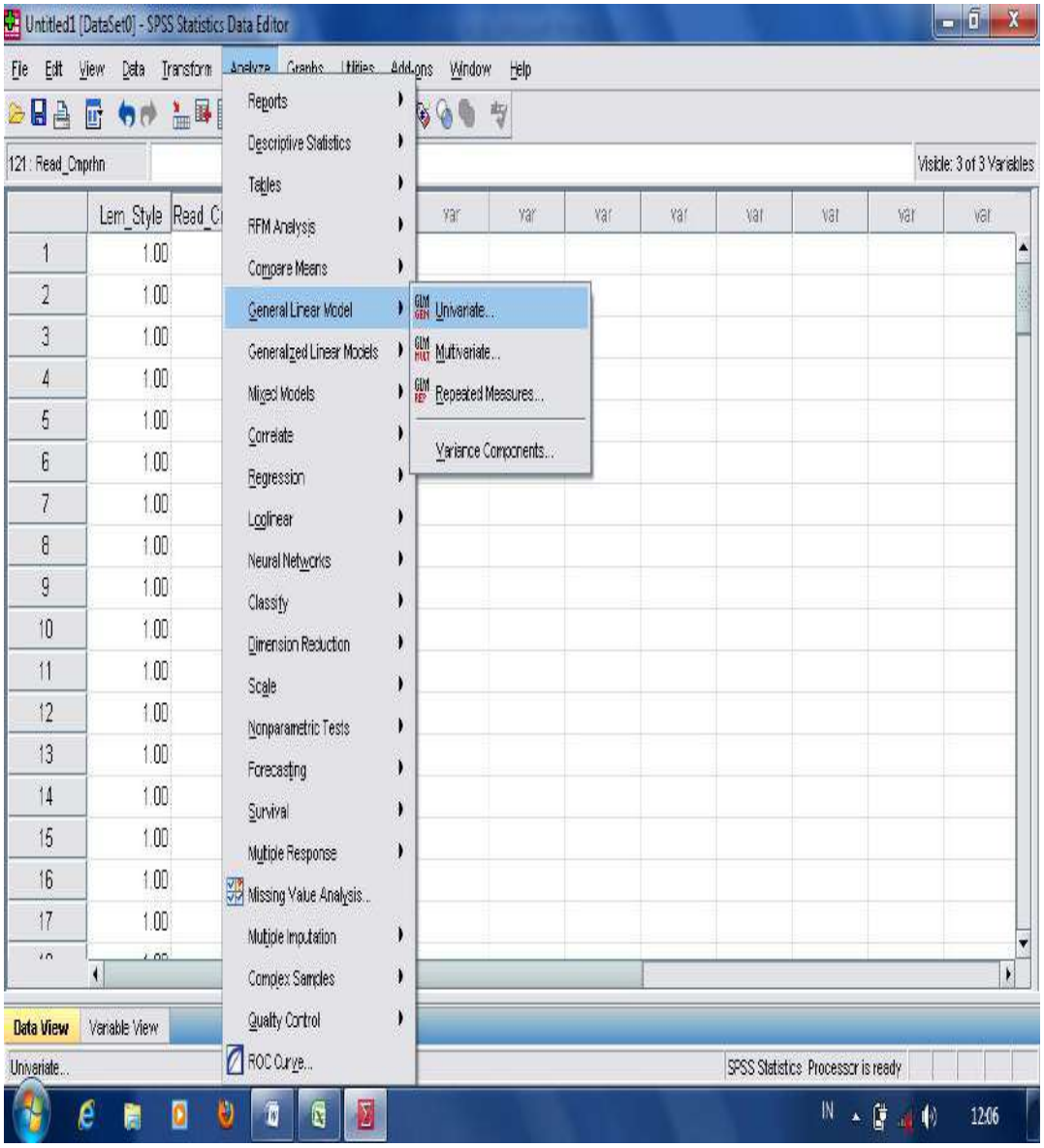
- In the value column, type 1 for KWLH strategy, and 2 for conventional strategy for the coloumn of reading strategy; Type 3 for learning styles and type each category for learning styles in the second column: 1 for visual, 2 for auditory, and 3 for kinesthetic;



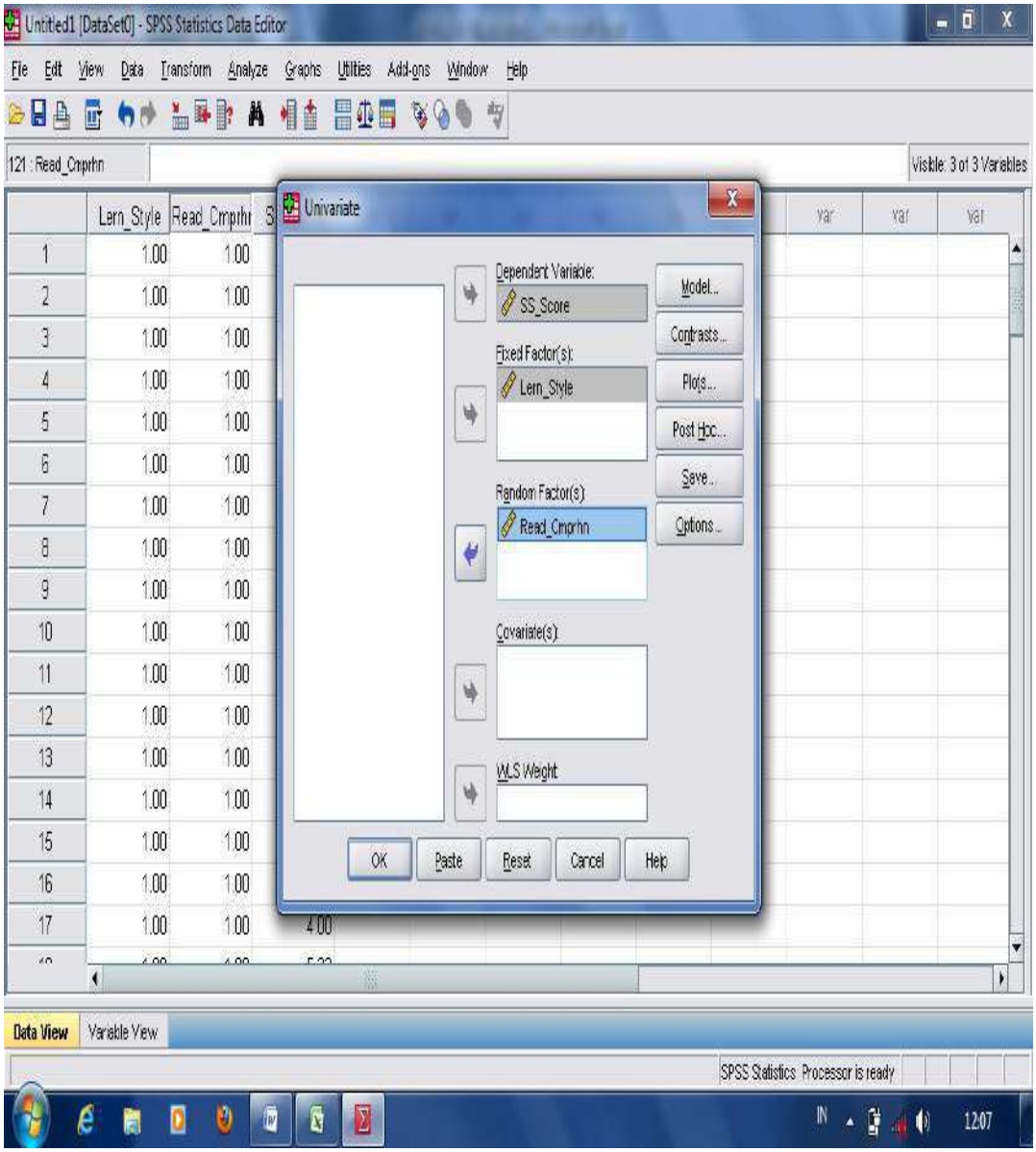
- Input the score in the first and second column, and each category in the column of "**Ss_ scores**";



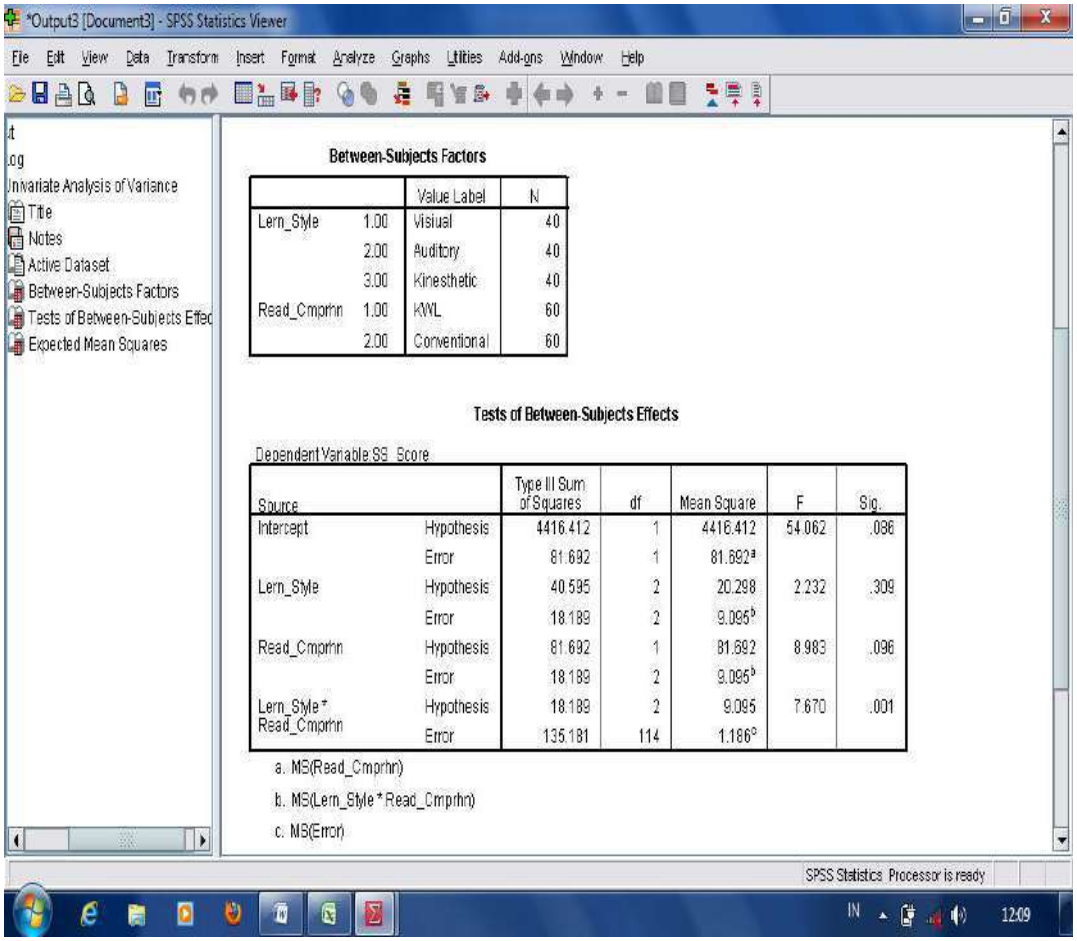
- Then, click "**analyze"**, "**general linear model**", and then choose "**Univerate**";



- Move "**Ss_Scores**" into dependent variable box, "**Lern_Styles**" into fixed factor(s) box, and "**Reading_Strgy**" into Random factor(s); Press OK.



- The result analysis in measuring significant different effects of students' learning styles on students' reading comprehension achievement taught using KWLH strategy using Analysis of Covarience (ANCOVA) is displayed in the following table.



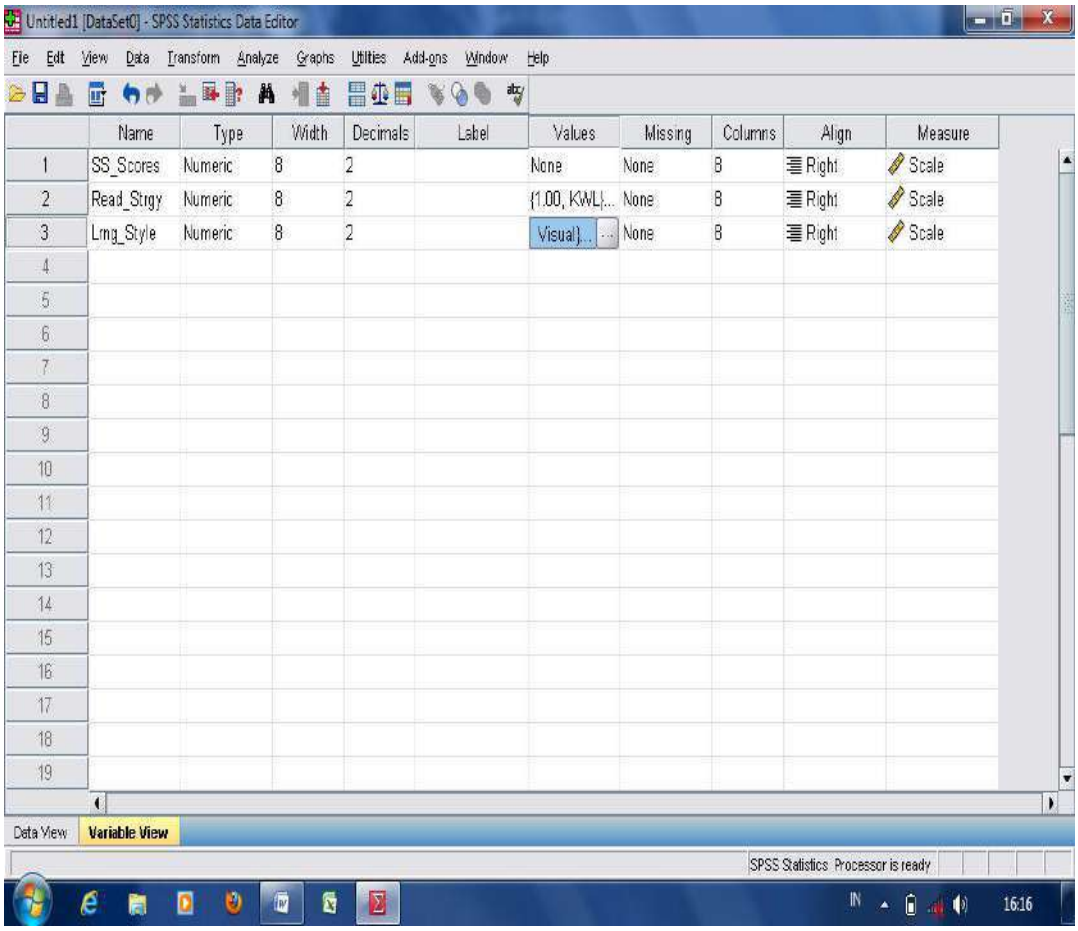
The result analysis in measuring **significant different effects** using ANCOVA found the p-output 0.001. From the score, it can be stated that there are significant different effects of students' learning styles on students' reading comprehension score taught using KWLH and conventional strategies since the p-output was lower than 0.05.

c. Two-way ANOVA

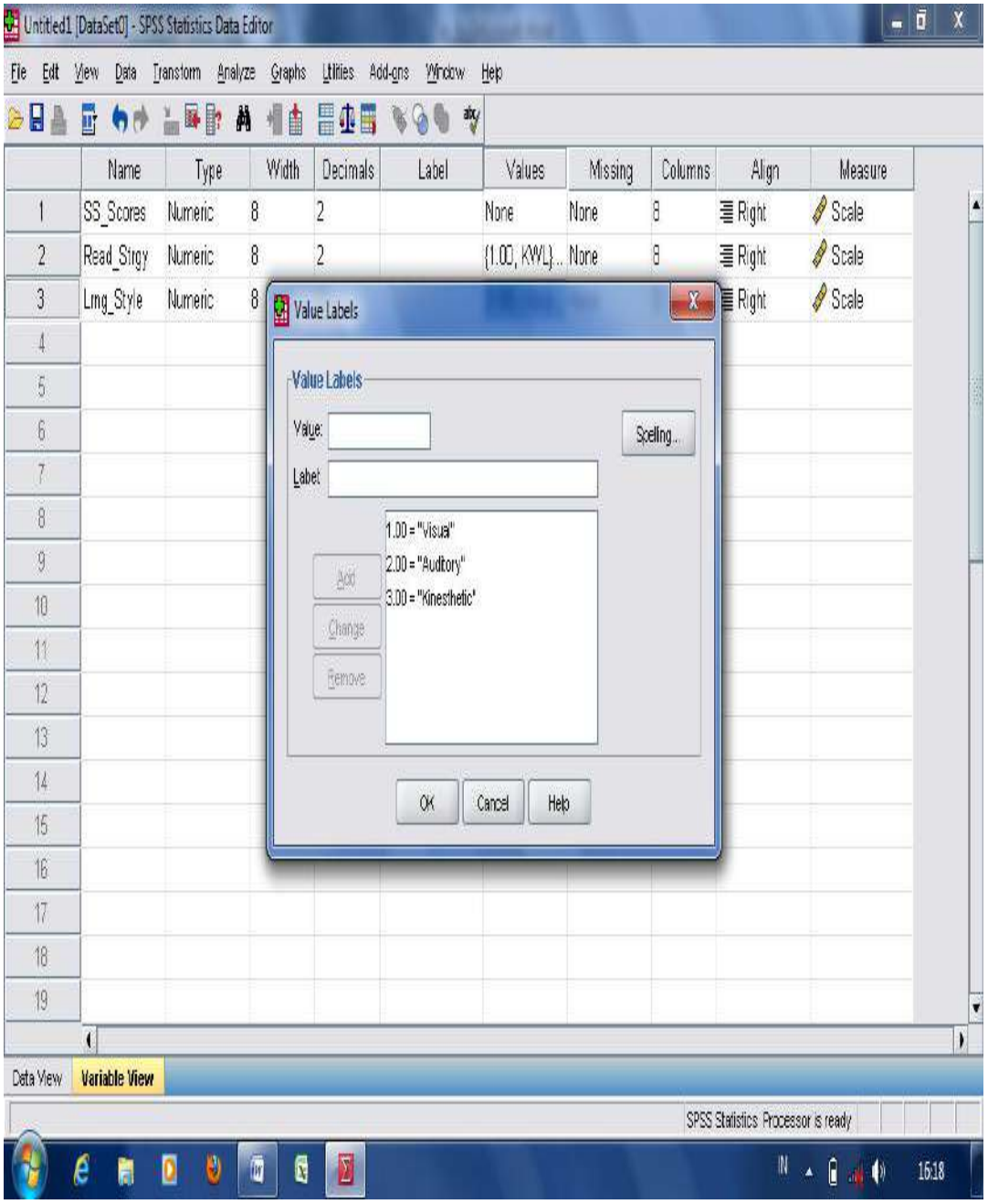
Two-way ANOVA is used to measure ***significant interaction effects*** which have more than three variables. For example: a research study finding significant interaction effects of students' learning styles in (visual, auditory, and kinesthetic) categories.on students' reading comprehension average scores taught using KWLH and conventional

strategies. Significant interaction effects are found whenever the p-output was lower than mean significant difference at 0.024 levels. The following are steps in measuring significant interaction effects using two-way ANOVA:

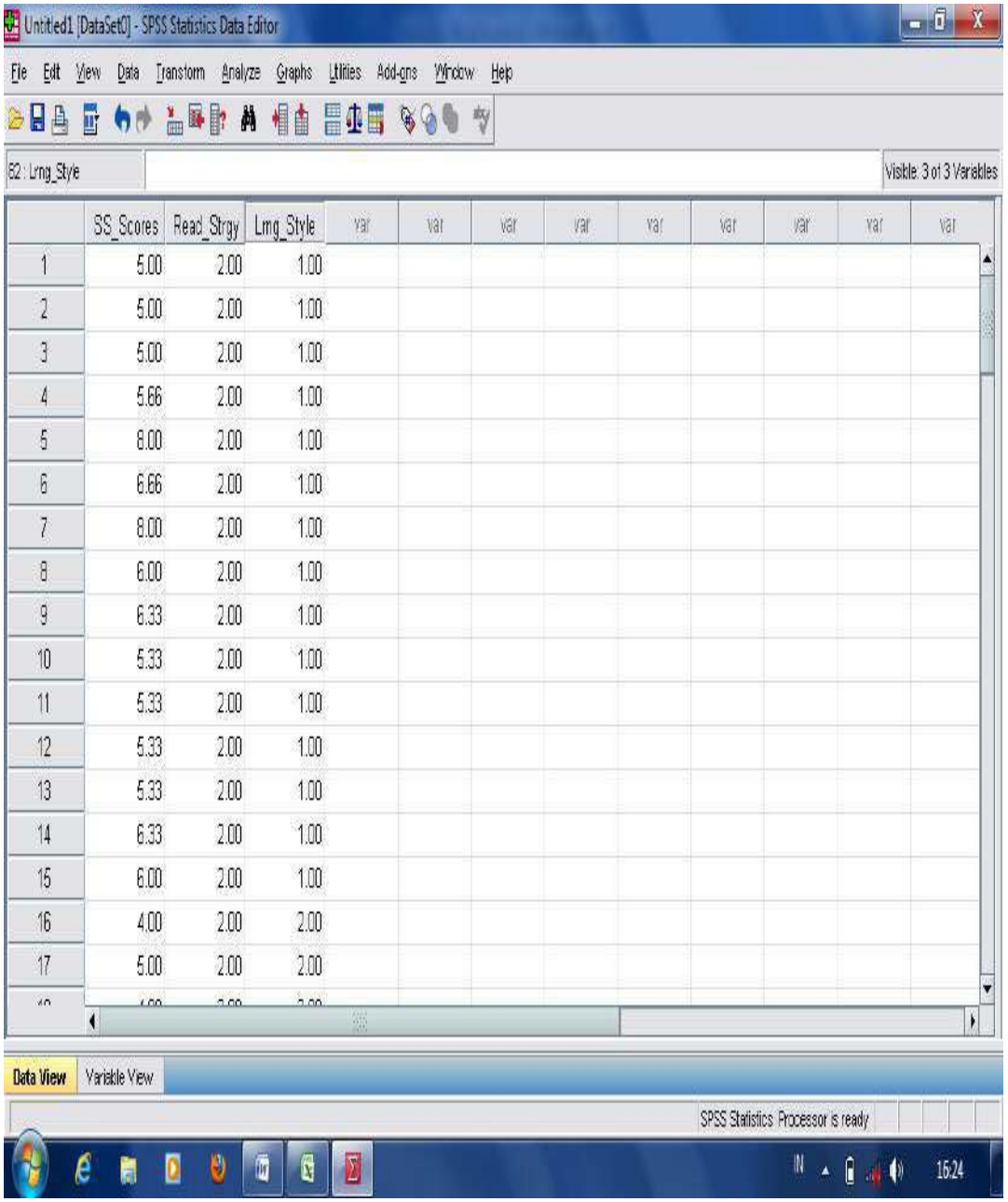
- Open the worksheet in the application of SPSS program;
- Move into "Variable View", then, type "**Ss_Scores**" in the first column, "**Reading Strategy**" in the second column, and "**Learning Styles**" in the third column.



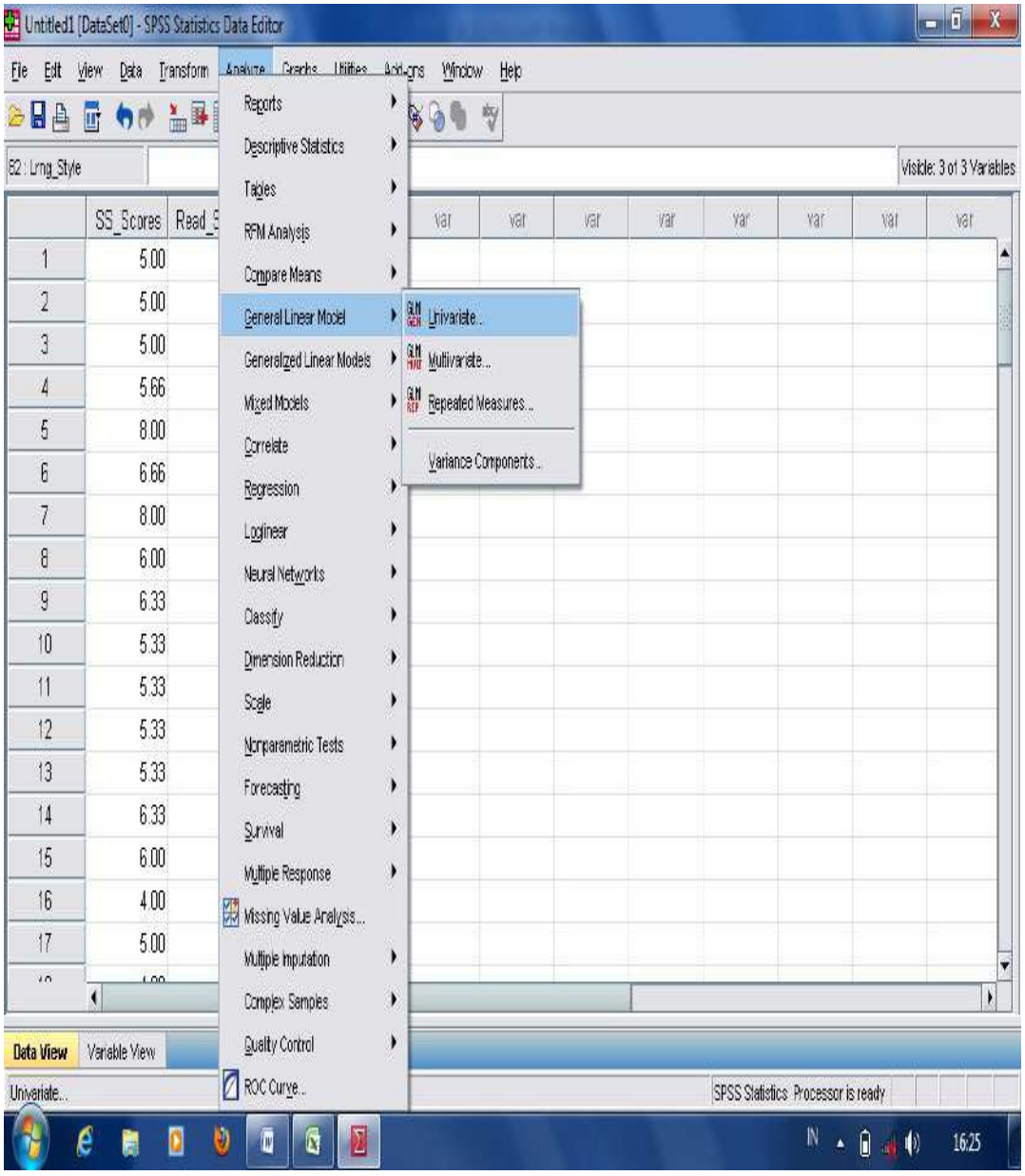
- Move into "value label" in column of Reading Strategy. Type 1 for KWLH, and 2 for Conventional, then press "add" and "OK". Then, in the third column, type "Learning Styles" with the three categories in "value label". Type 1 for Visual, 2 for Auditory, and 3 for Kinesthetic. Then, press "add" and "OK;



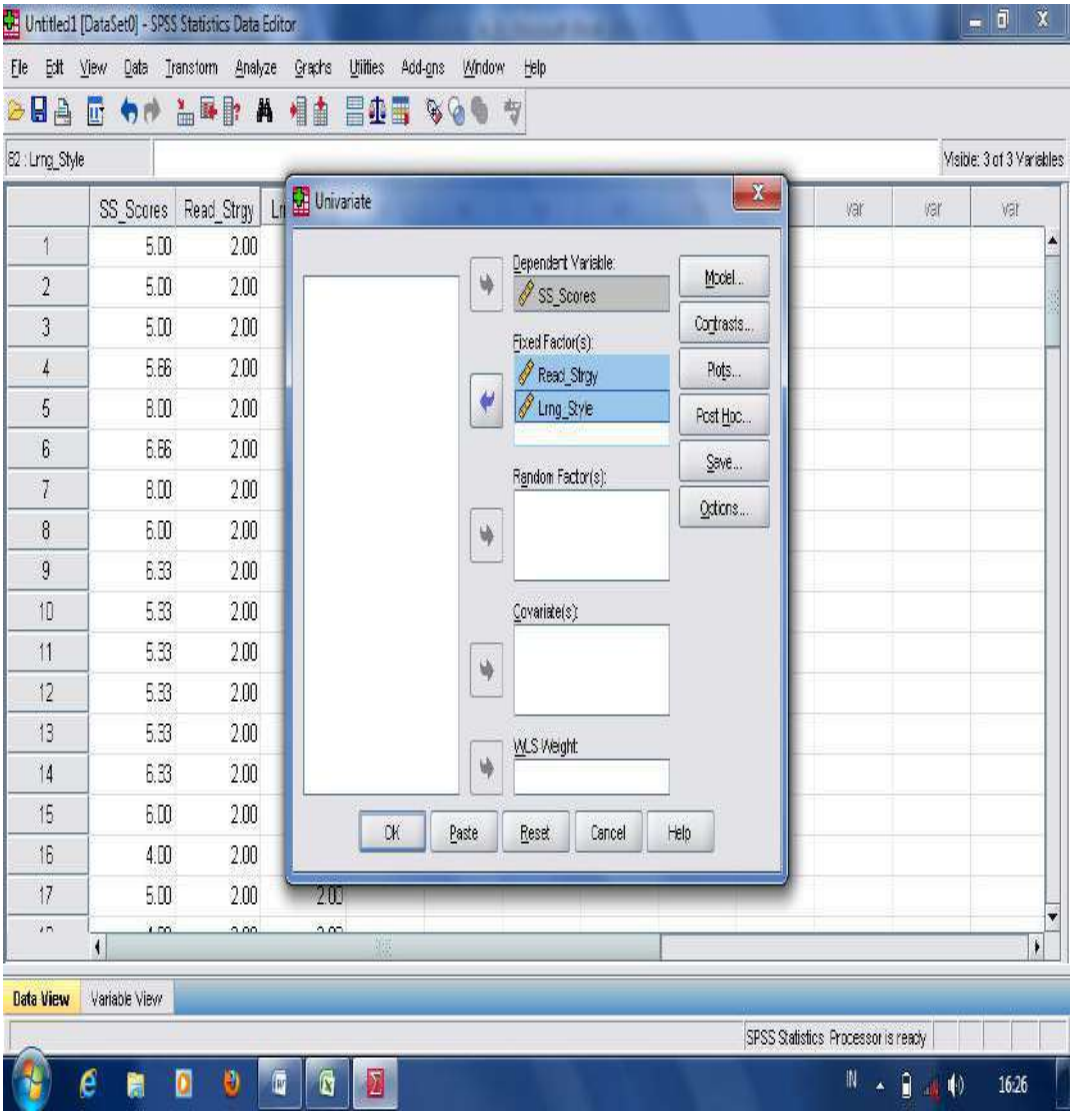
- Move into “data view” and input the obtained scores in each category;



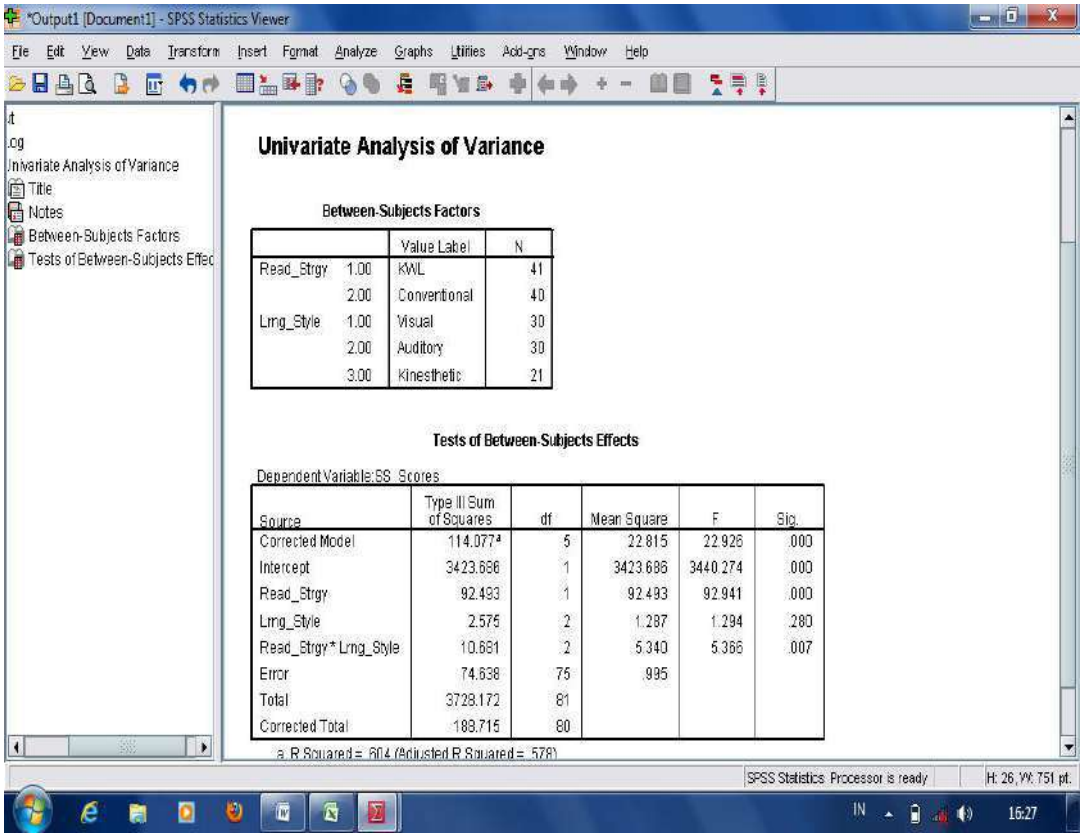
- Move to "data view", click "**Analyze"**, "**General Linear Model**", and then "**Univariate**";



- Move "**Ss_Scores**" into "dependent variable", then "**Reading Strategy**" and "**Learning Styles**" into Fixed Factor(s) column; then, press OK.



- The result analysis in measuring **significant interaction effects** using two-way ANOVA is displayed in the following table.



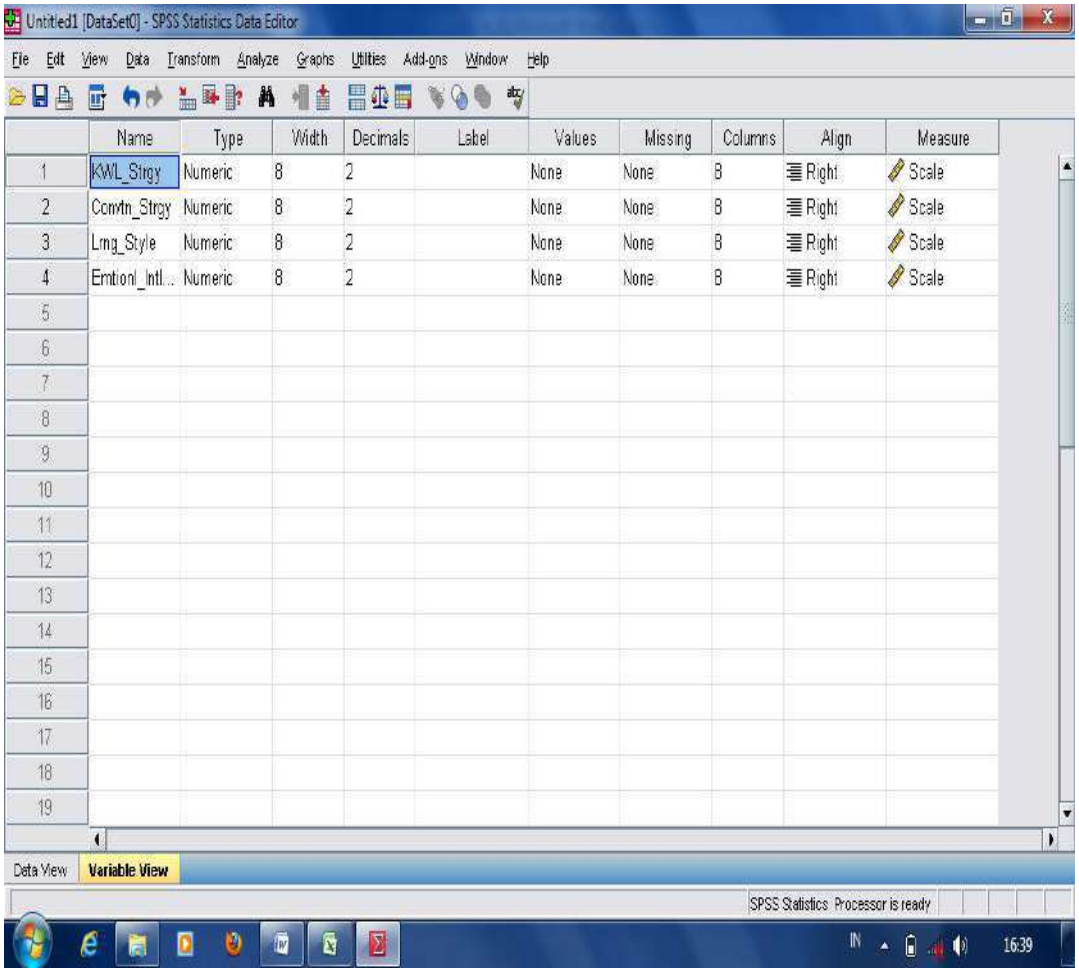
The statistical analysis in measuring significant interaction effects using two-way ANOVA found that the p-output is 0.007. From the p-output it can be stated that there are significant interaction effects of learning styles on students' reading comprehension achievement taught using KWLH and conventional strategies because the p-output was lower than 0.024.

d. Two-way MANOVA

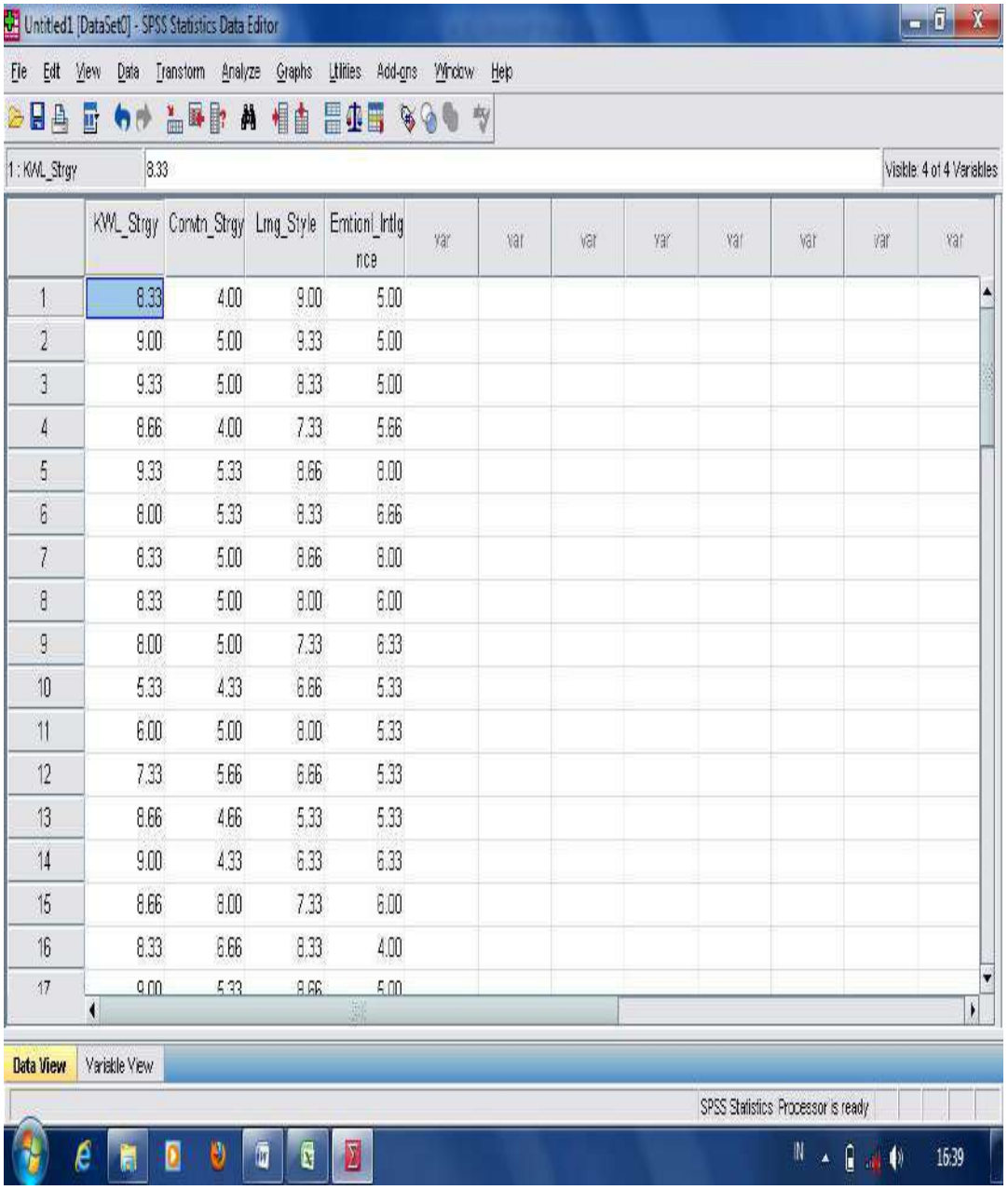
Two-way MANOVA is used to measure **significant main effects** which have more than four independent group variables**.** For example: a research study finding significant main effects of students' **learning styles** in (visual, auditory, and kinesthetic) categories and students' **emotional intelligence** in (high, middle, and low) categories on **students' reading comprehension average scores** taught using **KWLH** and **conventional** strategies. Significant main effects are found whenever the p-output was lower than mean

signficant difference at 0.05 levels. The following are steps in measuring significant main effects using two-way MANOVA:

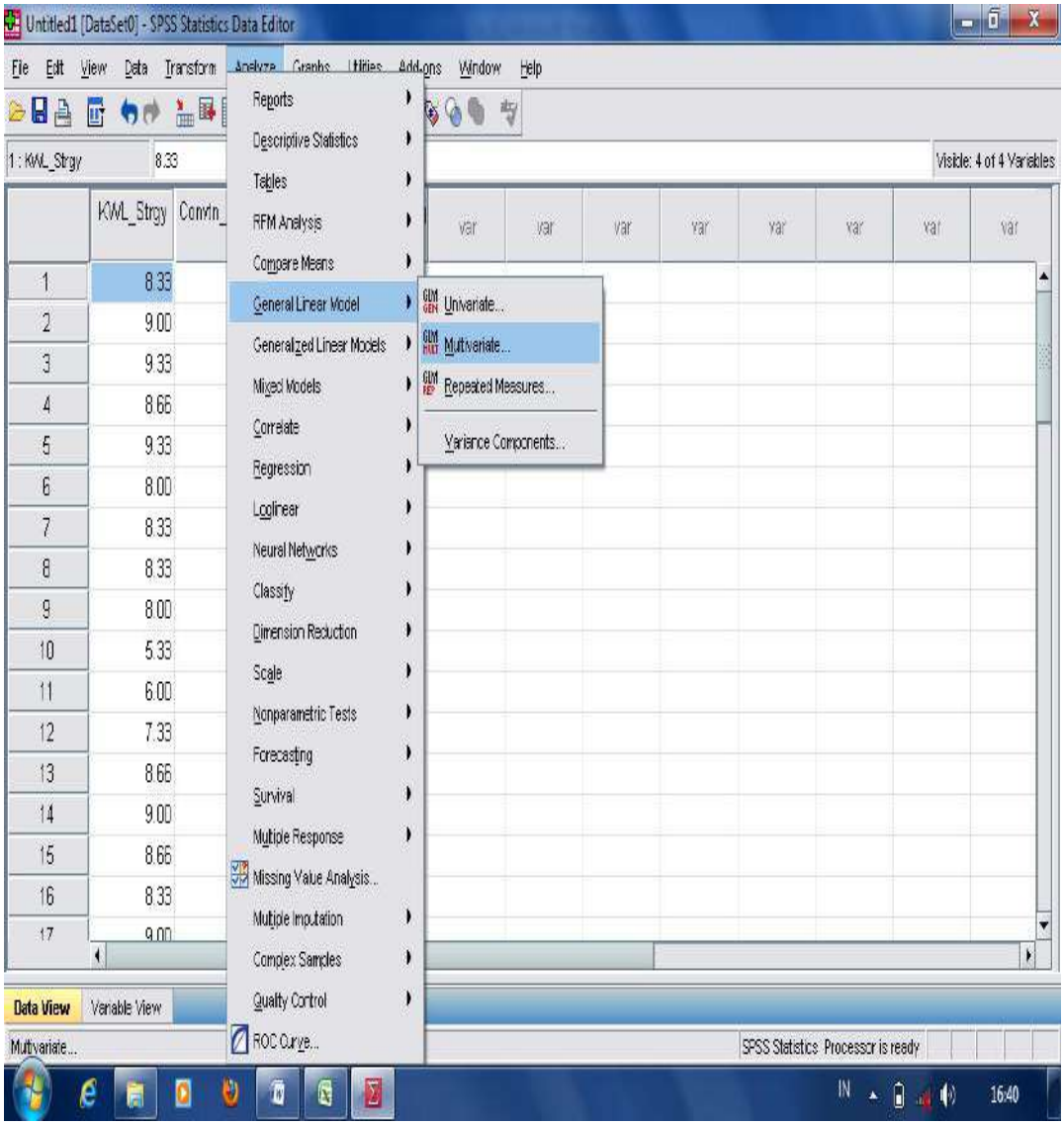
- Open the worksheet in the application of SPSS program;
- Move into "variable view" and type "**KWLH Strategy**" in the first column, "**Conventional Strategy**" in the second column, "**Learning Styles**" in the third column, and "**Emotional Intelligence**" in the fourth column;



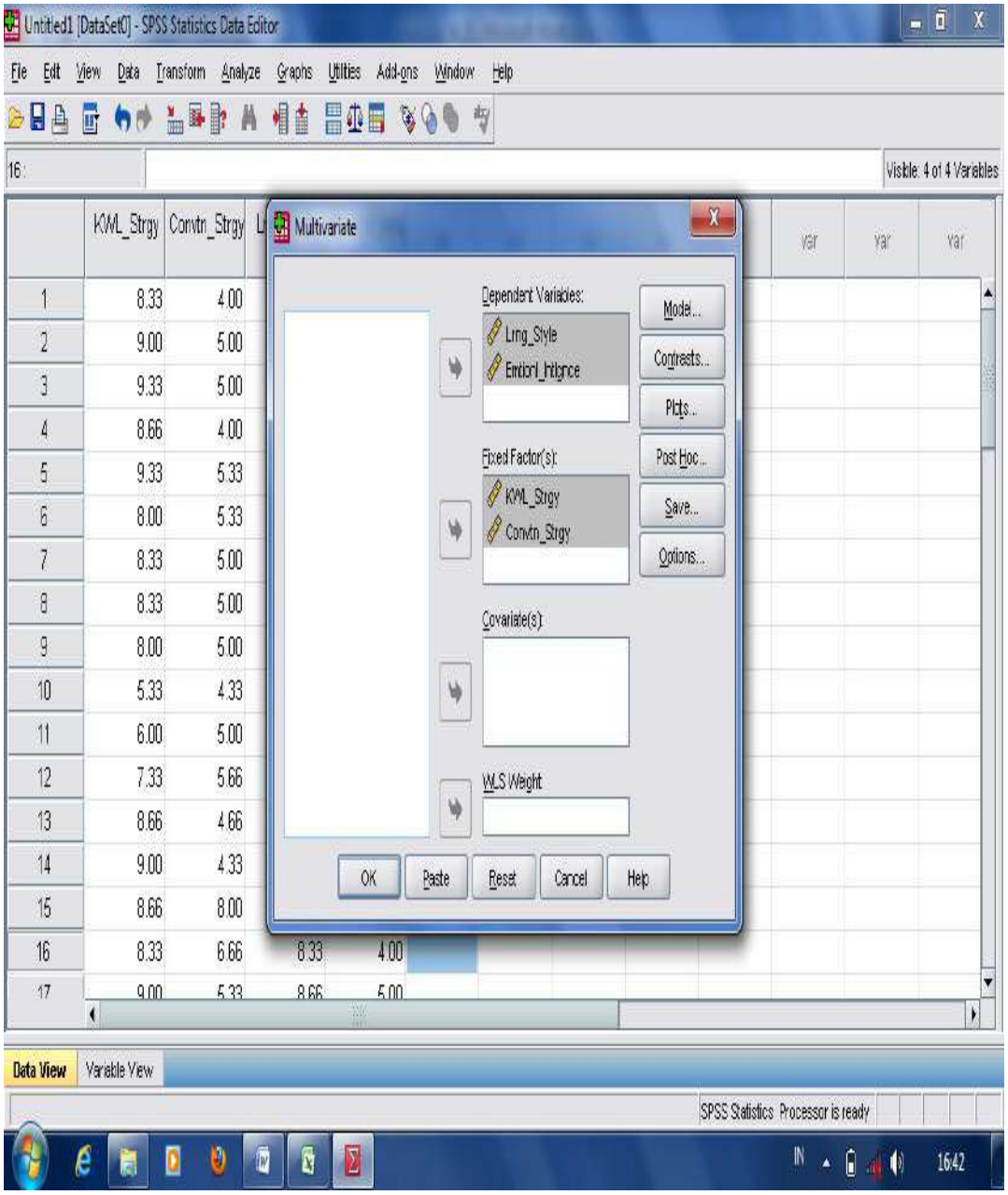
- Move the cursor into "data view", then, input the obtained score for each group of variable;



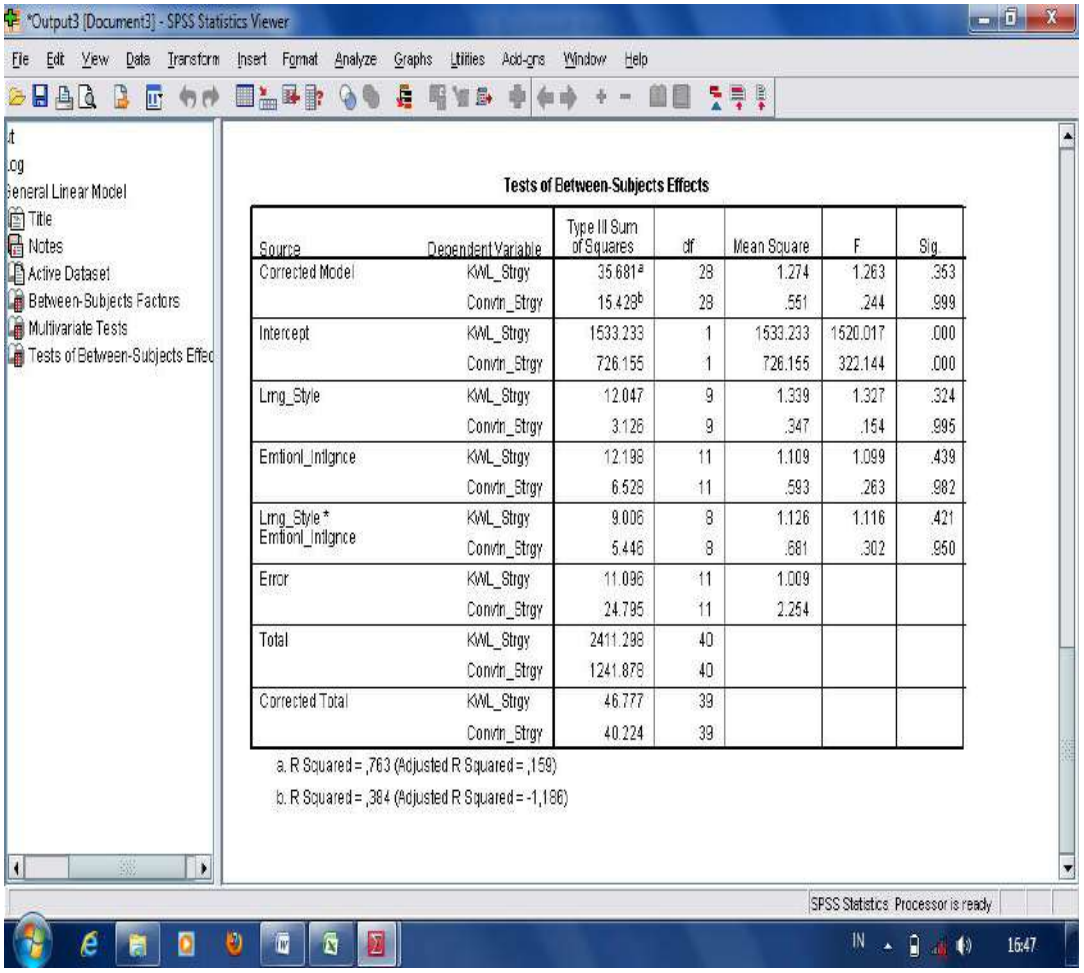
- Move into "data view", then click "**Analyze**", "**General Linear Model"**, and "**Multivariate**";



- Move the variable data "**Learning Styles**" and "**Emotional Intelligence**" into dependent variables. Then, "**KWLH**" and "**Conventional**" into the column of "Fixed Variables"; and press OK.



- The result analysis in measuring significant main effects using two-way MANOVA is displayed in the following table.



Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	KWL_Strgy	35.681[a]	28	1.274	1.263	.353
	Convtn_Strgy	15.429[b]	28	.551	.244	.999
Intercept	KWL_Strgy	1533.233	1	1533.233	1520.017	.000
	Convtn_Strgy	726.155	1	726.155	322.144	.000
Lrng_Style	KWL_Strgy	12.047	9	1.339	1.327	.324
	Convtn_Strgy	3.126	9	.347	.154	.995
Emtionl_Intlgnce	KWL_Strgy	12.198	11	1.109	1.099	.439
	Convtn_Strgy	6.528	11	.593	.263	.982
Lrng_Style * Emtionl_Intlgnce	KWL_Strgy	9.006	8	1.126	1.116	.421
	Convtn_Strgy	5.446	8	.681	.302	.950
Error	KWL_Strgy	11.096	11	1.009		
	Convtn_Strgy	24.795	11	2.254		
Total	KWL_Strgy	2411.298	40			
	Convtn_Strgy	1241.878	40			
Corrected Total	KWL_Strgy	46.777	39			
	Convtn_Strgy	40.224	39			

a. R Squared = ,763 (Adjusted R Squared = ,159)

b. R Squared = ,384 (Adjusted R Squared = -1,186)

The result of statistical analysis measuring significant main effects of learning style and emotional intelligence on students reading comprehension score taught using KWLH strategy found the p-output 0.421. While, significant main effects of learning styles and emotional intelligence on students' reading comprehension average scores who are taught using conventional strategy is 0.950. It means that there are no significant main effects of learning styles and emotional intelligence on students' reading comprehension taught using KWLH and conventional strategies since the scores of significance are higher than 0.05.

K. Two-Tailed and One-Tailed Tests of Significance

If a null hypothesis is proposed that there is no difference (other than in sampling error) between the mean of new and conventional strategies, we would be concerned only with a difference and not with the superiority or inferiority of either group, **a two-tailed test of significance** was applied (Best and Kahn, 1993, p.332). The following statement is an example in deteriming a two-tailed test.

- There is **no means significant difference** on students' reading comprehension achievements taught using H-5 and conventional strategies.
- There is **a means significant difference** on students' reading comprehension achievements taught using H-5 and conventional strategies.

However, when the null hypothesis is changed to indicate the superiority or inferiority of either group or a direction of difference rather than the mere existence of a difference, a **one-tailed test of significance** is used (Best and Kahn, 1993, p.332). It might be stated as in the following statements.

- Means paired differences on students' reading comprehension scores taught using KWLH strategy **is higher** than conventional strategy.
- Means paired differences on students' reading comprehension scores taught using KWLH strategy **is lower** than conventional strategy.

The following table is a large sample *t* critical values in determining a two-tailed test or a one-tailed test of significance for rejection or acceptance of the null hypothesis.

Table 8

Sample *t* Critical Values in Determining a Two-tailed Test or a One-tailed Test of Significance

Df	Two-tailed test Level of significance		One-tailed test Level of significance	
	0.05	*0.01*	*0.05*	*0.01*
1	12.706	63.557	6.314	31.821
2	4.303	9.925	2.920	6.965
3	3.182	5.841	2.353	4.541
4	2.776	4.604	2.132	3.747
5	2.571	4.032	2.015	3.365
6	2.447	3.707	1.943	3.143
7	2.365	3.499	1.895	2.998
8	2.306	3.355	1.860	2.896
9	2.262	3.250	1.833	2.821
10	2.228	3.169	1.812	2.764
11	2.201	3.106	1.796	2.718
12	2.179	3.055	1.782	2.681
13	2.160	3.012	1.771	2.650
14	2.145	2.977	1.761	2.624
15	2.131	2.947	1.753	2.602
16	2.120	2.921	1.746	2.583
17	2.110	2.898	1.740	2.567
18	2.101	2.878	1.734	2.552
19	2.093	2.861	1.729	2.539
20	2.086	2.845	1.725	2.528
21	2.080	2.831	1.721	2.518
22	2.074	2.819	1.717	2.508
23	2.069	2.807	1.714	2.500
24	2.064	2.797	1.711	2.492
25	2.060	2.787	1.708	2.485
26	2.056	2.779	1.706	2.479
27	2.052	2.771	1.703	2.473

28	2.048	2.763	1.701	2.467
29	2.045	2.756	1.699	2.462
30	2.042	2.750	1.697	2.457
40	2.021	2.704	1.684	2.423
60	2.000	2.660	1.671	2.390
120	1.980	2.617	1.658	2.358
∞	1.960	2.576	1.645	2.326

Source: Best and Kahn. (1993, p.399). *Research in Education (7th Edition*). Boston: Allyn and Bacon.

L. Experimental Research Designs

In this part, two kinds of experimental research designs are presented. They are ***true*** and ***quasi*** experimental designs. The following illustration generally describes the differences between ***true*** and ***quasi***-experiments:

- "In a ***true***-experiment, participants are randomly assigned to either the treatment or the control group, whereas they are not assigned randomly in a quasi-experiment". From previous the statement, it can be assumed that in a ***true***-experiment, the research samples are selected using a ***random sampling*** method, while in a ***quasi***-experiment, the samples are chosen using a ***non-random sampling*** method.
- "In a ***quasi***-experiment, the control and treatment groups differ not only in terms of the experimental treatment they receive, but also in other, often unknown or unknowable, ways. Thus, the researcher must try to statistically control for as many of these differences as possible". From the previous statement, it can be interpreted that there are two groups applied in a ***quasi***-experiment, they are control and experimental groups. Both groups are given different treatments. In control group, the treatment is given using a traditional/conventional teaching manner, while in experimental group, it is given using new or non-traditional/conventional teaching manner. And it compares the differences of the obtained scores in two groups.
- "Because control is lacking in ***quasi***-experiments, there may be several "rival hypotheses" competing with the experimental manipulation as

explanations for observed results". From the previous statement, it can be associated that in a ***quasi***-experiment, the hypotheses are constructed in two contrary statement of hypotheses. They are negative statement of hypothesis which is formulated in null hypothesis (H*o*), and positive statement of hypothesis formulated in alternative hypothesis (H*a*).

The following are further descriptions on the two kinds of experimental designs found in research study. They are explained further as follows.

1. **True-Experimental Design**

   Fraenkel and Wallen (1990, p.237) state that the essential ingredient of a true experimental design is that subject are randomly assigned to treatment group. In other words, true experiment can be defined as one in which the researcher manipulates the independent variable (or variables) to observe its effect on some behavior or cognitive process (the dependent variable) while using random assignment of participants to groups in order to control external factors from influencing the results. Furthermore, it is often claimed that true experiment is the only research method that can adequately measure the cause and effect relationship. From the previous statements, it can be classified that **correlational** and **regression** studies which estimate the causal relationship and the effects between two or more variables are categorized in a true experimental study. Then, it is also explained that there are three group designs in a true experimental study. They are randomized posttest-only control group design, randomized pretest posttest control group design, and factorial design.

   a. **Randomized Pottest-Only Control Group Design**

      The posttest only control group design involves two groups, both of which are formed by random assignment. One group receives the experimental treatment while the other does not and then both groups are given posttest on the dependent variable. A diagram of this group design is suggested by Fraenkel, et.at. (2012, p.271) are as follows.

Treatment :	**R**	**X**	**O**
Control	**: R**	**C**	**O**

Where:

R : Random Sampling Method

O : Measurement of the dependent variable

X : Treatment in experimental group with a new or non-traditional teaching strategy

C : Control group

In this design, each group of sample is selected randomly (R). The first group is given treatment (X), while the other group (C) is not. The treatment group is given only to the experimental group. Both groups are only given posttest. In this research of study, the effect of treatment is measured using t-test to obtain a mean significant difference.

b. Randomized Pretest-Posttest Control Group Design

The pretest-posttest control group design differs from the posttest only control group design. Two groups of subjects are used, with both groups being measured or observed twice. The first measurement serves as the pretest, the second one as the posttest. Random assignment is used to form the groups. The measurements or observations are collected at the same time for both groups. A diagram design applying **The Randomized Pretest-Posttest Control Group Design** is suggested by Fraenkel, eat.al., (2012, p.272) are as follows.

Treatment : R O X O
Control : R O C O

Where:

R : Random Sampling Method

O : Measurement of the dependent variable

X : Treatment in experimental group with a new or non-traditional teaching strategy

C : Treatment in control group with a traditional teaching strategy

Two groups are selected randomly, and then the two groups (control and experiment) are given pretest to know the effects before giving some treatments. Then, treatments are given to both groups (control and experiment) with two different teaching strategies. At the end, to know the effect after some treatements are done, posttest is given.

c. Factorial Design

Fraenkel and Wallen (1990, p.245) mention that factorial design extends the number of relationships that may be examined in an experimental study. This design is a modification of **the pretest-posttest control group design**. It involves one treatment and one control group, and a moderator variable having three levels ($Y1$, $Y2$ and $Y3$). For example: a research title findng the significance of emotional intelligence which has three category levels (high, middle, and low) on students' reading comprehension achievement taught using KWLH and conventional strategies. The diagram of factorial design is suggested by Fraenkel, et.at (2012, p.277) are as follows.

Control	**:**	**R**	**O**	**X**	**Y_1**	**O**
Treatment	**:**	**R**	**O**	**C**	**Y_1**	**O**
Control	**:**	**R**	**O**	**X**	**Y_2**	**O**
Treatment	**:**	**R**	**O**	**C**	**Y_2**	**O**
Control	**:**	**R**	**O**	**X**	**Y_3**	**O**
Treatment	**:**	**R**	**O**	**C**	**Y_3**	**O**

O : Measurement of the dependent variable
X : Treatment in experimental group with a new or non-traditional teaching strategy
C : Treatment in control group with a traditional teaching strategy
Y_1 : High emotional intelligence category
Y_2 : Middle emotional intelligence category
Y_3 : Low emotional intelligence category

All group samples are selected randomly, and then both groups are given pretest. After that, the treatments are given to both groups (experiment and control). At the end, posttest are given to know the effect of the given treatments (KWLH and conventional) reading strategies towards moderator variables on emotional intelligence questionnaire (high, middle, and low) categories. Other diagram related to factorial design is described in the following table.

Reading Strategy	Emotional Intelligence		
	High	Middle	Low
KWLH	-	-	-
Conventional	-	-	-

2. Quasi Experimental Design

Quasi experimental designs do not include the use of random in selecting sample of the research. In this design, the researchers use a non-random sampling method to select the sample of the study. In other words, it can be stated when it is not possible or practical to control all the key factors, so it becomes necessary to implement a quasi-experimental research design. Many research designs in quasi experiment are described as follows.

a. Matching Only Posttest-Only Control Group Design; the researcher matches the subject in the experimental and control groups on certain variables, but he or she has no assurance that they are equivalent on others. One group receiving one treatment are matched with individuals receiving the other treatments. A diagram of matching only posttest-only control group design is suggested by Fraenkel, et.at. (2012, p.275) are as follows.

Treatment : M X O

Control : M C O

Where:

M : Subjects in each group have been matched (on certain variables) but not randomly assigned to the groups.

O : Measurement of the dependent variable

X : Treatment in experimental group with a new or non-traditional teaching strategy

C : Treatment in control group with a traditional teaching strategy

In this design, each group of sample is selected with non random sampling method. Then, the subject in each group has been matched on certain variables. The first group is given treatment (X), while the other group (C) is not. The treatment group is given only to the experimental group. Both groups are only given posttest at the end of treatment. In this research of study, the effect of treatment is measured using t-test to obtain a mean significant difference.

b. Nonequivalent Groups Posttest-Only (Two or More Groups)
In the nonequivalent groups posttest-only design, one group (the experimental group) receives the intervention while the other group (the control group) does not, as depicted in the following design (Marczyk, et.al., 2005, p.138).

NR—X1—O
NR—X2—O

Where:

NR	:	Non-random sampling
X1	:	Treatment in experimental group
X2	:	Treatment in control group
O	:	Posttest

c. Matching Only Pretest-Posttest Control Group Design; the researcher matches the subject in the experimental and control groups on certain variables where both groups receive different treatment. The figure of Matching Only Pretest-Posttest Control Group Design is suggested by Fraenkel and Wallen (1990, p.243) as follows.

Treatment	**: O**	**M**	**X1**	**O**
Control	**: O**	**M**	**X2**	**O**

Where:

O : Measurement of the dependent variable

M : Subjects in each group have been matched (on certain variables) but not randomly assigned to the groups.

X1 : Treatment in experimental group with a new or non-traditional teaching strategy

X2 : Treatment in control group with a traditional teaching strategy

Two groups are selected with non-random sampling method. Then, the subjects in each group have been matched on certain variables after pretest is given. The first group is given treatment (X1), while the other group (X2) is given different treatment. Both groups are given posttest at the end of treatment. In this research of study, the effect of treatment is measured using paired sample t-test to obtain means significant improvement or means paired differences.

d. Pretest-Posttest Non-equivalent Group Design

One of the most commonly used quasiexperimental designs in educational research is pretest-posttest nonequivalent group design suggested by Cohen, et.al. (2007, p.283) as follows.

Treatment	**:**	**O1**	**X**	**O2**
Control	**:**	**O3**		**O4**

Where:

O1,3 : Pretest in experimental and control groups

X : Treatments in Experimental group

O2,4 : Posttest in experimental and control groups

------- : The dashed line separating the parallel rows in the diagram of the non-equivalent control group indicates that the experimental and control groups

have not been equated by randomization–hence the term ‘non-equivalent’.

e. Pretest-Posttest Nonequivalent-Groups Design

This design provides control of when and to whom the measurement is applied with a non-random assignment to experiement and control treatment. The pretest-posttest nonequivalent groups design is suggested by Best and Khan (1993, p.151) as follows.

O1	**X**	**O2**
O3	**C**	**O4**

Where:

O1,3 : Pretests
O3,4 : Posttest
X : Treatment in experimental group
C : Treatment in control group

f. Nonequivalent Groups Pretest-Posttest (Two or More Groups)

In the nonequivalent groups pretest-posttest design, the dependent variable is measured both before and after the treatment or intervention. The following is the design of nonequivalent groups pretest-posttest (two or more groups) suggested by Marczyk, et.al., (2005, p.139).

NR—O—X1—O
NR—O—X2—O

Where:

NR : Non-random sampling

O : Pretest and Posttest
X1 : Treatment in experimental group
X2 : Treatment in control group

M. Matching Variables

After selecting sample of the study using non-random sampling method, both groups are given pretest to formulate the existing sample into a group of matching variables. It is related to Fraenkel and Wallen (1990, p.243) that the subjects in each group have been matched on certain variables, but not randomly assigned to the groups. After matching variables have been done, the total number of selected sample is divided into two groups; the first group is treated as control group (taught using conventional strategy), and the second group is treated as experimental group (taught using non conventional or new teaching strategy). After some treatments are given to both groups with a different teaching strategy, posttest is given to know the effects of treatments in both groups.

Furthermore, Fraenkel, et.at., (2012, p.273) mention that matching variables may be done in either or both of two ways: mechanically or statistically. Both require a score for each subject on *each* variable on which subjects are to be matched.

a. **Mechanical matching;** is a process of pairing two persons whose scores on a particular variable are similar. Two girls, for example, whose mathematics aptitude scores and test anxiety scores are similar might be matched on those variables. After the matching is completed for the entire sample, a check should be made to ensure that the two groups are indeed equivalent on each matching.Figure 1 illustrates the process of mechanical matching variables from the existing sample with a non-random sampling method.

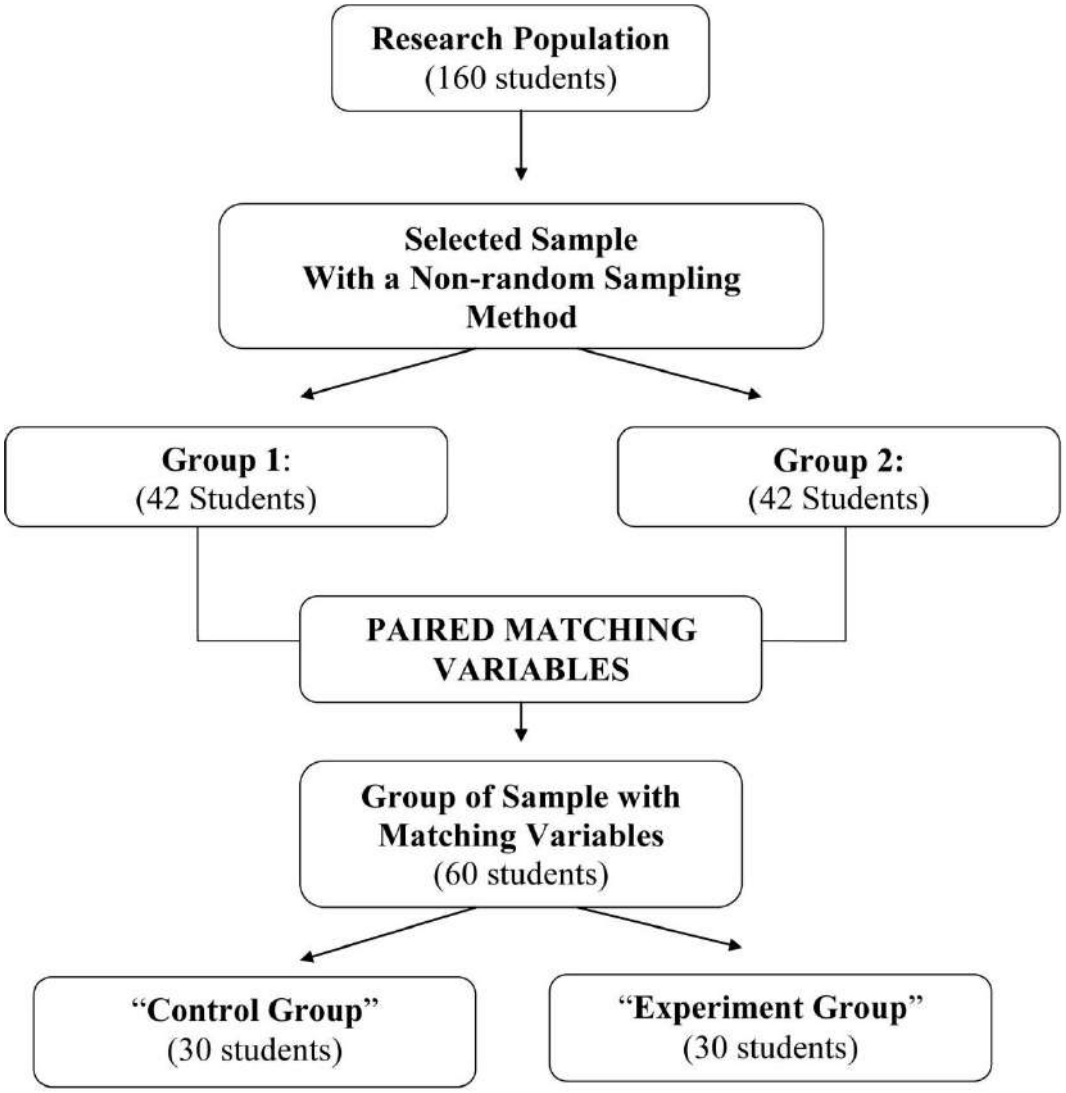


Figure 1:
Diagram of Paired Matching Variables

b. **Statistical matching;** on the other hand, does not necessitate a loss of subjects, nor does it limit the number of matching variables. Each subject is given a "predicted" score on the dependent variable, based on the correlation between the dependent variable and the variable (or variables) on which the subjects are being matched. The difference between the predicted and actual scores for each individual is then used to compare experimental and control groups.

N. Research Variables

There are three kinds of research variables. They are (a) dependent variable, (b) independent variable, and (c) moderator variable. From a research title: "The Significance of Students' Learning Styles in Teaching Reading Comprehension taught Using KWLH Strategy, it was found that the independent variable in the above title is "KWLH Strategy", the dependent variable is "Reading Comprehension", and the moderator variable is "Students' Learning Style". The further description about research variables is explained as follows.

a. Independent variable

Bell (2012, p.1) states that independent variable is a variable that is *manipulated* by the researcher. The independent variable is something that is hypothesized to influence the dependent variable. The researcher determines what level or condition of the independent variable that the participant in the experiment receives. Then, Fraenkel, et.al. (2012, p.111) mention that an independent variable is presumed to affect (at least partly cause) or somehow influence at least one other variable.

b. Dependent variable

Bell (2012, p.1) mentions that dependent variable is a variable that is simply *measured* by the researcher. It is the variable that reflects the influence of the independent variable. While, Fraenkel, et.al. (2012, p.111) state that the variable that the independent variable is presumed to affect is called a dependent variable. In commonsense terms, the dependent variable "depends on" what the independent variable does to it, how it affects it.

c. Moderator variable

Lani (2012, p.1) states that moderator variable is the independent qualitative or quantitative variable that affects the relationship of the dependent and independent variables. In correlation, a moderator is a third variable that affects the correlation of two variables. Then, Fraenkel, et.al. (2012, p.112) state that moderator variable is a special type of independent

variable. It is a secondary independent variable that has been selected for study in order to determine if it affects or *modifies* the basic relationship between the primary independent variable and the dependent variable.

O. Research Setting

In research setting, it describes the school's profile where the research study is held. In this part, the location of the school, the structure of school board organization including students, teachers, and staffs are illustrated in paragraph. If possible, it is also provided with a table which explains the illustration.

P. Operational Definition

Operational definition is a statement of the writer in giving definition on some terms related to his/her research title. The terms are important to define to avoid misunderstanding of some terms used. The definitions are given without quoting some experts' statements.

Q. Previous Related Study

In previous related study, it describes between the present and the previous studies whether or not the similarities and differences have been found in terms of method of research, population, sample, and result of hypothesis testing.

R. Population and Sample

1. Population

Fraenkel, et.al. (2012, p.122) mention that A sample in a research study is the group on which information is obtained. The larger group to which one hopes to apply the results is called the

population. It means that population is whole number of the research objects which are going to be investigated in a research study. There are four classes are set as population of the research. The total numbers of four classes in English education study program consist of 160 students. So, there are 160 students used as population of the research taken the whole number of second semester students of English education study program in academic year of 2012/2013. Then, population of the research is presented in the table of population as follows.

Table 9
Population of the Study

No	Class	Gender		Total
		Female	Male	
1	PBI 1	28	12	40
2	PBI 2	32	8	40
3	PBI 3	36	4	40
4	PBI 4	27	13	40
T o t a l		123	37	160

2. Sample

Fraenkel, et.al. (2012, p.122) state that sample is one of the most important steps in the research process is the selection of the sample of individuals who will participate (be observed or questioned) which refers to the process of selecting these individuals. Then, Trochim (2006) sates that sampling is the process of selecting units (e.g. people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they are chosen. Furthermore, Fraenkel, et.al (2012, p.134) mention that there are a few guidelines that is suggested with regard to the *minimum* number of subjects needed. They are: for **correlational studies**, a sample of at least **50** is deemed necessary to establish the existence of a relationship, and for

experimental and **causal comparative studies**, it is recommended that a minimum of 30 individuals per group is taken. Then, the selected number of sample is divided into two group samples.They are control and experimental groups. Samples are assigned to either the group that receives the treatment, known as the "experimental group" or "treatment group," or to the group which does not receive the treatment, referred to as the "control group". The following is example in determining a research sample from the above population. The sample is selected based on a questionnaire on students' learning styles which is distributed to all population. Then, a sample method is chosen using two-stage cluster random sampling method. From the result analysis of the questionnaire on students' learning styles given to all population, the students are grouped into three categories of learning styles. They are visual, auditory and kinesthetics. Then, the writer takes randomly 14 students who have visual learning styles category, 14 students who have auditory learning styles category, and 14 students who have kinesthetic learning styles category in each class. So the total samples of these three categories are 84 students. After that, the students' sample (84 students) is divided into two groups (experiment and control) where each group consists of 42 students. The experimental group is taught using new or non conventional teaching strategy, while another group (control) is taught using a conventional teaching strategy. Then, Table 9 shows an example of displaying sample of the study in a table.

Table 10

Sample of the Study

	Experimental Group	**Control Group**	**Total**
Visual Learners	14	14	28
Auditory Learners	14	14	28
Kinaesthetic Learners	14	14	28
T o t a l	42	42	84

Furthermore, Fraenkel and Wallen (1990, p.70) state that there are two kinds of methods in selecting the research sample. They are: random sampling method, and non-random sampling method.

a. Random Sampling Methods

The three commons ways of obtaining random sampling method are: simple random sampling, stratified random sampling, and cluster random sampling.

1. **Simple Random Sampling**; is one of method in which each and every member of the population has an equal and independent chance of being selected. If the sample is large, this method is the best way yet devised to obtain a sample representative of the population of interest. The key to obtaining a random sample is to ensure that each and every member of the population has an equal and independent chance of being selected. This can be done by using what is known as a **table of random numbers**—an extremely large list of numbers that has no order or pattern. Then, the diagram of selecting a simple random sampling is shown in Figure 1 (Fraenkel and Wallen, 1990, p.74).

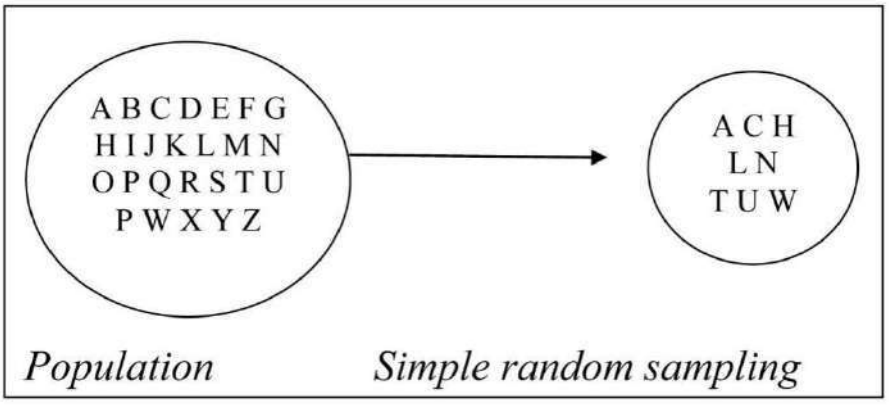


Figure 2:
Diagram of Simple Random Sampling

To select the sample using a simple random sampling, it can be chosen using the following manners.

a. Cointoss; it can be done if the population is small.
b. Lottery; is done by writing the names of the sample in the piece of paper. Then, it is rolled and put in a glass. The name which comes out of the glass is the name of the sample which is used as research sample.
c. Ranking class; is done by selecting the sample based on the students' raking in each class. For example: the class ranking is chosen from one to ten in each class. When there are eight classes in the population, using ranking class system, 80 students are obtained as the sample of the research.

2. **Stratified Random Sampling**; is a process in which certain subgroups, or strata, are selected for the sample in the same proportion as they exist in the population. A stratified sample is a probability sampling technique in which the researcher divides the entire target population into different subgroups, or strata, and then randomly selects the final subjects proportionally from the different strata. This type of sampling is used when the researcher wants to highlight specific subgroups within the population. The following is example to obtain a stratified sample of university students, the researcher would first organize the population by college class and then select appropriate numbers of freshmen, sophomores, juniors, and seniors. This ensures that the researcher has adequate amounts of subjects from each class in the final sample. Then, Fraenkel and Wallen, (1990, p.74) figure out the diagram of stratified random sampling as it is displayed in Figure 2 as follows.

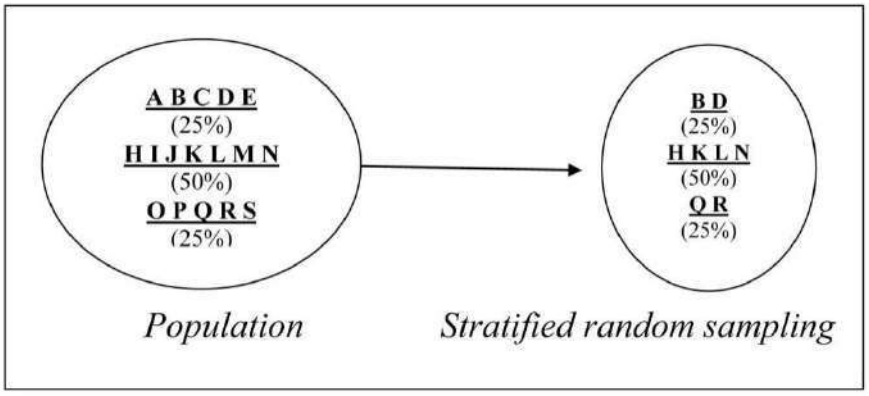


Figure 3:
Diagram of Stratified Random Sampling

3. **Cluster random sampling**; Fraenkel and Wallen (1990, p.72) state that there are times when it is not possible to select a sample of individuals from population due to administrative or other restriction, a researcher may include all of the subjects from the chosen clusters into the final sample, which is called a one-stage cluster random sampling. Or, one-stage cluster random sampling can be administered to the following example when a researcher is doing a research study on students' perception in learning English with a native speaker. There are nine classes in English education study program where the each of class consists of 30 students. These group classes are considered as population of the research. Then, four classes are randomly selected as sample of the research study. So, the total number of research samples is consisting of 120 students. Furthermore, Fraenkel and Wallen, (1990, p.74) figure out diagram of a cluster random sampling as it is displayed in Figure 3.

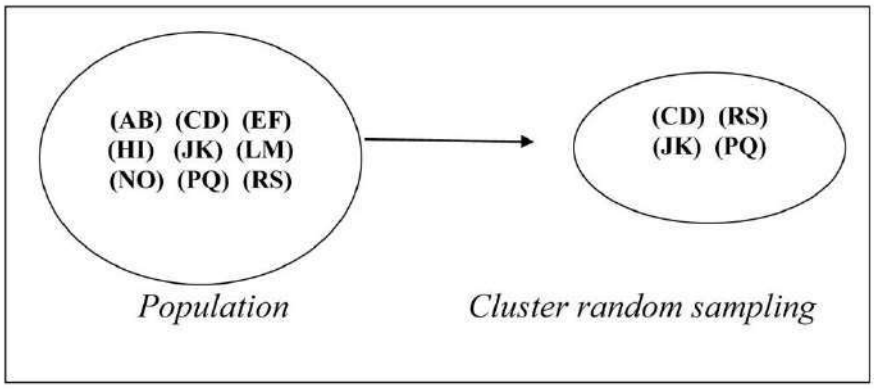


Figure 4:
Diagram of Cluster Random Sampling

4. **Two-Stage Cluster Random Sampling**; it is often useful to combine cluster random sampling with individual random sampling. Crossman (2013) states that a two-stage cluster sample is obtained when the researcher only selects a number of subjects from each cluster – either through **simple** random sampling or **systematic** random sampling. Using the above example of one-state cluster random sampling that there are 270 students in nine classes of English education study program selected as the population of the research where each class consists of 30 students. Then, four classes are randomly selected into a sub-group of sample consisting of 120 students. At last, the existing sample is then re-selected using a simple random sampling with lottery system into 70 students. Furthermore, Fraenkel and Wallen (1990, p.74) figure out diagram of two-stage cluster random sampling as it is displayed in Figure 4 as follows.

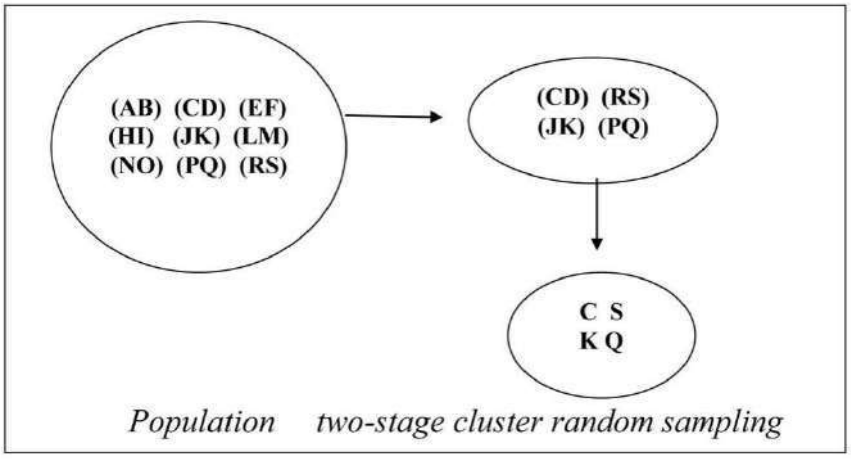
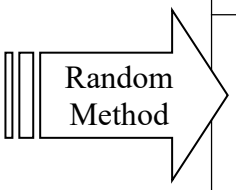


Figure 5:
Diagram of Two-stage Cluster Random Sampling

5. **Multi-Stage Cluster Random Sampling;** It refers to sampling plans where the sampling is carried out in stages using smaller and smaller sampling units at each stage. For example: the population of the study is the second semester students of English education study program in consists of four classes which have been already clustered. Each class consists of 40 students, so there are 160 students used as population of the study. From the population, a random sampling method is used to select the the class as the sample of the study. Then, the selected classes are grouped into two categories. One group is used as control class taught using conventional strategy, and other group is treated as experimental class taught using non-conventional strategy. Figure 6 shows a diagram of multistage cluster random sampling.

Population	Total
PBI.1	40
PBI.2	40
PBI.3	40
PBI.4	40
Total	**160**



Sample	Total
PBI1	40
PBI4	40
Total	**80**

Figure 6:
Diagram of multistage cluster random sampling

b. Nonrandom Sampling Methods

Fraenkel and Wallen, (1990, p.73) state that three types of methods are used in non-random sampling methods. They are: (1) systematic non-random sampling, (2) convenience non-random sampling, and (3) purposive non-random sampling.

1. **Systematic Non-random Sampling;** Sugiyono (2012, p.66) states that systematic non-random sampling is a technique for choosing sample based on the sequence of population member by giving a serial number. The procedure for selecting a systematic non-random sampling method is very easy and can be done manually. This process is much like an arithmetic progression. The main advantage of using systematic sampling is its simplicity. It allows the researcher to add a systematic element into the random selection of subjects, yet it is very easy to do. In selecting the sample using systematic non-random sampling, it is selected from odd or even number, and fold number of certain number from the population member. For example: 40 students are selected as research of the population. Then, the population member is arranged into a sequence of number. After that, the population member is selected using five-fold number. So, there are 9 members of sample is selected from number 1, 5, 10, 15, 20, 25, 30, 35, and 40. Sugiyono, (2012, p.67) illustrates the diagram

of systematic non-random sampling method as it is displayed in Figure 7.

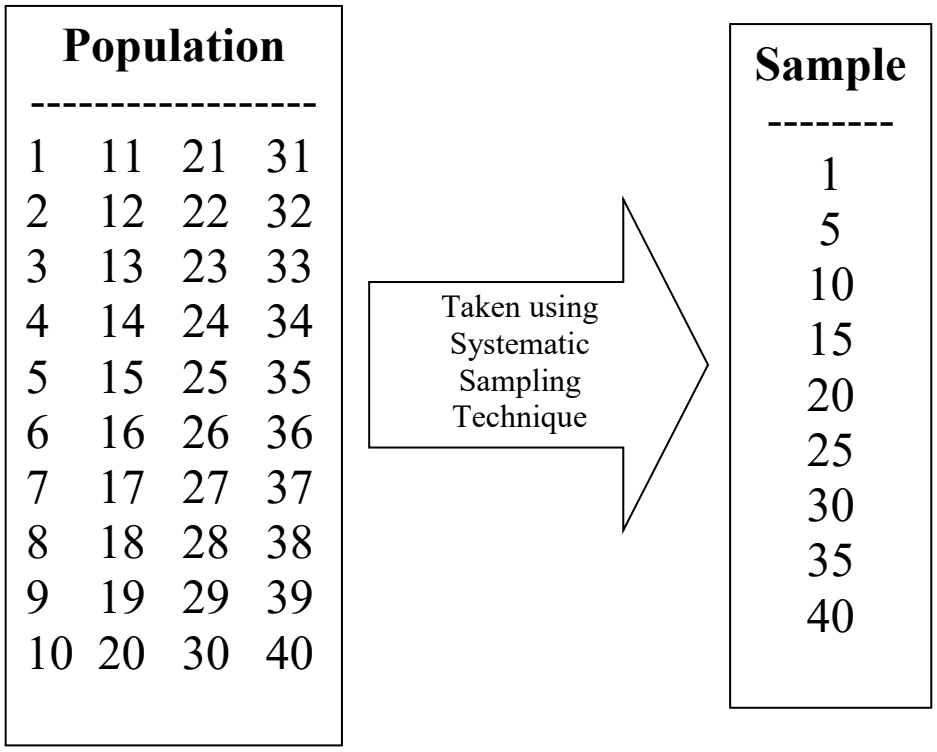


Figure 7:
Diagram of Systematic Non-random Sampling

2. **Convenience Non-random Sampling**; when it is extremely difficult or impossible to select either a random or systematic non-random sampling, a convenient sampling (a group of individuals who conveniently are available for study). Traditionally, experimental researchers have used convenience sampling to select study participants. However, as research methods have become more rigorous, and the problems with generalizing from a convenience sample to the larger population have become more apparent, experimental researchers are increasingly turning to random sampling. Furthermore, Fraenkel and Wallen

(1990:76) figure out the diagram of convenient sample in Figure 8 as follows.

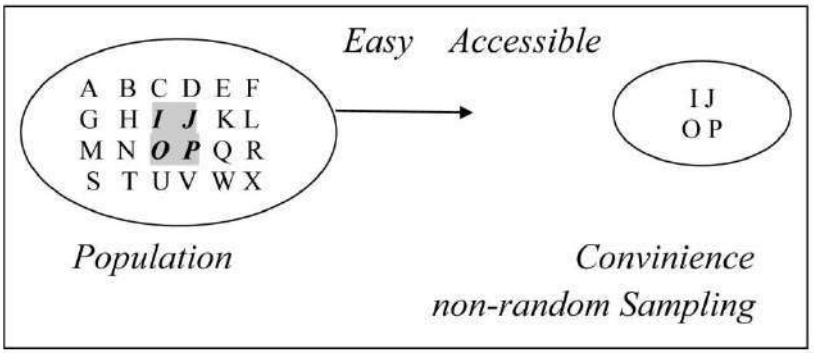


Figure 8:
Diagram of Convenience Non-random Sampling

3. **Purposive Non-random Sampling**; on occasion, purposive non-random sampling is selected based on previous knowledge of a population and the specific purpose of the research, investigators use "**personal judgment**" to select the sample. Furthermore, Arikunto (2010, p.183) states that purposive non-random sampling is done by taking the subject not based on the degree, random or scope but based on the specific purpose. So that researchers may assume that they can use their knowledge of the population to judge in selecting the sample for the specific purpose. For example: from member of research population, a personal judgement is achieved to select the sample for the specific purpose from a teacher(s) who teaches English subject in his or her class. He or she may recommend his or her students who have good English language proficiency or being qualified to be chosen as sample of the research. Fraenkel and Wallen (1990, p.76) figure out the diagram of purposive non-random sampling as it is displayed in Figure 9 as follows.

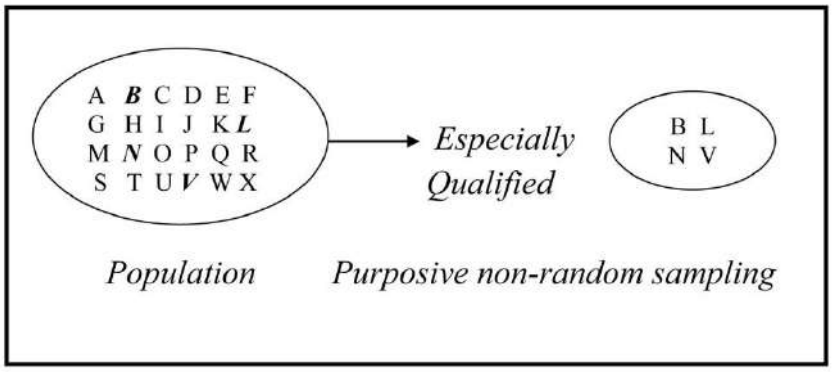


Figure 9:
Diagram of Purposive Non-random Sampling

S. Techniques for Collecting Data

In techniques for collecting the data, it presents tests, research treatments, and research instrument. They are further illustrated as follows.

1. **Tests;** Arikunto (2010, p.127) defines test as series of questions or exercises or other means of measuring skill, knowledge, intelligence, and capacities of an individual or a group. The purposes of administering a test are several, for example to diagnose a student's strength, weakness and difficulties, to measure achievement, to measure aptitude and potential, to identify readiness for a program. In collecting the obtained data, two kinds of test are done. They are: (a) **pretest**; it is done to know the students' achievements before research treatmens are given in both groups (control and experimental), and (b) **posttest**; it is done to know the effect after some treatments given in control and experimental groups.
2. **Research Instrument;** is test-question items designed for students' pretest and posttest activities. The test-question items which are used for students' pretest is the same as it is given for students' posttest activities. Before they are implemented as research instrument, it should be analyzed or checked for their validity, readability (reading instrument), and reliability tests.

a. Validity Test

Validity test is carried out to measure whether the instruments for pretest and posttest activities are valid or not. There are three kinds of validity test to be used. They are: (1) construct validity, (2) validity of each question item, and (3) content validity.

b. Reliability Test

Reliability test measures whether research instrument used for pretest and posttest activities is reliable or not. The scores of reliability are obtained from tryout analysis which is done twice using the same sample and instruments. The school where the tryout analysis is different from the school where the research study will be done. Fraenkel and Wallen (1990, p.136) state that the test score is considered reliable whenever the reliability coefficient of test score should be at least 0.70 and preferably higher. In this part, test and retest method is used to obtain the scores of tryout analysis and then, the result analysis are presented in a table score. The following is example of the scores obtained in a tryout using test and retest method.

Table 11
Scores of Students' Tryout
using Test-Retest Method

No	Students' Name	Tryout Scores	
		Test1	Test2
1	Septa Aljanati	65	70
2	M. Hasan	60	65
3	Arry Ardilla	70	70
4	Romiana Puspa	75	80
5	Mukhlis	70	75
6	M. Edward	60	70
7	Saidah Syarifah	70	80

8	Yulianto	65	60
9	Rizky Arif Afandi	85	90
10	Muammar Khadafi	75	80
11	Zulaifa Dewi Ariani	60	65
12	Raudah Tul Jannah	80	85
13	Juli Ardiansyah	65	70
14	Yovfita	55	65
15	Veby	70	80
16	Pebrianti	50	55
17	Sri Mahdalena	60	70
18	Arri Ardilla	80	75
19	Eko Saputra	60	55
20	Nani Melita	50	75
21	Tri Anggun Lestari	60	65
22	Anggung Prismadarti	70	85
23	Try Dina Marianti	80	85
24	Mawaddah Hidayati	55	60
25	Wiwin Ika Setiani	65	70

Then, from the result of tryout scores measuring reliability test is analyzed using Pearson correlation coefficient. The result analysis in measuring the reliability test using test-retest method is figured out in the following table.

Table 12
Result of Reliability Analysis
Using Pearson Correlation Coefficient

No	Number of Test	N	Pearson Correlation	Sig.	Result
1	Test 1	25	0.947	0.000	Reliable
2	Test 2	25			

From the above table analysis, it is found that the score of Pearson correlation is 0.947. From the score, it can be stated that the test is considered reliable since the score of Pearson correlation is higher than 0.70.

c. Inter-rater Reliability Test

To analyze the reliability test on students' writing achievement and speaking skill, it needs two or more raters (judges) which is called "**inter-rater reliability**".

Brown (2004, p.20) states that inter-rater reliability occurs when two or more scores yields inconsistent scores of the same test, possibly for lack of attention to scoring criteria, experience, inattention, or even preconceived biases. Then, Wang (2009, p.39) states that inter-rater reliability refers to the degree of similarity between different examiners: can two or more examiners, without influencing one another, give the same marks to the same set of scripts. From the statement, it can be inferred that two or more raters are required to avoid the inconsistent or the biases from scoring of the two language skills (speaking and writing). The following formula is used to analyze inter-rater reliability using ***Spearman Rank-Order Correlation (Rho)*** suggested by Hatch and Lazaraton, (1991, p.451).

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

Where:

ρ : Spearman Rank-Order Correlation

$\sum d^2$: The sum of the quared differences

N : Number of Sample

Table 12 shows the scorer result of inter-rater reliability using raking order method with sample (N) is 40 students. The students' writing achievements are evaluated using two raters with the five criteria level of contents. They are: content (13-30), organization (7-20), vocabulary (7-20), language use (5-25), and mechanics (2-5). Then, the scores in each rater are calculated to get its average score of each student. After that, the average scores are arranged from the highest into the lowest score to get its ranking. From the result calculation of students' writing achievement using ranking order method, it is found that the reliability score is 0.971523. From the score, it can be assumed that the test questions for pretest and posttest activities is reliable since the reliability score is higher than 0.70.

Table 13

Reliability Score of Students' Writing Achievement
Using Ranking Order Method

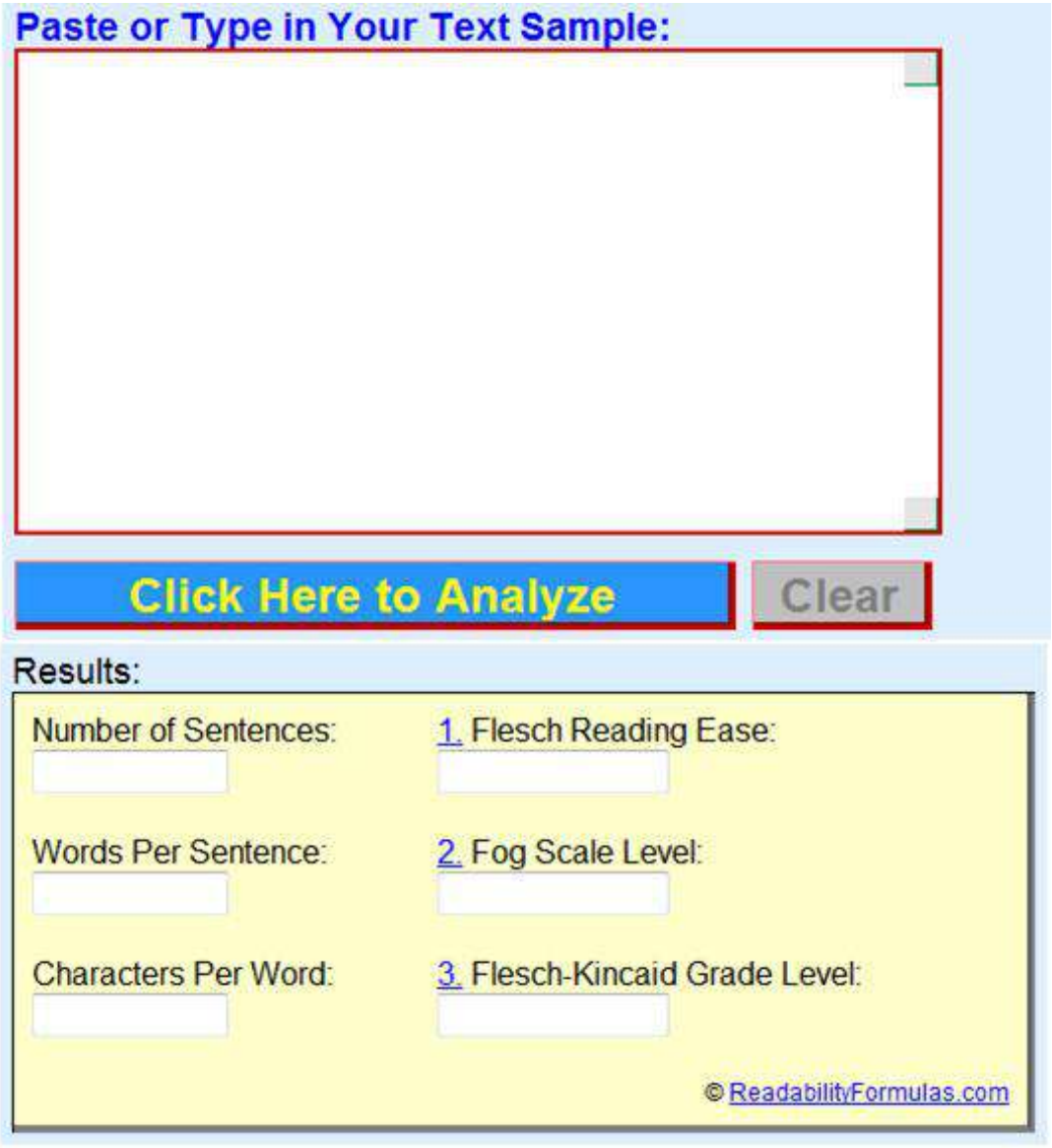
NO	RATER A							RATER B							TOT (P1, P2)	AVR (P1, P2)	D	D^2
	CNT	ORG	VOC	LANGU	MCH	TOT	RANK	CNT	ORG	VOC	LANGU	MCH	TOT	RANK				
	13-30	7-20	7-20	5-25	2-5			13-30	7-20	7-20	5-25	2-5						
1	15	15	10	10	5	55	8	25	15	15	10	4	69	8	124	62	0	0
2	20	15	20	10	5	70	1	10	20	10	20	4	64	2	134	67	-1	1
3	15	15	20	15	5	70	2	25	20	15	20	3	83	1	153	76,5	1	1
4	20	10	10	20	4	64	9	20	10	15	10	3	58	12	122	61	-3	9
5	10	20	10	15	4	59	9	10	15	15	15	5	60	7	119	59,5	2	4
6	20	15	10	15	5	65	2	25	10	10	20	5	70	7	135	67,5	-5	25
7	10	15	10	5	5	45	9	10	15	5	20	4	54	11	99	49,5	-2	4
8	20	15	20	5	4	64	3	25	20	10	15	2	72	4	136	68	-1	1
9	10	10	20	15	4	59	5	10	10	20	20	4	64	3	123	61,5	2	4
10	20	15	10	5	5	55	6	25	15	20	10	5	75	10	130	65	-4	16
11	10	10	10	5	5	40	11	15	15	20	15	3	68	9	108	54	2	4
12	15	20	20	15	4	74	3	20	10	15	20	5	70	5	144	72	-2	4
13	25	10	20	10	5	70	4	25	15	10	10	5	65	10	135	67,5	-6	36
14	10	10	10	15	5	50	8	10	20	10	20	4	64	9	114	57	-1	1
15	25	15	10	10	4	64	5	10	15	20	20	4	69	8	133	66,5	-3	9
16	15	10	10	5	4	44	10	25	15	10	20	4	74	8	118	59	2	4
17	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
18	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
19	20	15	15	5	4	59	7	15	15	15	15	5	65	7	124	62	0	0
20	25	10	10	10	5	60	9	25	15	20	20	4	84	9	144	72	0	0
21	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
22	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
23	25	10	20	10	5	70	4	25	15	10	10	5	65	10	135	67,5	-6	36
24	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
25	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
26	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
27	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
28	15	20	20	15	4	74	3	20	10	15	20	5	70	5	144	72	-2	4
29	25	10	20	10	5	70	4	25	15	10	10	5	65	10	135	67,5	-6	36
30	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
31	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
32	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
33	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
34	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
35	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
36	15	20	20	15	4	74	3	20	10	15	20	5	70	5	144	72	-2	4
37	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
38	25	15	10	15	4	69	7	10	15	15	15	5	60	5	129	64,5	2	4
39	10	15	15	15	4	59	5	25	15	15	20	3	78	6	137	68,5	-1	1
40	15	20	20	15	4	74	3	20	10	15	20	5	70	5	144	72	-2	4
∑						904							1366		2561	1280,5		128
M						22,6							34,15			32,0125		768
$R = 1-6(\sum D^2) / N(N^2-1)$																		26970
0,971523915																		

3. **Research Treatments:** treatments are designed at least for twelve meeting including pretest and posttest activities. The treatments are given in both groups (control and experiment) with similar teaching materials but different strategies. But, before treatments are done in both groups, readability test of teaching materials should be done.

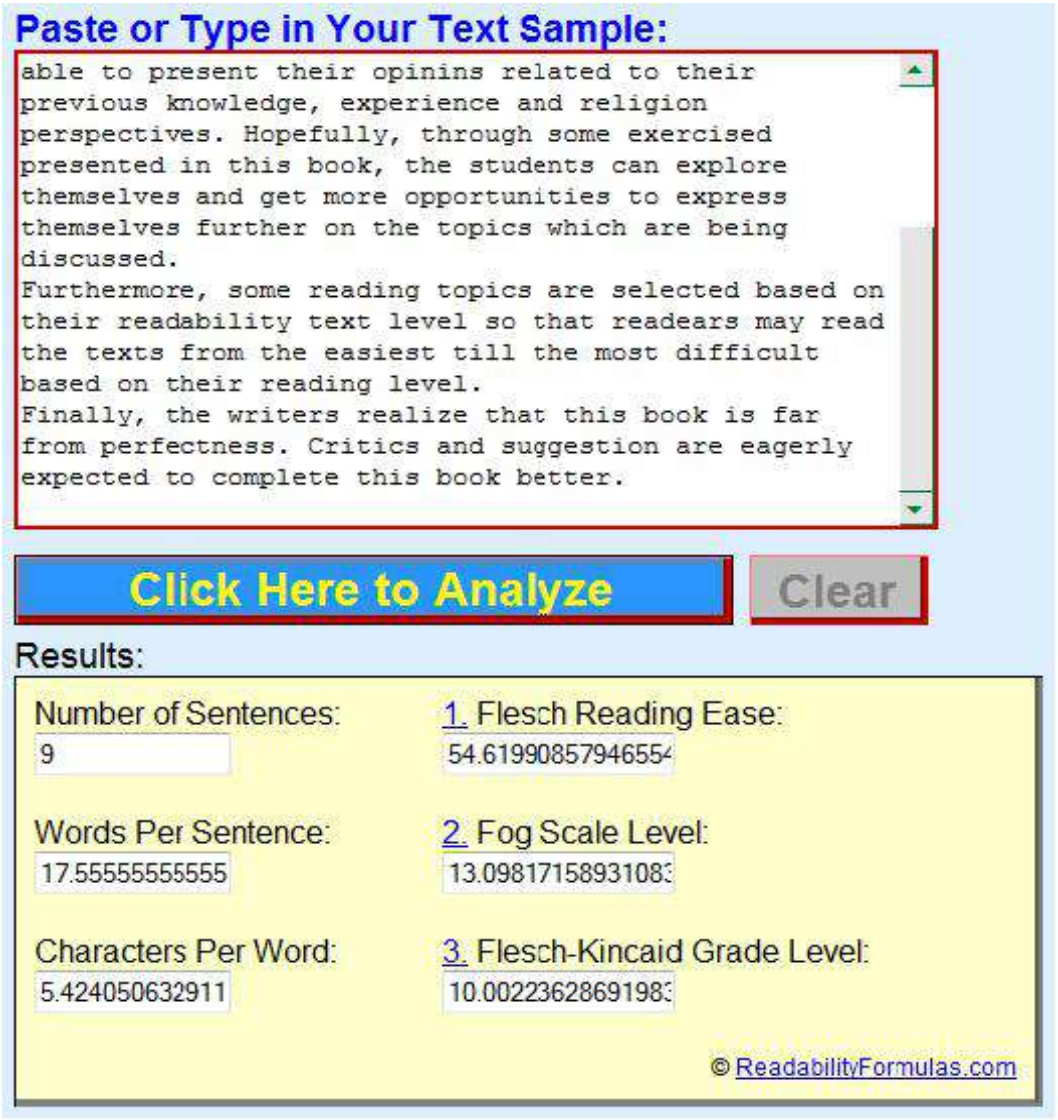
   a. **Readability Test**

      Readability test is done to know the appropriate level of reading texts for students' class level in comprehending the reading texts. It means that readability test is done to put the reading texts in an appropriate class meeting based on the difficulty level of each reading text during research treatments. Readability test is measured using **online readability test** which is accessed from *http//www.readabilityFormula.com*. There are seven categories in reading text level. They are: (a) ***very easy level*** whenever the result of flesh reading ease score is within 90-100, (b) ***easy text level*** whenever the result of flesh reading ease score is within 80-89, (c) ***fairly easy text level*** when the flesh reading ease score is within 70-79, (d) ***standard text level*** when the flesh reading ease score is within 60-69. (e) ***fairly difficult text level*** when the flesh reading ease score is within 50-59, (f) ***difficult text level*** when the flesh reading ease score is within 30-49, and (g) ***very confusing text level*** when the flesh reading ease score is within 0-29. The following are stepping procedures to estimate the reading texts for its readability using readability test online.

      - Open the website on ***http//www.readabilitytest.online***;
      - Then, the provided box in analyzing for the readability test will be seen as follows.



- Paste in a sample of text and click "ANALYZE." A sufficient sample size consists of 4-5 full sentences; approximately 100 - 250 words total.
- Then, the result analysis of readability assessment will be seen asin the following box:



After that, the categories of the texts are put in the description of the readability test table including: text title, kind of text, text statatistics: numbers of sentence, words per sentence, character per word, flesh reading ease score, and text level. All categories in readability test are presented in a table analysis of readability test. The following is example of the readability test result for research instruments as it is figured out in Table 14.

Table 14
Result of Readability Test for Research Instruments

No	Text Title	Text Statistics			Flesh-Kincaid Reading Ease Score	Text Category
		Character Per Word	Syllable Per Word	Words Per Sentence		
1	The Four Friends	4.0	1.2	13.9	88.4	Easy
2	The Jackal Who Saved the Lion	4.1	1.3	12.3	83.8	Easy
3	Public transportation should be free	4.5	1.5	13.8	69.7	Standard
4	Needle Exchange Program	4.6	1.4	20.2	64.5	Standard
5	Gawai Dayak	4.7	1.6	11.3	62.4	Standard
6	Should American be forced to be used public transportation?	4.3	1.4	27.2	59.1	Fairly difficult
7	The Queen of Adriatic	4.7	1.6	15.8	51.8	Fairly difficult
8	Thanksgiving days	4.8	1.7	21.5	44.1	Difficult
9	Five Trees should be Planted for Every Vehicle	4.8	1.7	22.3	39.2	Difficult
10	Washington DC	5.2	1.8	21.0	30.9	Difficult

b. Research Teaching Schedule

Teaching schedule presents the schedule when the treatment are carried out. In this part, the table is presented in terms teaching schedule, treatment materials, research meeting, and time allocation. The table of teaching schedule for research treatments is figured out in Table 15.

Table 15

Teaching Schedule for Research Treatments

No	Teaching Schedule For		Reading Material/Topic	Research Treatment Meeting	Time Allocation
	Experiment	Control			
1	1 Sept 2013	2 Sept 2013	The Four Friends	1st	2 x 45’
2	8 Sept 2013	9 Sept 2013	The Jackal Who Saved the Lion	2nd	2 x 45’
3	15 Sept 2103	16 Sept 2013	Public transportation should be free	3rd	2 x 45’
4	22 Sept 2013	23 Sept 2013	Needle Exchange Program	4th	2 x 45’
5	29 Sept 2013	30 Sept 2013	Gawai Dayak	5th	2 x 45’
6	6 Oct 2013	7 Oct 2013	Should American be forced to be used public transportations?	6th	2 x 45’
7	13 Oct 2013	14 Oct 2013	The Queen of Adriatic	7th	2 x 45’
8	20 Oct 2013	21 Oct 2013	Thanksgiving days	8th	2 x 45’
9	27 Oct 2013	28 Oct 2013	Five Trees should be Planted for Every Vehicle	9th	2 x 45’
10	3 Nov 2013	4 Nov 2013	Washington DC	10th	2 x 45’

T.Techniques for Analyzing Data

In analyzing the obtained data, it presents data descriptions, prerequisite analysis, and result of testing hypotheses.

1. Data Descriptions

In data descriptions, distribution of frequency data and descriptive statistics are illustrated from the obtained data of students' pretest and posttest scores in control and experimental groups.

a. Distributions of Frequency Data

In distributions of frequency data, the students' score, frequency, percentage are achieved. The distributions of frequency data are got from students' pretest scores in control group, students' posttest scores in control group, the students' pretest scores in experimental group, and students' posttest scores in experimental group. Then, the distribution of frequency data is displayed in a table analysis. The distributions of frequency data is figured out in Table 16.

Table 16

Frequency Data of Students' Pretest Scores in Control Group

Scores	Frequency	Percentage (%)
4.00	4	10.0
4.33	1	2.5
5.66	1	2.5
6.00	4	10.0
6.33	5	12.5
6.66	4	10.0
7.00	1	2.5
Total	**35**	**100.0**

b.Descriptive Statistics

In descriptive statistics, number of sample, the score of minimal, maximal, mean, standard deviation, and standard error of mean are obtained. Descriptive statistics are obtained from students' pretest and posttest scores in control and experimental groups. Then, descriptive statistic on students' pretest scores in control group is figured out in Table 17.

Table 17

Descriptive Statistics of Students' Pretest Scores in Control group

Students' Pretest Scores	N	Min	Max	Mean	Std. D
	20	27.00	35.00	31.5000	3.30824

2. Prerequisite Analysis

Before analyzing the obtained data, pre-requisite analysis should be done to see whether or not the data is normal and homogen. The following is the procedures in pre-requisite analysis.

a. Normality Test

Normality test is used to measure whether the obtained data is normal or not. The data can be classified into normal whenever the p-output is higher than 0.025 (Basrowi, 2007, p.85). In mesuring normality test, ***1-sample Kolmogronov Smrinov*** is used. The normality test is used to measure students' pretest and posttest scores in control and experimental groups. Then, the result analysis in measuring the normality test of students' pretest scores in control and experimental groups is further figured out in Table 18.

Table 18

Normality Test of Students' Pretest Scores in Control and Exprimental Groups Using 1-Sample Kolmogronov Smrinov Z

No	Students' Pretest	N	Kolmogronov Smrinov Z	Sig.	Result
1	Control Group	40	0.960	0.558	Normal
2	Experimental Group	40	0.316	0.915	Normal

From the above table analysis, it was found that the p-output from students' pretest in control and experimental groups are 0.558, and 0.915. From the two scores, it can be stated that the students' pretest score in control and experimental are considered normal since they are higher than 0.025.

b. Homogeneity Test

Homogeneity test is used to measure the obtained scores whether it is homogen or not. Basrowi (2007, p.106) states that the score is categorized homogen when the p-output was higher than mean significant difference at 0.05 levels. In measuring homogeneity test, **Levene Statistics** in SPSS is used. The homogeneity test is used to measure students' pretest scores and posttest score in control and experimental groups. Then, the result analysis in measuring the homogeneity test of students' pretest scores in control and experimental groups is further figured out in Table 19.

Table 19

Homogeneity Test of Students' Pretest Scores in Control and Experimental Groups Using Levene Statistics

No	Students' Pretest Scores	N	Levene Statistics	Sig.	Result
1	Control Group	20	0.609	0.795	Homogen
2	Experimental Group	20			

From the above table in measuring homogeneity test, it is found the p-output 0.795. From the score, it can be stated that the students' pretest scores in control and experimental groups are considered homogen since it was higher than 0.05.

3. Results of Hypothesis Testing

The results of hypothesis testing from statistical calculation using SPSS application program are described as follows.

a. Measuring Means Paired Differences

Means paired differences are used to compare the improvement from the scores in two groups (the scores of students' pretest to posttest in control and experimental groups) using a paired sample t-test. Then, to indicate whether means paired differeces are found or not, a table analysis of means paired differences of students' pretest to posttest scores taught using KWLH and conventional strategies is made as it is figured out in Table 20.

Table 20
Result Analysis in Measuring Means Paired Differences
on Students' Reading Comprehension Taught
Using KWLH and Teacher's Method

Conventional And KWLH Reading Strategy	Pretest to Posttest Scores in Control Group			Pretest to Posttest Scores in Experimental group			H*o*
	Mean Pretest	Mean Posttest	Mean Paired Difference	Mean Pretest	Mean Posttest	Mean Paired Difference	
	4.9250	6.3000	1.37500	5.8250	7.6500	1.82500	Rejected

From the result analysis measuring means paired differences, it is found that means paired differences in control group is 1.37500, while means paired differences in experimental group is 1.82500. From the two means difference scores, it can be stated that means paired differences in experimental group is higher than means paired differences in control group. So, it is concluded that the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.

b. Measuring Means Significant Improvement

Means significant improvement is used to find the improvement of the students' pretest to posttest scores taught using KWLH reading strategy or taught using conventional strategy. Means significant improvement is found whenever the p-output is lower than 0.25. The result analysis in measuring means significant improvement of students' pretest to posttest scores on students' reading comprehension achievement taught using KWLH strategy is figured out in Table 21.

Table 21

Result Analysis in Measuring Means Significant Improvement on Students' Reading Comprehension Achievement taught Using KWLH Strategy

KWLH Reading Strategy	Paired Sample T-Test			H*o*
	T	Df	Sig. (2-tailed)	
	-3.107	38	0.000	Rejected

From the result analysis above, it is found that there is means significant improvement from students' pretest to posttest scores taught using KWLH reading strategy since the p-output is lower than 0.025. So, it is concluded that the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is acccepted.

c. Measuring Means Significant Difference/Influence

In measuring means significant difference/Influence, there are two analyses to be done. They are as follows.

1. Using Independent Sample T-Test

Mens significant difference or influence is found from testing students' posttest scores in control and experimental groups using independent sample t-test. Means significant difference/influence is found whenever the p-output is lower than mean signficant difference at 0.05 levels. The result analysis in measuring means significant difference of students' reading comprehension achievment taught using KWLH and conventional strategies is figured out in Table 22.

Table 22

Result Analysis in Measuring Means Significant Difference on Students' Reading Comprehension Achievment taught Using KWLH and Conventional Strategies

KWLH and Conventional Reading Strategies	**Independent Sample T-Test**			H*o*
	T	Df	Sig. (2-tailed)	
	-3.107	38	0.004	Rejected

From the above table analysis, it is found the p-output 0.004. It can be stated that there is means significant difference on students' reading comprehension scores taught using KWLH and conventional reading strategies since the p-output is lower than 0.05. So, it is concluded that the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is acccepted.

2. Using One-way ANOVA

One-way ANOVA is used to measure means significant difference from more than two variables in one group. Significant difference is found whenever the p-output is lower than mean signficant difference at 0.05 levels. The result analysis in measuring means significant difference of learning styles on students' reading comprehension scores taught using KWLH strategy is figured out in Table 23.

Table 23

Resut Analysis in Measuring Means Significant Difference Using One-way ANOVA

learning styles on SS_reading comprehension scores taught using KWLH strategy	**One-way ANOVA**			H*o*
	Df	F	Sig. (2-tailed)	
	2	0.107	0.720	Accepted

From the above table measuring means significant difference of learning styles on students' reading comprehension scores taught using KWLH strategy, it is found the p-output 0.720. It can be stated that there is no means significant difference on students' visual, auditory, and kinesthetic categories taught using KWLH reading strategy since the p-output is higher than 0.05. So, it is concluded that the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

d. Measuring Regression

Regression analysis is used to predict the relationship of dependent variable towards independent variable (Sugiyono, 2012, p.260). For example: a research study finding the effects of motivation on students' reading comprehension average score. In this study, the formulation of the problem in analyzing regression analysis becomes "Is there any **significant linear relationship** between motivation and students' reading comprehension avarage scores?". The result analysis in measuring significant linear relationship using regression analysis between students' motivation and students reading comprehension average scores is shown in Table 24.

Table 24

Result Analysis in Measuring Significant Linear Relationship Using Regression Analysis

Ss_Motivation and Ss_reading comprehension Average Scores	**Regression Analysis**			H*o*
	Df	F	Sig. (2-tailed)	
	2	0.592	0.570	Rejected

From the above table measuring the effects of motivation on students' reading comprehension average score using regression analysis, it is found that the significant score is 0.570. From the p-output, it can be stated that there is no significant linear relationship between motivation and students' reading comprehension average scores since the significant

score is higher than 0.05. So, it is concluded that the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.

e. Measuring Correlation

Correlation analysis is used to describe whether or not there is strength of relationship between two variables (Arikunto, 2010, p.315). To measure the correlation between two or more variables, "**Pearson product-moment correlation coefficient"** is used. The correlation coefficient is found whenever the p-output was lower than 0.05. The following is example of a research study in finding the correlation between students' vocabulary and grammar scores on students' writing achievement. Table 25 shows the result in measuring correlation analysis using Pearson product moment correlation coefficient between vocabulary and grammar scores on students' writing achievement.

Table 25
Result Analysis in Measuring Correlation
Using Correlation Coefficient

Vocabulary and Grammar Scores on Students' Writing Achievement	**Correlation Coefficient**		H*o*
	Pearson Correlation	Sig. (2-tailed)	
	0.068	0.817	Accepted

From the above table, it shows that the significant correlation between students' vocabulary and grammar scores on students' writing achievement is 0.817. From the score it can be stated that there is no correlation between students' vocabulary and grammar scores on students' writing achievement since the correlation significant coefficience is higher than 0.05. So, it is concluded that the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

f. Measuring Significant Difference Effects

Analysis of Covarience (ANCOVA) is used to measure significant different effects from more than two or three independent variable. Significant different effects are found whenever the p-output is lower than mean signficant difference at 0.05 levels. The result analysis in measuring significant different effects of students' learning styles on students reading comprehension average scores taught using KWLH and conventional strategies is figured out in Table 26.

Table 26

Resut Analysis in Measuring Significant Different Effects
Using Analysis of Covariance (ANCOVA)

Learning styles on SS_reading comprehension scores taught using KWLH and conventional strategies	**Analysis of Covariance (ANCOVA)**			H*o*
	Df	F	Sig. (2-tailed)	
	2	0.107	0.720	Accepted

From the above table, it is found the p-output 0.720. From the score, it can be stated that there is no significant different effects of students' learning styles on reading comprehension average score taught using KWLH and conventional strategies since the p-output is higher than 0.05. So, it is concluded that the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

g. Measuring Significant Interaction Effects

Two-way ANOVA is used to measure significant interaction effects which have more than three variables. Significant interaction effect is found whenever the p-output is lower than 0.024 levels. The result analysis in measuring significant interaction effects of students' learning styles on students' reading comprehension scores taught using KWLH and conventional strategies is figured out in Table 27.

Table 27

Result Analysis in Measuring Significant Interaction Effects
Using two-way ANOVA

SS_ learning styles on SS' reading comprehension Scores taught using KWLH and conventional strategies	**Two-way ANOVA**			H*o*
	Df	F	Sig. (2-tailed)	
	2	0.811	0.043	Rejected

From the above table, it is found the p-output 0.043. From the score, it can be stated that there are significant interaction effects of students' learning styles on students' reading comprehension average score taught using KWLH and conventional strategies since the p-output is lower than 0.024. So, it is concluded that the null hypothesis (H*o*) is rejected, and the alternative hypothesis (H*a*) is accepted.

h. Measuring Significant Main Effects

Two-way MANOVA is used to measure significant main effects which have more than three variables. Significant main effects are found whenever the p-output was lower than mean signficant difference at 0.05 levels. The result analysis in measuring significant main effects between students' learning styles and emotional intelligence on students' reading comprehension average scores taught using KWLH and conventional strategies is figured out in Table 28.

Table 28

Result Analysis in Measuring Significant Main Effects
Using two-way MANOVA

SS' learning styles and emotional intelligence on SS' reading comprehension scores taught using KWLH and conventional strategies	**Two-way MANOVA**			H*o*
	Df	F	Sig. (2-tailed)	
	2	0.811	0.472	Accepted

From the above table, it is found that the p-output is 0.472. From the score, it can be stated that there is no significant main effects between students' learning style and emotional intelligence on students' reading comprehension average score taught using KWLH and conventional strategies since the p-output is higher than 0.05. So, it is concluded that the null hypothesis (H*o*) is accepted, and the alternative hypothesis (H*a*) is rejected.

U. Writing the References

In writing the references, APA (American Psychological Association) formatting and style is the most commonly used to cite sources within the social sciences where it should state: the **author's name**: first name, middle initial(s), and last name, do not use titles (Dr.) or degrees (Ph.D.), the **year** of the book published, the **book's title** (written in italic), the **city** of the book published, and the **publisher**. The reference should be written in one paragraph where the second line of the paragraph is put 5 inches inside. The following is example of writing the references based on APA formatting and style.

Sugiyono. (2006). *Metode Penelitian Kuantitaif, Kualitatif dan R&D*. Bandung: Alfabeta.

Fraenkel, Jack R and Norman Wallen. (1990). *How to Design and Evaluate Research.* New York: Mcgraw Hall, Inc.

Or when the sources are obtained from internet or website sources, the writing of references is as follows:

Allen, Janet. (2004). *Tools for Teaching Content Literacy*. Portland, Maine: Stenhouse Publishers. Retrieved on 13rd January 2013, from: http://www.pkwy.k12.mo.us/southsr/parkwaysouth/homepages/assignmentDetail.

Wilhelm, Jeff. (2012). *Understanding Reading Comprehension*. Retrieved on July 22nd, 2012, from: http://www.scholastic.com/teachers/article/understanding-reading-comprehension.

REFERENCES

Achelis, Steven B. (2013). *Correlation Analysis.* Retrieved, from: http://www.metastock.com. On: 22nd February 2013

Artha, Ratih Puspita. (2012). Pengembangan Multimedia Untuk Mata Pelajaran Ekonomi Kelas X di SMAN 1 Ngluwar. *Unpublished Thesis*: Pendidikan Ekonomi, Universitas Negeri Yogyakarta.

Arikunto, Suharsimi. (2010). *Prosedur Penelitian Suatu Pendekatan dan Praktik.* Jakarta: Bina Aksara.

Brad, Bell. (2012). *Independent Variable and Dependent Variable.* Retrieved on 4th February 2013, from: http://www.psychologyandsociety.com/index.html.

Basrowi and Soenyono. (2007). *Metode Analisis Data Sosial.* Kediri: CV. Jenggala Pustaka Utama.

Brown, H. Douglas. (2004). *Language Assessment Principle and Classroom Practice.* San Francisco: Longman.

Best, John W and James V. Khan. (1993). *Research in Education (7th Edition).* Boston: Allyn and Bacon.

Crossman, Ashley. (2013). *Cluster Sample.* Retrieved on 5th January 2013, from: http://sociology.about.com/od/Types-of-Samples/a/Cluster-Sample.htm

Cohen, Louis. Lawrence Manion, and Keith Morison. (2007). *Research Method in Education.* London and New York: Routledge.

Creswell, J.W. (1994). *Research Design: Qualitative & Quantitative Approaches.* London: SAGE Publications.

Cormack, D.S. (1991). *The Research Process*. Black Scientific: Oxford.

David, Nunan. (1992). *Research Methods in Language Learning.* UK: Cambridge Univesity Press.

Fels, Dawn.(2009). *How to Write a Reserach Questions.* Retrieved on 5th January 2013, from: http://writingcenter.gmu.edu.

Fraenkel, Jack R, Norman E. Wallen, and Helen H. Hyu. (2012). *How to Design and Evaluate Research in Education (8th-Ed)*. New York: Mcgraw Hall, Inc.

Fraenkel, Jack R and Norman Wallen. (1990). *How to Design and Evaluate Research.* New York: Mcgraw Hall, Inc.

Hatch, Evelyn and Anne Lazaration. (1991). *The Research Manual: Design and Statistics for Applied Linguistics.* Los Angles, University of California: Newbury House Publishers.

Hughes, Arthur. (1989). *Testing Language Teachers.* London: Cambridge University Press.

Hatch, Evelyn and Hossein Farhady. (1982). *Research Design and Statistics for Applied Linguistics.* Cambridge: Newbury House Publishers.

Lani, James. (2012). *Moderator Variable*. Retrieved on 5th January 2013, from: http://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/moderator-variable/.

Marczyk, Geoffrey, David Dematteo and David Festinger. (2005). *Essential of Research Design and Methodology.* New Jersey: John Wiley & Sons, Inc.

Nina, Schoolsher. (2000). *Independent Sample T-Test*. Retrieved on 22nd February 2013, from: http://www.wellesley.edu/Psychology/indepttest.html.

NCCP. (2012). *Research Connection*. Retrieved on 26th April 2013, from: http://www.researchconnections.org/childcare/datamethods/experimentsquasi.jsp.

Picciano, G Anthony. (2013). *Experimental Research.* Retrieved on 27th April 2013, from: http://www.hunter.cuny.edu/edu/apiccian/edstat09.html.

Sugiyono. (2012). *Statistika Untuk Penelitian.* Bandung: Alfabeta.

Sugiyono. (2010). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Sugiyono. (2006). *Metode Penelitian Kuantitaif, Kualitatif dan R&D*. Bandung: Alfabeta.

Santoso, Singgih. (2003). *Mengatasi Berbagai Masalah Statistik dengan SPSS Versi 15.* Jakarta: Penerbit PT. Elex Media Komputindo.

Trochim, William. (2006). *Sampling.* Retrieved on 6th January 2013, from: http://www.socialresearchmethods.net/kb/sampling.php

Wikipedia. (2013). *Quantitaive Research.* Retrieved on 21st February 2013, from: http://en.wikipedia.org/wiki/Quantitative_research

Wikipedia. (2013). *Analysis of Covariance*. Retrieved on 22nd February 2013, from: http://en.wikipedia.org/wiki/Analysis_of_covariance

Wang, Ping. (2009). *The Inter-rater Reliability in Scoring Composition*. English Language Teaching Journal, Vol.2 No.3 September 2009. Retrieved from: www.ccsenet.org/journal.html, on May 16th, 2013.

Wiegle, S. Cushing. (2002). *Assessing Writing.* London: Cambridge University Press.

>><<

EXPERIMENTAL RESEARCH OUTLINE

TITLE: ……

i. Introduction

A. Background
B. Problem of the Study
C. Objective of the Study
D. Significances of the Study
E. Hypotheses
F. Criteria of Testing the Hypotheses

ii. Literature Review

A. Theoretical Framework
   1. …….
   2. ……..
   3. ……..
   4. Teaching Procedures Using …..
   5. Teaching Procedures using Conventional Strategy
B. Previous Related Study
C. Research Setting

iii. **Research and Procedure**

A. Research Method
   1. Research Variables
   2. Operational Definitions
   3. Population and Sample
B. Techniques for Collecting Data
   1. Tests
   2. Research Instrument Analysis
      a. Validity Test
         1) Construct Validity
         2) Validity of Each Question Items
         3) Content Validity
      b. Reliability Test

3. Research Treatments
   a. Readability Test***
   b. Research Schedule

4. Techniques for Analyzing Data

A. Data Descriptions

1. Distributions of Frequency Data

a. Pretest Score in Control Group
b. Posttest Score in Control Group
c. Pretest Score in Experimental Group
d. Posttest Score in Experimental Group

2. Descriptive Statistics

a. Pretest Score in Control Group
b. Posttest Score in Control Group
c. Pretest Score in Experimental Group
d. Posttest Score in Control Group

C. **Pre-requisite Analysis**

1. Normality Test

a. Pretest Score in Control and Experimental Group
b. Posttest Score in Control and Experimental Group

2. Homogeneity Test

a. Pretest Score in Control and Experimental Group
b. Posttest Score in Control and Experimental group

D. **Hypothesis Testing**

a. **Measuring significant difference** of KWLH and conventional strategies using **independent sample t-test**
b. **Measuring significant influence** of KWLH strategy on students' reading comprehension scores using **independent sample t-test**

c. **Measuring effectiveness** of KWLH strategy on students' reading comprehension score using **independent sample t-test**

d. **Measuring means significant improvement** of KWLH strategy on students' reading comprehension scores using **paired sample t-test**

e. **Measuring means paired differences** of KWLH and conventional strategies on students' reading comprehension scores using paired sample t-test

f. **Measuring significant linear relationship** between vocabulary and grammar scores on students' descriptive writing achievement using **Analysis of Covariance (ANCOVA)**

g. **Measuring correlation** between motivation and students' reading comprehension scores using **correlation analysis**

h. **Measuring significant difference** of visual and auditory learning style category on students' reading comprehension taught using KWLH strategy using **one-way ANOVA**

i. **Measuring significant interaction effects** of students' learning style (visual, auditory, and kinesthetic) category on students' reading comprehension score taught using KWLH and conventional strategies analyzed using **two-ways ANOVA**

j. **Measuring significant main effects** of students' learning style (visual, auditory, kinesthetic) categories and students' motivation (high, middle, and low) categories on students' reading comprehension scores taught using KWLH and conventional strategies analyzed using **two-way MANOVA**

REFERENCES

STUDENTS' SCORES IN A GROUP SAMPLE SELECTED WITH A NON-RANDOM SAMPLING METHOD

NO	STUDENTS' NAME		STUDENTS' SCORES	
	GROUP 1	GROUP 2	GROUP 1	GROUP 2
1	Fariha	Nur Rahelmi	55	65
2	Dian Puspita	Rizky Andi Arif	60	60
3	Meggi Lestari	Rizka Razi Mona	62	67
4	Uliatul Fadhilah	Novelia PS	59	68
5	Marlen	Zuriah Kausari	63	67
6	Alit wigati	Siti Isrokah	60	62
7	Bena Yustia	M Khadafi	62	65
8	Julia Metasari	Mirrah Salsabillah	65	62
9	Ahlamia Utami	Santi Angraini	67	68
10	Mentari Rizky	Sani	66	66
11	Indah Purwati	Noprijaya	60	71
12	Mukhlis	Novia Fajrina	57	66
13	Mellyza	Wenny	72	70
14	Arie WP	Suparno	65	62
15	Darmawan	Tania Janesa	61	66
16	Ega Vansela	Sinta Pancarini	66	72
17	Irina Maharani	Selva Emalia	65	65
18	Ema Karmilasari	Raudatul Jannah	65	60
19	Joni Iskandar	Debby Farisa	64	69
20	Wahyudo	Rizky Amalia	62	71
21	Ella Ayu	Novika Astrilianti	65	60
22	Dian Utari	Zerta Pratiwi	71	69
23	Fariha	Septria Wulandari	73	60
24	Shohibul Khafi	Nur Rahelmi	65	65
25	Tuti Hendriyani	Yulinda	55	72
26	Wilta PS	Ushwatun Hasanah	62	60
27	Tria Novita	Winda Aulia	78	65

28	Sharil Rozi	Yoja Fitria	66	74
29	Tri Astuti	Tri Lestari	69	69
30	Yunica	Syarifah Sukaina	72	70
31	Siti Mardianti	Wilta PS	70	75
32	Sapria	Zakia Kurnia	59	66
33	Syarifah Salwa	Sondang	62	75
34	Tama Aprezky	Septi Aprianti	71	76
35	Surani Fitriana	Solihatun	73	67
36	Yuliasari	Zurnaila	72	60
37	Tiara Nita A	Tika Sari	65	76
38	Selly Marsela	Siti Sulistiani	70	74
39	Sri Ayu Anggaraini	Nistawati	55	68
40	Sonia Widiarti	Oktaria	62	70
41	Peti Tanjungsari	Ria Utami	59	66
42	Megaret Sari	Nura Bitaria	72	76

RESULT OF PAIRED MATCHING SCORES

NO	STUDENTS' NAME		PAIRED MATCHING SCORES
	GROUP 1	GROUP 2	
1	Dian Puspita	Rizky Andi Arif	60
2	Meggi Lestari	Siti Isrokah	62
3	Alit wigati	Raudatul Jannah	60
4	Bena Yustia	Mirrah Salsabillah	62
5	Julia Metasari	Nur Rahelmi	65
6	Ahlamia Utami	Rizka Razi Mona	67
7	Mentari Rizky	Sani	66
8	Indah Purwati	Septria Wulandari	60
9	Mellyza	Sinta Pancarini	67
10	Arie WP	Selva Emalia	65
11	Ega Vansela	Nistawati	65
12	Shohibul Khafi	Oktaria	55
13	Tuti Hendriyani	Ria Utami	62
14	Wilta PS	Nura Bitaria	78
15	Tria Novita	Winda Aulia	66
16	Ema Karmilasari	Yoja Fitria	69
17	Peti Tanjungsari	Tri Lestari	72
18	Wahyudo	Syarifah Sukaina	70
19	Ella Ayu	Septi Aprianti	59
20	Ega Vansela	Solihatun	62
21	Irina Maharani	Zurnaila	60
22	Ema Karmilasari	Tika Sari	65
23	Joni Iskandar	Sinta Pancarini	72
24	Julia Metasari	Wilta PS	60
25	Ahlamia Utami	Zakia Kurnia	65
26	Mentari Rizky	Sondang	74
27	Indah Purwati	Septi Aprianti	69

28	Tama Aprezky	Rizky Amalia	70
29	Surani Fitriana	Novika Astrilianti	78
30	Yuliasari	Zerta Pratiwi	66

STUDENTS' READING COMPREHENSION SCORES
IN CONTROL AND EXPERIMENTAL GROUPS

NO	STUDENTS' NAME		TREATMENTS' SCORES			
	CONTROL GROUP	EXPERIMENTAL GROUP	CONTROL		EXPERIMENT	
			PRE	POST	PRE	POST
1	Dian Puspita	Rizky Andi Arif	4,66	5,33	6	7
2	Meggi Lestari	Siti Isrokah	5,33	5,66	6	6,33
3	Alit wigati	Raudatul Jannah	5,33	6	6	8,33
4	Bena Yustia	Mirrah Salsabillah	4	4,66	5,66	7,33
5	Julia Metasari	Nur Rahelmi	4	8	5,33	8,33
6	Ahlamia Utami	Rizka Razi Mona	5,33	7	5,33	6
7	Mentari Rizky	Sani	5,33	6,66	4,33	6
8	Indah Purwati	Septria Wulandari	4,66	6	5	6,66
9	Mellyza	Sinta Pancarini	5,33	7	5	6
10	Arie WP	Selva Emalia	6	7,33	4	6
11	Ega Vansela	Nistawati	5,33	5	4,66	5,33
12	Shohibul Khafi	Oktaria	5	6	4,66	6,33
13	Tuti Hendriyani	Ria Utami	4	6	5,66	7
14	Wilta PS	Nura Bitaria	6,33	4,33	4	4,33
15	Tria Novita	Winda Aulia	4,66	5,66	7,33	6
16	Ema Karmilasari	Yoja Fitria	7,33	6	7	8,33
17	Peti Tanjungsari	Tri Lestari	4,66	5,33	8	8,66
18	Wahyudo	Syarifah Sukaina	6,66	5	4,66	6,66
19	Ella Ayu	Septi Aprianti	5	5,33	7,33	8
20	Ega Vansela	Solihatun	4	3,33	6	8
21	Irina Maharani	Zurnaila	6	7,33	5,33	6,66
22	Ema Karmilasari	Tika Sari	6	8	5,33	6,66
23	Joni Iskandar	Sinta Pancarini	6	8	7,33	8,33
24	Julia Metasari	Wilta PS	5	6	7	8
25	Ahlamia Utami	Zakia Kurnia	8,66	5	6,33	8,33
26	Mentari Rizky	Sondang	7,66	8,33	7	8
27	Indah Purwati	Septi Aprianti	5,33	6	6	7,33
28	Tama Aprezky	Rizky Amalia	5	6	7	7,33
29	Surani Fitriana	Novika Astrilianti	5,33	5,33	7,33	7,33
30	Yuliasari	Zerta Pratiwi	5	5,66	6,66	8

PEARSON'S CORRELATION COEFFICIENT r (Critical Values)

	Level of Significance for a One-Tailed Test										
	.05	.025	.01	.005	.0005		.05	.025	.01	.005	.0005
	Level of Significance for a Two-Tailed Test										
df=(N-2)	.10	.05	.02	.01	.001	df=(N-2)	.10	.05	.02	.01	.001
1	0.988	0.997	0.9995	0.9999	0.99999	21	0.352	0.413	0.482	0.526	0.640
2	0.900	0.950	0.980	0.990	0.999	22	0.344	0.404	0.472	0.515	0.629
3	0.805	0.878	0.934	0.959	0.991	23	0.337	0.396	0.462	0.505	0.618
4	0.729	0.811	0.882	0.971	0.974	24	0.330	0.388	0.453	0.496	0.607
5	0.669	0.755	0.833	0.875	0.951	25	0.323	0.381	0.445	0.487	0.597
6	0.621	0.707	0.789	0.834	0.928	26	0.317	0.374	0.437	0.479	0.588
7	0.582	0.666	0.750	0.798	0.898	27	0.311	0.367	0.430	0.471	0.579
8	0.549	0.632	0.715	0.765	0.872	28	0.306	0.361	0.423	0.463	0.570
9	0.521	0.602	0.685	0.735	0.847	29	0.301	0.355	0.416	0.456	0.562
10	0.497	0.576	0.658	0.708	0.823	30	0.296	0.349	0.409	0.449	0.554
11	0.476	0.553	0.634	0.684	0.801	40	0.257	0.304	0.358	0.393	0.490
12	0.457	0.532	0.612	0.661	0.780	60	0.211	0.250	0.295	0.325	0.408
13	0.441	0.514	0.592	0.641	0.760	120	0.150	0.178	0.210	0.232	0.294
14	0.426	0.497	0.574	0.623	0.742	8	0.073	0.087	0.103	0.114	0.146
15	0.412	0.482	0.558	0.606	0.725						
16	0.400	0.468	0.542	0.590	0.708						
17	0.389	0.456	0.529	0.575	0.693						
18	0.378	0.444	0.515	0.561	0.679						
19	0.369	0.433	0.503	0.549	0.665						
20	0.360	0.423	0.492	0.537	0.652						

1) Decide if you should use a One-Tailed or Two-Tailed Test: **(MSLS: 38.2)**
   a. One-Tail: if you have an *a priori*: hypothesis as to the sign (- or +) of the correlation.
   b. Two-Tail: if you have no *a priori*: hypothesis as to the sign of the correlation.
2) Calculate df (degrees of freedom) = N (sample size) - 2). **(MSLS: 31)**
3) Locate this df in the table.
4) Use this row of threshold values.
5) Read across this row from left to right until you find a value greater than your calculated r statistic.
6) The P–value for your observation is the P–value at the top of the first column to the left of your value.
   e.g. if r for df = 15 is 0.523, then $P < 0.025$ for a One-Tailed Test; if r is 0.599, then $P < 0.01$.
7) A $P < 0.05$ (or smaller) value indicates that you can reject the null hypothesis that the two variables are not correlated. In other words, you have evidence the variables are significantly related. If your r statistic value lies to the left of the 0.05 column, then your results are not significant (n.s. $P > 0.05$). You cannot reject the null hypothesis that the variables are unrelated.

CONTOH LEMBAR VALIDASI RPP

Petunjuk:

1. Anda akan diminta untuk memberikan penilaian atau validasi terhadap RPP.
2. Penilaian RPP ini dilakukan dengan memberikan tanda silang (X) dalam kolom skor sesuai dengan kemampuan praktikum dengan memperhatikan rambu-rambu skoring.
3. Kriteria penilaian: Kurang baik:1, Cukup baik: 2, Baik: 3, Sangat baik: 4.

No	Komponen Rencana Pembelajaran	Skor			
		1	2	3	4
I.	**Perumusan Indikator Belajar**				
	1. Kejelasan rumusan.				
	2. Kelengkapan cakupan rumusan indikator.				
	3. Kesesuaian dengan kompentensi dasar.				
	4. Kesesuaian dengan standar kompetensi.				
II.	**Pemilihan dan Pengorganisasian Materi Pembelajaran**				
	1. Kesesuaian dengan kompetensi yang akan dicapai.				
	2. Kesesuian dengan karakteristik perserta didik.				
	3. Keruntutan dan sistematika materi.				
	4. Kesesuaian materi dengan alokasi waktu.				
III.	**Pemilihan Sumber Belajar/Media Pembelajaran**				
	1. Kesesuaian sumber belajar/media pembelajaran dengan kompetensi (tujuan) yang ingin dicapai.				
	2. Kesesuaian sumber belajar/media pembelajaran dengan materi pembelajaran.				
	3. Kesesuaian sumber belajar/media dengan karakteristik peserta didik.				
IV.	**Skenario/Kegiatan Pembelajaran**				
	1. Kesesuaian strategi dalam metode pembelajaran dengan kompetensi (tujuan) pembelajaran.				
	2. Kesesuaian strategi dan metode pembelajaran dengan materi pembelajaran.				
	3. Kesesuaian strategi dan metode pembelajaran dengan karakteristik peserta didik.				

	4. Kelengkapan langkah-langkah dalam setiap tahapan pembelajaran dan kesesuaian dengan alokasi waktu.				
V.	**Penilaian Hasil Belajar**				
	1. Kesesuaian teknik penilaian dengan kompeensi yang ingin dicapai.				
	2. Kejelasan prosedur penilaian (awal, proses, akhir, tindak lanjut).				
VI.	**Penggunaan Bahasa Tulis**				
	1. Ketepatan ejaan.				
	2. Ketepatan pilihan kata.				
	3. Kebakuan struktur kalimat.				
	4. Bentuk huruf dan angka baku.				
VII.	**Penilaian Validasi Umum**				
	Penilaian atau validasi umum terhadap instrumen.	a	b	c	d

Keterangan:
a = dapat digunakan tanpa revisi, b = dapat digunakan dengan sedikit revisi, c = dapat digunakan dengan banhak revisi, d = belum dapat digunakan.

Catatan:.....
.....
.....

Palembang, ...
Validator II,

M. Holandyah, M.Pd
NIP.197405017201101 1 001

CONTOH LEMBAR VALIDASI INSTRUMEN TES

Petunjuk:

1. Anda akan diminta untuk memberikan penilaian atau validasi terhadap instrumen tes.
2. Pengisian instrumen validasi ini dilakukan dengan memberikan tanda cek (√).

No	U r a i a n	Validasi			
		1	2	3	4
I.	**Aspek Petunjuk**				
	a. Petunjuk tes, dinyatakan jelas.				
	b. Kriteria skor yang diberikan, dinyatakan jelas.				
II.	**Aspek Cakupan Tes Prestasi Kognitif**				
	a. Butir-butir pertanyaan pada prestasi kognitif, dinyatakan dengan jelas.				
	b. Pilihan jawaban pada tes prestasi, dinyatakan dengan jelas.				
	c. Pilihan Materi sesuai dengan kemampuan siswa				
	d. Pilihan topic pada soal menulis sesuai dengan materi dan dinyatakan dengan jelas				
III.	**Aspek Bahasa**				
	a. Menggunakan bahasa yang sesuai dengan kaidah bahasa yang benar.				
	b. Rumusan pernyataan komunikatif.				
	c. Menggunakan kalimat dan kata-kata yang mudah dipahami.				
IV.	**Penilaian Validasi Umum**	**a**	**b**	**c**	**d**
	Penilaian atau validasi umum terhadap instrumen.				

Keterangan:

1	=	Kurang baik	a	=	Dapat digunakan tanpa revisi.
2	=	Cukup baik	b	=	Dapat digunakan dengan sedikit revisi.
3	=	Baik	c	=	Dapat digunakan dengan banyak revisi.
4	=	Baik sekali	d	=	Belum dapat digunakan.

Catatan:

.....
.....
.....
.....
.....

Palembang, ...
Validator I,

M. Holandyah, M.Pd
NIP.197405017201101 1 001

CONTOH LEMBAR VALIDASI AHLI MATERI

Petunjuk:

1. Lembar evaluasi ini dimaksudkan untuk mengetahui pendapat Bapak/Ibu sebagai Ahli Materi tentang pembelajaran bahasa Inggris dengan menggunakan strategi pembelajaran yang saya gunakan.
2. Pendapat, kritik, saran, penliaian dan komentar Bapak akan sangat bermanfaat untuk memperbaiki dan meningkatkan kualitas program pembelajaran ini.
3. Komentar atau saran Bapak/Ibu mohon ditulis pada lembar yang disediakan. Apabila tidak mencukupi, mohon ditulis pada kertas tambahan yang disediakan.
4. Atas kediaan Bapak/Ibu untuk mengisi lembar evaluasi ini, diucapkan terima kasih.
5. Mohon beri tanda centang (√) pada angket dibawah ini, dengan keterangan sebagai berikut:
   Sangat baik = 5, Baik = 4, Cukup = 3, Kurang = 2, Sangat Kurang = 1.

No	Aspek Penilaian	Pertanyaan	Jawaban				
			1	2	3	4	5
1.	Aspek kesesuaian dengan tujuan pembelajaran	Kesesuaian dengan silabus					
		Adanya relevansi dengan kemampuan siswa					
		Kejelasan topik pembelajaran					
		Keruntutan materi yang tepat					
		Cakupan materi yang tepat					
		Ketuntasan materi					
		Adanya elevansi bagan dan gambar dengan materi					
2.	Aspek kemudahan siswa menyerap pelajaran melaui strategi pengajaran	Kemudahan siswa memahami materi melalui strategi pengajaran					
		Kesesuaian antara desain strategi pengajaran dengan tingkat berfikir siswa					
		Kemudahan interaksi guru menerapkan strategi pengajaran					

	Penilaian Validasi Materi		a b c

Kesimpulan:

1. Layak untuk diproduksi tanpa revisi.
2. Layak untuk diproduksi dengan revisi
3. Tidak layak produksi.

(Mohon beri tanda lingkaran pada nomor sesuai dengan kesimpulan)

Saran:

.....
.....
.....
.....
.....

Palembang, ...
.
Evaluator I,

M. Holandyah, M.Pd
NIP.197405072011011001

BIBLIOGRAPHY

The writer's name is **Muhammad Holandyah, S.Pd., M.Pd.** He is the seventh children in his family who was born in Ngulak village, the regency of Musi Banyuasin (MUBA) in South Sumatra province on May, 1974. He finished his S1 degree of English education study program on September, 1999 at Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) PGRI Palembang and his S2 degree in master of language education in English education study program at University of PGRI Palembang in 2009. He is married with Mellyza, SE in 2002 and now he is a father of his four children; Safhan Rakha, Davyna Dwi Amanda, and his twin son; Dava Avandindra Andika, Davy Avandindara Aditya.

His first teaching of English was started in 1996 at some English courses in Palembang. Since that time, on September, 1999, he stated his career in teaching of English as a permanent lecturer at STKIP PGRI Palembang. Besides that, he also taught English subject in some universities as his part-time lecturer in Universitas Kader Bangsa (UKB), Universitas Tamansiswa (UNITAS), Universitas Terbuka (UT), Akademi Bahasa Inggris dan Bisnis Bina Insan Indonesia (ABA-BII) Palembang, and Akademi Kebidanan (AKBID) Abdurrahman Palembang. Now, the writer is a permanent lecturer of English education study program in Tarbiyah faculty of Raden Fatah State Institution of Islamic Studies (RAFSIIS) Palembang.

>><<