

**ARCHETYPAL EFL READERS:  
PRELIMINARY EMPIRICAL EVIDENCE SUBSTANTI-  
ATED FROM SELECTED DISCRIMINATING VARI-  
ABLES**

**Gunadi H. Sulisty**

*Universitas Negeri Malang, Malang*

**Suharmanto**

*Universitas Negeri Malang, Malang*

**Abstract:** The study is an attempt to empirically examine factors that differentiate EFL readers with different EFL reading proficiency levels. Four selected factors believed to play a role in reading comprehension are considered, namely: linguistic knowledge, reading strategies, text structure knowledge, content and world background knowledge. Discriminant analysis was employed to scrutinize the data collected on these variables. The analysis wraps up, revealing that advanced EFL readers, while they are sufficiently equipped with formal schemata that necessarily embrace both linguistic knowledge and text structure knowledge, can be speculated to consist of three groups: those who fail to activate their relevant schemata, those who fail to use their reading strategies, and those who can function both content and world background knowledge and their reading strategies with ease.

**Key words:** EFL readers, variables, discriminating

Much has been attempted in exploratory searches, be they at conceptual and/or empirical levels, to reveal the secrets hidden in the factors that constitute efficient and effective readers and the mechanisms these factors actually come into play. In the conceptual level, for instance, exploratory searches have been exerted, yielding dissimilar conceptualizations regarding reading comprehension. These have been known as bottom-up, top-down, interactive, and interactive compensatory views (Rayner and Pollatsek, 1989). Several empirical studies have also been taken from

several perspectives. For instance, Grabe (1991) examines variables that play a role in reading comprehension. He identifies six component skills and knowledge areas that are believed to play a significant role in reading comprehension. These are: automatic recognition skills, vocabulary and structural knowledge, formal discourse structure knowledge, content and world background knowledge, synthesis and evaluation skills/strategies, and meta cognitive knowledge and skills monitoring. Nist and Mealey (1991) examine text types and comprehension. They confirm that readers generally recall information better from certain types of text-organization. Devine (1988) confirms that there is a significant correlation between gain in language competence and the use of effective reading strategies (as defined by both the type of oral reading strategies used and successful comprehension). More recently Suharmanto (2006) has examined patterns of relationship between several factors related with EFL reading comprehension. It was revealed in his study that five paths were empirically evidenced among linguistic knowledge, reading strategies, text structure knowledge, content and world background knowledge kept as independent variables and EFL reading comprehension as a dependent variable. These patterns are those associations (1) between content and world background knowledge and EFL reading comprehension; (2) between linguistic knowledge and EFL reading comprehension; (3) mediated by reading strategies, linguistic knowledge and EFL reading comprehension; (4) mediated by text structure knowledge, linguistic knowledge and EFL reading comprehension; and (5) mediated by both text structure knowledge and reading strategies respectively and EFL reading comprehension.

While these studies are useful in providing differing perspectives in understanding reading comprehension, the secrets are yet not satisfactorily uncovered. Several theories have been convincingly offered and enormous findings have been empirically revealed. Yet, their contributions to the successful EFL reading classes have not been significantly recognized to be as effective. Therefore, unrelenting attempts to understand factors in reading comprehension controlling reading comprehension practices in the classroom need to be endeavored.

As shown in the studies described previously, no single study has ever been made as an attempt to examine typical readers seen from important factors that are believed to play a role in reading comprehension. Examination of the readers' profile is without doubt useful. It is expected that such an attempt would constitute a valuable conduct of inquiry which provides empirical evidence as how some factors characterize readers of different levels of ability. Objective information concerning this aspect is projected to be beneficial on how to address learners' learn-

ing needs of reading comprehension more accurately. With respects to these, the current study is an attempt made to address several issues as follows: Do poor EFL readers, average EFL readers and good EFL readers really differ when they are observed from their linguistic knowledge, text structure knowledge, content and world background knowledge, and reading strategies? If so, which factor(s) determine(s) the distinction: their linguistic knowledge, text structure knowledge, content and world background knowledge, or reading strategies employed? Next, what is the established discrimination function like? Finally, how accurate is the function to classify EFL readers of different competencies?

### **THEORETICAL OVERVIEW**

Reading comprehension is a form of communication that involves a writer and a reader through written media of language. It is a kind of interaction between a writer and a reader mediated by means of written form of language. Viewed from this perspective, reading -- once considered a passive activity while it is actually a dynamic process residing in complicated physical and mental mechanisms (Nuttall, 1985), is a form of silent communication (Pugh, 1978) between a writer as a message sender and a reader as a message receiver by means of a written text. This way, on one side a writer encodes messages through a text as a medium of communication; on another side a reader processes the messages using decoding mechanisms.

In this mode of communication, actually a two-or probably three-level interaction can be identified: the first level is that between a writer and a text; the second is that between a reader and a text and the third between a writer and a reader. Reading as is commonly perceived refers to the second level. Communication in this regards i.e. conveying intended messages is considered accomplished completely when the messages contained there in the written text are successfully reconstructed fully on the part of the reader. A similar mode of information transfer is also discussed from the angle of cognitive processes (Garrod, 1999:409).

Seen on the part of a reader, reading also involves a number of successive mechanisms. Reading activities set out from physical and visual activities and come up to highly mental processes. The physical activities include eye movements. Mental mechanisms take in word recognition and mental processes that relate words and the experiences they represent. Eye movements in reading play the first substantial role. It is through the eyes that messages contained in the written text begin to be actively processed by the reader. Constraints that impede in this

stage embrace subsequent processes, which in an extreme context (Rayner, 1983:1). It can be stated that reading cannot be facilitated without the role of eyes. The next stage is word recognition. To recognize words, eye movements are essentially required (Smith, 1982). As such, word recognition relates to eye movements. In this stage in order to recognize words, first, eye movements are meant to explore forms of words. Subsequent processes follow when the condition at this stage is fulfilled. The requirement is the share between the language code in the text and that the reader has. This condition is inevitable. Further exploratory activities involve those analyzing the structure of the word, which is then followed with analyses on larger levels of language such as the structure of phrases and sentences. During this process, analyses are probably addressed to aspects such as accompanying illustrative figures or facts and contexts as well as vocabulary use.

The next stage is the stage at which words are associated with the readers' experiences with the concepts that represent the words. This implies that word recognition alone is not sufficient for comprehension in reading to take place. Efficient readers need to hang on to concepts, ideas and worldly experiences they have already held related to the words the reader encounters (Miller, 1971). Smith (1982:152) argues that immediate meaning identification of a word at a time is equally unrelated to the letter identification of a word. This means that in reading comprehension, word recognition and letter recognition play a role as a mediating agent from which the meaning is understood.

Seen from the theory of communicative language ability, reading comprehension is a language skill that involves a number of interacting factors to make it happen (Goodman, 1988). As stated previously, Grabe (1991) theorizes the crucial role of six components, skills and knowledge areas in reading comprehension: automatic recognition skills, vocabulary and syntactic knowledge, content and world background knowledge, formal discourse structure knowledge (formal schemata), synthesis and evaluation skills and strategies, and meta cognitive knowledge and skills monitoring.

Eskey (1988) emphasizes that good readers know the language. They can decode with occasional exceptions, and they do so, for the most part, not by guessing from context or prior knowledge of the world, but by a kind of automatic identification that requires no conscious cognitive effort. This automation makes it possible for the fluent readers to free up the mind of the language to think about and interpret what they are reading – that is, to employ higher-level, top-down strategies such as the use of schemata and other kinds of background knowledge. Poor readers, he further argues, are just like good readers. They both rely on the use of prior

knowledge in deciphering text. However, unlike the good readers who use this top-down strategy to interpret the text, the poor readers use the top-down strategy to interpret the language, namely word and syntax.

The discussion above suggests that at least there are several factors that play a substantial role in reading comprehension, namely: content and world background knowledge, text structure knowledge, reading strategies, and linguistic knowledge. These factors constitute the interest of the present study.

## METHOD

The current study is an *ex post facto* study. No attempt was made to affect the variables under investigation. Thus, the data obtainable from variables considered in the study: knowledge of the world, linguistic knowledge, text structure knowledge, reading strategies and EFL reading comprehension had already existed as they were when the data on these variables were collected. The data were basically those data re-analyzable from the data of Suharmanto's study (Suharmanto, 2006).

The target population of the current study was the students of the English department who had completed all requisite skills courses: listening, speaking, reading, and writing. The accessible population of the study was 53 (fifty three) students of the English Department, Faculty of Letters State University of Malang, who had finished all reading classes: Reading I, II, III and IV with the total of 14 credits. All these 53 (fifty three) students were then considered as the accessible population. Since all members of the accessible population were involved in the study, no sampling technique was applied. Thus, the current study employs a census, treating all these members as the sample for the current study.

To collect the data on variables of interest: four sets of data-collection instruments were utilized. The instruments were in the form of both tests and non-tests that closely corresponded to the nature of the data needed. Variables that follow: EFL reading comprehension, linguistic knowledge, content and world background knowledge, and text structure knowledge involve abilities in these areas because these variables require the activation of competences. These competences were latent and could be solicited by way of these tests and were reflected in their corresponding maximum performances of the subjects in order to be observed explicitly. Therefore, data-collecting instruments in the form of a test were required. The variable 'reading strategies' was concerned with typical performances of the subjects which were subject to temporary conditions. The state of these performances was then shown in the form of tendency, that is, the inclination of utilization

of reading strategies. To elicit the necessary data on this variable, a non-test technique called retrospective probing verbal report was employed.

Excepting partly the test of linguistic knowledge, all the other instruments were developed by Suharmanto (2006). The test of linguistic knowledge comprised two sub tests: test of vocabulary size and test of grammar. The test of vocabulary size had been adapted by Kweldju (2000) from a vocabulary size test known as Vocabulary Level Test (VLT) developed by Nation (1990) which was then employed in the current study; whereas the test of grammar had been developed by Suharmanto (2006).

The test of EFL reading comprehension consisted of two expository texts, each of which was followed with a set of comprehension questions in the multiple-choice format. The sub test of linguistic knowledge, i.e. the vocabulary sub test took a matching form, whereas the grammar sub test was in different formats such as identification, completion, rewriting and multiple-choice types depending on the aspects to be tested. The test of content and world background knowledge took the form of a paper-and-pencil question-and-answer format; the test of text structure knowledge was in the form of a question-and-answer format with short answers; whereas reading strategies were assessed using the paper-and-pencil form of questionnaire format with open-ended questions.

Prior to administration, all the instruments were informally trialed. The purpose was to examine the quality of the data-collection instrument. The results demonstrated that in general the instruments were good in terms of their reliability and validity. The reliability of the instruments used in the study that demonstrated each instrument used was considered sufficient as data-collection tools as was indicated by values of reliability coefficients as well as interscorer reliability estimates during the trial (Reliability coefficients: Reading Comprehension Test: K-R 20,  $r = .8900$ ,  $p = .05$ ; Vocabulary Test: Alpha,  $r = .7518$ ,  $p = .05$ ; Grammar Test: K-R 20,  $r = .8800$ ,  $p = .05$  and Alpha  $r = .6325$ ,  $p = .05$ ; Content and World Background Knowledge Test: Alpha,  $r = .8056$ ,  $p = .05$ ; Text Structure Knowledge Test: Alpha,  $r = .8121$ ,  $p = .05$ ; and Reading Strategies Test: K-R 20,  $r = .8500$ ,  $p = .05$ . Inter scorer reliability estimates: Content and World Background Knowledge: Pearson Correlation,  $r = .941$ ,  $p = .05$ ; Text Structure Knowledge:  $r = .779$ ,  $p = .05$ ; and Reading Strategies:  $r = .649$ ,  $p = .05$ .)

Similarly, analysis on the item validity of the tests used also revealed that most items in the tests demonstrated sufficient values for validity requirements (Reading Comprehension Test:  $r$  values ranging from .4355 to .6918; Vocabulary Test:  $r$  values ranging from .4090 to .6946 at  $p = .05$ ; Dichotomous-Item Grammar

Test: r values ranging from .3924 to .7165; Non-Dichotomous-Item Grammar Test: r values ranging from .4896 to .7058, two invalid items with  $r = .2251$  and  $r = .1803$  were revised; Content and World Background Knowledge Test: r values ranging from .4072 to .6552; Text Structure Knowledge Test: r values ranging from .4002 to .6707.)

The data was collected by administering all the data-collection instruments that had been validated, and the test parts which needed improvement were revised in line with the results of validation. The procedure was as follows. First, the pre reading test was administered first to tap the students' background knowledge of the text content (1 hour). Next, vocabulary and structure test was administered to get the data on the students' linguistic knowledge (45 minutes). Then, the administration of text structure test was performed to collect the data on the students' knowledge of the text structure (45 minutes). Finally, the reading comprehension test was administered (1 hour) followed by retrospective probing for reading strategies (1 hour). Between one administration to another there was a 15 minute break to allow students to alleviate their mental and physical tension arising due to data collection. The data collection started at 08.30. In all, the tests administration took effectively 4 hours 30 minutes to accomplish.

To answer the research questions, discriminant analysis was applied to the collected data. However, prior to main data analyses, the data were first analyzed to examine the fulfillment of the assumptions required for running discriminant analysis. The assumptions were concerned with normality, linearity, multicollinearity, and homoscedasticity (Hair, et al., 1998:259). These requirements should be satisfactorily met in order that the patterns of correlations of variables can be meaningfully interpreted. The results of the examination on these requirements revealed that, the data in each variable of interest were normally distributed: linguistic knowledge ( $z$  value = 1.576 significant at .025), content and world background knowledge ( $z$  value = 1.817 significant at .017), text structure knowledge ( $z$  value = 1.695 significant at .019), reading strategies ( $z$  value = 1.894 significant at .006), and EFL reading comprehension ( $z$  value = 1.842 significant at .010).

The testing of linearity of the data was performed by regressing each of the independent variables. The results indicated that there was evidence of linearity of the variables under study (linguistic knowledge: calculated coefficient of linearity ( $t$ ) = 9.091 significant at .000; content and world background knowledge: computed coefficient of linearity ( $t$ ) = 8.865 significant at .000; text structure knowledge: examined coefficient of linearity ( $t$ ) = 9.166 significant at .000; and reading strategies: observed coefficient of linearity ( $t$ ) = 10.137 significant at .000). The

next requirement was examination on multicollinearity among the independent variables. The analysis was aimed at examining whether there were correlations among independent variables. It was revealed that no multicollinearity was observed in the data (linguistic knowledge: tolerance = .372, VIF = 2.689; content and world background knowledge: tolerance = .448, VIF = 2.232; text structure knowledge: tolerance = .385; VIF = 2.596; reading strategies: .316, VIF = 3.168).

Finally, homoscedasticity of the data in each variable was also evidenced to be satisfactorily met as shown by the scatter plot of the data against residual standards in which the plots were scattered forming no patterns in each variable.

Data analyses were performed by using *SPSS for Windows Version 11.0*. The procedure was as follows. Students' performance in EFL reading comprehension was used as a basis to categorize them into three groups of abilities: Group I: Poor EFL Readers, Group II: Average EFL Readers, and Group III: Good EFL Readers. This is to say, students' EFL reading comprehension was established as the dependent variable; whereas the other four variables, linguistic knowledge, content and world background knowledge, text structure knowledge, and reading strategies, were put as the independent variables. The procedures of the analysis were carried out in the order as follows (Norusis, 1993; Tabachnick and Fidel, 1989:505-594): performing tests of equality of group means, estimating the discriminant function model using the simultaneous method, scrutinizing the strength of the function, examining the distance of the difference existing among the groups, observing the accuracy of the two discriminant functions, identifying function identity, analyzing values in discriminant mathematical equations, and examining the accuracy of classification.

## **FINDINGS**

The presentation of the findings is structured on the basis of the formulation of the research problems. The following section describes the results of the analyses in the order of the formulation of the problems.

### **Results of Analysis Addressing Problem 1**

To respond to question 1 inquiring whether poor EFL readers, average EFL readers and good EFL readers really differ when they are observed from their linguistic knowledge, text structure knowledge, content world background knowl-



edge, and reading strategies employed, equality of group means was examined. The examination aims at analyzing whether there were significant differences among the three groups viewed from the variables: linguistic knowledge, content and world background knowledge, text structure knowledge, and reading strategies from EFL reading comprehension. The results of the analysis were presented in Table 1.

**Table 1: Tests of Equality of Group Means**

Variables	Wilks' Lambda	F	df1	df2	Sig.
LANGUAGE	.562	19.514	2	50	.000
TXSTRUCT	.588	17.549	2	50	.000
BCKGROUN	.458	29.580	2	50	.000
STRATEGY	.481	26.962	2	50	.000

Note:

language = linguistic knowledge

txstruct = text structure knowledge

bckgroun = world background knowledge

strategy = reading strategies

As shown in Table 1, the observed values (Sig.) are much smaller than  $< .05$  (Sig.  $< .05$ ) in all variables examined, being all significance values = .000. This means that there is a significant difference between the three groups of EFL readers. In other words, good EFL readers differ from average EFL readers, and poor EFL readers, and average EFL readers differ from poor EFL readers.

### Results of Analysis Addressing Problem 2

The answer to question 1 was positive, meaning that differences were observed among good, average, and poor EFL readers. This answer leads to the next question: which factor(s) determine(s) the distinction: their linguistic knowledge, text structure knowledge, content world background knowledge, or reading strategies? To examine further whether variables linguistic knowledge, content and world background knowledge, text structure knowledge, and reading strategies can really constitute a discriminant function, a further analysis was performed. In the analysis, all these variables were included. This was performed because in the current study there is no attempt 'in seeing intermediate results based on only the most

discriminating variables'. Thus, the method used to estimate the discrimination model was simultaneous method. (Hair, et al., 1998:260).

The results of the analysis on the model estimation are summarized in Tables 2 and 3.

**Table 2: Independent Variables Analyzed**

Step	Entered	Removed	Min. D Squared					
			Statistic	Between Groups	Exact F		Sig.	
				Statistic	df1	df2		
1	TXSTRUCT		.971	2 and 3	8.711	1	50.000	4.804E-03
2	STRATEGY		1.320	1 and 2	5.499	2	49.000	7.005E-03
3	BCKGROUN		2.570	2 and 3	7.380	3	48.000	3.659E-04
4	TXSTRUCT		2.358	2 and 3	10.367	2	49.000	1.760E-04

At each step, the variable that maximizes the Mahalanobis distance between the two closest groups is entered.

- a Maximum number of steps is 8.
- b Maximum significance of F to enter is .05.
- c Minimum significance of F to remove is .10.
- d F level, tolerance, or VIN insufficient for further computation.

Table 2 indicates that four steps were needed to perform the analysis. This means that step 4 constituted the final step in which the discrimination model was established. It is also obvious from the table that the analysis included independent variables: text structure knowledge (step 1), reading strategies (step 2), world background knowledge (step 3), but excluded text structure knowledge (step 4).

Through these steps, surprisingly linguistic knowledge was excluded. Besides, each step yielded information concerning two closest groups when an independent variable was entered into the analysis. For instance, in step1, when independent variable text structure knowledge was entered, the two groups that were closest are Groups 2 and 3 (average EFL readers and good EFL readers respectively) and so on, all with significance (Sig. < .00). All this implies that the attention further on can be focused on the results of analysis performed in step 4.

Table 3 specifies the independent variables put in each stage of the analysis in relation with the groups in the dependent variables, groups of EFL readers. In step 4 particularly as shown in Table 3 reading strategies and world background knowledge constituted discriminant functions in the model. The independent variable

‘reading strategies’ was discriminating particularly for Groups 2 and 3 of EFL readers: average EFL readers and poor EFL readers respectively; whereas the independent variable ‘world background knowledge’ discriminated Groups 1 and 2 of EFL readers: good EFL readers and average EFL readers respectively.

**Table 3: Variables in the Analysis<sup>1</sup>**

Step	Variables	Tolerance	Sig. of F to Remove	Min. D Squared	Between Groups
1	TXSTRUCT	1.000	.000		
2	TXSTRUCT	.837	.075	.867	1 and 2
	STRATEGY	.837	.001	.971	2 and 3
3	TXSTRUCT	.829	.189	2.358	2 and 3
	STRATEGY	.596	.007	1.052	2 and 3
	BCKGROUND	.642	.001	1.320	1 and 2
4	STRATEGY	.648	.001	.336	2 and 3
	BCKGROUND	.648	.000	.867	1 and 2

The discrimination function of the model as shown in the previous tables was further confirmed in the next analysis to scrutinize the strength of the function. The results of further analyses on the model are summarized in Table 4. As shown in the table and consistent with the previous analyses, there were four steps taken in the analysis with the number variable(s) processed in each step. In step 4 two variables were held in the model, which, as shown in previous tables, were necessarily reading strategies and world background knowledge. Besides, along with the steps advanced, the significance of the discrimination function also increased from Sig. 1.684E-06 observed in step 1 to Sig. 7.171E-11 as the highest. This means that the independent variables in the model: reading strategies and world background knowledge were obviously discriminating for different types of EFL readers: poor, average, and good.

In addition to that, however, the values of Wilks' Lambda decreased relatively from a figure of .588 in step 1 to a figure of .340 in step 4. This means that with regards to step 1 in which the value of Wilks' Lambda for one independent variable, text structure knowledge, figures at .588, the amount of 58.8% variance existing in this model cannot be accounted for by the difference among types of EFL readers. This sizeable amount of unexplainable variance, however, decreased to 34% as

<sup>1</sup> Variables in *Not the Analysis* is presented in Appendix 1.

shown in the value of Wilks' Lambda for two independent variables: world background knowledge and reading strategies which figure at .340 occurring at step 4. This implies that 66% of variance in the model was explainable by the contrast in different types of EFL readers.

**Table 4: Statistics Confirming the Model**

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.588	1	2	50	17.549	2	50.000	1.684E-06
2	2	.433	2	2	50	12.743	4	98.000	2.175E-08
3	3	.317	3	2	50	12.404	6	96.000	2.674E-10
4	2	.340	2	2	50	17.510	4	98.000	7.171E-11

As has been demonstrated in the previous sections, after the fourth analytical step, differences among types of EFL readers were observed. Also, the model that included the variables characterizing different types of EFL had also been confirmed. The next step to perform then was to compare different types of EFL readers in terms of the underlying variables characterizing the model. In this analysis an examination on the distance of difference by pairing types of EFL readers was attempted to be sought. Through this analysis the extent to which different types of EFL readers in pair can exhibit differences was demonstrated.

Table 5 summarizes the analysis to examine the distance of the difference existing among different types of EFL readers. Like other previous analyses, this analysis also included all necessary steps taken to accomplish it. Thus, the results of the analysis examining the distance existing among different types of EFL readers included those generated from each step.

As shown in Table 5 the values indicating existence of distance that exhibited differences represented by F values as well as the level of significance (Sig.) are presented. More specifically, in stage 4 as the last step, it is obvious that again the differences in all pairs of comparison were significant (Sig. = .000). In addition, the widest distance was observed in a pair between poor EFL readers and good EFL readers ( $F = 33.135$ ; Sig. = .000) while the narrowest distance occurred in a pair between average EFL readers and good EFL readers ( $F = 10.367$ ; Sig. = .000). This means that good EFL readers differed markedly from poor EFL readers in terms of reading strategies and world background knowledge. The slightest difference was

observed between average EFL readers and good EFL readers. The difference between average EFL readers and poor EFL readers was fair.

**Table 5: Pairwise Group Comparison**

Step	NTILES of DCOMPR		Poor	Average	Good
1	Poor	F		8.358	35.067
		Sig.		.006	.000
	Average	F	8.358		8.711
		Sig.	.006		.005
	Good	F	35.067	8.711	
		Sig.	.000	.005	
2	Poor	F		5.499	31.006
		Sig.		.007	.000
	Average	F	5.499		10.522
		Sig.	.007		.000
	Good 3	F	31.006	10.522	
		Sig.	.000	.000	
3	Poor	F		10.672	24.334
		Sig.		.000	.000
	Average	F	10.672		7.380
		Sig.	.000		.000
	Good	F	24.334	7.380	
		Sig.	.000	.000	
4	Poor	F		15.266	33.135
		Sig.		.000	.000
	Average	F	15.266		10.367
		Sig.	.000		.000
	Good	F	33.135	10.367	
		Sig.	.000	.000	

a 1.50 degrees of freedom for step 1. b 2.49 degrees of freedom for step 2.

c 3.48 degrees of freedom for step 3. d 2.49 degrees of freedom for step 4.

The present analysis put EFL readers into three groups: poor EFL readers, average EFL readers and good EFL readers. This follows that two discriminant functions were established: Discriminant Function 1 and Discriminant Function 2. Discriminant functions were useful to examine the closeness between an individual case with the group in which the case can be appropriately put. With Discriminant

Function 1 and Discriminant Function 2, this necessitates that the former was to put a case into membership of Group 1 and the latter put a case into membership of Group 2. The results of the analysis in this respect are summarized in Table 6.

**Table 6: Summary of Analysis on Canonical Discriminant Functions**

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.360	84.7	84.7	.759
2	.246	15.3	100.0	.444

a First 2 canonical discriminant functions were used in the analysis.

As indicated in Table 6, Discriminant Function 1 has a strong correlation coefficient (.759); Discriminant Function 2 has a moderate correlation coefficient (.444) (cf. Fraenkel and Wallen, 1993:296). This means that Discriminant Function 1 can be used to predict the membership of individual cases into their appropriate group better than Discriminant Function 2. However, both functions are necessarily useful for classifying these individual cases into their appropriate group.

To observe the accuracy of the two discriminant functions that were established: Discriminant Function 1 and Discriminant Function 2, that is, whether they can be used to analyze the mean difference of the cases in each group, a further analysis, test of discriminant functions, was performed. The results of this analysis are presented in Table 7.

**Table 7: Wilks' Lambda Test of Discriminant Functions**

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1 through 2	.340	53.385	4	.000
2	.803	10.885	1	.001

Table 7 presents that in the heading 'test of functions 1 through 2' there was a significant difference in the means established by the two discriminant functions (Chi-square = 53.385; Sig. = .000). This implies that the characteristics of the three types of EFL readers: poor EFL readers, average EFL readers and good EFL readers were distinct. In other words, reading strategies and content and world background knowledge of the three types of EFL readers characterized the groups of EFL readers in different modes as has been evidenced in Table 2: the independent variable 'reading strategies' was discriminating particularly for Groups 2 and 3 of

EFL readers: average EFL readers and poor EFL readers respectively; whereas the independent variable ‘world background knowledge’ discriminated Groups 1 and 2 of EFL readers: good EFL readers and average EFL readers respectively. In addition, with test function 2, the statistics indicated a similar trend. This time, however, the focus was merely put to examine the difference between average EFL readers and good EFL readers. The figures in test of function 2 show that there was a significant difference in the means established by this discriminant function (Chi-square = 10.885; Sig. = .001). This again implies that the characteristics of the two types of EFL readers: average EFL readers and good EFL readers were distinct. As both functions demonstrate their significance, these two functions were potential to be used to classify cases of types of EFL readers into their appropriate groups.

### Results of Analysis Addressing Problem 3

The next analysis was aimed at identifying the function identity. In this analysis, membership of independent variables according to their respective discriminant function (s) was determined. This is demonstrated by the value that indicates the relationship between the independent variables in the analysis and the discriminant functions thus established, i.e. Discriminant Function 1 and Discriminant Function 2.

In the previous presentations it has been demonstrated that two independent variables were considered, and two discriminant functions had been established. The results of the analysis identifying the function identity are summarized in Table 8: Structure Matrix.

The information in Table 8 shows that two independent variables were consistently considered, namely: world background knowledge and reading strategies. The other two independent variables are not analyzed: language knowledge and text structure knowledge.

**Table 8: Structure Matrix**

	Function	
	1	2
BCKGROUN	.917*	-.398
STRATEGY	.865*	.503
LANGUAGE <sup>a</sup>	.534*	.049
TXSTRUCT <sup>a</sup>	.395*	.123

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.

\* Largest absolute correlation between each variable and any discriminant function

<sup>a</sup> This variable not used in the analysis.

Besides, figures showing the values of relationship between the independent variables and discriminant functions indicate that both reading strategies and world background knowledge (note that: language knowledge and text structure knowledge were not put into analysis) had higher correlation coefficients (.917 and .865 respectively) with Discriminat Function 1 than the correlation coefficients with Discriminat Function 2 (-.398 and .503 respectively). This means that both independent variables: reading strategies and world background knowledge were put into Discriminant Function 1.

When the identity of the discriminant functions was established, the next step was analyzing values in Discriminant Functions 1 and 2 for the purpose of creating components that constitute discriminant mathematical equations on the basis of the discriminant Z score as follows:

$$Z \text{ score Discriminant Function 1} = C + (r \times \text{CWBK}) + (r \times \text{RS})$$

$$Z \text{ score Discriminant Function 2} = C + (r \times \text{CWBK}) + (r \times \text{RS})$$

where:

C = constant

r = canonical coefficient

CWBK = content and world background knowledge

RS = reading strategies

These discriminant mathematical equations can be known when the information related with canonical discriminant function coefficients is available. The results of the canonical correlation analysis for the purpose are presented in Table 9.



**Table 9: Canonical Discriminant Function Coefficients**

	Function	
	1	2
BCKGROUND	.156	-.269
STRATEGY	.049	.113
(Constant)	-8.444	2.168

Unstandardized coefficients<sup>2</sup>

Based on the data presented in Table 9, the discriminant mathematical equations created are formulated as follows:

Z score Discriminant Function 1 =  $-8.444 + (.156 \text{ CWBK}) + (.049 \text{ RS})$

Z score Discriminant Function 2 =  $2.168 + (-.269 \text{ CWBK}) + (.113 \text{ RS})$

These two mathematical discriminant functions were used to classify cases into their appropriate groups. The mathematical discriminant function f (Z score Discriminant Function 1) =  $-8.444 + (.156 \text{ CWBK}) + (.049 \text{ RS})$  was used to classify EFL readers into poor or average category; while the mathematical discriminant function f (Z score Discriminant Function 2) =  $2.168 + (-.269 \text{ CWBK}) + (.113 \text{ RS})$  was used to classify EFL readers into average or good category.

The points in which cases in each group of EFL readers tend to cluster are called centroids. As there were three groups of types of readers and two discriminant functions, there were three centroids. Based on the analysis, the centroids for each group according to respective discriminant functions are known from the values as shown in Table 10.

Table 10 shows that the centroids of the three types of EFL readers as seen from Discrimination Function 1 and Discrimination Function 2 are located at different places at territorial map<sup>3</sup>. The centroid of poor EFL readers is at -1.511 and .280; the centroid of average EFL readers is at .134 and -.699, and the centroid of good EFL readers is at 1.232 and .375 as viewed from Discrimination Function 1 and Discrimination Function 2 respectively.

<sup>2</sup> Standardized coefficients are presented in Appendix 2

<sup>3</sup> Territorial map is the projection of scores in the form of plots for all cases in each group that indicates areas of membership for each case, including the centroids of the existing groups.

**Table 10: Functions at Group Centroids**

Group	Discriminant Function	
	1	2
1	-1.511	.280
2	.134	-.699
3	1.232	.375

Unstandardized canonical discriminant functions evaluated at group means

As shown in Figure 1, cases classified as poor EFL readers occupy the areas indicated with number 1; cases classified as average EFL readers reside in the areas indicated with number 2; and cases classified as good EFL readers take up the areas indicated with number 3. Centroids for each group indicate the area around which most cases in a corresponding group will crowd together. In Figure 1 it is shown by an asterisk (\*).

Territorial map is useful for putting an individual case of an EFL reader in his/her appropriate group. For example, an EFL reader gets a score of 57 points at the independent variable content and world background knowledge (CWBK) and earns a score of 90 points at the independent variable reading strategies (RS). The reader's coordinating point in the map can be known by applying the discriminant mathematical equations that are already created as follows:

$$\text{Z score Discriminant Function 1} = -8.444 + (.156 \text{ CWBK}) + (.049 \text{ RS})$$

$$\text{Z score Discriminant Function 2} = 2.168 + (-.269 \text{ CWBK}) + (.113 \text{ RS})$$

Accordingly, based on these equations, the Z scores of each function can be calculated:

$$\text{Z score Discriminant Function 1} = -8.444 + (.156 \times 57) + (.049 \times 90)$$

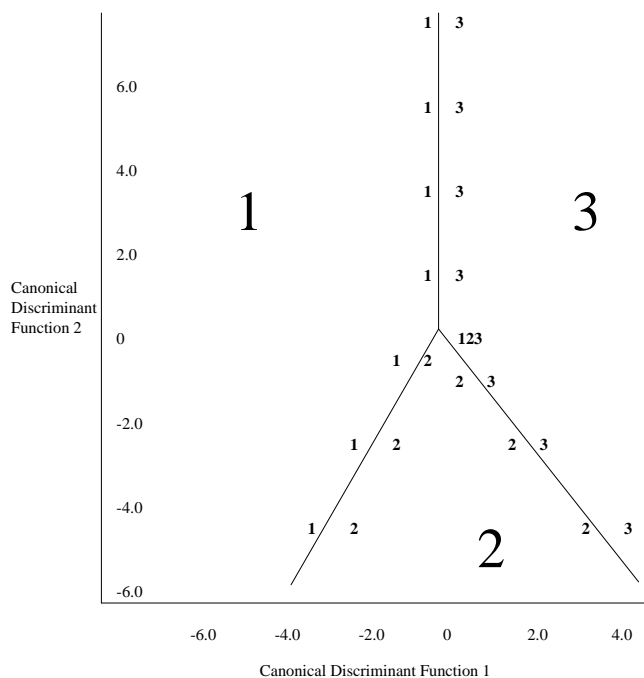
$$\text{Z score Discriminant Function 2} = 2.168 + (-.269 \times 57) + (.113 \times 90),$$

resulting in the following:

$$\text{Z score Discriminant Function 1} = 4.786$$

$$\text{Z score Discriminant Function 2} = -2.995$$

**Figure 1: Territorial Map**



Symbols used in territorial map

Symbol	Group	Label
1	1	Poor
2	2	Average
3	3	Good
*		Indicates a group centroid

Thus, the reader's coordinate in the territorial map is (4.786, -2.995). When this coordinate is plotted onto the territorial map, they will meet at a point in an area of grouping. This area belongs to area 3. This means that the reader is classified as a good EFL reader.

### Results of Analysis Addressing Problem 4

Problem 4 is concerned with the accuracy of the discriminant function that was established to classify cases of individual EFL readers of different competencies according to their appropriate groupings. The results of the analysis to address the problem are presented in Table 11.

**Table 11: Classification Results**

		NTILES of RDCOMPR	Predicted Group Membership			Total
			Poor	Average	Good	
Original	Count	Poor	12	3	2	17
		Average	1	13	3	17
		Good	0	5	14	19
	Percentage	Poor	70.6 %	17.6 %	11.8 %	100.0 %
		Average	5.9 %	76.5 %	17.6 %	100.0 %
		Good	.0 %	26.3 %	73.7 %	100.0 %
Cross-validated	Count	Poor	10	5	2	17
		Average	1	13	3	17
		Good	1	7	11	19
	Percentage	Poor	58.8 %	29.4 %	11.8 %	100.0 %
		Average	5.9 %	76.5 %	17.6 %	100.0 %
		Good	5.3 %	36.8 %	57.9 %	100.0 %

a Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b 73.6% of original grouped cases correctly classified.

c 64.2% of cross-validated grouped cases correctly classified.

As shown in Table 11, of 53 respondents in the present study, 17 were classified into the group of poor EFL readers; the other 17 into the group of average EFL readers; and the rest (19) were classified into the group of good EFL readers. In the original data, the number of members of poor EFL readers was predicted to be 13 (summing up the figures 12 + 1); the number of members of average EFL readers to be 21 (3 + 13 + 5); and the number of members of good EFL readers is predicted to be 19 (2 + 3 + 14). When cross validated, however, these figures change a bit, with poor EFL readers = 12 (10 + 1 + 1); average EFL readers = 25 (5 + 13 + 7); and good EFL readers = 16 (2 + 3 + 11). This change is confirmed by the figure

showing the percentage of classification (in point b. 73.6% or rounded up to 74%), which is considered to be highly accurate for correctly classifying original cases into their appropriate groups. The percentage figure in cross validation (point b. = 64.2%), which is close to the figure showing the percentage of classification of original data (in point b. 73.6%) further confirms the accuracy of the classification. In short, it can be stated that the discriminant functions that are established are accurate measures and thus they can be used to classify cases of individual EFL readers of different competencies according to their appropriate groupings with confidence.

## **DISCUSSION**

The results of the analysis yield an answer to each formulated problem. In short it can be stated that poor EFL readers, average EFL readers and good EFL readers are empirically shown to be substantially different. The factor that differentiates poor EFL readers and average EFL readers was their content and world background knowledge; whereas the factor that differentiated average EFL readers and good EFL readers was their reading strategies. Two discrimination functions were established: Discriminant Function 1 =  $-8.444 + (.156 \text{ CWBK}) + (.049 \text{ RS})$  and Discriminant Function 2 =  $2.168 + (-.269 \text{ CWBK}) + (.113 \text{ RS})$ . Finally, it is revealed that the functions established were (74%) accurate for classifying EFL readers of different competencies into their respective groups.

These findings are interesting to be discussed. First, the context of classification of EFL readers in the present study was higher learning with a particular reference to students of the English Department, Faculty of Letters State University of Malang. These students had finished all reading classes: Reading I, II, III and IV with the total of 14 credits when their data was collected on the variables under interest: EFL reading comprehension, linguistic knowledge, text structure knowledge, content and world background knowledge, and reading strategies. In addition, based on the catalog of the English Department, Faculty of Letters, State University of Malang (2005), the subjects of the study were programmed to have completed also all English skill courses other than reading courses as well as English grammar courses. With this in mind, the subjects were assumed to have advanced mastery of language components of English, and as readers they were supposed to have been beyond advanced level (Katalog Jurusan Sastra Inggris, 2005:18-19). In terms of their reading abilities, they were taken for granted (at least in the present

study) to have been at the level of reading beyond the lines (cf. McRae and Boardman, 1988). The subjects, thus, by design were advanced EFL readers.

These conditions presumably constitute some of the factors that can explain why the subjects' linguistic knowledge and text structure knowledge as advanced EFL readers did not make up a variable in the equation that plays a significant role in their EFL reading comprehension. Linguistic knowledge to them, which was measured by using grammar and vocabulary tests, had become a part of their mastery of English. Linguistic aspects of the texts seemingly did not pose a substantial problem to them. As such when they were posed to reading materials, their language proficiency processes the language aspects of the texts as automatic mechanisms, just like 'fluent native [readers who may activate formal schemata] automatically (Eskey, 1988:96). This follows then that the sufficient mastery of language proficiency smoothed the progress of the activation of text structure knowledge, hence constituting a basis for EFL comprehension (Suharmanto, 2006). When this was the case, then, the challenges remaining the advanced EFL readers face were content and world background knowledge they had and reading strategies they employed. This argument seems to be parallel and supported by the results of the study conducted by Hudson (1988:197) revealing that 'advanced level L2 readers in English apparently do have more facile or robust networks for fitting meaning than do lower level readers.'

The finding uncovering that the factor differentiating poor EFL readers from average EFL readers was their content and world background knowledge is also worth addressing. One possible explanation to this can be referred to the theory of schema (Rumelhart, 1980:33; Carrell and Eisterhold, 1983). Carrell (1988:105) argues that '... schema availability alone is not a sufficient condition for adequate comprehension.' Activation of the relevant schema is necessary (Carrell and Eisterhold, 1983 as quoted in Carrell (1988:105). It is argued that these poor readers in the present study, rather than lacked content schemata or/and formal schemata, seemed to fail to activate their relevant schemata as measured in the content and world background knowledge test. Several studies have revealed similar findings (Carrell, 1983, and Carrell and Eisterhold, 1983 quoted in Carrell, 1988:105).

Finally, it is also interesting to discuss the finding that the factor that differentiated average EFL readers and good EFL readers was their reading strategies. While using the same data analyzed quite differently statistically, the finding reflects Suharmanto's findings. Suharmanto (2006) reveals that two possible paths that EFL comprehension ultimately takes place are those as follows: one which initiates with formal schemata as basis for reading strategies to further process EFL

comprehension and the other one in which formal schemata is a requisite to reading strategies prior to EFL comprehension.

Reading strategies conceptually involve cognitive strategies and meta cognitive strategies (Grabe,1991). Grabe (1991) argues that fluent reading is flexible, implying the utilization of a wide range of strategies employed by the readers to read effectively and efficiently. Block (1986) envisions reading strategies as how readers conceive a task, what textual cues they attend to, how they make sense of what they read, and what they do when they do not understand. This follows that, as Langer (1982) sees it, reading strategies reveal a reader's resourcefulness for understanding. In this perspective, synthesis and evaluation skills and strategies are nonetheless but reading strategies when these skills and strategies are so defined as the readers' skills in synthesizing and evaluating information found in the passage with other information especially related to the readers' content background knowledge as their strategies to understand the text being read. Seen in this way, reading strategies conceptually include complicated mechanisms more closely related to the handling of the message or information in the text. This implies that this reading strategy mechanism goes beyond the mastery of readers' content and world background knowledge, which in the present study is revealed to differentiate poor from average EFL readers. If this argument can hold correctly, then, good EFL readers are necessarily those who can process all these mechanisms effectively and efficiently, implying a support for interactive processes in reading.

Other explanations pertaining to the findings relate to data-collection instruments. The concepts text structure knowledge, reading strategies and content and world background knowledge as they are referred to in the present study are latent, and thus necessitate exacting conceptualization of their constructs and their dimensions as well as empirical construct validation (Cronbach and Meehl, 1955). While evidence in the informal trial shows that the data collection instruments in these variables were valid in terms of inter-item and reliable, construct validation was not sufficiently attempted in terms of design and involvement of subjects. This may become a source of potential bias for the constructs of text structure knowledge, reading strategies and content and world background knowledge to be accurately measured.

## **CONCLUSION AND RECOMMENDATION**

This present study has answered all issues encompassing the variables putative to typically characterize advanced EFL readers of different levels of profi-

ciency. While substantially they were different, the variables that characterized the differences were substantiated to consist of content and world background knowledge and reading strategies. Empirical and conceptual discussions on the findings lead to the following conclusion. Advanced EFL readers, while they are sufficiently equipped with formal schemata embracing both linguistic knowledge and text structure knowledge, then can be speculated to consist of three groups: those who fail to activate their relevant content schemata, those who fail to use their reading strategies, and those who can function both content and world background knowledge and their reading strategies with ease. This finding is thus in favor for interactive views of reading processes.

In the teaching of reading to advanced EFL readers, in particular, attention needs to be paid carefully in activating their content schemata as well as reading strategies. To address the former, pre-reading activities may be attempted in which discussions on aspects related to the topic of the text can be introduced as advance organizers. To attend to the latter, activation of types of reading strategies during whilst reading activities is recommended.

The present study involves university students who are taken for granted to have sufficient language proficiency. In order for subjects to be really representative embracing low and high levels of reading proficiency, more homogeneous subjects need to be involved in further research. With this, the variables that really constitute different levels of EFL comprehension can be more accurately examined. Also, data-collection instruments seem to be of utmost importance because a study like the present one is concerned with latent attributes such as reading strategies, which requires careful and accurate handling of providing evidence of quality data-collection instruments. Therefore, the development of data-collection instruments seems to need exercising in appropriate procedural fashions commonly applied to instrument construction through construct validation processes.

## REFERENCES

- Block, E. 1986. 'The comprehension strategies of second language readers' *TESOL Quarterly*, (20) pp. 463-494.
- Carrell, P.L. 1983. 'Three Components of Background Knowledge in Reading Comprehension' *Language Learning* 33 (2), pp.183-203.
- Carrell, P.L. 1988. 'Introduction: Interactive Approaches to Second Language Reading' in Carrell, P.L., J. Davine and D.E. Eskey (Eds.), *Interactive Ap-*



- proaches to Second Language Reading*. Cambridge: Cambridge University Press. pp.1-7.
- Carrell, P.L. and Eisterhold, J.C. 1983. 'Schema theory and ESL reading pedagogy' in Carrell, P.L., J. Davine and D.E. Eskey (Eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp.73-92.
- Cronbach, L.J. and Meehl, P.E.1955. 'Construct Validation in Psychological Tests' in *Classics in the History of Psychology: An internet Resource* developed by Christopher D. Green, York University, Toronto, Ontario. Downloaded on 15 May 2004 from <http://psychclassics.yorku.ca/Cronbach/construct.htm>.
- Devine, J. 1988. 'The relationship between general language competence and second language reading proficiency: implications for teaching.' in Carrel, P.L., J. Davine, and D.E. Eskey (Eds.), *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 260-278.
- Block, E. 1986. 'The Comprehension Strategies of Second Language Readers' *TESOL Quarterly*, (20). pp. 463-494.
- Eskey, D.E. 1988. 'Holding in the Bottom: An Interactive Approach to the Language Problems of Second Language Readers' in Carrell, P.L., J. Davine and D.E. Eskey (Eds.), *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 93-100.
- Fraenkel, J.R and N.E. Wallen, 1993. *How to Design and Evaluate Research in Education* 2<sup>nd</sup> Edition. New York: McGraw-Hill, Inc.
- Garrod, S. 1999. 'The Challenge of Dialog for Theories of Language Processing' in Simon Garrod dan Martin Pickering (eds.) *Language Processing*. Hove, East Sussex, UK: Psychology Press, p. 409.
- Goodman, K. 1988. 'The reading process' in Patricia L. Carrell, Joanne Devine, dan David E. Eskey (ed.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 11-20.
- Grabe, W. 1991. Current Development in Second Language Reading Research. *TESOL Quaterly* 25(3). pp. 375-405.
- Hair, J.F., Anderson, R.E., Tatham, R.L., dan Black, W.C. 1998. *Multivariate Data Analysis 5<sup>th</sup> Edition*. Upper Saddle River, NJ 07458: Prentice Hall.
- Hudson, T. 1988. 'The effect of induced schemata on the "short circuit" in L2 reading: non-decoding factors in L2 reading performance' in in Patricia L. Carrell, Joanne Devine, dan David E. Eskey (ed.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press. pp. 183-205.

- Katalog Jurusan Sastra Inggris. 2005. Malang: Fakultas Sastra, Universitas Negeri Malang.
- Kweldju, S. 2000. Developing Vocabulary Size Tests to Measure Graduate Students' Vocabulary Size. *Jurnal Teknologi Pembelajaran: Teori dan Penelitian* 8(1).
- Langer, J. 1982. 'The Reading Process' in Berge, A. and H.A. Robinson (Eds.), *Secondary School Reading: What research reveals for classroom practice. ERIC Clearinghouse on Reading and Communication Skills*. pp. 39-52.
- McRae, J and Boardman, R. 1988. *Reading between the Lines: Integrated Language and Literature Activities*. Cambridge: Cambridge University Press.
- Miller, W.H. 1971. *Identifying and Correcting Reading Difficulties in Children*. New York: The Center for Applied Research in Education, Inc.
- Nation, P. 1990. *Teaching and Learning Vocabulary*. Boston, Massachusetts: Heinle & Heinle Publisher.
- Nist, S., and Mealey, D. 1991. 'Teacher-directed comprehension strategies' in R. Flippo and D. Caverly (Eds.) *Teaching reading and study strategies*. pp: 165-185. Hillsdale, NJ: Lawrence Erlbaum.
- Nuttall, C. 1985. *Teaching Reading Skills in Foreign Language*. Oxford: Heinemann International.
- Norusis, M. 1993. *SPSS: SPSS for Windows Professional Statistics Release 6.0*. Chicago: SPSS Inc.
- Pugh, A.K. 1978. *Silent Reading: An Introduction to Its Study and Teaching*. London: Heinemann Educational Books.
- Rayner, K, and Pollatsek, A. 1989. *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rayner, K. 1983. *Eye Movements*. Chichester: John Willey and Sons.
- Rumelhart, D.E., 1980. 'Schemata: The building block of cognition' in R.J. Sapiro, B.C.Bruce, and W.F.Brewer (Eds.). *Theoretical Issues in Reading Comprehension*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers. pp. 123-156.
- Suharmanto. 2006. Patterns of relationship among some selected factors affecting students' EFL reading comprehension. Unpublished Dissertation. Graduate Program. State University of Malang.
- Smith, F. 1982. *Understanding reading. 3<sup>rd</sup> Edition*. New York: Holt, Rinehart and Winston.
- Tabachnick, B.G and Fidell, L.S. 1989. *Using Multivariate Statistics 2<sup>nd</sup> Edition*. New York: HarperCollins Publishers.