

9-19-2007

# The Effects of Semantic and Thematic Clustering on Learning English Vocabulary by Saudi Students

Sameer S. Al-Jabri

*Indiana University of Pennsylvania*

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

---

## Recommended Citation

Al-Jabri, Sameer S., "The Effects of Semantic and Thematic Clustering on Learning English Vocabulary by Saudi Students" (2007). *Theses and Dissertations (All)*. 914.  
<http://knowledge.library.iup.edu/etd/914>

This Dissertation is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact [cclouser@iup.edu](mailto:cclouser@iup.edu), [sara.parme@iup.edu](mailto:sara.parme@iup.edu).

THE EFFECTS OF SEMANTIC AND THEMATIC CLUSTERING ON  
LEARNING  
ENGLISH VOCABULARY BY SAUDI STUDENTS

A Dissertation

Submitted to the School of Graduate Studies and Research  
in Partial Fulfillment of the  
Requirements for the Degree  
Doctor of Philosophy

Sameer S. Al-Jabri  
Indiana University of Pennsylvania  
December 2005

© by Sameer S. Al-Jabri

All Rights Reserved

Indiana University of Pennsylvania  
The School of Graduate Studies and Research  
Department of English

We hereby approve the dissertation of

Sameer Saleh Al-Jabri

Candidate for the degree of Doctor of Philosophy

August 9, 2005

Signature on File

Dr. Jeannine Fontaine  
Associate Professor of  
English, Advisor

August 9, 2005

Signature on File

Dr. Michael M. Williamson  
Professor of English

August 9, 2005

Signature on File

Dr. Gary Dean  
Professor of Adult and Community Education

ACCEPTED

Signature on File

August 10, 2005

Ms. Michele S. Schwietz  
Assistant Dean for Research  
The School of Graduate Studies and Research

Title: The Effects of Semantic and Thematic Clustering on Learning English Vocabulary by Saudi Students

Author: Sameer Saleh Al-Jabri

Dissertation Chair: Dr. Jeannine M. Fontaine

Dissertation Committee Members: Dr. Michael M. Williamson  
Dr. Gary Dean

The present study aimed to compare the effects of semantic and thematic clustering on learning English vocabulary by Saudi students. It investigated whether thematic grouping or the use of meaningful context facilitates vocabulary learning. It was also conducted to shed light on how the use of context might be combined with clustering in order to facilitate learning.

The study consisted of two parts. In the first part, the quantitative stage, 160 participants studied four lists of English words representing semantic clustering, unrelated grouping, thematic clustering, and contextualized presentation. They were tested twice; immediately after the study phase and a week later, on recall for the words in each list. Tests were English-to-Arabic and Arabic-to-English. In the second part, the participants' reflection, four participants from each level representing the two participants who learned the greatest number of words and the two who learned the smallest number of words were asked individually some questions about their reflection on the word lists.

An Analysis of Variance (ANOVA) was used to analyze the quantitative data collected. In analyzing the data, measurement of the dependent variable has been achieved by simply counting the number of recalled words from each eight-word set. Results of Arabic-to-English and English-to-Arabic tests showed that participants

recalled more words from the thematic list than from the semantic list. Words from the semantic list were the least to be recalled by all participants. The difference was significant in some results while they were insignificant in others.

Participants' answers to the questions I asked showed that all participants claim that they use repetition as their technique for learning new vocabulary. It was unanimous. Along with repetition, some Highest participants claim using other techniques such as the Keyword method and the use of sentences. All participants claim that teachers never tried to provide them with different techniques to learn vocabulary. When it comes to which list was easier to learn, two out of six participants talked with prefer the unrelated list, two prefer the semantic list, and two have no preference of one over the other. On the other hand, most of the participants who studied the thematic list and the context show a clear preference for the thematic presentation for new words.

## Dedication

I dedicate this work to my parents, brothers, sisters; and to  
my wife and unborn child

## Acknowledgements

First, all praises and thanks be to Allah, the Almighty, for helping me finish my graduate studies and this research.

Professionally, I would like to offer my appreciation and gratitude to my advisor, Jeannine Fontaine, for her brilliance and amazing insight. Her support, encouragement, and patience helped me during my study at IUP. I am grateful for her helpful advice and guidance during the creation of this dissertation.

I would also like to thank my two committee members, Dr. Michael Williamson and Dr. Gary Dean, for their valuable and helpful critiques of my work. My sincere thanks also go to Dr. Tom Short, the head of Applied Research lab, for his advice and feedback on the statistical analysis for this project.

Personally, special thanks and love to my parents, brothers, and sisters, for their steadfast support during my stay at the States. The continual encouragement and prayers of my parents have been extremely instrumental in motivating me to pursue the studies that led to this achievement.

Words can not even begin to express the debt of gratitude and admiration I feel for my wife for her support and sacrifice during my study. Her impact on me has been greater than I could say. I am truly grateful.



## TABLE OF CONTENTS

CHAPTER		Page
ONE	INTRODUCTION .....	1
	Statement of the Problem .....	3
	Purpose of the Study .....	4
	Hypotheses of the Study .....	5
	Significance of the Study .....	5
	Definitions of Terms .....	6
TWO	REVIEW OF LITERATURE .....	8
	Semantic Clustering .....	8
	Justification for Semantic Clustering .....	10
	Semantic Fields .....	14
	Evidence against the Use of Semantic Clustering .....	19
	Interference Theory .....	19
	Retroactive interference .....	22
	Proactive interference .....	24
	Intralist interference .....	26
	The Distinctiveness Hypothesis .....	28
	Thematic Clustering .....	30
	Justification for Thematic Clustering .....	31
	Frames .....	31
	Evidence Supporting the Use of Thematic Clustering .....	35
	Schema Theory .....	35
	Semantic Clustering Vs Thematic Clustering .....	41
	Conclusion .....	43
	Context .....	44
	Studies Guessing from Context with Second Language Learners .....	47
	Causes of Poor Guessing .....	47
	Memory .....	48
THREE	METHODOLOGY .....	57
	Introduction .....	57
	Quantitative Stage .....	60
	Subjects .....	60
	Materials .....	61
	English Words .....	65
	Pilot Study .....	66
	Procedure .....	66
	Immediate Recall Tests .....	66
	Delayed Recall Tests .....	69

	Data Analysis .....	69
	Informal Interviews.....	70
FOUR	FINDINGS AND DISCUSSION .....	72
	Quantitative Findings .....	72
	Arabic to English Translation Direction .....	73
	Level 1(Immediate Tests) .....	73
	Level 1(Delayed Tests) .....	78
	Level 2(Immediate Tests) .....	80
	Level 2 (Delayed Tests) .....	83
	Level 3(Immediate Tests) .....	85
	Level 3(Delayed Tests) .....	87
	Within Conditions .....	90
	Semantic Clustering (Immediate Tests) .....	90
	Semantic Clustering (Delayed Tests) .....	92
	Unrelated Clustering (Immediate Tests) .....	94
	Unrelated Clustering (Delayed Tests) .....	97
	Thematic Clustering (Immediate Tests) .....	99
	Thematic Clustering (Delayed Tests) .....	101
	Context Clustering (Immediate Tests) .....	103
	Context Clustering (Delayed Tests) .....	106
	English to Arabic Translation Direction .....	108
	Level 1 (Immediate Tests) .....	108
	Level 1 (Delayed Tests) .....	112
	Level 2 (Immediate Tests) .....	116
	Level 2 (Delayed Tests) .....	118
	Level 3 (Immediate Tests) .....	122
	Level 3 (Delayed Tests) .....	125
	Within Conditions .....	128
	Semantic Clustering (Immediate Tests) .....	128
	Semantic Clustering (Delayed Tests) .....	130
	Unrelated Clustering (Immediate Tests) .....	132
	Unrelated Clustering (Delayed Tests) .....	134
	Thematic Clustering (Immediate Tests) .....	137
	Thematic Clustering (Delayed Tests) .....	139
	Context Clustering (Immediate Tests) .....	141
	Context Clustering (Delayed Tests) .....	143
	Interaction Effects .....	145
	Immediate Tests .....	145
	Delayed Tests .....	149
	Participants' Reflection.....	154
	Findings from the Interviews .....	154
	Participants' Response Patterns.....	160
	Level 1(Phonological Errors) .....	160
	Level 1(Semantic Errors) .....	161
	Level 1 (Translation Errors) .....	162
	Level 2 (Phonological and Spelling Errors) .....	163
	Level 2 (Semantic Errors) .....	164
	Level 2 (Translation Errors) .....	164
	Level 3 (Phonological Errors) .....	165

	Level 3 (Semantic Errors) .....	166
	Level 3 (Translation Errors) .....	167
	Discussion .....	168
	Semantic Clustering Vs. Thematic Clustering .....	168
	Backward and Forward Translations .....	169
	Use of context .....	170
	Summaries of Major Findings .....	170
	Findings from the Quantitative Part .....	170
	Findings from the Participants' Reflection .....	176
	Findings from Analysis of Answers Sheets .....	178
	Results Compared with the Research Hypotheses .....	178
<b>FIVE</b>	<b>SUMMARY, IMPLICATIONS, AND SUGGESTIONS FOR FUTURE STUDY .....</b>	<b>180</b>
	Introduction .....	180
	Summary .....	180
	Importance of the Study .....	182
	Implications and Recommendations for Teaching .....	183
	Limitations of the Study .....	186
	Suggestions for Future Study .....	186
<b>REFERENCES</b>	.....	<b>189</b>
<b>APPENDICES</b>	.....	<b>201</b>
	Appendix A – Semantic List .....	202
	Appendix B – Arabic-to-English Test (Semantic List) .....	204
	Appendix C – English-to-Arabic Test (Semantic List) .....	206
	Appendix D – Unrelated List .....	208
	Appendix E – Arabic-to-English Test (Unrelated List) .....	210
	Appendix F – English-to-Arabic Test (Unrelated List) .....	212
	Appendix G – Thematic List .....	214
	Appendix H – Arabic-to-English Test (Thematic List) .....	216
	Appendix I – English-to-Arabic Test (Thematic List) .....	218
	Appendix J – Context .....	220
	Appendix K – Arabic-to-English Test (Context) .....	222
	Appendix L – English -to- Arabic Test (Context) .....	224
	Appendix M – Consent Form .....	226

## LIST OF TABLES

Table		Page
1	Means and Standard Deviation of Recalled Words (Level 1–Immediate test– A-to-E) .....	73
2	Analysis of Variance of Number of Recalled Words (Level 1–Immediate test–A-to-E) .....	75
3	Post Hoc Tests (Level 1 – Immediate test – A-to-E) .....	77
4	Means and Standard Deviation of Recalled Words (Level 1 – Delayed test – A-to-E) .....	78
5	Analysis of Variance of Number of Recalled Words (Level 1 – Delayed test– A-to-E) .....	80
6	Means and Standard Deviation of Recalled Words (Level 2– Immediate test – A-to-E) .....	80
7	Analysis of Variance of Number of Recalled Words (Level 2–Immediate test–A-to-E) .....	82
8	Post Hoc Tests (Level 2 – Immediate test – A-to-E) .....	83
9	Means and Standard Deviation of Recalled Words (Level 2 – Delayed test – A-to-E) .....	84
10	Analysis of Variance of Number of Recalled Words (Level 2 – Delayed test – A-to-E) .....	85
11	Means and Standard Deviation of Recalled Words (Level 3–Immediate test – A-to-E) .....	86
12	Analysis of Variance of Number of Recalled Words (Level 3–Immediate test–A-to-E) .....	87
13	Means and Standard Deviation of Recalled Words (Level 3 – Delayed test – A-to-E) .....	88
14	Analysis of Variance of Number of Recalled Words (Level 3 – Delayed test – A-to-E).....	89
15	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Semantic List – A-to-E).....	90
16	Analysis of Variance of Number of Recalled Words	

	(All Levels – Immediate test – Semantic List – A-to-E) .....	92
17	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Semantic List – A-to-E).....	92
18	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Semantic List – A-to-E).....	94
19	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Unrelated list – A-to-E).....	94
20	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Unrelated list – A-to-E).....	96
21	Post Hoc Tests (All Levels – Immediate test – Unrelated list – A-to-E).....	97
22	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Unrelated list – A-to-E).....	98
23	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Unrelated list – A-to-E).....	99
24	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Thematic list – A-to-E) .....	100
25	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Thematic list – A-to-E).....	101
26	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Thematic list – A-to-E) .....	102
27	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Thematic list – A-to-E).....	103
28	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Context list – A-to-E) .....	104
29	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Context list – A-to-E).....	105
30	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Context list – A-to-E).....	106
31	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Context list – A-to-E) .....	108
32	Means and Standard Deviation of Recalled Words (Level 1– Immediate test – E-to-A).....	108

33	Analysis of Variance of Number of Recalled Words (Level 1 – Immediate test – E-to-A) .....	110
34	Post Hoc Tests (Level 1 – Immediate test – E-to-A) .....	111
35	Means and Standard Deviation of Recalled Words (Level 1 – Delayed test – E-to-A) .....	112
36	Analysis of Variance of Number of Recalled Words (Level 1 – Delayed test – E-to-A).....	114
37	Post Hoc Tests (Level 1 – Delayed test – E-to-A).....	115
38	Means and Standard Deviation of Recalled Words (Level 2– Immediate test – E-to-A) .....	116
39	Analysis of Variance of Number of Recalled Words (Level 2–Immediate test–E-to-A) .....	117
40	Means and Standard Deviation of Recalled Words (Level 2 – Delayed test – E-to-A).....	118
41	Analysis of Variance of Number of Recalled Words (Level 2 – Delayed test – E-to-A).....	120
42	Post Hoc Tests (Level 2 – Delayed test – E-to-A).....	121
43	Means and Standard Deviation of Recalled Words (Level 3– Immediate test – E-to-A) .....	122
44	Analysis of Variance of Number of Recalled Words (Level 3–Immediate test–E-to-A).....	124
45	Post Hoc Tests (Level 3 – Immediate test – E-to-A) .....	124
46	Means and Standard Deviation of Recalled Words (Level 3 – Delayed test – E-to-A) .....	125
47	Analysis of Variance of Number of Recalled Words (Level 3 – Delayed test – E-to-A) .....	127
48	Post Hoc Tests (Level 3 – Delayed test – E-to-A) .....	127
49	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Semantic list – E-to-A) .....	128
50	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Semantic list – E-to-A).....	130
51	Means and Standard Deviation of Recalled Words	

	(All Levels – Delayed test – Semantic list – E-to-A) .....	130
52	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Semantic list – E-to-A).....	132
53	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Unrelated list – E-to-A) .....	132
54	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Unrelated list – E-to-A).....	134
55	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Unrelated list – E-to-A).....	134
56	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Unrelated list – E-to-A).....	136
57	Post Hoc Tests (All Levels – Delayed test – Unrelated list – E-to-A).....	136
58	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Thematic list – E-to-A).....	137
59	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Thematic list – E-to-A).....	138
60	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Thematic list – E-to-A).....	139
61	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Thematic list – E-to-A).....	141
62	Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Context list – E-to-A) .....	141
63	Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Context list – E-to-A) .....	142
64	Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Context list – E-to-A).....	143
65	Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Context list – E-to-A) .....	144
66	Means and Standard Deviation of Recalled Words (Three Levels – Four conditions – Two Translation Directions – Immediate Tests) .....	145
67	Three-way ANOVA Model for Immediate Tests .....	147

68	Post Hoc Tests (All Levels – Immediate tests – Two Translation Directions).....	147
69	Means and Standard Deviation of Recalled Words (Three Levels – Four conditions – Two Translation Directions – Delayed Tests).....	149
70	Three-way ANOVA model for Delayed Tests .....	150
71	Post Hoc Tests (All Conditions – Delayed tests – Two Translation Directions).....	151
72	Post Hoc Tests (All Levels – Immediate tests – Two Translation Directions).....	152



## LIST OF FIGURES

Figure	Page
1 Ebbinghaus' Forgetting Curve .....	50
2 Forgetting Rate in Bahrick's Study.....	51
3 Basic Forgetting Curve.....	52
4 Box-plot diagram of score in relation to types of clustering (level 1 – immediate test – A-to-E).....	74
5 Box-plot diagram of score in relation to types of clustering (level 1 – delayed test– A-to-E).....	79
6 Box-plot diagram of score in relation to types of clustering (level 2 – immediate test – A-to-E).....	81
7 Box-plot diagram of score in relation to types of clustering (level 2 – delayed test – A-to-E) .....	84
8 Box-plot diagram of score in relation to types of clustering (level 3 – immediate test – A-to-E) .....	86
9 Box-plot diagram of score in relation to types of clustering (level 3 – delayed test – A-to-E) .....	88
10 Box-plot diagram of score in relation to level of participants (all levels – immediate test – semantic list – A-to-E).....	91
11 Box-plot diagram of score in relation to level of participants (all levels – delayed test – semantic list – A-to-E).....	93
12 Box-plot diagram of score in relation to level of participants (all levels – immediate test – unrelated list – A-to-E).....	95
13 Box-plot diagram of score in relation to level of participants (all levels – delayed test – unrelated list – A-to-E).....	98
14 Box-plot diagram of score in relation to level of participants (all levels – immediate test – thematic list – A-to-E).....	100
15 Box-plot diagram of score in relation to level of participants (all levels – delayed test – thematic list – A-to-E).....	102
16 Box-plot diagram of score in relation to level of participants (all levels – immediate test – context list – A-to-E).....	104

17	Box-plot diagram of score in relation to level of participants (all levels – delayed test – context list – A-to-E).....	107
18	Box-plot diagram of score in relation to types of clustering (level 1 – immediate test – E-to-A).....	109
19	Box-plot diagram of score in relation to types of clustering (level 1 – delayed test – E-to-A).....	113
20	Box-plot diagram of score in relation to types of clustering (level 2 – immediate test – E-to-A).....	117
21	Box-plot diagram of score in relation to types of clustering (level 2 – delayed test – E-to-A).....	119
22	Box-plot diagram of score in relation to types of clustering (level 3 – immediate test – E-to-A).....	123
23	Box-plot diagram of score in relation to types of clustering (level 3 – delayed test – E-to-A).....	126
24	Box-plot diagram of score in relation to level of participants (all levels – immediate test – semantic list – E-to-A).....	129
25	Box-plot diagram of score in relation to level of participants (all levels – delayed test – semantic list – E-to-A) .....	131
26	Box-plot diagram of score in relation to level of participants (all levels – immediate test – unrelated list – E-to-A).....	133
27	Box-plot diagram of score in relation to level of participants (all levels – delayed test – unrelated list – E-to-A) .....	135
28	Box-plot diagram of score in relation to level of participants (all levels – immediate test – thematic list – E-to-A) .....	138
29	Box-plot diagram of score in relation to level of participants (all levels – delayed test – thematic list – E-to-A).....	140
30	Box-plot diagram of score in relation to level of participants (all levels – immediate test – context list – E-to-A).....	142
31	Box-plot diagram of score in relation to level of participants (all levels – delayed test – context list – E-to-A) .....	144
32	Interaction plot .....	148
33	Interaction plot .....	153

## CHAPTER ONE

### INTRODUCTION

The mastery of vocabulary is central and essential in the process of second / foreign language learning. It facilitates comprehension, one of the primary factors that lead to good progress in second language learning (Lynch, 1996). The role of vocabulary is one of the first aspects of method design to receive attention in second language teaching programs (Richards & Rodgers 2001 p.37). One of the challenges facing the second language learner is how to master a large vocabulary in order to speak, listen to, read and write the target language effectively, and thus communicate successfully and appropriately with others. But vocabulary building has often been downgraded (Judd, 1978), while grammatical and phonological structures have been given more emphasis and considered the starting point in the learning process. This low status for vocabulary building results from the adoption of language teaching approaches based on the American linguistic theories dominant during the 1940s - 1960s (Decarrico, 2001). Teaching vocabulary has not been a central goal of second language English instruction during the very active decades of the mid-twentieth century, nor was it considered a priority in the larger context of language teaching and learning at that time. As a result of this view, learners of English have often faced communication barriers in various situations which require control over a large variety of vocabulary items rather than a narrow range of syntactic structures.

However, this dominant view has been challenged since the late 1970s and early 1980s. More emphasis and considerable attention have been directed to vocabulary building since that time. Educational researchers and psychologists began, even early in this period, to produce a number of word frequency studies in different

languages in response to the increasing need for vocabulary control in language courses (Stern, 1983).

As a result of the growing interest in vocabulary building by these researchers, various techniques have been introduced and used. However, there are still problems. Examining the current ESL (English as a second language) textbooks, we find that new vocabulary items are typically presented to ESL/EFL students in semantically related sets. Gairns and Redman (1986) call such sets "lexical sets," while Marzano and Marzano (1988) use the term "semantic clusters," since the sets are tightly-knit collections of words selected from semantic fields. In simple terms, these sets are composed of words whose syntactic class and meaning are closely related. For example, Costinett (1987) clusters *bed, sofa, chair, table, and dresser* together in the text *Spectrum, 2*, while Franklin and Meyers (1991) cluster *single, married, divorced, separated, and widowed* together in *Crossroads, 1*. Such sets are indeed semantic clusters, and words within these sets share a common superordinate (headword) such as [furniture].

Course designers, teachers, and writers have made the largely unexamined assumption that grouping new vocabulary items in related sets facilitates learning. As justification for this approach, curriculum developers say that related words help learners see how knowledge is organized (Dunbar, 1992), and the assumption is made that learning this way does not require more effort. However, educators' dedication to such an argument as this rests on personal methodology rather than on empirical support or theoretical orientation.

Despite the lack of empirical or theoretical basis for these assumptions, teaching systems have quite typically relied on semantic grouping to present vocabulary. For example, the situational approach, developed by the British linguists

Palmer and Hornby and introduced in the 1950s and 1960s (Stern, 1993), considers grammatical structures and word lists its basic components. Textbooks based on this approach are still used worldwide. Richards and Rodgers (2001) provide an example of how vocabulary items are presented in the situational approach:

This is ..... [*book – pencil, ruler, desk*].

*[chair, picture, door, window]*

Again, empirically, there is little if any direct evidence that such lexical clustering facilitates learning. According to Tinkham (1994), presenting students with new words grouped in semantic clusters is not motivated by empirical support or theoretical concerns. Rather, the writers' loyalty to a specific methodology, whether it be language-centered or more learner-centered, tends to determine the approach they follow in second language development.

#### Statement of the Problem

Teaching methods in Saudi Arabia are no exception to the general trend discussed above. English textbooks in use there present vocabulary items grouped in semantic clusters. In the first intermediate level English textbook, curriculum writers select the new English words that fit specific situations and tasks or express different notions, and they present these words in semantic clusters. For example, in a lesson titled 'Organs of the Face,' the following words are introduced: forehead, nose, chin, mustache, beard, mouth, ear, eye. Within a unit titled 'Kinds of Fruit,' the following words appear in the text: banana, apple, orange, watermelon, mango (Ministry of Education 2002, p. 37). It appears that ESL program designers and textbook authors assume that the presentation of semantically and syntactically related lexical items facilitates learning.

A growing body of research indicates that this widely accepted way of presenting new vocabulary items does not facilitate learning. Rather, it makes learning more difficult and interferes with the learning of similar words. Evidence in support of concepts such as the Interference Theory and the Distinctiveness Hypothesis, discussed further in Chapter 2, strongly suggest that semantic clustering may actually impede rather than facilitate learning. Another concern, involves the documented need to present new vocabulary in meaningful context.

### Purpose of the Study

The purpose of the present study is to compare the effects of semantic and thematic clustering on learning English vocabulary by Saudi students. The two labels are intended to differentiate between two different methods of organizing lexical items. Semantic clustering is based on grouping words that share various semantic and syntactic characteristics, Thematic clustering is based on psychological associations between clustered words and a shared thematic concept. The terms mother, father, daughter, son provide an example of a semantic cluster. In contrast, a cluster perceived as thematically related would include terms like frog, pond, swim, and green; note that these terms do not refer to semantically similar concepts; however, they cluster around the concept of a pond, and might come to mind when a speaker is thinking about a story involving a pond and its inhabitants.

The present study is motivated by the desire to examine the effect of meaningful thematic and contextual grouping on the learning of vocabulary items in sets. The goal was to investigate whether thematic grouping or the use of meaningful context facilitates vocabulary learning.

### Hypotheses of the Study

Given the trends in recent research, I approach this study with an intuitive sense that the presentation of new vocabulary in semantically clustered sets may actually impede learning (even as compared with random word lists), while both thematically organized or contextually related word groups may be more readily learned. Hence, the hypotheses made at the outset are the following:

Hypothesis One: Saudi students learn more unrelated words than semantic clusters of new English words.

Hypothesis Two: Saudi students learn more thematic clusters of new English words than semantic clusters or unrelated English words.

Hypothesis Three: Saudi Students find the semantic related sets the most difficult to learn.

Hypothesis Four: Saudi Students find the thematic sets embedded in a meaningful context the easiest to learn.

Hypothesis Five: The use of context facilitates learning thematic sets.

Hypothesis Six: Saudi students with higher levels might be less affected by semantic or thematic clusterings when learning English words.

### Significance of the Study

Research about the effects of clustering on L2 vocabulary learning is limited and indirect (Tinkham, 1994), although two methods of clustering are currently identifiable and employed in L2 vocabulary instruction; these are semantic and thematic clustering.

Research generated by the interference theory has been concerned with general learning structures; this research explores the similarities between a stimulus

and its associated response, and hypothesizes about the possibly harmful effect of similarities between sets of stimuli on learning and memory. Only few writers have extended their research to study the effect of clustering on the learning of L2 vocabulary (Higa, 1963; Tinkham, 1993, 1994; Waring, 1997). The present study will be among the efforts to continue research into this field. This study examines clustering of two types, as it is an important factor in learning vocabulary in a second/foreign language. It also aims to shed light on how the use of context might be combined with clustering in order to facilitate learning. The results of the study might be of great value to writers and planners of ESL/EFL textbooks, in their plans to introduce vocabulary in the course of their lessons. Moreover, English teachers might find this study helpful as they seek to improve or modify the teaching methods they use in order to gain the best results in the learning process.

#### Definitions of Terms

Distinctiveness Hypothesis: This deals with the ease with which distinctive information is learned. It "relates ease of learning to the distinctiveness (non-similarity) of the information to be learned" (Waring, 1997, p. 373). It states that "the most important factor in recognition memory is the extent to which the test-trial encoding contains information that is unique to the study-trial encoding." (Eysenck, 1979)

Interference Theory: For much of the last century, this has been the dominant theory regarding forgetting. It provides evidence connecting learning difficulties to similarities between targeted and interfering materials. It states that "when words are being learned at the same time, but are too "similar" or share too many common elements, then these words will interfere with each other thus impairing retention of



them. The degree of interference increases with the degree to which the interfering material becomes more similar to the material already learned" (Waring, 1997, p. 261-262).

Lexical set: One word or vocabulary unit is commonly called a lexical item, or a lexeme. When groups of words share "certain formal or semantic features," they are called lexical sets (Crystal, 1997, p.221).

Schema: a data structure for representing the generic concepts stored in memory

Semantic Clustering: a method of grouping words that share semantic and syntactic characteristics. An example is the group *arm*, *leg* and *hand*, which are all body parts; often the term "lexical sets" is also used (Tinkham, 1997, p. 138)

Thematic Clustering: another method of grouping words based upon psychological associations between clustered words and a shared thematic concept

Unrelated sets: words that do not share semantic or syntactic characteristics

## CHAPTER TWO

### REVIEW OF LITERATURE

This chapter will discuss the research literature on the main areas that inform this study, or whose insights can be helpful in interpreting the study's results. First, a detailed overview of semantic clustering, the justifications offered for such clustering and the underlying theoretical concept of semantic fields will be presented. Next, a body of research will be reviewed that claims that semantic clustering is actually detrimental, rather than being helpful to language learning. Subsequent sections will focus on the distinctness hypothesis, thematic clustering and justifications for thematic grouping, schema theory and context, and finally memory.

#### Semantic Clustering

Organizing or grouping words according to their semantic or syntactic attributes is a cognitive strategy (Mitchell and Myles, 1998). This type of clustering is widely known and frequently appears in general ESL textbooks. The following are examples of this type of clustering:

- *Interchange, 3* (Richards, 1998, p. 28) groups *anxious, nervous, suspicious, worried, depressed, embarrassed, calm, comfortable, confident* and *uncertain* as adjectives that describe how people feel when they live in a foreign country in a unit titled, "Crossing Cultures".
- *Side by Side, 1* (Molinsky and Bliss, 1989, p.43) groups *daughter, husband, mother, brother, aunt, uncle, cousin, grandmother and grandfather* in a unit titled, "My Favorite Photographs".

- *In Contact, 2* (Lavie, Briggs, Raht and Denman, 1991, p. 55) provides the cluster *can, cup, glass, dish* as examples of containers in a unit titled “Are you hungry?”

Similarly, semantic clusters appear in ESL textbooks dedicated to vocabulary development. The following selective list provides examples:

- *American Vocabulary Builder, 1* (Seal, 1990, p. 23) presents the cluster *thumb, middle finger, palm, ring finger, wrist, fingernail, index finger, fingertip, pinky* in a unit titled “Parts of the Hand”.
- *Making Sense of Vocabulary* (Digby & Myers, 1991, p. 14) clusters *towel, carpet, tablecloth, napkin* as “soft furnishings” in a unit titled “Home Life”.
- *English for Saudi Arabia* (Ministry of Education, 2002, p.16) groups *ruler, pencil, board, desk, chair* in a unit titled “My Classroom”. In a unit titled “What is it for?” the same text presents a vocabulary list of tools, including *scissors, screw-driver, pliers, hammer, saw, and tin-opener.*(p. 76)

Specialists in vocabulary development find this method helpful, especially for beginning students, as they feel it deepens understanding of a given context (Decarrico, 2001). Authors of curriculum who follow a structure-centered approach also find semantic clusters helpful, since the items in such lists fit into “slots” within structures targeted by exercises such as substitution drills (Tinkham, 1994). This allows students to change the meaning of the sentences produced within a repetitive framework. *Side by Side, 1* (Molinsky and Bliss, 1989: p.80) includes the names of all seven days of the week and seven nationalities for the slots within the following structure: “On ----- he cooks -----food”. Another example occurs in *Spectrum, 2* (Costinett, 1987, p.56), which asks students to substitute *bed, sofa, chair, table, dresser* in exercises within a unit under the title “Going Shopping.”

Tinkham (1994) claims that authors and planners of ESL programs following a more learner-centered approach select vocabulary items based on the communicative needs of the learners, and then organize their programs into units to reflect situations in which students have to use English (e.g., mailing a letter in a post office). However, even in this approach, the vocabulary items needed to express notions (e.g., expressions of time) and functions (e.g., requests) and to fit tasks tend to be presented as they are in the structure-centered approach; that is, they are grouped into semantic clusters. Ramires (1995), for instance, lists the following vocabulary items as terms students can use when ordering lunch in a restaurant: *meat, ham, beef, steak, chicken, and fish*.

It would seem then, that semantic clusters fit into most ESL programs and are commonly used in an attempt to make learning meanings easier regardless of the language learning approach used. Tinkham (1997) believes that this type of clustering is compatible with the audio-lingual methodology prescribed by the structure-centered approach; it also fits well into the situational syllabi prescribed by the learner-centered approach. Students are presented vocabulary items grouped in semantic clusters in both approaches.

#### *Justification for Semantic Clustering*

Most authors of the above mentioned ESL textbooks have not mentioned their rationale for presenting new vocabulary items in semantic clusters. The exception is Seal (1991), the author of *American Vocabulary Builder, 1* who provides two reasons for his use of semantic clusters. First, he claims that they give students the sense of structure they need. However, he does not consider whether this sense of structure might be achieved just as effectively by grouping words with a shared theme. Second,

he feels that this organization may help students guess the meaning of new words within the lexical set; of course, where one can easily see that a word's class membership might be clear from its inclusion in a semantic set, it is difficult to see how the specific meaning could be 'guessed' from such membership. Nation (2000) mentions the following justifications provided by numerous writers for teaching words in semantically related lexical sets:

- 1- It requires less effort to learning to learn words in a set.
- 2- It is easier to retrieve related words from memory.
- 3- It helps learners see how knowledge can be organized.
- 4- It reflects the way such information is stored in the brain
- 5- It makes the meaning of words clearer by helping students to see how they relate to and may be differentiated from other words in the set.

Further justification for semantic clusters may be found in 'notional syllabi'. The notional Syllabus is an idea proposed by Wilkins in his discussion of the functional view of language for syllabus design (Richards & Rodgers, 2001). Wilkins (1976) provides justification for semantic clusters through such notional syllabi, which focus on what speakers communicate through language. The basic idea is that content supersedes form. Therefore, Wilkins suggests a number of notional categories and lists expressions which would fit within each category. Once again, as with thematically inspired syllabi, the expressions grouped in notional syllabi tend to form semantic clusters. For example; *confirm, corroborate, endorse, support, assent, acquiesce, agree, concur, consent, ratify*, and *approve* are listed under the category "agreement." According to Wilkins "it is probably necessary to establish a number of themes around which semantically related items can be grouped and from which in constructing a notional syllabus an appropriate selection can be made" (p. 76). Once

the idea of a notional syllabus became popular and commonplace in second language development, it became the norm to use semantic clusters in ESL textbooks based on this approach.

Several pragmatic arguments have also been advanced to support the use of semantic clusters in second language vocabulary acquisition (Tinkham, 1994). First, Gairns and Redman (1986) believe that presenting L2 words grouped in semantic clusters helps the learner “to understand the semantic boundaries: to see where meaning overlaps and learn the limits of use of an item” (p. 32). Thus, semantic clustering is thought to help the learner see the distinctions between semantically related words. They also believe that this method gives coherence to the lesson, and therefore gives the students the sense that the language is organizable. Building on evidence that lexical items are stored in the human mind in semantic sets, (Tulving, 1962), Gairns and Redman (1986) say that semantic clusters form useful “building blocks” and can be revised and expanded as students progress; The authors go on to claim that this grouping can provide “a clear context for practice” (p. 69) and can also “speed up the learning process and facilitate learning” (p. 89). Similarly, Seal (1991) mentions that items to be taught should come from the same lexical domain, and lists several advantages. “First, by learning items in sets, the learning of one item can reinforce the learning of another. Second, items that are similar in meaning can be differentiated. Third, students may more likely feel a sense of tangible progress in having mastered a circumscribed lexical domain” (p. 300-301).

Studies on first language vocabulary learning also provide indirect evidence supporting semantic clustering. Beck and her colleagues (Beck, Perfetti, and McKeown, 1982; McKeown, Beck, Omanson, and Pople, 1985) designed a program of “rich” vocabulary instruction to study the effect of this method of vocabulary

instruction on reading comprehension. Beck, Perfetti, and McKeown, (1982) taught 27 grade four children a corpus of 104 words over a five-month period. Words were taught in semantic groups such as “People” (virtuoso, novice, hermit, rival, etc.). Children who received instruction and performed tasks involving single-word semantic decisions, simple sentence verification and memory for connected text outperformed the control subjects on comprehension tests. Also, subjects with such training were able to improve their reading comprehension of texts containing the newly learned words. These results were presented as showing that the semantic type of word grouping is beneficial to learning.

One study sometimes taken as affirmation of semantic grouping actually did not contribute directly to this question. To determine the relative contribution of the nature of instruction and the frequency of instructional encounters in improving verbal processing skill, McKeown, Beck, Omanson, and Pople (1985) provided students with three kinds of instruction. They found that subjects who received instruction on word meaning combined with vocabulary enrichment activities such as Word Wizard outperformed subjects who received only instruction on word meaning in story comprehension. Furthermore, while four encounters with a word did not improve reading comprehension, 12 encounters did.

Stahl, Burdge, Machuga, and Stecyk (1992) point out that the Beck and McKeown studies did not employ a group that studied unrelated words, and thus all participants in the study learned words in semantically related sets. Therefore, it seems that semantic clustering was assumed to contribute to their programs success, and was not tested in comparison to other kinds of grouping. In fact, Stahl et al. (1992) conducted a study in which the results for one group, which received rich instruction with words grouped semantically, was compared to the results for another

group, which received the same instruction using unrelated words. Results indicated that semantic grouping had no effect on vocabulary learning. They found that “if the instruction is extensive enough, students will form their own knowledge of the hierarchical relations between the new words and already known words” (p. 33).

Despite the lack of evidence for such a claim, many ESL program administrators and textbook designers continue to assume that the presentation of semantically and syntactically related lexical items reflects the way information is stored in the mental lexicon, and that this organization, in turn, reflects the best presentation system for language learners. The term “mental lexicon” refers to the collection of words that a speaker of a language knows within that language (Aitchison, 1987). To develop a better understanding of the mental lexicon, lexical semanticists have been addressing questions such as whether the mental lexicon is organized within the speaker’s mind, and what that organization might be. Thus, the proponents of lexical grouping base their concept of semantic clusters on the well-known psycholinguistic concept of “semantic fields,” which will be explored in the next section.

### *Semantic Fields*

The semantic field is a set of lexical items in which words applicable to a conceptual domain are organized by a number of relationships. In terms of affinity and contrast, two defining characteristics of semantic fields (Kittay and Lehrer, 1992), the relationships of synonymy, antonymy, and hyponymy are easily understood. Tinkham (1994) describes two different approaches taken by lexical semanticists in their attempts to provide an analysis of semantic fields. First, some writers take an intuitive approach. Grandy (1992), following this approach, emphasizes “contrast



sets,” which are sets (S) of contrasting terms organized under one covering term (T) that show contrast relations (R). These contrast sets “are fundamental to explicating the grander idea of semantic fields” (p. 104). They must be “defined in terms of common linguistic beliefs that competent speakers have about the contrasts and inclusions” (p. 105). He defines the contrast set in terms of beliefs rather than knowledge because it happens that someone might believe that the matters which are semantic a priori are not true. For example, speakers tend to feel that *Johns’ brother cannot be his uncle* is a semantic a priori principle. But it is in principle possible that John’s older brother could marry his mother’s younger sister. Such examples reveal that “almost all adult speakers are competent in their use, but there is also a large (indeed, larger) number of contrast sets where speakers are less competent. . . . Speakers have beliefs about the general field relations among the terms without having precise or accurate knowledge of the specifics of the relation” (p. 107).

A simple description of a basic and simple contrast set is that it includes monolexic terms, has a simple relation, and the contrasting relation between any two terms holds also between all contrasting terms. A clear example of a contrast set occurs when we analyze the color terms of English. The covering term of this set would be "color," the contrasting terms would be "red, yellow, blue, brown, white, black, green . . .etc." and the contrast relation would be "different color than." Complex relations can be cyclical, such as the "day after" relation of the contrast set "day," or multidimensional, such as the relations within the contrast set of family terms in English. These relations can be dyadic, as between male and female, or generational, as between parent and child.

Grandy (1992) defines the semantic field overall as "a set including one or more contrast sets and possibly also including permutation relations such that:

1. At most one covering term does not occur as an element of a contrast set in the semantic field.
2. Except for the covering term mentioned in (1), any expression that occurs in a contrast set with an element of the semantic field is also in the field." (p. 109).

Thus, a semantic field can be a single contrast set or a group of such sets. For example, the contrast set (animal: dog, cat...) is a semantic field. If we have to include "German Shepherd", then we have to include all the other contrasting terms such as (corgi, poodle,...).

In the same way, Kittay (1992) prefers to rely upon her intuitions regarding most speakers' "understanding" of "the distinctions that mark out the boundaries of a concept" (p. 236), or "individuations" (Kittay's term) which take place between a concept and another related concept. She believes that the way an item is differentiated from other related terms in the language defines the understanding of the term. As a result, semantic fields are clusterings of lexicalized concepts and they "not only group together semantically close terms, they also encode the differentiations that individuate concepts and terms"(p. 230). Within the semantic field, the relatedness of terms and the interrelations are the result of defined relations of contrast and affinity such as graded antonymy, scalar relations, hyponymy, contraries, and so on. Kittay (1992) concludes with the comment that "a semantic field, consisting as it does of a level of content and a level of expression, is, and must be, publicly accessible; that is, it must be available to either the whole of, or some significant part of, the linguistic community" (p. 246).

Another approach to defining semantic fields is to provide analytical descriptions of the fields. Tversky (1977) attempts to analyze the linguistic "similarity" upon which semantic clusters are based in terms of the semantic features

of the words within the clusters. In the process, he advances a new set-theoretical approach to similarity. In this approach, objects are represented as collections of features, and therefore, similarity is considered as a process of feature-matching. He states that "the similarity of objects is expressed as a linear combination, or a contrast, of the measures of their common and distinctive features" (p. 338). This representation of similarity between objects is called the contrast model. Two constructs are necessary in this model: the contrast rule and the scale ( $f$ ). The contrast rule is used in the assessment of similarity between objects while the scale ( $f$ ) is used to reflect the prominence or salience of different features. This means that " $f$  measures the contribution of any particular (common or distinctive) feature to the similarity between objects." (p. 332).

In his study, 12 vehicles, *bus – car – truck – motorcycle – train – airplane – bicycle – boat – elevator – cart – raft – sled*, served as stimuli, and 48 subjects rated the similarity between 66 pairings of these vehicles on a scale ranging from 0 (no similarity) to 20 (maximal similarity). Another 40 subjects listed the characteristic features of each vehicle. To predict the similarity between vehicles based on the features, measures of the vehicles' common and distinctive features were defined. Also, the number of common and distinctive features was counted to get the simplest measure. Tversky (1977) found that

- (i) it is possible to elicit from subjects detailed features of semantic stimuli such as vehicles...;
- (ii) the listed features can be used to predict similarity according to the contrast model with a reasonable degree of success; and
- (iii) the prediction of similarity is improved when frequency of mention and not merely the number of features is taken into account. (p. 339)

Wierzbicka (1992) focuses on the empirical investigations rather than abstract discussions in the study of the lexicon. Her basic assumptions are:

1. The lexicon of any language can be divided into two parts: a small set of words (or morphemes) that can be regarded as indefinable, and a large set of words that can be regarded as definable and that in fact can be defined in terms of the words from the set of indefinables.
2. For any language, its indefinables can be listed and the other words of this language can be defined in terms of these language-specific indefinables.
3. Although the set of indefinables is in each case language specific, one can hypothesize that each set realizes, in its own way, the same universal and innate "alphabet of human thoughts."... Consequently, the number of indefinables is probably the same in all languages, and the individual indefinables can be matched cross-linguistically. (p. 209-210)

She presents the following indefinables as "semantic primitives" which are employable in the analysis of semantic fields:

*I, you, someone (who), something (what), this, the same, two, all, think, say, know, want, feel, do, happen, good, bad, big, small, can, place (where), time (when), after, under, kind of, part of, like (how), because, if (imagine), more, very, and no ( I don't want).*

These primitives help by showing us "how to distinguish nonarbitrary semantic groupings from arbitrary ones; and how to distinguish discrete, self-contained groupings from open-ended ones" (p. 211). The meaning of a specific word does not depend on the meaning of other words in the lexicon, but is considered a configuration of the indefinables known as "semantic primitives". Therefore, to establish the meaning of a word, one has to compare this word to meanings of

intuitively related words. Once the comparison is complete and the meaning is established, one can compare them more precisely to identify the elements that are different. Wierzbicka emphasizes methodical semantic analysis as a means of establishing “category membership,” an alternative to consulting proficient speakers of the language. An example she provides to clarify her idea is that of basic color concepts, which she describes as something people think of when they see or think about a given object. These concepts are based on human experiences and human environment, and the distinction between them has to do with the presence or absence of light. This means that light is associated with day, sun, and fire. Therefore, “macro-whites” include white, yellow, and red. When fire is thought of as a separate prototype, “macro-red” emerges. On the other hand, the absence of light is associated with the night, the absence of sun, and the absence of fire. “Macro-black” includes black and dark colors.

Wierzbicka proposes that basic colors words have an analyzable semantic structure, which is based on a semantic scheme. Therefore, when one thinks of X one can think of Y “the sky” (for *blue*), things grow out of the ground (for *green*), the sun (for *yellow*), etc.

### *Evidence against the Use of Semantic Clustering*

#### *Interference Theory*

In quite sharp contrast to the assumptions discussed above, the “Interference Theory,” formulated by McGeoch (1942), can be evoked to argue that presenting L2 learners with vocabulary items grouped in semantic clusters actually impedes vocabulary learning rather than acting as a support to learning. The term “interference” first appeared in the literature on the psychology of learning, and is

originally derived from behaviorist theories (Gass & Selinker, 2001). Behaviorists define interference as “[t]he use of the first language (or other languages known) in a second language context when the resulting second language form is incorrect” (p.455). The interference theory has been the dominant theory of forgetting for much of the 19<sup>th</sup> century. It is traceable to the work of Muller and Pilzecker (1900, cited in McGeoch, 1942) to refer to the decrease in retention because of a learning activity that interpolates between original learning and later recall. The theory’s hypothesis is that new knowledge loss or retention is influenced by the nature of the subsequently acquired knowledge. Its main goal is to explain why people forget information they knew. This theory of interference works both ways: retroactive interference/inhibition and proactive interference / inhibition. Retroactive interference refers to the type of interference when newly-learned information inhibits previously-learned information, while the second one refers to the type of interference that occurs when previously-learned information disrupts the learning or recall of subsequent material.

Many attempts have been made to discuss the idea of similarity since it plays a central role within the interference theory (Tinkham, 1994). Related research identifies different types of similarity including visual, verbal, and acoustic similarity. Verbal similarity has been further divided by Underwood, Ekstrand, and Keppel (1965) into formal similarity, for instance repeated letters in a list, meaningful similarity, which refers to the degree of synonymity between words,, and conceptual similarity, which refers to the extent to which words in a list belong to a specific category.

Several studies provide evidence that the kind of similarity involved in semantic clustering affects L2 learners negatively. Kroll and Stewart (1994) conducted a study in which Dutch-English bilingual subjects had to translate Dutch

words to English and vice versa. Two lists were constructed. One list included words grouped by semantic category, while the other list was randomized. Results indicated that there is category interference when bilingual translation is performed in the context of semantically organized lists, since subjects were slower when translating Dutch words into English with categorized words than when words were randomized. However, this result did not hold true for L2-L1 translation.

Altarriba and Mathis (1997) stated that there are semantic interference effects in L2-L1 translation tasks. In their first experiment, English monolinguals and bilingual speakers of English and Spanish had a list of Spanish words paired with three types of English words: the correct translation, an orthographically similar word, and an unrelated word (for example: *cama – bed – beg – owe*). After hearing the word pairs, subjects completed a written matching test. Results showed that subjects had longer response times for orthographically related words than for unrelated words. With expert bilinguals, interference effects were smaller. The second experiment was similar to the first one, except that each Spanish word was paired with the English translation, a semantically similar English word, and an unrelated English word (for example: *cama: bed – sleep – block*). Results of this experiment showed that subjects needed more response time with semantically related words than with unrelated words. The effects of interference were greater with bilingual subjects in this case.

Finkbeiner and Nicol (2003) studied the effect of presenting L2 vocabulary in semantic sets. Monolingual English speakers learned 32 new L2 labels paired with pictures of familiar concepts and then completed recognition and translation tasks in both directions. Results showed that “translation times were significantly slower for words learned in semantic sets versus in random order” and the researchers concluded that “presenting semantically grouped L2 words to learners has a deleterious effect on

learning” (p. 376). Similar results regarding semantic interference effects appear in studies by La Heij, Hooglander, Kerling, and Van der Velden (1996).

*Retroactive interference.* The examples given so far involve the more common kind of interference, called ‘proactive’ to distinguish it from ‘retroactive’ interference. This latter kind of interference occurs when newly learned information inhibits previously learned information. McGeoch and McDonald (1931) conducted two experiments to examine the effects of similarity on learning lists of adjectives. In the first experiment, subjects learned a list of 11 two-syllable adjectives, then subsequently learned other lists, consisting of several kinds of material: adjectives judged to be synonymous, antonymous, or unrelated to the corresponding adjectives in the original list; nonsense syllables; or 3-place numbers. Subjects had five presentations of the original list, ten presentations of the interpolated lists, and a test on the original list. Results showed that there was an increase in the number of adjectives remembered from the original list as the similarity between the interpolated and the original lists decreased. In the second experiment, subjects had the same original list, and three interpolated lists of synonyms reflecting three degrees of closeness. Subjects in this study again had five presentations of the original lists, ten presentations of the interpolated lists, and a test of the original materials. McGeoch and McDonald found that retroactive interference decreased as the similarity between the two materials decreased.

Johnson (1933) conducted a single word study. Subjects had a list of 21 abstract nouns as original material. The interpolated materials were three lists of synonyms for the original nouns reflecting three degrees of similarity of meaning. The fourth condition, the control group, had no interpolated material. Subjects learned the



original list, the interpolated lists, and then they took a recall and a relearning test of the original list. Results showed that subjects had fewer trials for relearning and recalled more items as the similarity between the two materials decreased, supporting the findings of McGeoch and McDonald. She concluded that “on the basis of this fact, similarity of meaning is judged to be a determining factor in retroaction” (p. 386).

McGeoch and McGeoch (1937) also conducted research to study retroactive interference by using an interlist of paired associate words. Subjects learned a list of ten paired unrelated adjectives as original material. The interpolated materials were also lists of paired adjectives of three types: one in which the first item in the pair, the stimulus, was synonymous to its counterpart in the original list; another in which the second item in the pair, the response, was synonymous to its counterpart in the original list; and a third type where both stimulus and response items were either synonymous or unrelated to their counterparts from the original list. Subjects then had to recall and relearn the original list. Results revealed that the original list was recalled most easily when the interpolated list had adjectives unrelated to the original list and was done least easily when the two members of the pair in the interpolated list were synonymous with the members of the original list. An interesting finding is that similarity of stimulus members, not response members, of the pairs was crucial in establishing retroactive interference.

Bugelski and Cadwallader (1956) conducted an interlist paired associate study to examine the influence of similarity of interpolated materials on the retention of original learning. 144 college students learned 15 sets of 13 associates with different degrees of similarity. Stimulus items varied between identical, similar, less similar, and neutral. Response items were adjectives varying between identical, similar, neutral and opposed. Subjects learned all of the words, had a rest for two minutes,

learned the interpolated materials, had another two-minute rest, and then recalled the original list. Bugelski and Cadwallader found that negative effect increases as similarity of stimuli increases in the interpolated materials.

Pollak (1969) studied the same issue. Subjects learned a list of 10 flower names and a list of 10 girls' names as original material. Then they learned a list of girls' names, or a list of flower names, or a list of names that can be used to refer to a girl or a flower, e. g., "Rose", or a list of animal names, as interpolated materials; a control group received no interpolated list. . Results showed that items of the original list which were conceptually similar to the interpolated lists tended to be more often forgotten. Those who learned the list of animal names, and who thus did not have semantically similar interpolated lists, forgot fewer words than those who learned the list of girl-flower names.

*Proactive interference.* As stated earlier, the typical case of interference involves what was originally termed 'proactive' interference. This type of interference occurs when previously learned information disrupts the learning or recall of subsequent material. Gibson (1941) identified this concept when he conducted a paired associate study. Subjects first learned the original material, which consisted of paired associate lists. The stimuli were irregular geometric forms which were judged for degree of similarity (similar – dissimilar) while the responses were nonsense syllables and were not manipulated for similarity. Then subjects learned the interpolated list in which stimulus forms corresponded to the positions of their counterparts on the original list. Finally, subjects had recall and relearning tests of the original list. The number of recalled words from the original list decreased as the similarity between the two lists increased. This result showed that retroactive

interference occurs and that subjects had difficulty learning the interpolated list because of its increased similarity to the original list. This showed that proactive interference is a function of similarity between two learned lists. Gibson also found that subjects had more difficulty learning the interpolated lists since there were higher levels of similarity between the original lists and the interpolated lists.

Melton and Von Lackum (1941) provided a comparison of retroactive and proactive inhibition when the formal similarity of interpolated and original lists is high or low. There were six conditions: Condition I was the retroactive inhibition control group. Condition II was the retroactive inhibition group which learned a list of 10 three-consonant groups constructed from a list of nine consonants. Then subjects learned the interpolated list of these different consonant trigrams constructed from the same set of consonants. After a 20-minute rest period, subjects relearned the original list. Condition III was similar to Condition II except that the two lists were constructed from lists of different consonants. Condition IV was the proactive inhibition control group; these subjects learned only the interpolated list and then relearned it after a 20-minute rest. Conditions V and VI were similar to Condition II, in that subjects relearned the interpolated list. Results showed that the groups which had the two similar lists recalled the least, the groups which had the two dissimilar lists recalled more, and the groups that had no proactive or retroactive materials recalled the most. Melton and Von Lackum (1941) concluded that “the amount of inhibition of both lists is greater when the two lists are similar than when the two lists are dissimilar” (p. 173). Therefore, proactive interference is related to the similarity between the items to –be learned and the proactive materials.

The studies discussed above show interference in interlist designs. That is, interference across lists. In the following section, I provide studies which include intralist design. That is, they show interference within lists.

*Intralist interference.* As mentioned in Baddeley (1990), an experimental procedure was introduced by Brown (1958) and Peterson and Peterson (1959). In this procedure, the subject is presented with a single item to be remembered. Then he is given a certain task to prevent rehearsal. Finally, he is given a recall test of the item. This procedure is intended to study forgetting, which reflects the decay of short-term memory. But the study of Keppel and Underwood (1962) showed that forgetting resulted from proactive interference. This kind of procedure is a measurement of short-term memory because the time between learning and testing is less than 30 seconds. This is different from the procedure developed earlier to examine interlist interference, in which there is a longer time between the learning and the testing and which thus yields long-term memory measurements.

The studies of Underwood, Ekstrand and Keppel (1965) deals with intralist interference, conceptual similarity, and paired association learning. In the first experiment, subjects were presented with eight 12-pair lists of paired associates. The first member of the pair, the stimulus, was a noun e.g. *horse*. The second member of the pair, the response, was a double letter (e.g. AA). There were four conditions. The first condition had the stimulus members as high frequency names from a single category (animals) such as *dog* and *cat*. The second condition used low frequency nouns such as *lynx* and *weasel*. The third condition had the stimulus members as unrelated high frequency nouns such as *leg* and *sofa*. The fourth condition used unrelated low frequency nouns such as *spleen* and *ottoman*. Then subjects received 15

anticipation trials during which they were supposed to give correct responses per trial. Results showed that subjects learned the unrelated words more easily than the conceptually similar words regardless of the frequency. Therefore, the researchers concluded that conceptual similarity impedes paired associate learning. In the second experiment, Underwood et al. (1965) replicated the first trial except that they changed the items in two lists, just to make sure the results were not affected by the peculiarities of any item in one or both of the lists. The results were similar to the first experiment's results. In a third experiment, Underwood et al. (1965) varied the degree of conceptual similarity among stimulus items. In the five lists used in this experiment, there were 1, 2, 3, 6, or 12 different categories. This means that words were similar in the list with 1 category while those in the 12 categories were not. The first list was drawn from a single category (high similarity). The second list, with two categories, had 6 words representing each category. With three categories in the third list, there were 4 words representing each category. With 6 categories in the fourth list, there were 2 words representing each category. The last list was drawn from 12 categories (no similarity). Subjects had 20 anticipation trials. Results indicated that as conceptual similarity increased, the intralist interference increased and "as number of concepts increases (similarity decreases)..., [the] amount learned increases" (p. 458).

In sum, a significant body of literature shows that similarity between the to-be-learned information and information learned before or after the critical information leads to interference (proactive, retroactive and intralist interference), which in turn leads to learning difficulties.

### *The Distinctiveness Hypothesis*

Although the behaviorist approach to learning no longer dominates the field, the effect of the similarity of stimuli on learning is still a matter of concern for many psychologists. The distinctiveness hypothesis was developed as an alternative to the “depth of processing” theory developed by Craik and Lockhart (1972). According to Craik and Lockhart’s theory, there are series of processing stages, and information semantically processed is better remembered than information processed without attention to meaning (e.g. orthographically or phonetically) because of the greater depth of semantic processing. The Distinctiveness Hypothesis, which received considerable attention during the 1980s, considers the ease with which distinctive information is learned. It states that

the most important factor in recognition memory is the extent to which the test-trial encoding contains information that is unique to the study-trial encoding. In the case of an item that is phonemically encoded at input and at test, there would appear to be substantial encoding overlap. (Eysenck, 1979, p.111)

The claim is that people remember distinct items better than they remember those that are nondistinct. Hunt and Mitchell (1982) say “the distinctiveness hypothesis focuses on the utility of encoding information in reconstructing to-be-remembered information. In general, information will be more useful in reconstruction or retrieval if it is unique to the to-be-remembered item” (p. 81).

Research demonstrates that, as this hypothesis predicts, distinctiveness of information facilitates memory. Hunt and Elliott (1980) studied orthographic distinctiveness and retention rates for distinct and common words in a study of

six experiments. Subjects had a list of 20 orthographically distinct words such as khaki and afghan and another list of 20 orthographically common words such as kennel and airways. Then they had free-recall tests. Results showed that subjects remembered more orthographically distinct words than they did orthographically common words. Hunt and Elliott concluded that “nonsemantic information is useful, and perhaps essential, in long-term memory” (p. 71).

Expanding the study of the effects of distinctiveness to include conceptual distinctiveness, Hunt and Mitchell (1982) studied the effects of orthographic and conceptual distinctiveness on recall. Subjects learned a 20 item word list consisting of four critical words and 16 surrounding background words in four conditions. In all conditions, all of the background words were orthographically common words from the same conceptual category. In the first condition, the critical words were from the same conceptual category but orthographically distinctive (such as the animal items *hyena* and *lynx*). In the second condition, they were orthographically common but from a different conceptual category (such as *accordion*, *mandolin*). In the third one, they were both distinctive and from a different category (such as *ukulele*, *cymbals*). In the last one, they were orthographically common and from the conceptual category of the background words (such as *raccoon*, *bison*). Results of free-recall tests showed that category and orthographic isolation had positive effects on recall.

In a two-experiment study employing recognition and recall tests, Schmidt (1985) investigated the learning of conceptually distinctive words. In the first experiment, subjects learned either a list of 24 words with four belonging to one conceptual category while the other 20 belonged to another category, or a list of 24 words all belonging to the same conceptual category. After that, subjects took a recognition test

in which they decided whether words belonged to the original list. Results showed that “conceptually distinctive words were better recognized than the same words from the homogeneous lists” (p. 570). In the second experiment, subjects learned either a list of 24 conceptually mixed words or a 24-word list that was conceptually homogeneous. As a measurement of retention, subjects completed a recall test. Results indicated that “the distinctive items are very likely to be retrieved and are retrieved as a group. In contrast, background items are less likely to be recalled from lists containing distinctive targets” (p. 574).

### Thematic Clustering

Lexical semanticists, when investigating the way speakers organize words in their mental lexicons, propose that speakers subconsciously organize words in “frames” or “schemas” with reference to the speaker’s background knowledge rather than in semantic fields (Fillmore, 1985). A cluster of words drawn from such a frame or schema might include *frog*, *pond*, *hop*, *swim*, *green*, and *slippery*; words of different parts of speech that are all closely associated with a common thematic concept (in this case, *frog*). Such words reflect the schemata that English speakers share for a word (Celce-Murcia & Olshtain, 2000). Based on associative strength, clusters of this sort are cognitively rather than linguistically derived, and consequently would appear to fit most easily into learning-centered second language acquisition programs, which are more concerned with learning processes than with linguistic analysis.

Thematic clustering depends upon psychological associations between clustered words and a shared thematic concept. *Haunted*, *ghost*, *yell*, *moonlight*, and *groan*, for instance are said to be thematically related, as they are all words drawn



from a haunted house schema. Neither the Interference Theory nor the Distinctiveness Hypothesis attempted to predict the effect of thematic clustering. Although researchers have been concerned with similar words in studies of interference, word clusters such as *frog, green, swim, and slippery* have not been their concern when seeking evidence for interference. Similarly, sets of words such as *car, raceway, team, champion, and drive*, which are not similar, have not attracted researchers of the Distinctiveness Hypothesis to study their learnability.

### *Justification for Thematic Clustering*

#### *Frames*

Within “frame semantics,” as labeled by Fillmore (1985), “speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes. Within such an approach, words or word senses are not related to each other directly, word to word, but only by way of their links to common background frames and indications of the manner in which their meanings highlight particular elements of such frames” (Fillmore & Atkins, 1992, p. 77). An example of such a frame that Fillmore and Arkins (1992) provide includes the verbs *buy, sell, charge, spend, pay, and cost*. Those verbs are linked with nouns such as *buyer, seller, goods, and money* in a Commercial Transaction frame.

Frames, like semantic fields, may be described intuitively or analytically. When Fillmore and Atkins (1992) tried to account for how words fit within frames, they relied on their intuitive understanding of “valence description,” which specifies the interrelations between words and their surrounding semantic and syntactic contexts. They believe that words are not directly linked to each other. Rather, they

are related through their links to common background frames, experiences, beliefs, and practices. In the example of the Commercial Transaction frame, a person acquires possession of something from another person by paying a certain amount of money. The common background of this frame "requires an understanding of property ownership, a money economy, implicit contracts, and a great deal more. This schema incorporates ('inherits') many of the structural properties of a simple exchange frame, but it adds to that base a number of further specifications regarding ownership, contractual acts, and the trappings of a money economy" (p. 78).

Barsalou (1992) presents a system for the analytical analysis of frames. He begins by demonstrating the problems with feature list representations of categories in knowledge theories. A feature list of birds might include the words *feathers*, *wings*, *tail*, *claws*, *flies*, *beak*, *head*, and so forth. These features are produced by speakers for the category and are intercorrelated with either excitatory or inhibitory relations as they do or do not co-occur in connectionist models. According to Barsalou, theories in psychology assume that

representations contain more than feature list. In many cases, however, the additional structure – which tends to be frame-like – remains implicit theoretically and receives little attention empirically. As a result, these representations essentially reduce to feature lists. Consider work on artificial category learning in cognitive psychology. Typically, this work assumes the presence of frames in category representations.... A frame includes a co-occurring set of abstract attributes that adopt different values across exemplars. (Barsalou, 1992, p. 23)

Barsalou goes on to explain the difference between features and the attribute-value sets he proposes: namely, that features are independent representational

components which constitute a single level of analysis, while attribute-value sets are interrelational sets of representational components which constitute at least two levels of analysis. Barsalou (1992) goes on to review literature to establish that people encode characteristics of exemplars as values based on attributes rather than as features, that they are cognizant of the relations between representational components and that they do not store these independently of one another.

As Barsalou explains, there are three components of frames. The first of these is the attribute-value set. The core of a frame is the attribute that might adopt different values. The frame of CAR, for example, has the following as attributes: fuel, engine, transmission, driver, and wheels. Of course these are not all of the attributes that might exist. The FUEL attribute might have the following values: gasoline, diesel, and gasohol. ENGINE has 4 cylinder, 6 cylinder, and 8 cylinder values. Thus, attribute can be defined as "a concept that describes an aspect of at least some category members" (p. 30) while 'value' refers to the subordinate concepts of the attribute.

The second component of a frame is its 'structural invariants,' a term which refers to correlational relations between a frame and its attributes. For example, people understand that there is an operational relation in which the driver controls the speed of an engine. These structural invariants can include spatial relations (such as the relation between *seat* and *back* in the frame of *chair*), temporal relations (such as the relation between *eating* and *paying* in the frame of *dining out*), causal relations (such as the relation between *fertilization* and *birth* in the frame of *reproduction*), and intentional relations (such as the relation between *motive* and *attach* in the frame of *murder*).

The third component Barsalou defines is ‘constraints,’ which also involve relations between a frame’s attributes. "Rather being normative, constraints produce systematic variability in attribute values. The central assumption underlying constraints is that values of frame attributes are not independent of one another. Instead, values constrain each other in powerful and complex manners" (p. 37). Barsalou (1992) describes these relations in the following types of constraints:

1. Attribute Constraints: These are the general rules that constrain attribute values globally. In a *transportation* frame, there could be two attribute constraints:

- a negative attribute constraint (-) between *speed* and *duration*:

(As a form of transportation becomes faster, its duration becomes shorter.)

- a positive attribute constraint (+):

(As a form of transportation becomes faster, its cost becomes higher.)

While some of these attribute constraints are logical or empirical truths, others indicate personal preferences and statistical patterns.

2. Value Constraints: These are the specific rules which relate specific sets of values locally. For example, there is an *enabling* relation between *Rockies* and *Surfing* in the frame of VACATION. In this relation, there is a certain value of the *location* attribute that constrains a certain value of the *activity* attribute.

The same can be said about the relation of *enabling* between *San Diego* and *surfing*. There are more complex value constraints such as the one in the relation of requirement between *surfing* and *ocean beach*. It is considered complex because it crosses levels within the frame.

3. Contextual Constraints: These constraints occur when one aspect of a situation constrains another, such as physical constraints in nature. For

example, *speed* of transportation constrains its *duration* over a fixed distance. Similarly, the activity of *surfing* requires an *ocean beach*. Contextual constraints also reflect cultural conventions. For example, people's *income* and the *taxes* they pay may bear a relationship to one another. In general, the various aspects of a particular situation are not independent of one another. Instead, physical and cultural mechanisms place constraints on combinations of compatible attribute values.... Contextual constraints can either be attribute constraints or value constraints. (p. 39)

4. Optimizations: These are the constraints that reflect the goals of an agent. For example, the goal of an agent of *short travel* constrains the value of *duration* in the *transportation* frame to be *short*. Optimization constraints can be both attribute and value constraints, and they require that one value excel beyond all others.

To conclude, Barsalou (1992) states: "Frames contain attribute-value sets. Attributes are concepts that represent aspects of a category's members, and values are subordinate concepts of attributes" (p. 43). These different values are adapted across exemplars or category members. In addition, "because frames contain attribute-value sets and relations, they provide natural solutions to the problems of feature lists" (p. 28).

### *Evidence Supporting the Use of Thematic Clustering*

#### *Schema Theory*

This theory explains how old information possessed by the learner influences the learning of new information. It aims to explain the way different types of knowledge are learned and people's interpretation of the world from a psychological

perspective. Schema theory is among the most intellectually exciting areas in cognitive psychology (Brewer & Nakamura, 1984). The theory was developed by Bartlett in his book *Remembering* (1932). Bartlett defined the term "schema" as "an active organization of past reactions, or of past experiences, which must always be supposed to be operating in any well-adapted organic response" (cited from Brewer & Nakamura, 1984, p. 120-121). This means that he considered schemas as unconscious mental processes. Furthermore, he hypothesized that schemas are complex unconscious knowledge structures organized into generic cognitive representation. But this view has a problem as a general way of representing memories. When an individual tries to recall an image of an office, he recalls chairs, table, fax machine, typewriter, computer, and so forth, and all of these pieces of information are generic. The theory, as developed by Brewer & Nakamura (1984), cannot account for the recall of nongeneric information such as the detail that the table was wooden or that the computer was a DELL.

The idea that schemas are unconscious was rejected by psychologists and philosophers who claim that psychology data are restricted only to conscious rather than unconscious phenomena. Behaviorists also rejected the same idea, claiming that the data of psychology are restricted to observations of overt behaviors. But after much discussion, the idea has now been universally accepted (Brewer & Nakamura, 1984).

Based on the definitions provided by Bartlett, a schema has two basic properties; it is organized and composed of old knowledge and past experiences. Modern schema theory goes back to Minsky (1975) and Rumelhart (1975, 1984). As Brewer & Nakamura (1984) state,

Molinsky introduces the construct of the frame. A frame has fixed "nodes" that provide its basic structure. It has "slots" that can be filled by specific information from the environment. This provides additional structure, since a slot will only accept a particular class of instances. If there is no information to the contrary the slots are filled with "default assignments. (p. 132-133).

This means that slots are either filled with compulsory values (the cat is an animal), or with default values (the cat has four legs), or are empty until certain values are instantiated from a specific situation (the cat's color is grey).

Rumelhart (1984) defines schema as "a data structure for representing the generic concepts stored in memory. There are schemata representing our knowledge about all concepts: those underlying objects, social situations, events, sequences of events, actions and sequences of actions" (p. 163). Rumelhart & Ortony (1977) follow Minsky in the idea that in schemas there are variables with values, and once these variables are assigned, then schemas are instantiated. It is these instantiated schemas that are stored in memory. When one recalls generic schemas, he/she uses information to interpret a specific memory from the instantiated schemas. This idea becomes clear when one thinks of the reading process as one where a "top down" perceptual process interacts with a "bottom up" data-driven process.

If a reader arrives at the schema intended by the author the text has been correctly comprehended. If the reader can find no schema to accept the text information the text is not comprehended. If the reader finds a schema, but not the one intended by the author, the text is misinterpreted. (Brewer & Nakamura, 1984. p. 134).

Generally speaking, schemas help us by organizing our knowledge, assisting with recall, and guiding our behavior. They help us make sense of current experiences and interpret situations. An example of a schema that every student has in mind is that of the library. Students have schemata for the way books are arranged, for people who work in the library and their functions, and for things you expect to see in the library, (carrels, and so on).

Much research has been done studying the schema theory to show that information which is schema-related is recalled better than schema-unrelated information. More research has been done concentrating on full sentences rather than nonsense syllables. Smith, Adams, and Schorr (1978) presented their subjects with two pairs of unrelated sentences like these:

*Herb produced sour notes*

*Herb realized the seam was split*

Then they presented subjects with a third sentence. This sentence provided integration with the previous two sentences, (called a theme / integrated sentence):

*Herb played a damaged bagpipe*

Alternatively, the third sentence provided no integration (called theme-control/ unintegrated):

*Herb painted an old barn.*

Results showed that subjects could learn sentences which showed integration and were schema-related information with the first two sentences more easily than they could learn the theme-control sentences. And “with unintegrated facts, the time to recognize a test sentence increased with the number of facts learned about the person mentioned in that sentence; with integrated facts, no such increase was found” (p. 438).



Similarly, Goetz, Anderson, and Schallert (1981) provided subjects with two types of sentences—integrated and unintegrated. Two native speakers of American English made judgments of integratedness. In integrated sentences which describe ordinary situations, the actors, actions, goals, and objects “go together.” For example:

*The pilot flew from New York to Los Angeles*

*The customer wrote the company a complaint*

Unintegrated sentences describe unlikely but possible situations. For example:

*The pony danced from the theater to the church.*

*The comedian supplied glassware to he convicts.*

After judging the sentences, subjects took a free recall test on words and sentences. Results showed that subjects recalled more words from integrated sentences than from unintegrated sentences. Also, they recalled more integrated sentences than they did unintegrated sentences, but the difference in this case was not significant.

There is a great body of research studying schema theory and L1 reading. Steffensen, Joag-dev, and Anderson (1979), for example, in research demonstrating the connection between the reader’s schemata and his comprehension of the text, studied how two groups of people with different cultural heritages comprehended two written texts. American and Indian subjects were presented two letters about an Indian wedding and American wedding, and then they were tested using a vocabulary test and free recall. Results indicated that American subjects read the American wedding letter faster than they read the Indian letter. The Indian subjects read the Indian wedding letter faster than they read the American one. Moreover, the Americans recalled more ideas units from the letter about an American wedding than from the other letter, and the reverse was also true. Steffensen et al. concluded that “the

schemata embodying background knowledge about the content of a discourse exert a profound influence on how well the discourse will be comprehended, learned, and remembered” (p. 19).

Attention has been paid to the L2 learner’s background knowledge and reading comprehension. Carrell (1987) studied the effects of content schemata, which refers to the reader’s knowledge about the text’s content domain, and formal schemata, the reader’s knowledge about a text’s rhetorical structures, on ESL reading comprehension. Two groups of students with high-intermediate level English as a Second Language proficiency and different cultural backgrounds were given two texts to be read by each group. The first text had culturally familiar content while the second one had culturally unfamiliar content. With each group, half of the subjects read the text in a well-organized rhetorical format, while the other half read it in an unfamiliar format. After reading, both groups completed a multiple-choice test and a debriefing questionnaire. Results showed that subjects used both content and formal schemas and that content schema affected reading comprehension in mixed conditions (familiar rhetoric form, unfamiliar content; familiar content, unfamiliar rhetoric form).

Nagy and Scott (1990) studied the relationship between schemas and L1 vocabulary acquisition. Seventh-grade, tenth-grade, and undergraduate students rated 96 definitions based on the likelihood that the definitions represented real definitions for the words. The study demonstrated that “people possess both general and specific word schemas and that their knowledge about words increases from junior high school through college” (p. 124-125).

### Semantic Clustering Vs Thematic Clustering

When it comes to using sets of related words, several studies provide evidence that semantic clustering affects L2 learners negatively. One of the first studies in this field, which was generated by the interference theory and which extends the research on similarity between stimulus and response to the learning of L2 vocabulary, was done by Higa (1963). He used seven paired-associate lists with seven kinds of meaning relationships to examine whether a list of high-association words is more difficult than a list of low-association words. He found that the lists of strong associates are difficult to learn.

Tinkham (1993) conducted a study showing that presenting L2 vocabulary items grouped in semantic sets inhibits learning. He provided subjects with lists of semantically related (*shirt, jacket, sweater*) and unrelated words (*rain, car, frog*) associated with artificial words. Subjects were tested by recall trials-to-criterion tests. Tinkham found that subjects learned the semantically related English words more slowly and with more difficulty than they learned the unrelated English words.

Kroll and Stewart (1994) conducted a study in which Dutch-English bilingual subjects had to translate Dutch words to English and vice versa. Two lists were constructed. One list included words grouped by semantic category, while the other list was randomized. Results indicated that there is category interference when bilingual translation is performed in the context of semantically organized lists, since subjects were slower translating Dutch words into English with categorized words than when words were randomized. But this result did not hold for L2-L1 translation.

Waring (1997) replicated the study of Tinkham (1993). In two experiments, Waring provided native-speaking Japanese subjects with six Japanese word-pairs, including three semantically related words sharing a common concept of “clothes”

and three unrelated words as stimuli. Responses were Japanese artificial words. In the second experiment, subjects were required to learn two separate sets of six Japanese word-pairs; semantically related words (types of fruit) and semantically unrelated words (such as *mountain, television, sky, mouse*). Results of trials-to-criterion showed that subjects learned the related word-pairs more slowly than they learned the unrelated word-pairs and that “presenting new words that share a common superordinate in a set of words to learn does interfere with learning” (p. 267).

Comparing semantic clustering and thematic clustering, Tinkham (1997) conducted a study with two experiments. In the first experiment, subjects were presented two six-pair sets of associate pairs. The first list included three semantically related English words and three unrelated English words paired with artificial words. The second list included three thematically associated English words and three unassociated English words paired with artificial words. Subjects studied the lists in two modalities (oral and written) and had recall and recognition tests. In the second experiments, there were four six-pair lists; semantically related, semantically unrelated, thematically associated, and thematically unassociated words paired with artificial words. Results showed that semantic clustering of L2 vocabulary items was a detriment to learning, while thematic clustering was a facilitator. Moreover, artificial words paired with semantically related words were learned with more difficulty than those paired with thematically related words. On the other hand, artificial words paired with thematically related words were learned more easily than those paired with unassociated words.

Schneider, Healy, and Bourne (1998) found that grouping words by category (e.g. body parts) facilitates initial learning. But on long-term retention, sets of unrelated words were again easier to retain than sets of related words.

Hippner-Page (2000) compared the effects of semantic and thematic clustering on the learning of English vocabulary items by third, fourth, and fifth graders. Two groups of students of English as a second language were given instruction and treatment about one of the methods and then were tested. Results showed that one method of clustering was helpful to one group while the other method was helpful to the second group. To justify this finding, the researcher cited “task-learning effects”: some students might know some words from the semantic set or theme; there might also be prototypicality effects, as well as differences in subjects’ learning styles.

To study the effect of presenting L2 vocabulary in semantic sets, Finkbeiner and Nicol (2003) had monolingual English speakers learn 32 new L2 labels paired with pictures of familiar concepts. Then subjects had recognition and translation tasks in both directions. Results showed that “translation times were significantly slower for words learned in semantic sets versus in random order” and the researchers concluded that “presenting semantically grouped L2 words to learners has a deleterious effect on learning” (p. 376). The same results for semantic interference effects appeared in studies done by La Heij, Hooglander, Kerling, and Van Der Velden (1996).

### Conclusion

To conclude, empirically, there is little if any strong, direct evidence that grouping vocabulary in semantic clusters facilitates learning. Rather, a growing body of research indicates that this way of presenting new vocabulary makes learning more difficult and interferes with learning similar words. Interference Theory and the Distinctiveness Hypothesis predict that semantic clustering will inhibit learning vocabulary. The interference theory states that when words which are similar or share many common elements are being learned, they interfere with each other and impair

their retention. Researchers interested in the Distinctiveness Hypothesis collected data which strongly suggest that distinctiveness is a very crucial factor in the learning of new information, and that as the distinctiveness (non-similarity) of the to-be-learned information increases, so does the ease with which that information is learned.

However, the results of research guided by the interference theory and the distinctiveness hypothesis should not suggest that all types of clustering inhibit learning L2 words. Grouping words in thematic sets has been shown to facilitate learning vocabulary. This result is consistent with Schema theory, which indicates that learning information is easier when it is related to background knowledge.

### Context

The use of context to explain the meaning of words is seen as a good method for vocabulary acquisition. It is considered the best strategy for learning low frequency words which learners encounter infrequently. When they meet such words, "it is better to use context clues to infer their meaning than to spent time on learning the words themselves" (Na & Nation, 1985, p. 33). The importance of context is supported by the well-known Schema Theory covered earlier. The learner's background knowledge of the passage plays a critical role in guessing meanings successfully. Without relevant schemata, the learner cannot perceive, learn, or recall the new information (Kang, 1995). Thus, both old and new information are used to incorporate the new information.

Various studies have investigated this issue, examining the effectiveness of contextual learning and focusing on ways readers guess the meanings of unfamiliar words. The question "What is a context?" has been studied by many researchers attempting to reach a common definition of this term. Muller (1970) (mentioned in

Engelbar & Theuerkauf (1999) differentiates between "verbal context" which is equated with situation (speaker, time, location) and "nonverbal context" which is seen as the real context (words, sentences, etc.). Beheydt (1987) also describes two types of context. The first is "pregnant context," a type of "nonverbal context" which provides a description of the new word and evokes the prototypical scene to which the word belongs. The second, "pregnant semantic context," is a synonym for "verbal context" consisting of morphological, syntactic, and collocational information. Nation and Coady (1988) provided "context within a text" and "general context" as two types of context. The first one, "context within a text," refers to the morphological, syntactic, and discourse information included in a text. The second type, "general context," refers to the background knowledge on the topic. Engelbar and Theuerkauf (1999) differentiated between "verbal context" and "nonverbal context" based on a review of literature on this issue. According to their definition, "verbal context" is the linguistic environment of an unknown word. This environment includes morphological, syntactic, phonetic, and semantic information and each, in turn, includes a number of clues. "Nonverbal context," on the other hand, is the content-oriented environment. This environment includes the situative context (speaker, location, time, matter, and acting person) surrounding the unknown word and the learner's world knowledge and expertise.

When it comes to learning from context, two types of learning appear in relevant research. Incidental vocabulary learning and direct intentional learning of vocabulary are seen by Nation (2001) as complementary activities, each enhancing the learning that comes from the other. He writes:

Learning from context is taken to mean the incidental learning of vocabulary from reading or listening to normal language use while the main focus of the

learner's attention is on the message of the text. The text may be short or long. Learning from context thus includes learning from extensive reading, learning from taking part in conversations, and learning from listening to stories, films, television or the radio. (p. 232).

Inferring the meanings of unfamiliar words from context is an important process that involves the use of relevant schemata in order to capture the message in the unknown words. Several studies have emphasized the importance of context in learning vocabulary. Nist and Olejnik (1995) studied their subjects' abilities to learn and remember new vocabulary depending on the strength of context and adequacy of definition. They found out that there was no interaction between levels of context and levels of definitions. Adequate definitions positively influenced performance on vocabulary measures regardless of the strength of the context provided. Moreover, context helped learners' performance when they saw a word in context and then looked at its definition on a multiple-choice test.

Prince (1996) looked at weak and advanced learners when examining two conditions, L1 translation (learning words in pairs) and context use, in terms of the performance of learners in accessing and using the learned materials. Translation condition subjects examined 44 English words accompanied by their French equivalents, while context condition subjects looked at sentences in English with one unknown word per sentence. Thus, context condition provided no translation or definition. Learners were tested in two ways: by translating isolated words in the L2-L1 direction and by filling the blank in sentences. Results revealed that translation learning was superior in terms of quantity (number of correct answers) for both weak and advanced learners. Weak learners outperformed advanced ones where learning by



translation was tested by translation, while the advanced learners who learned through context did slightly better on the sentence completion test than on the translation test.

### *Studies Guessing from Context with Second Language Learners*

Various studies have investigated second language learners' guessing from context. Bensoussan and Laufer (1984), for example, studied second language learners' levels of proficiency and word guessability from the context. Sixty first-year students were given a list of 70 words to translate into their first language. A week later, they were given the same list with a text containing all 70 words followed by comprehension questions. Bensoussan and Laufer conclude that level of proficiency has no effect on the ability to guess meanings of unfamiliar words. They also concluded that word guessability is less a function of using context than of applying "preconceived notions" about the meanings of words.

### *Causes of Poor Guessing*

The form of the word that needs to be guessed is a major difficulty facing learners. Bensoussan and Laufer (1984) found that some words were easily confused with other words that sound or look similar, such as implication / application. Second language learners produced wrong guesses because of this resemblance. In a study with second language learners, Li (1988) examined the effects of cue adequacy on inferring and remembering the meanings of new words in discrete, semantically disconnected sentences. He compared the effects of cue adequacy in both reading and listening contexts. He found that subjects receiving cue-adequate sentences reported

greater ease in inferring and remembering the meanings of unfamiliar words in context, in contrast to those receiving cue-inadequate sentences. Also, subjects reading the sentences, in contrast to those listening to them, scored higher in both inferring and remembering contextual meanings of unfamiliar words. Na and Nation (1985) saw that words in a high density text (1 unknown word in 10) were more difficult to guess than words in a low density text (1 word in 25).

### Memory

Since the present study includes two tests of memory, one immediate and one delayed, a brief review of recent work on memory is appropriate. Baddeley (1990) defines human memory as "a system for storing and retrieving information, information that is, of course, acquired through our senses" (p.13). This means that the memory records everything experienced through the senses either by seeing, smelling, hearing, tasting, or touching a thing. Therefore, to know how the human memory works, we need to know how stimuli, either visual or auditory, are processed.

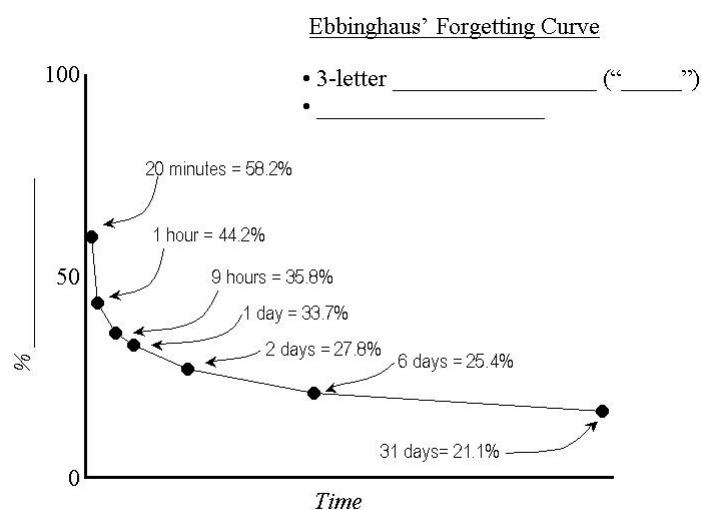
When something is seen or heard, it is stored first in short-term memory. This type of memory is temporary and only stores a limited amount of information. This stage lasts for milliseconds but further manipulation takes place in this type of memory, leading information to be held for seconds. When the attention of the person is diverted to another task, the information is no longer available in the short-term memory. After this stage, sensory-based information is integrated "with information from other sources through the operation of the limited capacity working memory system..... Such information is also fed into long-term memory, which although relying heavily on coding in terms of meaning, is also able to store more specifically sensory characteristics such as those involved in memory for faces and scenes, voices

and tunes" (Baddeley, 1990. p. 38). During the 1950s, there was virtually no interaction between researchers concentrating on these two types of memory. Short-term memory was extensively studied in Britain, and most of this research used information processing approaches. In contrast, long-term memory was studied by Americans who were interested in the verbal learning paradigms within the framework of the interference theory. In the early 1960s, researchers started distinguishing between these two types of memory (Baddeley, 1986). Long-term memory is assumed to be very much larger and more durable than short-term memory. Moreover, long-term memory storage is associative, in that it relates items to one another. Also the time needed for storing a new memory trace in long-term memory is estimated to be longer, about ten seconds (Ericsson & Kintsch, 1995).

In retaining information across a long period of time, one depends on two forms of long-term memory: episodic memory and semantic memory. Episodic memory holds personal experiences that took place at a certain time. Semantic memory, on the other hand, involves the storage of facts and general information. A third type, proposed by Schacter (2001), "intervenes between the moment of perception and eventual establishment of long-lasting episodic or semantic memories"(p.27-28). This is called working memory. Baddeley (1986) defines this as "a system for the temporary holding and manipulating of information during the performance of a range of cognitive tasks" (p. 34).

When it comes to forgetting, researchers have noticed that people start forgetting seconds after they see or hear an event. The first researcher to study memory scientifically was Hermann Ebbinghaus, in the late 1880s. He studied human memory by learning and forgetting artificial concepts. He studied 13 meaningless syllables and tested himself at six different times ranging from 20 minutes to one

complete month after his learning trial. He noticed that his rate of forgetting reached 60% of the nonsense syllables as early as nine hours after studying. The forgetting rate increased to 75% a month later. Ebbinghaus' main conclusion was that most forgetting takes place early after learning and then slows down. In other words, the rate of forgetting is nonlinear in that forgetting occurs rapidly at the beginning and then slows down.



*Figure 1*  
Ebbinghaus' forgetting curve.

In very long studies, Bahrick (1984) and Bahrick and Philps (1987) studied the Spanish vocabulary retention of subjects who learned Spanish in high school and college. In the first study, and in order to provide normative data regarding long-term retention of semantic memory content, Bahrick (1984) tested the retention of Spanish learned in school for 733 individuals. He tested subjects at different times ranging from immediately after learning the Spanish vocabulary to fifty years later. He found out that the amount of forgetting was rapid during the first three years, then retention remains unchanged for periods of up to 30 years. After that, a final decline took place. Bahrick found out that a huge part of the information individuals had was accessible regardless of the fact they had never used it before. In other words, materials that

subjects learned could be recallable for more than 25 years if materials were not lost in the first 5 years after learning.

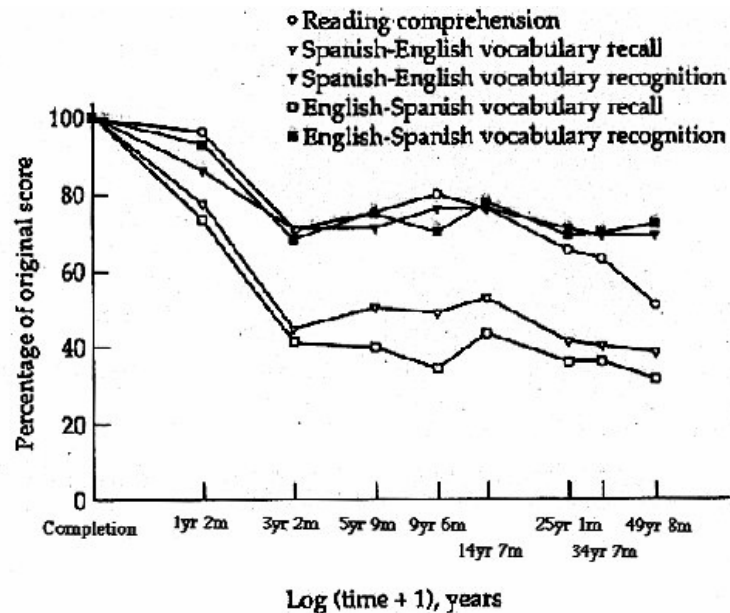


Figure 2  
Forgetting rate in Bahrick's study.

In the second study, Bahrick and Philips (1987) tested 35 individuals who had to learn 50 English-Spanish word pairs for recall and recognition after 8 years. Results indicated that rapid forgetting takes place in the beginning followed by an unchanged period. Results also showed that 10% of vocabulary learned in six to eight sessions could be recalled after 8 years.

Other studies have focused on autobiographical incidents. In a study that shows the importance of cues to retention, Wagenaar (1986) kept a diary that included his personal memories every day for four years. He kept writing details of events such as what happened, who was involved, where the event happened, and so forth. During that period, he did not review the diary. After four years he tested himself, probing his memory with a number of cues such as when, where, how, and who. Wagenaar found

that he could remember more details of the events for which he had provided more cues.

In another study, Linton (1975) studied her own long-term memory by collecting data over a period of five years. In the study, events were recorded, scheduled for test, and tested. At the end of each month, she selected a sample of cards on which events were written on one side and a short description was written on the other. She read the description of 3006 items or events and tried to construct the date of each event. If she failed to remember the date, that event was excluded. In the first ten months, she failed to recognize 8 items. During the second 10 months, 90 more items were not recognized. Over a period of 20 months, 80% of lost items were forgotten after the first 6 months, 20% in the second 6 months, and none during the last 8 months.

Other studies have focused on complex skills such as riding a bicycle, cardio-pulmonary resuscitation, or first aid procedures. Baddeley and Longman (1978), Glendon, McKenna, Blaylock, and Hunt (1987), and McKenna and Glendon (1985) all found that a rapid drop-off in the memory rate takes place in the first years in these areas as well.

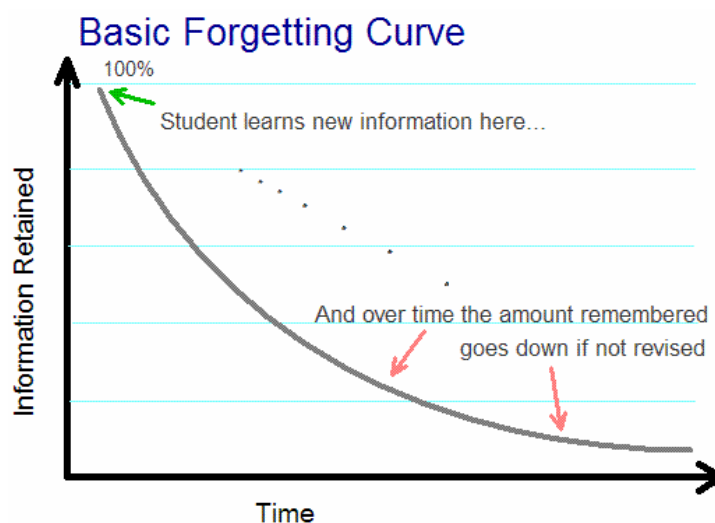


Figure 3  
Basic forgetting curve.

Ebbinghaus proposed a number of explanations for forgetting. The first states that old or earlier pieces of information are covered by newer ones and therefore become submerged and lost. This explanation, mentioned above, is called Interference Theory. Another explanation is that memory traces are worn away by time. This explanation is currently called the Trace Decay Theory. The difference between the two is that, in the interference theory, forgetting occurs as a result of disruption of the memory trace by other traces, while in the trace decay hypothesis it occurs because the memory trace naturally fades with time.

Both theories assume that a trace is left in the brain as a result of learning. This trace could take the form of neural activity, in which case the trace decay theory would claim that this activity dies eventually and would be no longer available if it is not activated by repeated presentation of the event or information held in memory. In other words, forgetting is due to a spontaneous fading of the neural memory trace over time. An advantage of this theory is that it describes how certain memories such as the feeling of an auditory sound may wear away. But it is difficult to use it to test verbal learning because learners typically try to strengthen the learning trace by rehearsing the learned material many times. If learners were prevented from rehearsing by interpolated activities, then their loss or drop-off could be due to the interpolated activity rather than the decay of the memory trace. Jenkins and Dallenbach (1924) had two subjects learn lists of nonsense syllables. They tested subjects after different delays during which they were awake or asleep. They found that more forgetting happened when subjects were awake. One possible interpretation for this result is that forgetting cannot be explained by the trace decay hypothesis but could be due to the interference theory, since the trace decay hypothesis would predict similar amounts of forgetting. Another factor affecting this result is that subjects

might have practiced rehearsing the material before sleeping. Hockey, Davies, and Gray (1972) found no evidence that sleeping reduced the forgetting rate.

Two of the most basic evaluations of the trace decay hypothesis are Hebb (1949) and Solso (1995). Hebb (1949) says that trace decay hypothesis can be applied to short-term memory. He believes that there is a correspondence between the long-term memory and the neural change caused by a repeated neural activity. Solso (1995) on the other hand, believes that forgetting from the long-term memory is not due to the trace decay hypothesis.

Whatever the explanation or theoretical model used, researchers have long agreed that human memory is frail. As one researcher put it, "The time at which we are aware of our memory is when it fails. The effect of this is often trivial but irritating; how often have you claimed to have a terrible memory?" (Baddeley, 1990, p.233). This basic pattern needs to be kept in mind in a study of this kind, whose main focus is on memory of newly learned words.

Before discussing which memories live and which vanish, it is necessary to know the stages of information processing. Memory consists of three stages of information processing: encoding, storage, and retrieval. In the first stage, encoding, information is rendered into a form that can be retained in memory. This stage involves the use of knowledge already existing in memory in the interpretation of new information. Human memory, as a result, is considered a scaffold, meaning that information organized in memory forms a framework that can be used for interpreting new information. Thus, "[t]he more scaffolding there is, the greater the capacity to attach (encode) new information" (Richardson-Klavehn & Bjork, 2002, p. 1102).

The second stage is storage, in which information is held in memory over a long period of time. There are various distinctions between forms of memory, as



mentioned above. The third stage is the retrieval stage, in which information is accessed from storage and used to perform a specific task. Information retrieval has two types; recognition and recall. In recognition retrieval, information is used as a cue for retrieving knowledge stored in memory. In recall retrieval, information is reproduced from memory with no cues provided.

When it comes to encoding and retrieval, pre-existing world knowledge plays a critical role. The prior knowledge one has about an event is represented by cognitive structures or Schemas which affect our recall of events. Freeman, Romney, and Freeman (1987) proposed that two factors affect our recall of an event; how the person's schema is organized for that event and how typical the event is. Memory will be better if the schema is better organized and the event is more typical. A wealth of research emphasizes the importance of schemas in remembering events from memory, specifically called "Schematic Memory" (Stilwell and Markman, 2003). Bransford and Johnson (1972) emphasize the role of schema in text comprehension. In a series of experiments, subjects had to read an ambiguous text alone or accompanied by two drawings, one of which included elements organized in a contextually sensible way. Results showed that comprehension ratings and recall scores increased when subjects were provided with relevant information, while lower comprehension ratings and recall scores decreased when the information was provided subsequent to the text.

Schustack and Anderson (1979) found out that schemas related to the theme facilitate recall. Showing that recall is a schema-based process, Brewer and Treyns (1981) had subjects wait in the experimenter's office for 35 seconds and then move to another room. They were then asked to recall the office. Everyone recalled the expected things such as a chair, a desk, and walls. Twenty-five percent, or one in four, recalled unexpected or salient items such as a skull. In contrast to Brewer and Treyns

(1981), there is evidence that memory is better for unexpected or inconsistent items (Lampinen, Copeland, and Neuschatz 2001). In a meta-analysis, Rojahan and Pettigrew (1992) found out that the difference between the number of studies concluding that consistent or expected items lead to better memory performance and studies supporting the conclusion that inconsistent or unexpected items show better memory performance is small. The variation, as the researchers proposed, is due to variation in the methodology, time of exposure to the environment, and the number of items in the environment.

One of the common aspects of the above mentioned studies is that schemas are already present in the background knowledge of participants. But in the study of Stilwell and Markman (2003) schemas were created and then participants were exposed to events that remind them of it. In a two-experiment study, they wanted to verify whether memory is better for schema-relevant or for schema-irrelevant information. Results supported findings of previous studies in that memory for schema-relevant information is better than that for schema-irrelevant information, and that information in a schema intrudes on recall for related information.

## CHAPTER THREE

### METHODOLOGY

#### Introduction

The current study used a quantitative method in addition to participants' reflection to examine the effects of using semantic and thematic clusterings on English vocabulary learning by Saudi students. In the first part, the quantitative stage, data was collected from 160 participants studying in the English Language Department at Umm Al-Qura University; the participants were 60 freshmen, 60 sophomores, and 40 juniors. Participants studied four lists of English words representing semantic clustering, unrelated grouping, thematic clustering, and contextualized presentation. In the first part of the study, they were tested twice; immediately after the study phase and then a week later, on recall for the words in each list. Data from the numbers of recalled words was used to make comparisons demonstrating the effect of each type of clustering on vocabulary learning.

In the second part of the study, participants' reflection, twelve participants representing the six participants who learned the greatest number of words and the six who learned the smallest number of words were questioned briefly and individually. In brief, semi-structured interviews, specific questions and their sequence were determined in advance. Subjects were allowed to speak in either English or Arabic, whichever allowed them to express their ideas more clearly.

Quantitative research is a research approach used to quantify the relationships between variables, which are the items the researcher measures with regard to his participants. There are many strategies associated with quantitative research. Creswell (2003) states

During the late 19<sup>th</sup> century and throughout the 20<sup>th</sup>, strategies associated with the quantitative research were those that invoked the postpositivist perspectives. These include the true experiments and the less rigorous experiments called quasi-experiments and correlational studies ... and specific single-subject experiments.... More recently, quantitative strategies involved complex experiments with many variables and treatments (e.g., factorial designs and repeated measures designs). They also included elaborate structural equation models that causal paths and the identification of the collective strength of multiple variables (p. 13-14).

In quantitative research, the researcher uses positivist claims, including cause and effect, specific variables and hypotheses, etc., to develop and collect information or data that leads to statistical results. In other words, quantitative research investigates how many people react to a limited set of questions, and measures that reaction. This facilitates data comparison and statistical aggregation, giving broad and generalizable findings (Patton, 1987). One of the advantages to quantitative research is that "Quantitative measures are succinct, parsimonious, and easily aggregated for analysis: they are systematic, standardized, and easily presented in a short space" (Patton, 1987, p. 11).

In the present study, quantitative methods were the ideal approach, as the questions involved the interaction of dependent and independent variables. Quantitative method was the right choice for my main test since I am seeking to know how participants are likely to respond or perform in the test. Moreover, the results of the quantitative research are an accurate representation of the population being studied and provide valuable insight into the nature of a population.

Interviewing is one of the basic ways for collecting data. It is used to elicit the participants' views, opinions, and evaluations of the topic being discussed. The interview is considered a fascinating source for a wealth of research data. Interviews are also useful for elaborating on the results of quantitative tests to gain insights into interesting or unexpected findings; this was the function of the brief interviews held with selective participants in the present study. The key element of the interview is "the verbal give-and-take between two people with the questions and answers providing its form" (Sommer & Sommer, 1997, p. 106). Merriam and Simpson (1995) state that

An advantage of the interview technique is its effectiveness in surveying special populations and gaining in-depth information. Interviewing is particularly useful in gathering data from "hard-to-reach" populations. Also, when it is not possible to anticipate all that the researcher may need to know in preparing the schedule of questions before meeting the research participants, an unstructured interview may yield more reliable data than a written questionnaire. A personal, face-to-face interview is recommended to develop rapport and gain the widest range of data (p. 71).

Emphasizing the importance of interviews and mentioning ways of conducting them, Sommer and Sommer believe that "their intrinsic interest stems from the personal interaction that is the core of the procedure. Modern technology has led to a broadening of the face-to-face concept to include interviews by telephone, video, and other extended means of communication" (p. 106).

There are many types of interviews. Guba and Lincoln (1981) list the following types: team and panel interviewing, covert or overt interviewing, oral history interviewing, and structured and unstructured interviewing. For the present

study, a simple, brief structured interview was used to gather participant responses about their experience, both with the experimental test and with vocabulary learning generally.

## Quantitative Stage

### *Subjects*

One hundred sixty subjects participated in the study. 60 were freshmen, 60 were sophomores, and 40 were juniors. All are male Saudi students enrolled at the Department of English, Umm Al-Qura University, Makkah, Saudi Arabia, and are native speakers of Arabic. They were randomly selected and randomly assigned to either a semantic or thematic group.

I talked to a number of professors at the department of English about the study and the number and levels of participants needed for the study. They showed willingness to help by giving me classes and through talking to their students about the nature and importance of my study. The whole study took about eight weeks. At the beginning of each class, the professor introduced me to the participants and explained to them the nature of the study. I told them from the beginning that their participation in the study is voluntary and their identity will be kept confidential. After they have finished the tests, they signed the consent forms (see Appendix M).

For Level 1 and Level 2 participants, 30 took tests on the semantic/unrelated sets while the other 30 took tests on the thematic/context sets. For Level 3 participants, 20 took tests on the semantic/unrelated sets while the other 20 took tests on the thematic/context sets. Participants had immediate and delayed recall tests on the words in the lists they studied, both from Arabic-to-English and from English-to-Arabic.

### *Materials*

The purpose of each study was to compare the learnability of semantic, unrelated, and thematic sets of associate pairs composed of English and Arabic words. Moreover, the effect of using context on learning words was examined by comparing the learnability of a relatively unrelated list of words learned in isolation and with the aid of a context. Four lists were used, each consisting of eight English words. The first list was related by semantic clustering and included eight English (L2) words for types of flowers. These words were accompanied by Arabic (L1) equivalents (see Appendix A). The second list contained unrelated words and included 8 English words accompanied by Arabic equivalents (see Appendix D). The third list had eight English words with Arabic equivalents, and they were categorized under the theme of fishing-related terms (see Appendix G). The last set was composed of unrelated words, but these were presented as underlined words in the context of a paragraph (see Appendix J). The text consisted of 260 words in nine sentences describing a family's trip to a forest. Unlike other studies conducted on this issue, Arabic (L1) words were paired with real, but infrequently used English words; though the meaning assigned to these English words differed in some cases from their original meaning. For the semantic category, the names of flowers were used and paired with well-known Arabic equivalents. For the other three lists, infrequent words were drawn from the first four volumes of the Dictionary of American Regional English (D.A.R.E., Cassidy, 1985). The goal was to present words whose orthographic and phonological features were those of plausible English words, and every attempt was made to ensure that the resulting lists were equivalent in length, structure and overall level of difficulty since they are all low frequent words.

Other studies have used artificial words (Tinkham, 1994, 1997; Waring, 1997); however, the choice of actual words for this study was motivated by the desire to design a study task that would imitate the natural language learning experience as closely as possible. The study also differed from previous studies in its inclusion of a contextualized list, which is expected to further minimize the effect of interference (Waring, 1997). The inclusion of the context condition presented certain problems, in particular, how to adjust the time allowed for initial exposure to the lists, to take into account the additional time needed to process the context and relate the words to it.

Condition 1: semantic set

Iris	سوسن
Lily	زنبق
Tulip	خزامى
Daffodil	النرجس البري
Pansy	زهرة الثالوث
Daisy	زهرة الربيع
Aster	زهرة النجمة
Crocus	زعفران



Condition 2: unrelated set

Batean	قارب ركاب
Capsheaf	متطرف
Bollix	يخلط
Doty	ناعم
Hamper	سبت
Pismire	نملة
Portico	رواق المنزل
Pothole	حفرة في الطريق

Condition 3: thematic set:

Leister	رمح
Reel	بكرة
Dory	زورق مسطح
Lure	طعم
Cast	يلقي الصنارة
Shoal	ضحل
Pompano	سمك البنبان
Angling	الصيد بالصنارة

Condition 4: words in a context

*One day Jack felt donsie because he has eaten so much at a party.*

متوئك

*His wife and his two kids wanted to spend that day out in the open*

*and to go for a walk. When they left home they realized that it was*

*airish outside since it is the beginning of the winter season so they*

بارد

*put jackets on. Jack drove to an area out of town where he saw a*

*running bayou. The water was moving slowly with a quiet sound.*

نهر صغير

*His kids loved the view of the running water and threw dornickets*

أحجار صغيرة

*in the water to make the frogs jump. The water looked gaumy so*

ملوث

*he did not drink from it fearing of becoming sick of Malaria.*

*Suddenly his wife saw a doty and bright-colored object that*

ناعم

*looked strange in the ground. Jack got a rastus, which*

مجرفة

*he always keeps in his car to use on his farm, and started digging.*

*What they found was a very pretty little metal box, with punky*

مهترىء

*wooden handle “It must be very old,,: said Jack.*

Words used for the lists were chosen based on the following conditions:

**Semantic set:** words which share the same semantic and syntactic characteristics, grouped under a common concept (here, flower names).

**Unrelated set:** words that do not share semantic and syntactic characteristics, and are not associated with a shared thematic concept.

**Thematic set:** words that are grouped together based on a shared thematic concept (here, fishing), and thus are cognitively associated.

**Words presented in context:** words that do not relate closely to a thematic concept, but are presented in a reading passage, and thus can be learned in relation to a context rather than in isolation.

### *English Words*

All of the English words in the sets were selected by the researcher to fit into sets of eight, as described above. All of the English words chosen were considered low frequency based on *The Dictionary of American Regional English* (Cassidy, 1985). Using this dictionary, the researcher originally identified 102 words which are considered low frequency, based on the maps provided. Some words were excluded because they were compounds, such as *Barn burner*, or slang such as *fizzle* and *ornery*. Other excluded words were of clearly borrowed origin, such as French *beaucoup* and *armoire*, and Spanish *arroyo*. The main criterion for choosing low-frequency words was that these had a high probability of being unknown to the participants, while being real English words.

Participants did not know the meaning of the English words in the Arabic language, although they were familiar with the Arabic words, nor had they established connections between the English words and their equivalents. For example, participants had not encountered the English word *iris*, and did not know that (سوسن) is the counterpart of the English word *iris* although they had acquired (سوسن) before as a part of their first language; in other words, they had not established the connection between the concept they had acquired and the English term. Given the possibility that the learnability of a word is affected by its form-class (Rodgers, 1969), every effort was made to make the sets equivalent in that respect. The words in the thematic clusters and in the unrelated sets were varied, and included nouns, verbs, and

adjectives. However, given the nature of semantic classes, word class was necessarily held constant in the semantic set, with all eight terms being nouns. For variety, words in each set included monosyllabic, bisyllabic, and trisyllabic forms.

### *Pilot Study*

A pilot study was conducted, in an attempt to determine the amount of time needed to learn lists of this type of list, and to identify any problems with the research design. Although the pilot participants, 30 English graduate students from diverse linguistic backgrounds, were very different from the main study participants, it was felt that they might yield insights into some aspect of the study's design. In fact, some changes were made as a result of the pilot study. First, the pilot study showed that participants need to be informed that there would be delayed tests a week after the study phase. Second, it showed that participants need more time than the pilot participants had to study lists of words and even extra more time to read the context.

### *Procedure*

#### *Immediate Recall Tests*

Participants within each level were divided into two groups corresponding to the two learning conditions. Thirty of the 60 Level 1 participants took the semantic list test in class. These 30 subjects were divided into two groups of 15. The first 15 participants were given a handout of two pages. The first page consisted of the list of eight semantically related English words accompanied by their Arabic equivalents. Subjects were required to study them for a total of four minutes, that is, 30 seconds per item. After four minutes, the participants were instructed to stop referring to the lists, and the immediate recall phase took place. Each subject was required to turn to

the following page, which contained the eight English words, and to provide their Arabic equivalents. In order to eliminate any chance of memorizing the list as a whole rather than learning words, the participants were informed that the words on the test were arranged in a different sequence.

Translation in the L2-L1 rather than L1-L2 direction is efficient (Nation, 2001). Also, Prince (1996) states that using translation in the L2-L1 direction improves recall performance. He found that weak learners even significantly outperformed advanced learners when translation was in L2-L1 direction. Studies of bilingual translation (Kroll & Curley, 1986; Kroll & Stewart, 1989) found that learners were faster when translating into their first language than into their second language. The reason behind this is provided by Kroll and Stewart (1994):

For most bilinguals, even those who are relatively fluent, more words are known in the native than in the second language. Lexical associations from L2 to L1 are assumed to be stronger than those from L1 to L2 because L2 to L1 is the direction in which second language learners first acquire the translations of new L2 words. (p. 158).

Therefore, it should be easier for an Arabic speaker to retrieve an Arabic word from memory when given the English word than to retrieve an English word when given the Arabic word (Ellis & Beaton, 1993).

Trials to criterion, in which participants say or write responses as they hear or read stimuli correctly on two consecutive trials, have not been used as in Tinkham's study (1997) because subjects might perform better on the second test due to the effects of task-learning (Waring, 1997). This technique is not realistic since learners do not use it in classroom or in real life communicating with others. Moreover, it is not clear that this type of measurement is straightforward, since subjects in Waring's

study (1997) had fewer learning trials in experiment 2 than in experiment 1. Also, the use of trials to criterion (producing correctly all the words in a set) could be employed by subjects as a guessing task rather than as a learning task. Waring (1997) found that

In the first experiment it was often found that one set would be successfully provided (say the unrelated words ) and as the learner was trying to learn the other set, some of the first set, which had already been checked by the researcher as learned, were forgotten temporarily. (p. 268).

There was no time limit for this recall test. Once the students had all finished, they were given another handout containing another list of eight English words, but these were unrelated words accompanied by their Arabic equivalents. They followed the same procedure they had followed with the semantic list.

The other 15 participants had the same tests but with a different order. That is, they viewed the unrelated word list first and then the semantic list.

The remaining 30 participants were divided into two groups of 15. The first 15 participants were given a handout containing a list of eight thematically grouped English words accompanied by their Arabic equivalents. They were given four minutes to study them. Next, they moved to the recall test page which had the English words, and they were asked to provide the Arabic words. Words were arranged in a different sequence to eliminate any chance of memorizing the list rather than studying the individual words. As with the semantic and unrelated lists, this test had no time limit. Next, participants were given a text which contained a set of underlined words. Participants were instructed to read the text and to learn the meanings of the underlined words, and they were given 5 minutes to learn the words, the time being adjusted to allow for the extra cognitive effort involved in consulting and integrating the brief story context. In the context recall test, participants were required to provide

the Arabic equivalents to the English words. With the context condition, again, there was no time limit for completion of the test. The second 15 participants had the same two tests but in the reverse order, that is, they took the context section first and then moved to the thematic word list. This same procedure was followed with sophomores and junior subjects.

### *Delayed Recall Tests*

One week later, participants were given the tests without word lists to study. The order of the words in the test was changed. There was no time limit for completion of the tests.

### *Data Analysis*

An Analysis of Variance (ANOVA) was used to analyze the data collected. A 4 (Type of clustering: semantic / unrelated / thematic / contextual) by 3 (Level: Level 1 / Level 2 / Level 3) by 2 (Direction of translation: English to Arabic / Arabic to English) repeated measure factorial design was used. Analysis of Variance is a statistical technique that makes full use of the interval/ratio level of dependant variables like vocabulary items recalled measures in numbers. It is used to compare two or more than two means. The goal of the ANOVA was to explain the variance in the dependent variable in terms of the variance in the independent variables. In this study the dependent variable was the vocabulary test scores, while the independent variables were types of clustering, levels of students, and direction of translation. Measurement of the dependent variable was the number of recalled words. ANOVA examines differences among means by decomposing the total variance in the dependant variable into variance that occurs within independent variable groups and

variance that occurs between independent variable groups. It is called repeated measure when the production of the participants is measures before and after the treatment or when there are immediate and delayed measurements. ANOVA is called factorial design when there are two or more than two independent variables.

### Informal Interviews

Interviewing is one of the basic ways used by researchers to collect data. It is used to elicit the participants' views, opinions, and evaluations of the topic being discussed. Although the nature of this study was not primarily qualitative, I found it useful as a follow-up to the quantitative tests, to speak briefly with participants in order to gain insights into any interesting or unexpected findings about the participants' experiences.

The twelve subjects who scored the highest six scores and the lowest six scores on each test were questioned after completing each test session. These were not formal interviews. Rather, they were like discussions or informal ways of collecting data about vocabulary learning and strategies used. Participants were asked individually to give them more freedom to talk. Moreover, they were allowed to talk either in English or Arabic, whichever allowed them to formulate and express their ideas better. A set of questions were asked to elicit data about how subjects learned the lists and which they preferred. The questions were:

- 1- Which sets were difficult? Why was X more difficult than Y?
- 2- Which sets were easier? Why was X easier than Y?
- 3- How did you learn English words in previous levels?
- 4- Why do you think this task was easy?



- 5- Do you think that providing the context makes learning easier? (For thematic / context subjects)?
- 6- How were words presented to you?
- 7- How were lessons organized?
- 8- How did you try to memorize words?
- 9- What specific words do you recall confusing or forgetting often, and why?

## CHAPTER FOUR FINDINGS AND DISCUSSION

### Quantitative Findings

An Analysis of Variance (ANOVA) was used to analyze the data collected. A 4 (Type of clustering: semantic / unrelated / thematic / contextual) by 3 (Level: Level 1 / Level 2 / Level 3) by 2 (Direction of translation: English to Arabic / Arabic to English) repeated measure factorial design was used. The goal of the ANOVA is to explain the variance in the dependent variable in terms of the variance in the independent variables. In this study the dependent variable is the vocabulary test scores while the independent variables are types of clustering, levels of students, and direction of translation.

A note of caution should be inserted here as to the types of clustering. While the semantic, unrelated and thematic conditions were treated as expected, the ‘contextual’ list did not function quite as I had foreseen in setting up the study’s design. The intention was that the learners would attend to the story, and would actively integrate these words’ meanings into the passage as they learned them. In fact, as I observed the participants, I noted that many seemed to ignore the story context entirely, and simply turn the unfamiliar words and their meanings into another, virtually ‘unrelated,’ set. I will return to this point later; but it is worth mentioning here as a caution to the reader in interpreting the figures given in this and the coming sections.

In analyzing the data, measurement of the dependent variable has been achieved by simply counting the number of recalled words from each eight-word set.

*Arabic to English Translation Direction*

*Level 1(Immediate Tests)*

In this section, the results for the Arabic-to-English task are reported. That is, these results relate to the condition where participants were given the Arabic meaning as prompt, and asked to recall the English word. The Arabic-to-English results will be presented in the next section

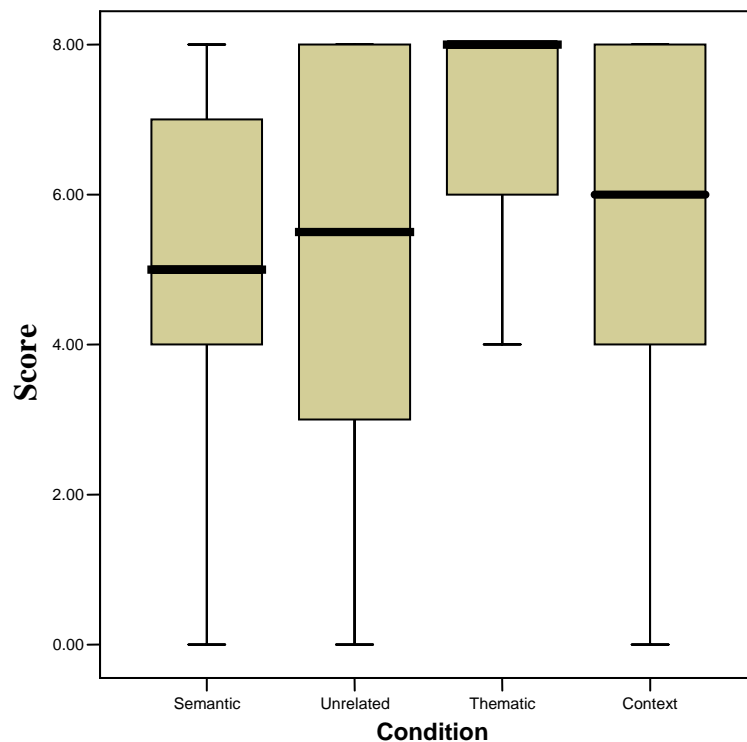
Table 3 reports the means and standard deviations within level 1 only as associated with types of clustering and for the test administered immediately after the students studied the list.

Table 1  
*Means and Standard Deviation of Recalled Words (Level 1–Immediate test– A-to-E)*

Condition	N	Mean	Std. Deviation
Semantic	30	4.93	2.21
Unrelated	30	5.00	2.59
Thematic	30	7.03	1.33
Context	30	5.63	2.36

As Table 1 clearly shows, Level 1 participants recalled more words from the thematic list on average than words from the semantic, unrelated, or the context lists. In turn, more words from the context list were recalled on average than words from the semantic and unrelated lists. Also, the means for semantically grouped and unrelated word lists are the lowest. The low standard deviation for the thematic condition is caused in part by ceiling of 8 possible correct answers.

These results are in the expected direction, particularly as the guiding hypothesis of this study is that thematically clustered words will evoke a cognitive scaffolding and will thus be easier to learn and retain.



*Figure 4*

Box-plot diagram of score in relation to types of clustering (level 1 – immediate test – A-to-E).

The box-plot diagram in Figure 4 provides a visual display for the distribution of scores for this same test and group in terms of percentiles. A brief explanation is given here for the reader unfamiliar with this type of display. The median of a group is considered the midpoint. Of each box, the lower edge indicates the 25<sup>th</sup> percentile of the data set while the upper edge indicates the 75<sup>th</sup> percentile. The term "Inter-quartile range" refers to the range of the middle two quartiles. The thick line in the box indicates the median value. That is, half the values are equal to or greater than the median and the other half the values are equal to or less than the median. The box-plot is considered a five number summary, as it displays five figures clearly: the low value, the 25<sup>th</sup> percentile, the median (50<sup>th</sup> percentile), the 75<sup>th</sup> percentile, and the high value.

In figure 4, the median score for recalled words score for each type of vocabulary clustering is represented by the horizontal line in the middle of each box. The inter-quartile range for unrelated clustering is larger than the others because the variation is larger. The inter-quartile range for context clustering is larger than that of the semantic and thematic word groups, with the thematic clustering the smallest of all. It appears that 50% of Level 1 participants recalled all eight words correctly in the thematic list, while only 25% of the participants recalled all eight words in the unrelated and context lists. In other words, half of the beginning students actually remembered the full list of thematically related words, while only one in four remembered the whole list for unrelated and context lists. The lower score is four in the thematic list, while it is zero in the other lists. One last thing to mention is that the whiskers of all types of clustering are left skewed toward lower values. This means that most values are 'large', but there are a few exceptionally small ones.

Table 2 displays the results of applying the ANOVA to this same set of data, from the Level 1 participants.

Table 2  
*Analysis of Variance of Number of Recalled Words (Level 1–Immediate test–A-to-E)*

<b>ANOVA</b>					
Immediate					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	85.500	3	28.500	6.035	.001
Within Groups	547.800	116	4.722		
Total	633.300	119			

Table 2 reports an F of 6.035, which with 3 and 116 degrees of freedom is statistically significant at the .001 level ( $P < 0.05$ ). These figures indicate that the differences across independent variable categories are significant at a very low probability. Therefore, we can feel comfortable in rejecting the "Null Hypothesis" of

no difference between the four types of vocabulary grouping, at least here in the immediate test for the population of Level 1 participants. Eta Square ( $E^2$ ), the correlation ratio, is a proportion that is used to express how much we reduce errors in guessing dependent variable scores. It is also a measure of association; it indicates how strongly the dependent variable is related to the independent variable. Its value is obtained by dividing the Between Groups Sum of Square by the Total Sum of Squares. In Table 2 above,  $E^2$  is 0.14 for this set of results. This means that the four types of vocabulary grouping explain about 14 percent of the variation in number of recalled words. We could say that this relationship is weak but statistically significant.

Table 3 shows the Post Hoc Test results for the same data. Post Hoc Tests are used to show the relationships between the clustering types one at a time. In this case, LSD refers to Fisher's Least Significant Difference Multiple Comparison Procedure, which was chosen because it clearly displays the importance of the results for each pair of clustering conditions.

Table 3  
*Post Hoc Tests (Level 1 – Immediate test – A-to-E)*  
**Multiple Comparisons**

Dependent Variable: Immediate  
 LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-.06667	.56109	.906	-1.1780	1.0447
	Thematic	-2.10000(*)	.56109	.000	-3.2113	-.9887
	Context	-.70000	.56109	.215	-1.8113	.4113
Unrelated	Semantic	.06667	.56109	.906	-1.0447	1.1780
	Thematic	-2.03333(*)	.56109	.000	-3.1447	-.9220
	Context	-.63333	.56109	.261	-1.7447	.4780
Thematic	Semantic	2.10000(*)	.56109	.000	.9887	3.2113
	Unrelated	2.03333(*)	.56109	.000	.9220	3.1447
	Context	1.40000(*)	.56109	.014	.2887	2.5113
Context	Semantic	.70000	.56109	.215	-.4113	1.8113
	Unrelated	.63333	.56109	.261	-.4780	1.7447
	Thematic	-1.40000(*)	.56109	.014	-2.5113	-.2887

\* The mean difference is significant at the .05 level.

Table 3 shows that there is a statistically significant difference between the number of words recalled from the semantic and thematic lists for Level 1 participants taking the immediate test, since the mean of the second is higher. A particular comparison is statistically significant if the Sig. value is 0.05 or below.

The most important section of this chart is the third, which displays the comparisons with the thematic set. The differences between the numbers of words recalled from the thematic list as compared with all three other lists are statistically significant in favor of the thematic list (the thematic mean of 7.03 compares with 4.93, 5.00, and 5.63 respectively for the others, as given earlier in Table 1). Again, the thematic condition stands out in being significantly different from all others; in contrast, the semantic, unrelated, and context conditions do not appear to be significantly different from each other when compared in pairs, as the difference between these is not statistically significant.

There is no significant difference when we compare the number of words recalled from the semantic list to the number of words recalled from the unrelated or context lists. When it comes to the unrelated list, again the only statistically significant difference appears with the thematically clustered words. Table 3 reveals that the difference between the number of words recalled from the semantic, context, and unrelated lists is statistically insignificant.

*Level 1(Delayed Tests)*

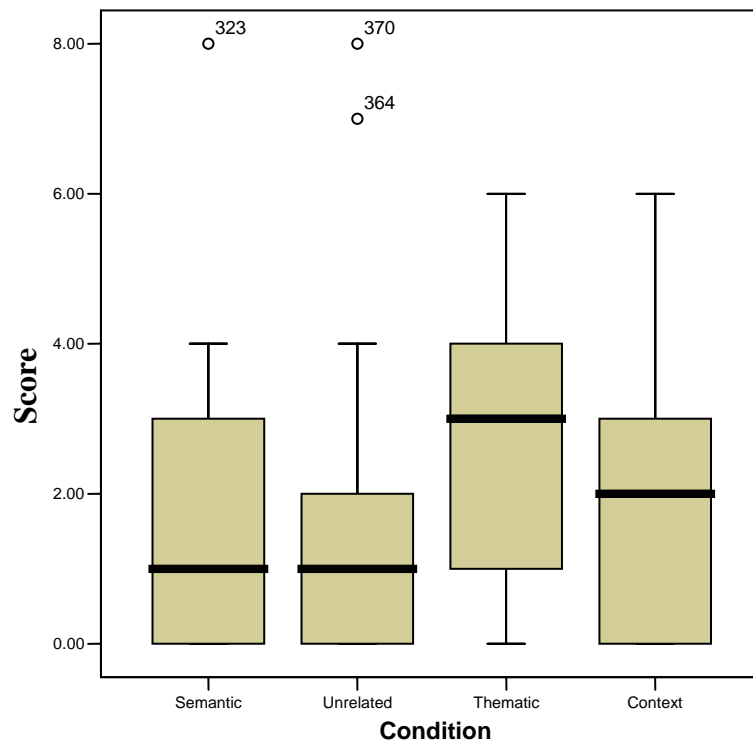
Table 4 lists the means for number of words recalled by Level 1 participants in the delayed test, by cluster, with the standard deviation for each group.

Table 4  
*Means and Standard Deviation of Recalled Words (Level 1 – Delayed test – A-to-E)*

Condition	N	Mean	SD
Semantic	30	1.53	1.83
Unrelated	30	1.77	1.92
Thematic	30	2.67	1.54
Context	30	1.97	1.75

These results are similar to those in Table 1. Level 1 participants recalled more thematically clustered words than semantically, unrelated, or contextually clustered words. Although the means are close to each other, the thematic list comes first with a mean of (2.66) followed by context, unrelated, and semantic lists respectively. The difference between the means of semantic, unrelated and context conditions is about 0.2, a very small difference.





*Figure 5*

Box-plot diagram of score in relation to types of clustering (level 1 – delayed test– A-to-E).

The figure above shows that the semantic and the unrelated lists have the lowest medians (and lowest means) for delayed recall for the Level 1 participants. On the other hand, the median of the thematic list is the highest for this group of participants. While 25% of the recalled semantic, unrelated and context clustered words appear to be 0, the 25% of the scores is one in the thematic list. The boxes, high values, and low values indicate that the boxes are right skewed toward higher values for all the four types of clustering. The boxes for semantic, thematic, and context clustering have the same height and they are larger than the box for the unrelated clustering. This means that the variation is smaller in the unrelated clustering while the other types appear to have about the same amount of variation. The numbers [44, 50, and 3] on the diagram indicate "outliers" or numbers that lay far beyond the maximum.

Table 5 displays the results of the ANOVA performed on the means of Level 1 student delayed test results for recalled words by cluster.

Table 5  
*Analysis of Variance of Number of Recalled Words (Level 1 – Delayed test– A-to-E)*  
**ANOVA**

Delayed					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21.500	3	7.167	2.294	.082
Within Groups	362.467	116	3.125		
Total	383.967	119			

With  $F(3, 116) = 2.294$ ,  $P = 0.082$ . This result indicates that the difference is on the borderline of significance.

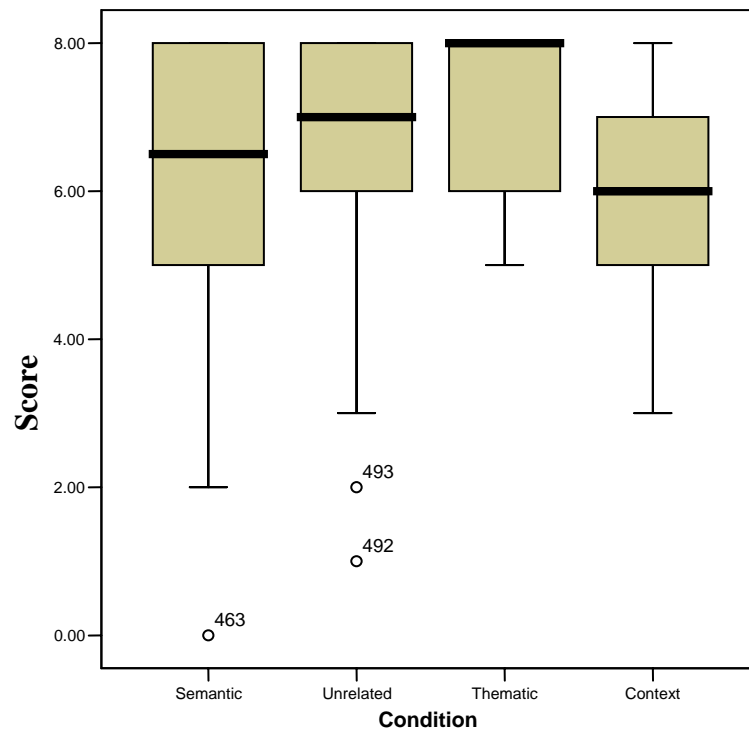
#### *Level 2 (Immediate Tests)*

Table 6 lists the means for number of words recalled by Level 2 participants on the delayed test, by cluster, with the standard deviation for each cluster group.

Table 6  
*Means and Standard Deviation of Recalled Words (Level 2– Immediate test – A-to-E)*

Condition	N	Mean	SD
Semantic	30	6.00	2.07
Unrelated	30	6.53	1.83
Thematic	30	7.10	1.09
Context	30	5.73	1.68

It is clear from Table 6 that Level 2 participants recalled more words from the thematic list than from any of the other lists. Following the thematic, unrelated clustered words were recalled second, followed by the semantically clustered words. Words that appeared in context were the least well recalled.



*Figure 6*

Box-plot diagram of score in relation to types of clustering (level 2 – immediate test – A-to-E).

It appears from Figure 6 that the median of the recalled thematically clustered words is the highest, while the median of the recalled contextual clustered words is the lowest. The boxes are left skewed toward lower values for the semantic, unrelated, and thematic clustering while it looks symmetrical (bell shaped) for the context clustering. Twenty-five percent of Level 2 participants recalled all the eight words of the semantic and unrelated lists correctly, while 50% of them were able to recall all the words in the thematic list correctly. This means that half of the second-year participants remembered the full list of thematically related words, while only one in four remembered the whole list for the semantic and unrelated lists.

Table 7 shows the results of the ANOVA performed to compare means for number of words recalled by Level 2 participants in the immediate test, grouped by cluster, in an effort to eliminate the null hypothesis for this tested group.

Table 7  
*Analysis of Variance of Number of Recalled Words (Level 2–Immediate test–A-to-E)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.958	3	10.986	3.770	.013
Within Groups	338.033	116	2.914		
Total	370.992	119			

With  $F(3, 116) = 3.770$ ,  $P = 0.013$  ( $P < 0.05$ ). This result indicates that the difference is statistically significant. Thus, we can once again feel comfortable in rejecting the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the immediate tests in the population of Level 2 participants.

The  $\eta^2$  value of 0.09 tells that despite the statistical significance, the type of vocabulary clustering explains only about 9 percent of the variation in number of words. This is a weak relationship, but is still significant.

Table 8 compares means of the cluster groups, by pairs, for the immediate test administered to Level 2 students.

Table 8  
*Post Hoc Tests (Level 2 – Immediate test – A-to-E)*

**Multiple Comparisons**

Dependent Variable: Immediate  
 LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-.53333	.44076	.229	-1.4063	.3397
	Thematic	-1.10000(*)	.44076	.014	-1.9730	-.2270
	Context	.26667	.44076	.546	-.6063	1.1397
Unrelated	Semantic	.53333	.44076	.229	-.3397	1.4063
	Thematic	-.56667	.44076	.201	-1.4397	.3063
	Context	.80000	.44076	.072	-.0730	1.6730
Thematic	Semantic	1.10000(*)	.44076	.014	.2270	1.9730
	Unrelated	.56667	.44076	.201	-.3063	1.4397
	Context	1.36667(*)	.44076	.002	.4937	2.2397
Context	Semantic	-.26667	.44076	.546	-1.1397	.6063
	Unrelated	-.80000	.44076	.072	-1.6730	.0730
	Thematic	-1.36667(*)	.44076	.002	-2.2397	-.4937

\* The mean difference is significant at the .05 level.

The difference between the mean of recalled thematic clustered words (7.10) and the mean of recalled semantic clustered words (6.00) is statistically significant, as Table 8 clearly shows. Also the difference between the mean of recalled thematically clustered words (7.10) and the mean of context-embedded words (5.73) is statistically significant. On the other hand, the difference between the means for recalled thematically and unrelated clustered words is not statistically significant.

*Level 2 (Delayed Tests)*

In Table 9, means and standard deviations for numbers of recalled words are displayed by cluster for the delayed test administered to Level 2 participants.

Table 9  
*Means and Standard Deviation of Recalled Words (Level 2 – Delayed test – A-to-E)*

Condition	N	Mean	SD
Semantic	30	1.83	1.72
Unrelated	30	2.40	2.22
Thematic	30	2.53	1.93
Context	30	1.60	1.61

The results in Table 9 show that Level 2 participants recalled more words from the thematic list than from other lists. The mean number of recalled words from the unrelated list ranks second, followed by semantically and contextually clustered words.

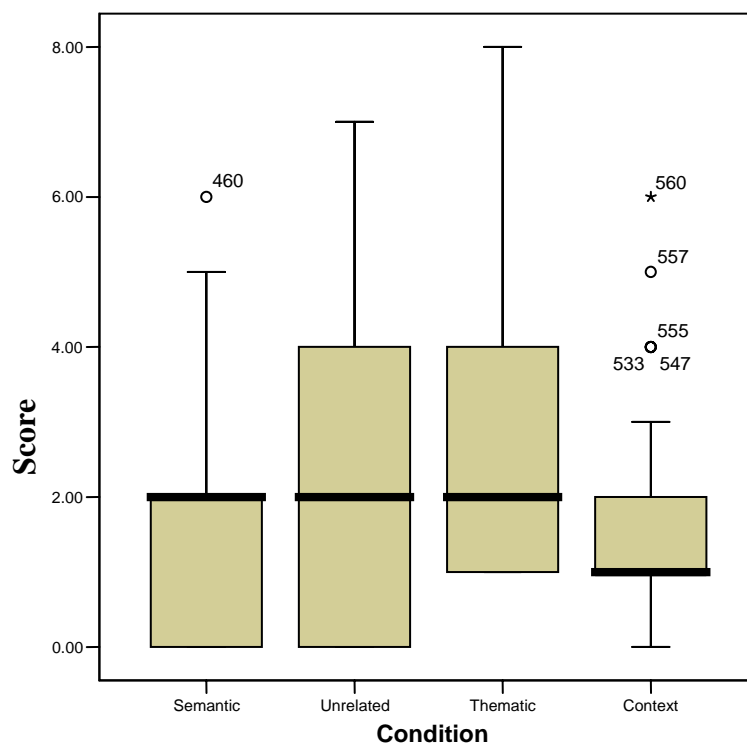


Figure 7  
 Box-plot diagram of score in relation to types of clustering (level 2 – delayed test – A-to-E).

Figure 7 above shows that the medians of semantic, unrelated and thematic lists are the same. Several outliers represent numbers that lay far above the maximum.

All boxes are left skewed toward lower values except for the box of the context list, which is symmetrical. Twenty-five percent of Level 2 participants could not recall any word from the semantic and unrelated lists, and 25% of Level 2 participants recalled only one word from the thematic list. On the other hand, 50% of Level 2 participants recalled only one word from the context list.

Table 10 compares means for recalled words by cluster group, for the delayed test given to Level 2 participants.

Table 10  
*Analysis of Variance of Number of Recalled Words (Level 2 – Delayed test – A-to-E)*  
**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	17.958	3	5.986	1.685	.174
Within Groups	412.033	116	3.552		
Total	429.992	119			

The ANOVA results, reported in Table 10, indicate that the differences among means yield an  $F(3, 116) = 1.685$ ,  $P = 0.174$  ( $P > 0.05$ ). This indicates that no particular type of vocabulary clustering was superior to the others in recalling words in the delayed tests. Thus, we have to accept the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the population of Level 2 participants in the delayed tests. Therefore, there is no need to present the table of Post Hoc Tests.

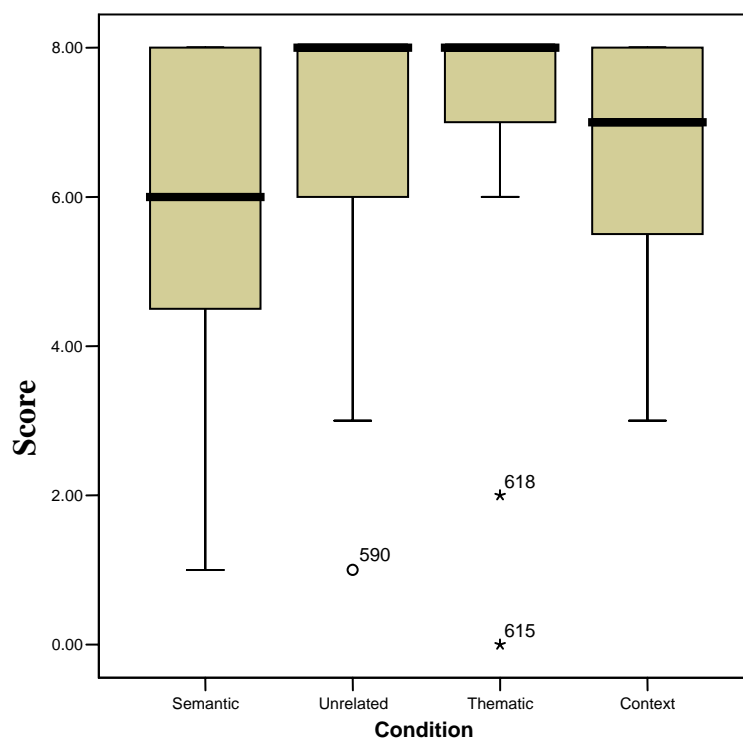
### *Level 3 (Immediate Tests)*

Table 11 presents means and standard deviations for words recalled by Level 3 participants on the immediate test, arranged by cluster.

Table 11  
*Means and Standard Deviation of Recalled Words (Level 3–Immediate test – A-to-E)*

Condition	N	Mean	SD
Semantic	20	5.90	2.05
Unrelated	20	6.65	1.98
Thematic	20	7.05	2.16
Context	20	6.50	1.61

As Table 11 clearly shows, Level 3 participants recalled more thematically clustered words than words grouped semantically, at random, or contextually. Words grouped semantically were the fewest recalled. It appears from the table that the means for recalled unrelated and contextual words are almost the same (6.65 – 6.50).



*Figure 8*  
 Box-plot diagram of score in relation to types of clustering (level 3 – immediate test – A-to-E).

The figure above shows that 25% of Level 3 participants recalled eight words correctly from the semantic and context lists in the immediate test while 50% of the



participants recalled all eight words from the other two lists correctly. This means that half of Level 3 participants remembered the full lists of thematically and unrelated clustered words, while only one in four remembered the whole list for the semantic and context lists. The inter-quartile range for the thematic list is the smallest while the inter-quartile range for the semantic list is the largest, illustrating the greater range of scores for recall of semantically clustered words. The median of the semantic list is the lowest, with 6 words recalled. All box plots are left skewed toward lower values.

In Table 12, the ANOVA compares mean numbers of recalled words by cluster on Level 3 participants' immediate test, to explore the possibility of rejection of the null hypothesis.

Table 12  
*Analysis of Variance of Number of Recalled Words (Level 3–Immediate test–A-to-E)*  
**ANOVA**

Immediate					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.650	3	4.550	1.183	.322
Within Groups	292.300	76	3.846		
Total	305.950	79			

The ANOVA reports an  $F$  of 1.183, which with 3 and 76 degrees of freedom is not statistically significant ( $P > 0.05$ ). The "Null Hypothesis" is accepted here, of no difference between the four types of vocabulary clustering in the immediate tests in the population of Level 3 participants.

#### *Level 3 (Delayed Tests)*

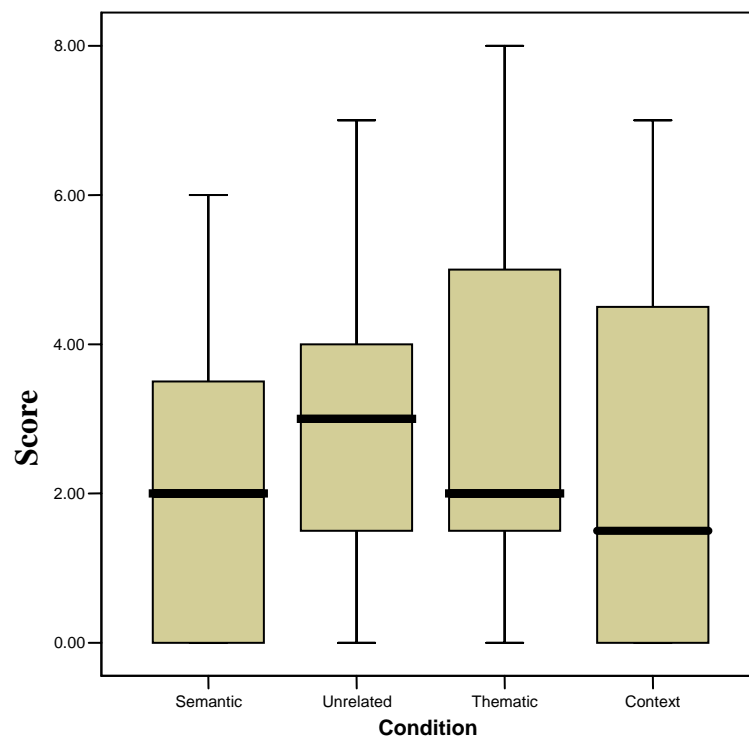
Tables 13 and 14 report data for numbers of words recalled by Level 3 participants on the delayed test. Table 13 displays mean number of recalled words and standard deviation for each of the four cluster groups.

Table 13

*Means and Standard Deviation of Recalled Words (Level 3 – Delayed test – A-to-E)*

Condition	N	Mean	SD
Semantic	20	2.10	2.00
Unrelated	20	3.00	1.92
Thematic	20	3.35	2.48
Context	20	2.30	2.49

The results of Table 13 are similar to the results of the immediate tests for this group. As Table 13 clearly shows, Level 3 participants recalled more thematically grouped words than words from any of the other clusters. Fewer unrelated words were recalled than thematically clustered words. The mean for the recalled semantically clustered words is the smallest. The order of the means for the delayed tests is identical to the order of means for the immediate tests.



*Figure 9*

Box-plot diagram of score in relation to types of clustering (level 3 – delayed test – A-to-E).

The inter-quartile range for context clustering shown in Figure 9 is larger than the range for the other clusters because the variation within the middle 50% of scores is larger. A smaller variation in numbers of recalled words makes the inter-quartile range for the unrelated clustering the smallest. The box plots of the unrelated and thematic lists are more symmetric. The median of the unrelated list is the highest while that of the context list is the lowest. Moreover, 25% of Level 3 participants could not recall any word from the semantic and context lists in the delayed test. This means that one in four could not remember any word from the semantic and context lists. One last thing to mention is that the boxes for all cluster groups are right skewed toward higher values.

Table 14 displays results of the ANOVA used to evaluate the accuracy of the null hypothesis for the Level 3 participants' delayed test.

Table 14  
*Analysis of Variance of Number of Recalled Words (Level 3 – Delayed test – A-to-E)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.638	3	6.879	1.374	.257
Within Groups	380.550	76	5.007		
Total	401.188	79			

The ANOVA reports an  $F$  of 1.374, which with 3 and 76 degrees of freedom is not statistically significant. The difference is much higher than the significance level ( $P > 0.05$ ). Thus, the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the delayed test in the population of Level 3 participants is accepted.

*Within Conditions*

*Semantic Clustering (Immediate Tests)*

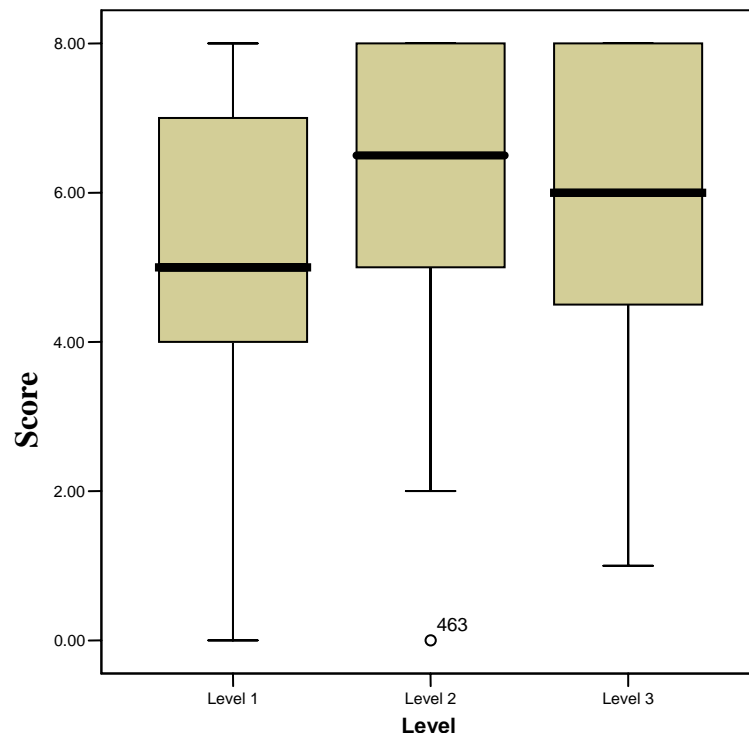
Table 15 lists the standard deviation and the mean number of semantically clustered words recalled accurately on the immediate test by each level of participants.

Table 15

*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Semantic List – A-to-E)*

Level	N	Mean	SD
Level 1	30	4.93	2.21
Level 2	30	6.00	2.07
Level 3	20	5.90	2.05

As Table 15 clearly shows, Level 2 participants and Level 3 participants recalled more semantically clustered words than Level 1 participants did, with very little difference between Levels 2 and 3. Whether the difference is significant is shown in Table 16 below.



*Figure 10*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – semantic list – A-to-E).

Figure 10 illustrates the inter-quartile ranges for the immediate test on semantically clustered words, by level. The median number of recalled words by Level 2 participants is the highest, while the median number for Level 1 participants is the lowest. The inter-quartile range for level 3 is larger than the others because the variation in number of words recalled is larger. The inter-quartile range for level 1 and 2 are of the same length. Twenty-five percent of Level 2 participants and the same percentage of Level 3 participants recalled all eight words in the semantic list on the immediate test.

Table 16 displays ANOVA results for the data displayed in Table 15, comparing means for number of words recalled from the semantic cluster list for the three levels who took the immediate test.

Table 16  
*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Semantic List – A-to-E)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.883	2	9.942	2.215	.116
Within Groups	345.667	77	4.489		
Total	365.550	79			

With  $F(2, 77) = 2.215$ ,  $P = 0.116$  ( $P > 0.05$ ). This result indicates that the difference is not statistically significant. The difference is clear between the means for Level 1 participants and Level 2 participants regarding the recalled semantically clustered words, but the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding semantically clustered words is accepted.

*Semantic Clustering (Delayed Tests)*

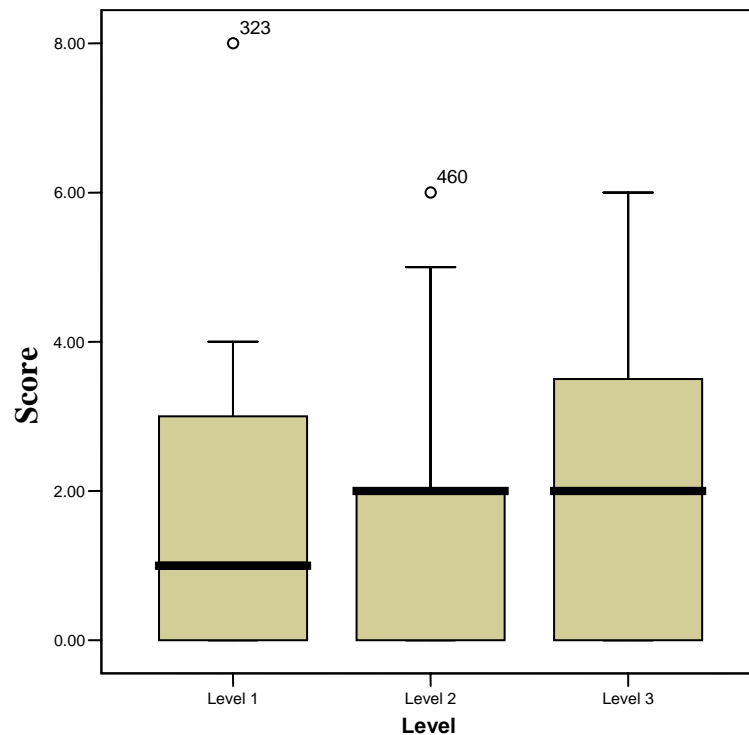
Like Table 15, Table 17 displays test results for semantically clustered words, including mean number of recalled words for each of the three levels and standard deviations. Table 15 reports results for the delayed test.

Table 17  
*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Semantic List – A-to-E)*

Level	N	Mean	SD
Level 1	30	1.53	1.83
Level 2	30	1.83	1.72
Level 3	20	2.10	2.00

Results of Table 17 are similar to results of Table 15 in that the mean for the recalled semantically clustered words by Level 1 participants is the lowest. The two tables differ in that in Table 15 Level 2 participants remembered more semantically

clustered words than Level 3 participants did, while in Table 17, the opposite holds true. The table below shows whether this difference is significant.



*Figure 11*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – semantic list – A-to-E).

Figure 11 shows that the medians for Levels 2 and 3 are the same. The interquartile range for level 3 is larger than the range of the other level because the variation is larger. Twenty-five percent of all participants could not recall any words from the semantic list in the delayed test. The highest scores are four for Level 1, five for Level 2, and six for Level 3, with one outlier from Level 1 and one from Level 2.

Table 18 shows ANOVA results for the delayed test on the semantic cluster, comparing mean number of recalled words for the three levels of participants.

Table 18

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Semantic List – A-to-E)*

## ANOVA

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.954	2	1.977	.587	.559
Within Groups	259.433	77	3.369		
Total	263.388	79			

Table 18 reports an F of 0.587, which with 2 and 77 degrees of freedom is not statistically significant. The significance value is 0.559 ( $P > 0.05$ ). This is expected since the differences between the means in Table 17 are small. The "Null Hypothesis" of no difference between the three levels in the delayed tests regarding semantically clustered words is accepted.

*Unrelated Clustering (Immediate Tests)*

Table 19 displays means and standard deviations for the number of words from the unrelated list recalled accurately by each level of participants on the immediate test.

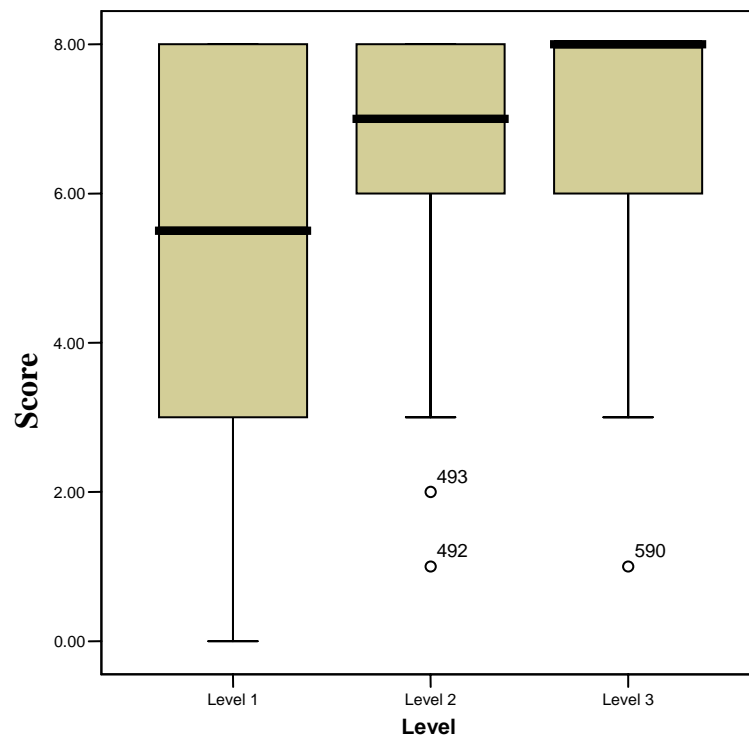
Table 19

*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Unrelated list – A-to-E)*

Level	N	Mean	SD
Level 1	30	5.00	2.59
Level 2	30	6.53	1.83
Level 3	20	6.65	1.98

It is clear from Table 19 that as we move to a higher level, the mean number of accurately recalled words goes up. The difference between Levels 2 and 3 is tiny, but a greater difference exists between first-year students and the other groups.





*Figure 12*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – unrelated list – A-to-E).

A large variation in numbers of words recalled led the inter-quartile range for Level 1 to be the largest. The inter-quartile ranges for Level 2 and Level 3 have the same length. The median goes higher as the participants' level increases. In Level 1 and Level 2, 25% of participants correctly recalled all eight words from the unrelated list in the immediate test, as did 50 % of Level 3 participants. Boxes for all levels of participants are left skewed toward lower values.

Table 20 displays ANOVA results for recalled words from the unrelated list, based on the immediate test scores for the three levels.

Table 20

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Unrelated list – A-to-E)*

## ANOVA

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	46.971	2	23.485	4.941	.010
Within Groups	366.017	77	4.753		
Total	412.988	79			

Table 20 reports an F of 4.941, which with 2 and 77 degrees of freedom is statistically significant at the .010 level ( $P < 0.05$ ). By this point, it is not surprising that differences across independent variable categories are significant at a very low probability. Therefore, we can feel comfortable in rejecting the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding unrelated clustered words. Based on Table 20 above,  $E^2$  is 0.11. This means that the differences between the three levels of participants explain about 11 percent of the variation in number of accurately recalled words from the unrelated list. We could say that this relationship is weak but statistically significant.

Table 21 displays results of the Post Hoc test performed on data from Table 20 namely results by level for number of accurately recalled unrelated words, as measured on the immediate test.

Table 21  
*Post Hoc Tests (All Levels – Immediate test – Unrelated list – A-to-E)*

**Multiple Comparisons**

Dependent Variable: Immediate  
 LSD

(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Level 1	Level 2	-1.53333(*)	.56294	.008	-2.6543	-.4124
	Level 3	-1.65000(*)	.62938	.011	-2.9033	-.3967
Level 2	Level 1	1.53333(*)	.56294	.008	.4124	2.6543
	Level 3	-.11667	.62938	.853	-1.3699	1.1366
Level 3	Level 1	1.65000(*)	.62938	.011	.3967	2.9033
	Level 2	.11667	.62938	.853	-1.1366	1.3699

\* The mean difference is significant at the .05 level.

The difference between the mean of accurately recalled unrelated clustered words by Level 2 participants (6.53) and the mean of recalled unrelated clustered words by Level 1 participants (5.00) is statistically significant ( $P = 0.008$ ) as Table 21 clearly shows. Also the difference between the mean of recalled unrelated clustered words by Level 3 participants (6.65) and the mean of recalled unrelated clustered words by Level 1 participants (5.00) is statistically significant since  $P = 0.011$ . On the other hand, the difference between the means for correctly recalled unrelated clustered words by Level 2 and Level 3 participants is not statistically significant since the means are close.

*Unrelated Clustering (Delayed Tests)*

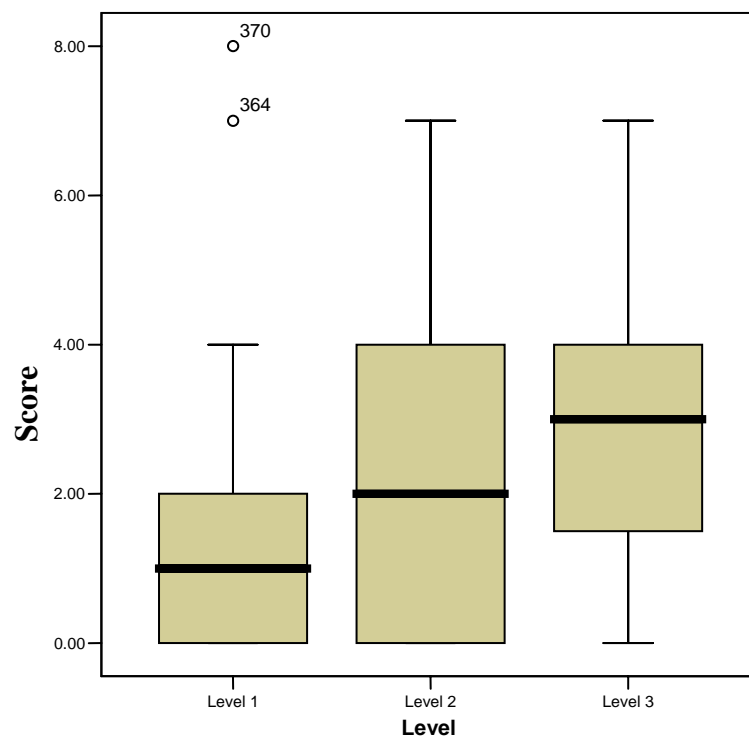
Table 22 consists of mean number of accurately recalled words for the delayed test on unrelated words, comparing the three levels of participants.

Table 22

*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Unrelated list – A-to-E)*

Level	N	Mean	SD
Level 1	30	1.77	1.92
Level 2	30	2.40	2.22
Level 3	20	3.00	1.92

With results similar to those of Table 20, Table 22 shows that the means for the number of correctly recalled unrelated clustered words by Level 3 participants in the delayed tests is the highest while the mean accurately recalled words by Level 1 participants in the same tests is the lowest. The table below shows whether the differences between the means are statistically significant.



*Figure 13*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – unrelated list – A-to-E).

Figure 13 shows that the inter-quartile range for unrelated words recalled by Level 2 students is the largest, showing large variation, while low variation the smallest range for Level 1 participants. The medians go higher as the participants' level increases. Twenty-five percent of Level 1 participants and the same percentage of Level 2 participants could not remember any word from the unrelated list on the delayed test. The boxes for Level 1 and Level 2 are right skewed toward higher values, while that of Level 3 is symmetric.

Table 23 displays ANOVA results for the mean number of unrelated words accurately recalled by participants on the delayed test, grouped by Level.

Table 23

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Unrelated list – A-to-E)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18.621	2	9.310	2.236	.114
Within Groups	320.567	77	4.163		
Total	339.188	79			

With  $F(2, 77) = 2.236$ ,  $P = 0.114$  ( $P > 0.05$ ), indicating that the difference is not statistically significant. Thus, we accept the "Null Hypothesis" of no difference between the three levels in the delayed tests regarding unrelated clustered words.

*Thematic Clustering (Immediate Tests)*

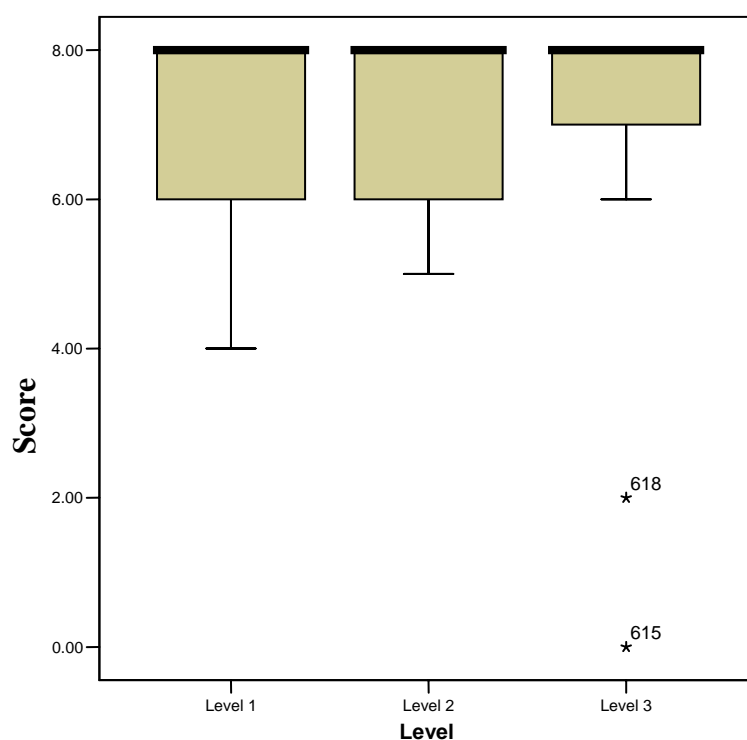
Table 24 shows mean and standard deviation for accurately recalled words from the thematic cluster, based on the immediate test for the three levels of participants.

Table 24

*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Thematic list – A-to-E)*

Level	N	Mean	SD
Level 1	30	7.03	1.33
Level 2	30	7.10	1.09
Level 3	20	7.05	2.16

Means for number of recalled thematically clustered words by participants of all levels are close. The difference between the highest mean (7.10) and the lowest one (7.03) is only 0.07 indicating that student level had no significant effect on recall for this test.



*Figure 14*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – thematic list – A-to-E).

As Figure 14 shows, the median number of accurately recalled words for this cluster and group at all levels is 8. A small variation appears in Level 3, explaining

why its inter-quartile range is the smallest. The boxes for all the levels of participants are left skewed toward lower values. Fifty percent of all participants recalled all eight words from the thematic list on the immediate test. The lowest median score is that of Level 1, with four accurately recalled words, then five words for Level 2, and six for Level 3, with two outliers. This figure also reveals that with thematic clustering, Level 1 participants did as well as those of Level 3. This is a strong argument for the use of thematic clustering with beginners.

Table 25 displays ANOVA results for the thematic cluster test given to participants immediately after studying the words. Participants are grouped by level.

Table 25

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Thematic list – A-to-E)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.071	2	.035	.016	.985
Within Groups	174.617	77	2.268		
Total	174.688	79			

As a result of the tiny differences between the means in Table 24, The ANOVA reports an  $F$  of 0.016, which with 2 and 77 degrees of freedom is not statistically significant ( $P > 0.05$ ). The high  $P$  value in Table 25 indicates that we should accept the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding thematically clustered words.

*Thematic Clustering (Delayed Tests)*

Table 26 displays mean number of words recalled on the delayed test of thematically clustered words, by level.

Table 26

*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Thematic list – A-to-E)*

Level	N	Mean	SD
Level 1	30	2.67	1.54
Level 2	30	2.53	1.93
Level 3	20	3.35	2.48

Like the results of table 24, the means in Table 26 are close. The difference between the highest mean (3.35) for thematic words accurately recalled by Level 3 participants, and the lowest one (2.53), for those recalled by Level 2 participants, is only 0.82.

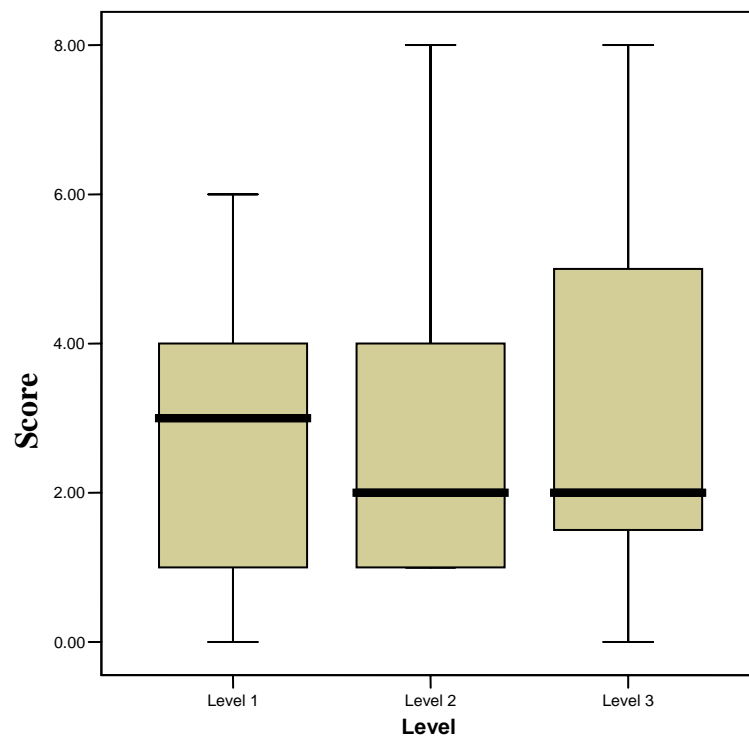


Figure 15

Box-plot diagram of score in relation to level of participants (all levels – delayed test – thematic list – A-to-E).

The whiskers of Level 1 and level 3 look symmetric while whiskers of level 2 are right skewed toward higher values. The median number of correctly recalled



words for Level 1 is 3, while it is 2 for Levels 2 and 3. Twenty-five percent of Level 2 participants recalled only one word correctly from the thematic list on the delayed test, and 25% recalled from 4 to 8 words.

In Table 27, ANOVA results compare the means of accurately recalled thematically clustered words on the delayed test for Levels 1, 2, and 3.

Table 27

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Thematic list – A-to-E)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.704	2	4.352	1.145	.324
Within Groups	292.683	77	3.801		
Total	301.388	79			

Table 27 reports an F of 1.145, which with 2 and 77 degrees of freedom is not statistically significant ( $P > 0.05$ ). The "Null Hypothesis" of no difference between the three levels in the delayed tests regarding thematically clustered words is accepted.

*Context Clustering (Immediate Tests)*

Table 28 offers the mean number of recalled words for the contextually clustered list. Results for the three participant levels on the immediate test are compared.

Table 28

*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Context list – A-to-E)*

Level	N	Mean	SD
Level 1	30	5.63	2.36
Level 2	30	5.73	1.68
Level 3	20	6.50	1.61

Table 28 reveals that Level 3 participants recalled more contextual clustered words than all other participants. Level 1 participants were the fewest to recall this type of vocabulary. However, the difference between the means for recalled contextual clustered words by Level 1 participants and Level 2 participants is only 0.1 which is a very tiny difference.

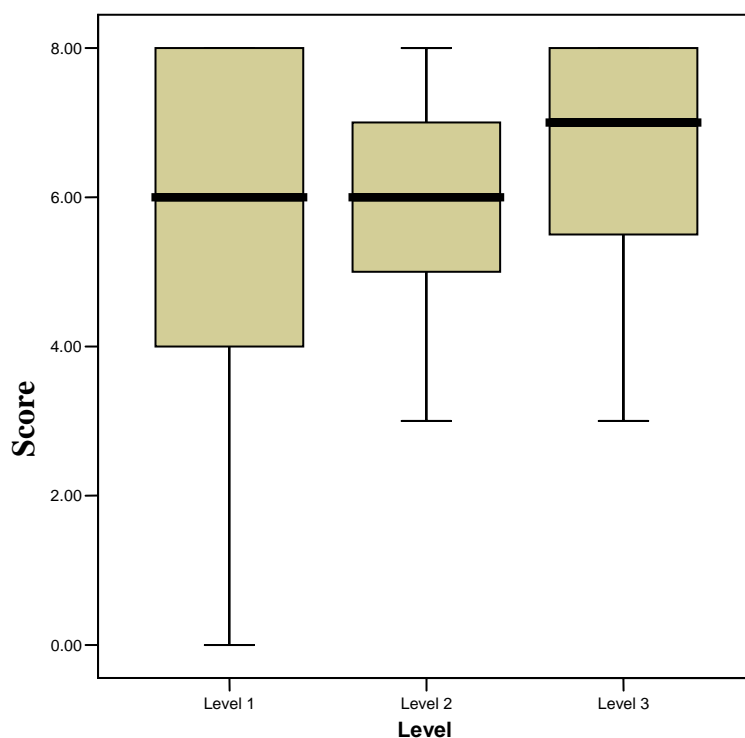


Figure 16

Box-plot diagram of score in relation to level of participants (all levels – immediate test – context list – A-to-E).

Figure 16 above indicates that the inter-quartile range for level 1 is the largest because the variation in number of contextually clustered words recalled correctly is larger than the variation in other groups. The whisker of Level 2 is symmetric while the other whiskers are left skewed toward lower values. The median of Levels 1 and 2 is 6, while that of level 3 is 7. Twenty-five percent of Level 1 participants and Level 3 participants recalled eight words from the context list in the immediate test. The lowest number of recalled words by Level 1 participants is 0, while 3 was the lowest score for Level 2 and Level 3 participants.

In Table 29, ANOVA results compare the means of accurately recalled, contextually clustered words on the immediate test for Levels 1, 2, and 3.

Table 29

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Context list – A-to-E)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.154	2	5.077	1.340	.268
Within Groups	291.833	77	3.790		
Total	301.987	79			

The ANOVA reports an  $F$  of 1.340, which with 2 and 77 degrees of freedom is not statistically significant ( $P > 0.05$ ). The "Null Hypothesis" of no difference between the three levels in the immediate tests regarding contextually clustered words is accepted.

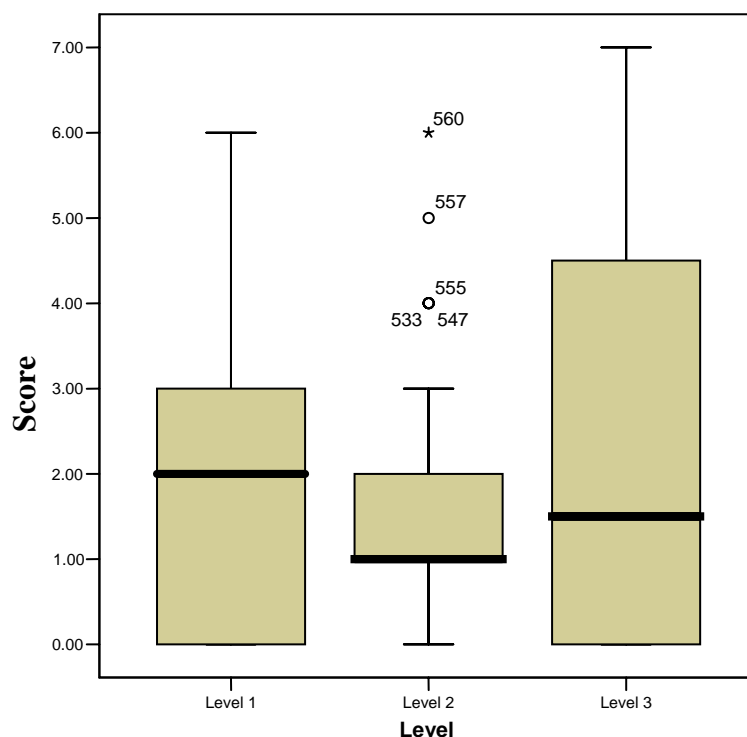
*Context Clustering (Delayed Tests)*

Table 30 presents means and standard deviations for the number of accurately recalled words on the delayed test, for the contextual cluster. Participants are grouped by level.

Table 30  
*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Context list – A-to-E)*

Level	N	Mean	SD
Level 1	30	1.97	1.75
Level 2	30	1.60	1.61
Level 3	20	2.30	2.49

Table 30 is similar to Table 28 in that Level 3 participants recalled more contextually clustered words than other participants. This table differs from the previous one in that Level 1 participants recalled more of these words than Level 2 participants. The issue of significance for these means is illustrated in the next table.



*Figure 17*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – context list – A-to-E).

In the figure above, the inter-quartile range for Level 3 is the largest because the variation in number of words recalled is larger than that of the other Levels. The whisker of Level 2 is symmetric, while the other whiskers are right skewed toward higher values. Level 2 has the lowest median, which is 1 accurately recalled word, while the median of Level 1 is 2 and that of level 3 is 1.5. Twenty-five percent of Level 2 participants recalled one word from the context list in the delayed test. Also, 25% of Level 1 participants and the same proportion of Level 3 participants could not recall any word from the list.

Results presented in Table 31 compare the means of accurately recalled contextually clustered words for the three levels of participants on the delayed test.

Table 31  
*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Context list – A-to-E)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6.021	2	3.010	.821	.444
Within Groups	282.367	77	3.667		
Total	288.387	79			

With  $F(2, 77) = 0.821$ ,  $P = 0.444$  ( $P > 0.05$ ). This result indicates that the difference is not statistically significant. Thus, we accept the "Null Hypothesis" of no difference between the three levels in the delayed tests regarding contextual clustered words.

*English to Arabic Translation Direction*

*Level 1 (Immediate Tests)*

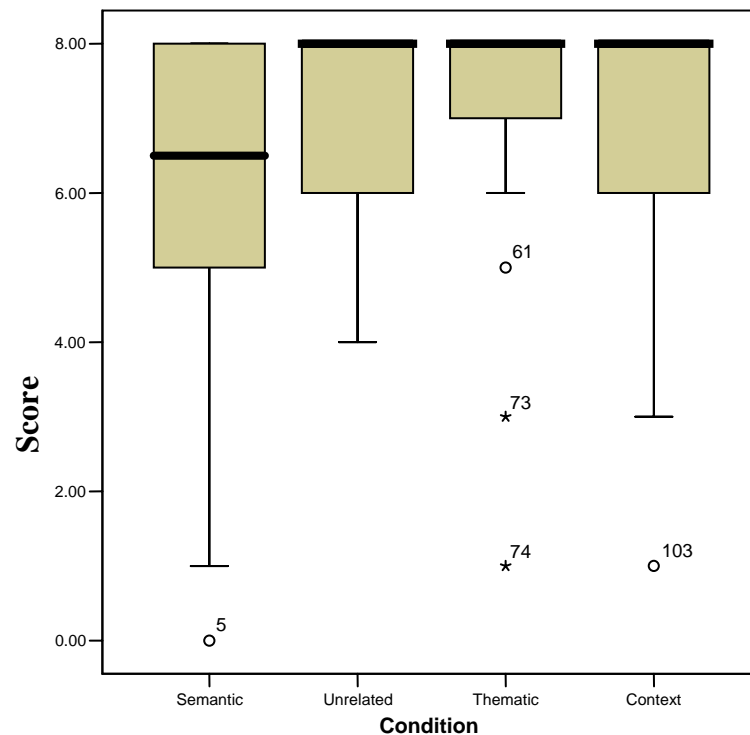
In Table 32, means and standard deviations for the number of accurately recalled words in the thematic cluster are displayed for the immediate test administered to Level 1 participants.

Table 32  
*Means and Standard Deviation of Recalled Words (Level 1– Immediate test – E-to-A)*

Condition	N	Mean	SD
Semantic	30	5.77	2.43
Unrelated	30	6.97	1.30
Thematic	30	7.03	1.63
Context	30	6.87	1.85

As Table 32 clearly shows, Level 1 participants recalled more thematically clustered words than words from the other three lists. More unrelated clustered words were recalled than the semantically and contextually clustered words. Also, this table

shows that the means for semantically and contextually clustered recalled words are the lowest.



*Figure 18*

Box-plot diagram of score in relation to types of clustering (level 1 – immediate test – E-to-A).

The inter-quartile range for recalled semantic clustered words is larger than the others because the variation is larger. A relatively small variation makes the inter-quartile range for recalled thematically clustered words the smallest. Variation in the inter-quartile ranges for the recalled unrelated and contextual clustered words is the same. The medians of the recalled unrelated, thematically, and contextual clustered words are higher than the median of the recalled semantic clustered words. Fully 75% of participants recalled 7-8 thematically clustered words, with several outliers represented below the inter-quartile ranges.

ANOVA results in Table 33 compare the means for Level 1 participant accurate word recall on the immediate test, grouped by the four types of vocabulary clusters.

Table 33  
*Analysis of Variance of Number of Recalled Words (Level 1 – Immediate test – E-to-A)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.225	3	10.742	3.140	.028
Within Groups	396.767	116	3.420		
Total	428.992	119			

Table 33 reports an F of 3.140, which with 3 and 116 degrees of freedom is statistically significant ( $P = 0.028$ ). By this point, we can feel comfortable in rejecting the "Null Hypothesis" of no difference between the four types of vocabulary grouping in the immediate test in the population of Level 1 participants.  $E^2$  is 0.07. This means that the four types of vocabulary grouping explains about 7 percent of the variation in number of recalled words. We could say that this relationship is weak but statistically significant.

Table 34 illustrates the Post Hoc tests performed on pairs of variables to further specify the significant relationship described for Level 1 participants in Table 33.



Table 34  
*Post Hoc Tests (Level 1 – Immediate test – E-to-A)*

**Multiple Comparisons**

Dependent Variable: Immediate  
 LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-1.20000(*)	.47752	.013	-2.1458	-.2542
	Thematic	-1.26667(*)	.47752	.009	-2.2125	-.3209
	Context	-1.10000(*)	.47752	.023	-2.0458	-.1542
Unrelated	Semantic	1.20000(*)	.47752	.013	.2542	2.1458
	Thematic	-.06667	.47752	.889	-1.0125	.8791
	Context	.10000	.47752	.834	-.8458	1.0458
Thematic	Semantic	1.26667(*)	.47752	.009	.3209	2.2125
	Unrelated	.06667	.47752	.889	-.8791	1.0125
	Context	.16667	.47752	.728	-.7791	1.1125
Context	Semantic	1.10000(*)	.47752	.023	.1542	2.0458
	Unrelated	-.10000	.47752	.834	-1.0458	.8458
	Thematic	-.16667	.47752	.728	-1.1125	.7791

\* The mean difference is significant at the .05 level.

Table 34 shows that there is a statistically significant difference between the numbers of recalled semantically clustered words on one hand and the number of recalled unrelated, thematically, and contextual clustered words on the other hand. When examining Tables 32 and 34, we find out that when we compare number of accurately recalled words for any other type of clustering with results for the semantic cluster, Level 1 participants recalled more words from all other types of clustering than from the semantic clustering. The strongest (P) value appears when we compare semantic and thematic clusterings (0.009), suggesting that the difference between the means for the accurately recalled words from the thematic list and those from the semantic list is larger than the differences between the number of accurately words recalled from any other list and the semantic list. On the other hand, Table 34 reveals that the difference between the number of recalled unrelated, thematic, and contextual clustered words is not statistically significant.

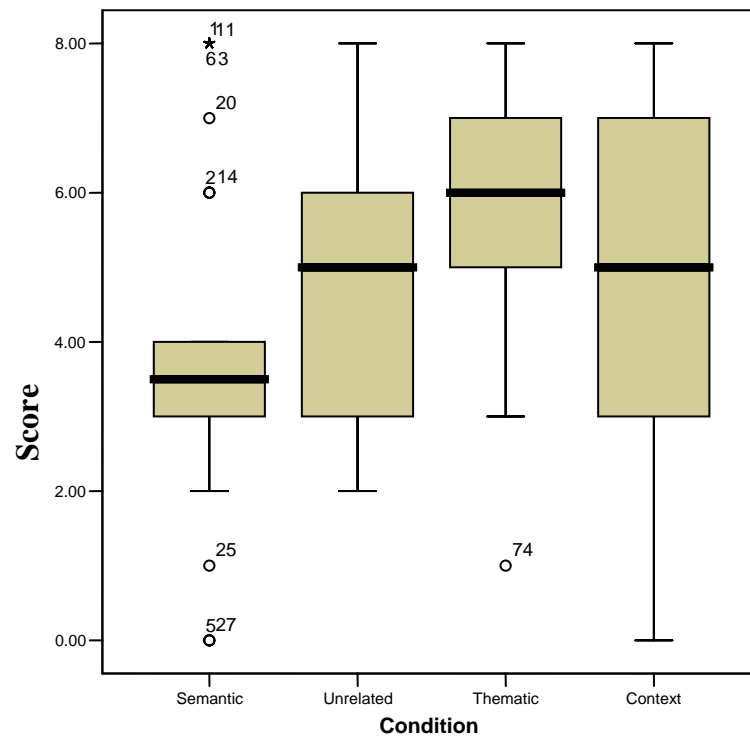
*Level 1 (Delayed Tests)*

Table 35 reports mean number of accurately recalled words and standard deviations for the delayed test taken by Level 1 participants.

Table 35  
*Means and Standard Deviation of Recalled Words (Level 1 – Delayed test – E-to-A)*

Condition	N	Mean	SD
Semantic	30	3.87	2.24
Unrelated	30	5.03	1.96
Thematic	30	5.83	1.88
Context	30	5.00	2.32

The results that appear in Table 35 are similar to those in Table 32. Level 1 participants accurately recalled more thematically clustered words than words from the semantic, unrelated, or contextual clusters. The difference between the means for unrelated and context conditions is very small (0.03). The semantic condition comes last, with a small mean of only (3.87). The difference between the means for the semantic and thematic conditions is large (2.04), in favor of the thematic condition. This finding suggests that there is a strong possibility that the differences between the means are statistically significant.



*Figure 19*  
Box-plot diagram of score in relation to types of clustering (level 1 – delayed test – E-to-A).

In the figure above, the thematic and context conditions got the highest 75% of correctly recalled words, with a score of 7. A very small variation in the number of recalled semantically clustered words shows that all participants, except some outliers, recalled between 2 and 4 words. The lowest number of recalled words from the thematic list is 3, except for one outlier with a value of 1. The box for the semantic list is left skewed toward lower values, while the other boxes are more symmetric.

Table 34 reports ANOVA results for Level 1 participants on the delayed test, grouped by cluster type.

Table 36  
*Analysis of Variance of Number of Recalled Words (Level 1 – Delayed test – E-to-A)*  
**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	58.867	3	19.622	4.423	.006
Within Groups	514.600	116	4.436		
Total	573.467	119			

The ANOVA reports an  $F$  of 4.423, which with 3 and 116 degrees of freedom is statistically significant ( $P = 0.006$ ) ( $P < 0.05$ ). Thus, we can feel comfortable in rejecting the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the delayed tests in the population of Level 1 participants.

$E^2$  of 0.10 tells that despite the statistical significance, the type of vocabulary clustering explains only about 10 percent of the variation in number of words. This is a weak relationship but still significant.

Table 37 explores the above result of significance through Post Hoc tests for the Level 1 delayed test, examining relationships between scores for all cluster types.

Table 37  
*Post Hoc Tests (Level 1 – Delayed test – E-to-A)*  
**Multiple Comparisons**

Dependent Variable: Delayed  
 LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-1.16667(*)	.54383	.034	-2.2438	-.0896
	Thematic	-1.96667(*)	.54383	.000	-3.0438	-.8896
	Context	-1.13333(*)	.54383	.039	-2.2104	-.0562
Unrelated	Semantic	1.16667(*)	.54383	.034	.0896	2.2438
	Thematic	-.80000	.54383	.144	-1.8771	.2771
	Context	.03333	.54383	.951	-1.0438	1.1104
Thematic	Semantic	1.96667(*)	.54383	.000	.8896	3.0438
	Unrelated	.80000	.54383	.144	-.2771	1.8771
	Context	.83333	.54383	.128	-.2438	1.9104
Context	Semantic	1.13333(*)	.54383	.039	.0562	2.2104
	Unrelated	-.03333	.54383	.951	-1.1104	1.0438
	Thematic	-.83333	.54383	.128	-1.9104	.2438

\* The mean difference is significant at the .05 level.

The results displayed in Table 37 are exactly identical to those of Table 34. The difference between the semantic and unrelated conditions is significant in favor of the unrelated condition ( $P = 0.034$ ). When we compare the thematic condition to the semantic condition, we find out that the thematic condition outperformed the semantic condition with a  $P$  value of 0.000. The same can be said regarding the semantic and contextual condition, since the second one outperformed the first, with a  $P$  value of 0.039.

Table 37 also shows that the differences between the thematic, unrelated, and contextual conditions are not statistically significant. One more point to mention is that we see a  $P$  value of 0.951 when we compare the unrelated condition to the contextualized one. This high value is expected, since the difference between means for both conditions is a very small number (0.03). In other words, Level 1 participants performed very similarly when tested on unrelated and contextualized words.

*Level 2 (Immediate Tests)*

Table 38 shows mean number of words accurately recalled and standard deviations for each cluster, on the Level 2 immediate test.

Table 38

*Means and Standard Deviation of Recalled Words (Level 2– Immediate test – E-to-A)*

Condition	N	Mean	SD
Semantic	30	6.53	1.66
Unrelated	30	7.13	1.53
Thematic	30	7.13	1.14
Context	30	7.40	1.10

The results of table 38 shows that the unrelated, thematic, and context conditions outperformed the semantic condition since the mean of the recalled semantically clustered words is the lowest (6.53). The means for the recalled thematically and unrelated clustered words are identical (7.13). One last and probably the most interesting finding is that Level 1 participants recalled more contextually clustered words than all the other cluster types. This result never appeared elsewhere in the Arabic-to-English translation direction, or in the English-to-Arabic translation direction with Level 2 participants.

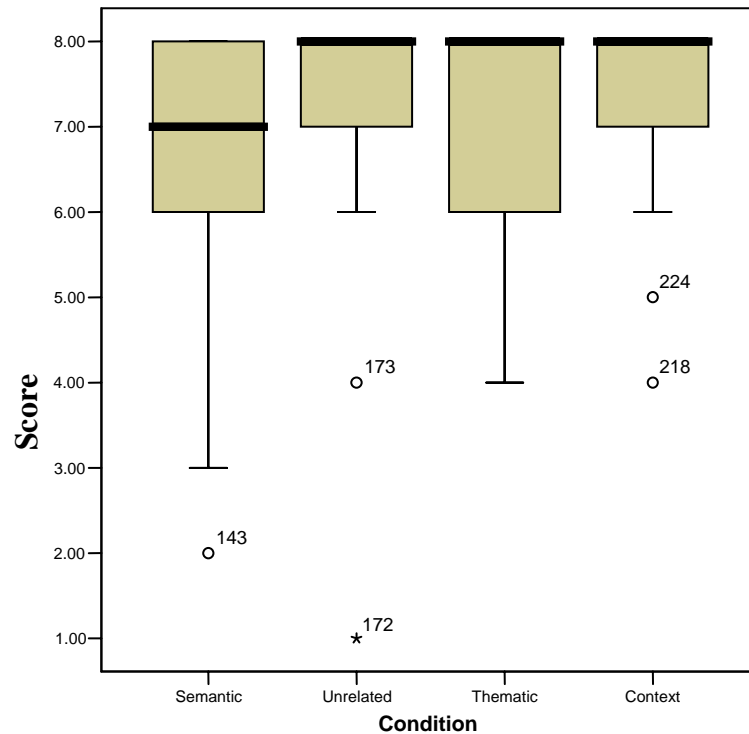


Figure 20

Box-plot diagram of score in relation to types of clustering (level 2 – immediate test – E-to-A).

The medians of the recalled words from the unrelated, thematic, and context lists are higher than that of the semantic list. All the boxes are left skewed toward lower values. Fifty percent of the participants recalled eight words correctly from the unrelated, thematic, and context lists, while 25% recalled the same number correctly from the semantic list.

The ANOVA results in Table 39 compare means of accurately recalled words for the Level 2 immediate test, by cluster group.

Table 39

*Analysis of Variance of Number of Recalled Words (Level 2–Immediate test–E-to-A)*  
ANOVA

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.100	3	4.033	2.131	.100
Within Groups	219.600	116	1.893		
Total	231.700	119			

An ANOVA, reported in Table 39, indicates that the differences among means yield an  $F(3, 116) = 2.131$ , with  $P = 0.100$  ( $P > 0.05$ ). This indicates that no particular type of vocabulary clustering was superior to the others in recalling words in the immediate tests for Level 2 participants. Thus, we have to accept the "Null Hypothesis" of no difference between the four types of vocabulary clustering for this group of participants, on the immediate tests. Therefore, there is no need to present the table of Post Hoc Tests.

#### *Level 2 (Delayed Tests)*

Table 40 displays mean and standard deviations for accurately recalled words in each cluster group, on the delayed test for Level 2 participants.

Table 40  
*Means and Standard Deviation of Recalled Words (Level 2 – Delayed test – E-to-A)*

Condition	N	Mean	SD
Semantic	30	4.30	2.07
Unrelated	30	6.23	1.83
Thematic	30	4.83	1.93
Context	30	4.90	1.56

Table 40 shows a number of interesting results. First, Level 2 participants recalled more unrelated cluster words than words from any other type of clustering, and the difference between the mean of the unrelated condition and all the other three conditions is quite large. This is the first occurrence for this result, since there was no similar finding in the immediate tests. Second, there is a big decline in the means for recalled words. The mean of recalled semantically clustered words goes from (6.53) in the immediate test to (4.30) in the delayed test for this group, showing that retention of semantically related words for this population was rather poor. The mean



of recalled unrelated clustered words goes from (7.13) in the immediate test to (6.23) in the delayed test. The mean of recalled thematically clustered words goes from (7.13) in the immediate test to (4.83) in the delayed test. The biggest decline appears in the contextualized cluster, since the mean of recalled contextually clustered words goes from (7.40) on the immediate test to (4.90) on the delayed test. Third, as seen with other tested groups, the mean for the recalled semantically clustered words is the lowest of the cluster groups, indicating that Level 2 participants had difficulty recalling words from the semantic list.

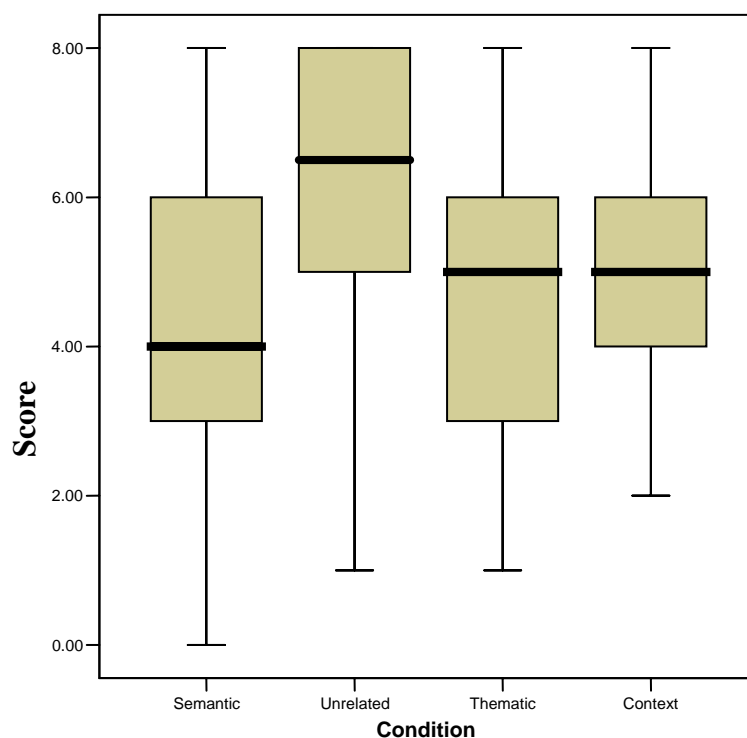


Figure 21

*Box-plot diagram of score in relation to types of clustering (level 2 – delayed test – E-to-A).*

The box for the unrelated list looks left skewed toward lower values while the other boxes are symmetric. The median of recalled words from the unrelated list is the highest, while that of the semantic list is the lowest. Twenty-five percent of this group

accurately recalled all 8 words from the unrelated list. Furthermore, 25% of participants recalled from 6 to 8 words from each of the other lists.

The ANOVA in Table 41 shows results for comparison of mean scores for correctly recalled words on all Level 2 delayed tests, grouped by cluster.

Table 41  
*Analysis of Variance of Number of Recalled Words (Level 2 – Delayed test – E-to-A)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	60.933	3	20.311	5.882	.001
Within Groups	400.533	116	3.453		
Total	461.467	119			

With  $F(3, 116) = 5.882$ ,  $P = 0.001$  ( $P < 0.05$ ). This result indicates that the difference is statistically significant at the 0.001 level. Thus, we are comfortable in rejecting the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the delayed tests in the population of Level 2 participants.

An  $E^2$  of 0.13 reveals that despite the statistical significance, the type of vocabulary clustering explains only about 13 percent of the variation in number of words. This is a weak relationship but still significant.

Table 42 explores the above result of significance through Post Hoc tests for the Level 2 delayed test, examining relationships between mean scores for all cluster types.

Table 42

*Post Hoc Tests (Level 2 – Delayed test – E-to-A)*

**Multiple Comparisons**

Dependent Variable: Delayed  
LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-1.93333(*)	.47978	.000	-2.8836	-.9831
	Thematic	-.53333	.47978	.269	-1.4836	.4169
	Context	-.60000	.47978	.214	-1.5503	.3503
Unrelated	Semantic	1.93333(*)	.47978	.000	.9831	2.8836
	Thematic	1.40000(*)	.47978	.004	.4497	2.3503
	Context	1.33333(*)	.47978	.006	.3831	2.2836
Thematic	Semantic	.53333	.47978	.269	-.4169	1.4836
	Unrelated	-1.40000(*)	.47978	.004	-2.3503	-.4497
	Context	-.06667	.47978	.890	-1.0169	.8836
Context	Semantic	.60000	.47978	.214	-.3503	1.5503
	Unrelated	-1.33333(*)	.47978	.006	-2.2836	-.3831
	Thematic	.06667	.47978	.890	-.8836	1.0169

\* The mean difference is significant at the .05 level.

Table 42 reveals that the difference between the means for the recalled semantically and unrelated clustered words is significant. The P value is 0.000. It shows also that the difference between the means for the recalled unrelated and thematically clustered words is significant, since the P value is 0.004. Moreover, the difference between the means for the recalled unrelated and contextually clustered words is significant, since the P value is 0.006. These P values are very small, indicating the large difference between the means for all the four conditions in favor of the unrelated condition. Finally, the small difference between the means for the thematic and context conditions, (4.83) and (4.90) respectively, explains the high P value (0.890) that appeared when we compared the two conditions. This result suggests that the difference is not statistically significant.

*Level 3 (Immediate Tests)*

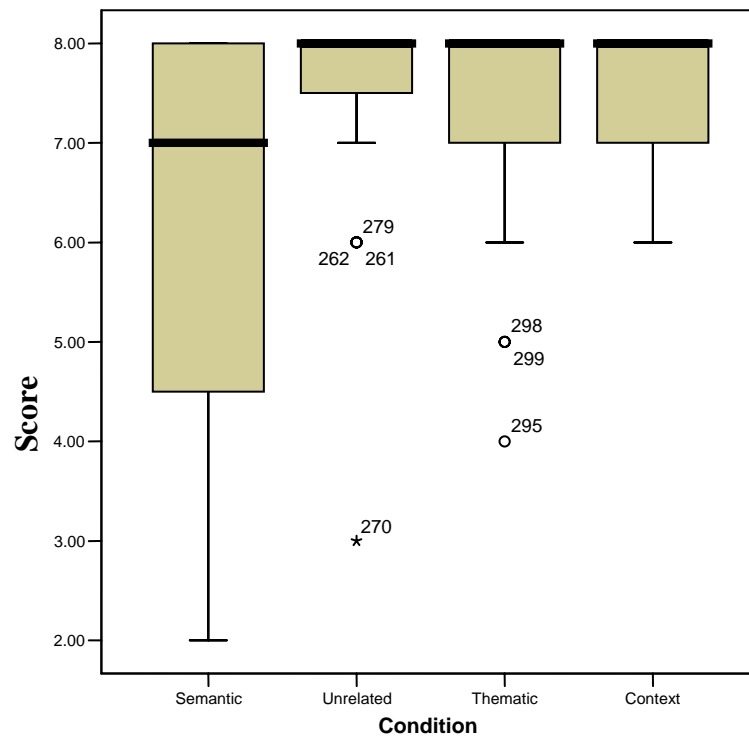
Table 43 presents means and standard deviations for words recalled correctly by Level 3 participants on the immediate test, arranged by cluster.

Table 43

*Means and Standard Deviation of Recalled Words (Level 3– Immediate test – E-to-A)*

Condition	N	Mean	SD
Semantic	20	5.95	2.11
Unrelated	20	7.40	1.27
Thematic	20	7.25	1.25
Context	20	7.55	0.76

As Table 43 clearly shows, the words recalled with greatest accuracy in the immediate tests are the contextually clustered words. The unrelated clustered words were second in accuracy of student recall. The mean of recalled semantically clustered words is the lowest one, and the means for the unrelated, thematic, and context conditions are close to each other. The low standard deviation for the context condition is caused in part by the ceiling of 8 possible correct answers.



*Figure 22*

Box-plot diagram of score in relation to types of clustering (level 3 – immediate test – E-to-A).

The inter-quartile range for semantic list is the largest because the variation is larger. Fifty percent of Level 3 participants recalled all eight words correctly. This means that half of Level 3 participants remembered the full list of unrelated, thematically, and contextually related words. Seven is the lowest number of recalled words from the unrelated words, while the corresponding number is six from the thematic and context lists.

Table 44 reports ANOVA results for the Level 3 test administered immediately after participants had studied the word lists. Means for the number of words recalled accurately are compared by cluster type.

Table 44

*Analysis of Variance of Number of Recalled Words (Level 3–Immediate test–E-to-A)*  
**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.437	3	10.812	5.252	.002
Within Groups	156.450	76	2.059		
Total	188.888	79			

The ANOVA reports an  $F$  of 5.252, which with 3 and 76 degrees of freedom is statistically significant at the 0.002 level ( $P < 0.05$ ). Therefore, we reject the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the immediate tests in the population of Level 3 participants. The  $E^2$  of 0.17 demonstrates that despite the statistical significance, the type of vocabulary clustering explains only about 17 percent of the variation in number of words. The significant relationship described in Table 46 above is further explored in Table 45 via a series of Post Hoc Tests performed on all cluster types.

Table 45

*Post Hoc Tests (Level 3 – Immediate test – E-to-A)*

**Multiple Comparisons**

Dependent Variable: Immediate  
 LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-1.45000(*)	.45371	.002	-2.3536	-.5464
	Thematic	-1.30000(*)	.45371	.005	-2.2036	-.3964
	Context	-1.60000(*)	.45371	.001	-2.5036	-.6964
Unrelated	Semantic	1.45000(*)	.45371	.002	.5464	2.3536
	Thematic	.15000	.45371	.742	-.7536	1.0536
	Context	-.15000	.45371	.742	-1.0536	.7536
Thematic	Semantic	1.30000(*)	.45371	.005	.3964	2.2036
	Unrelated	-.15000	.45371	.742	-1.0536	.7536
	Context	-.30000	.45371	.510	-1.2036	.6036
Context	Semantic	1.60000(*)	.45371	.001	.6964	2.5036
	Unrelated	.15000	.45371	.742	-.7536	1.0536
	Thematic	.30000	.45371	.510	-.6036	1.2036

\* The mean difference is significant at the .05 level.

When we compare the number of recalled unrelated clustered words to that of words semantically clustered, we find that the difference is statistically significant (0.002). The difference between the thematic and semantic clustering is significant in favor of the former (0.005). Likewise the context and semantic conditions are significant, in favor of the contextual (0.001). The differences between the means for the unrelated, thematic, and context conditions are small, as shown in Table 43, and therefore, not statistically significant.

### *Level 3 (Delayed Tests)*

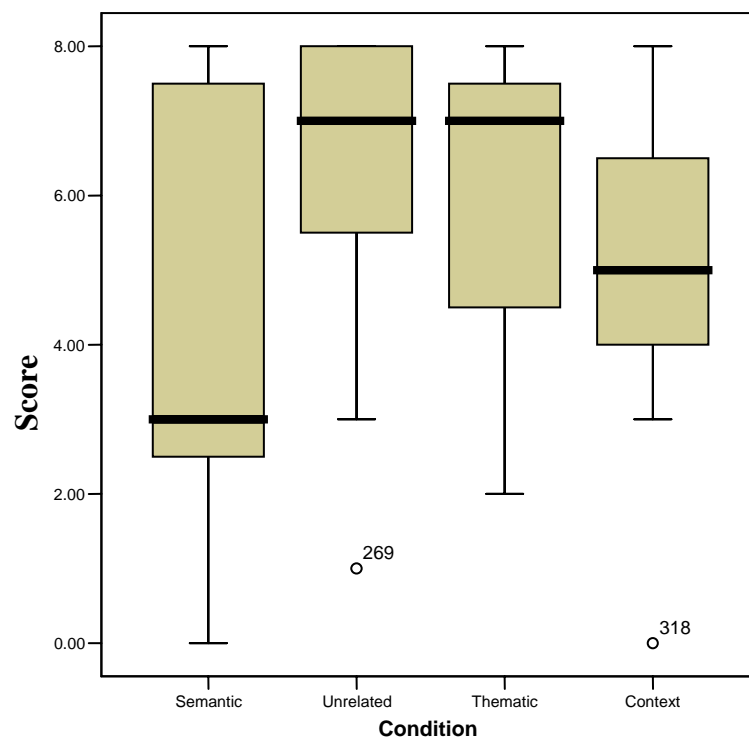
In Table 46, means and standard deviations for words recalled accurately on the Level 3 delayed test are displayed.

Table 46

*Means and Standard Deviation of Recalled Words (Level 3 – Delayed test – E-to-A)*

Condition	N	Mean	SD
Semantic	20	4.20	2.69
Unrelated	20	6.30	1.92
Thematic	20	5.90	2.00
Context	20	5.15	2.03

The results of Table 46 are not different from results of Table 43 in that the mean of the recalled semantically clustered words is the lowest (4.20). They differ in that the mean of the recalled unrelated clustered words is the highest (6.30), followed by the mean of the recalled contextual clustered words (5.15) which comes before the mean of the recalled thematically clustered words. It appears that there are differences between the means but whether these differences are significant will appear in the following table.



*Figure 23*

Box-plot diagram of score in relation to types of clustering (level 3 – delayed test – E-to-A).

A large variation appears in the inter-quartile range of the semantic list. The medians of the recalled words from the unrelated and thematic lists are similar to each other and different from the others. 50% of Level 3 participants recalled between seven and eight words correctly, while the other 50% of participants recalled between two to seven words from the thematic list. The lowest number of recalled words is 0 from the semantic list and 2 from the thematic list. The whisker is symmetric for the context list, while the others are left skewed toward lower values.

Table 47 shows results for the ANOVA performed on Level 3 delayed test scores, comparing the mean number of words recalled accurately for each cluster group.



Table 47

*Analysis of Variance of Number of Recalled Words (Level 3 – Delayed test – E-to-A)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	51.237	3	17.079	3.588	.017
Within Groups	361.750	76	4.760		
Total	412.988	79			

AN ANOVA, reported in Table 47, indicates that the differences among means yield an  $F(3, 76) = 3.588$ ,  $P = 0.017$  ( $P < 0.05$ ). This indicates that one or more particular type of vocabulary clustering was superior to the others in recalling words in the delayed tests. Thus, we reject the "Null Hypothesis" of no difference between the four types of vocabulary clustering in the population of Level 3 participants in the delayed tests. The post hoc test results displayed in Table 48 compare means for cluster group accuracy scores, on the data listed in Table 47, above.

Table 48

*Post Hoc Tests (Level 3 – Delayed test – E-to-A)*

**Multiple Comparisons**

Dependent Variable: Delayed  
LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-2.10000(*)	.68992	.003	-3.4741	-.7259
	Thematic	-1.70000(*)	.68992	.016	-3.0741	-.3259
	Context	-.95000	.68992	.173	-2.3241	.4241
Unrelated	Semantic	2.10000(*)	.68992	.003	.7259	3.4741
	Thematic	.40000	.68992	.564	-.9741	1.7741
	Context	1.15000	.68992	.100	-.2241	2.5241
Thematic	Semantic	1.70000(*)	.68992	.016	.3259	3.0741
	Unrelated	-.40000	.68992	.564	-1.7741	.9741
	Context	.75000	.68992	.280	-.6241	2.1241
Context	Semantic	.95000	.68992	.173	-.4241	2.3241
	Unrelated	-1.15000	.68992	.100	-2.5241	.2241
	Thematic	-.75000	.68992	.280	-2.1241	.6241

\* The mean difference is significant at the .05 level.

As Table 48 shows, the difference between the unrelated and the semantic conditions is statistically significant (0.003). The difference between the thematic and semantic conditions is also statistically significant (0.016). On the other hand, the difference between the context and semantic conditions is not significant, nor are the differences between unrelated and thematic, unrelated and context, or thematic and context.

### *Within Conditions*

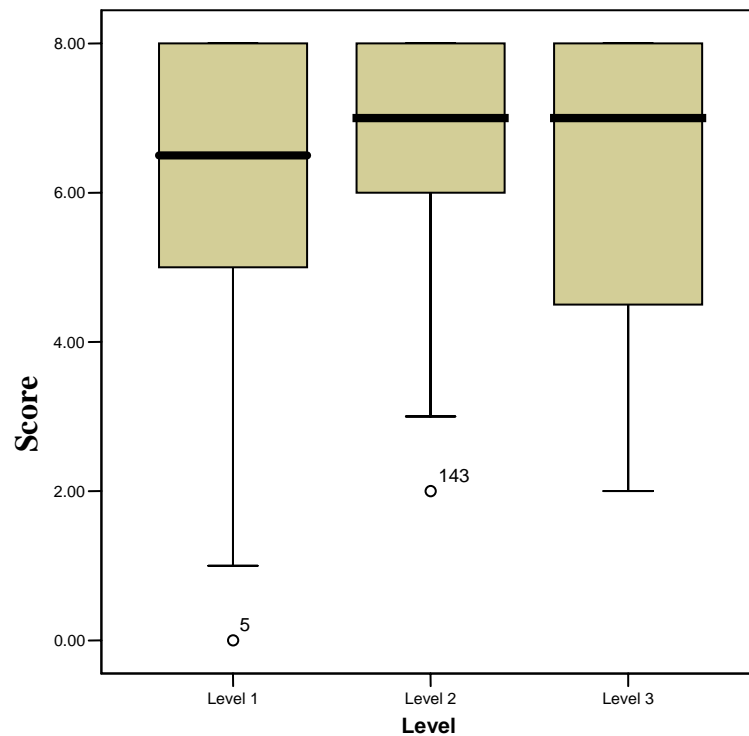
#### *Semantic Clustering (Immediate Tests)*

Table 49 reports mean number of words accurately recalled by participants at all levels on the immediate test for the semantic cluster.

Table 49  
*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Semantic list – E-to-A)*

Level	N	Mean	SD
Level 1	30	5.77	2.43
Level 2	30	6.53	1.66
Level 3	20	5.95	2.11

Table 49 reveals that Level 2 participants recalled more semantically clustered words than Level 3 participants did, and both groups outperformed Level 1 participants in the immediate tests. The difference between the means for the recalled words by Level 1 participants and Level 3 participants is small. The low standard deviation for Level 2 participants is caused in part by the ceiling of only eight possible correct answers.



*Figure 24*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – semantic list – E-to-A).

A similar median for all levels is the main finding of the above figure, and 25% of Level 1, Level 2, and Level 3 participants recalled eight words correctly from the semantic list. Fifty percent of Level 2 participants and Level 3 participants recalled seven to eight words correctly. All boxes are left skewed toward lower values.

Table 50 shows ANOVA results that compare mean scores for correctly recalled words from the semantic cluster on the immediate test, grouped by participant level.

Table 50

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Semantic list – E-to-A)*

## ANOVA

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.417	2	4.708	1.080	.345
Within Groups	335.783	77	4.361		
Total	345.200	79			

With  $F(2, 77) = 1.080$ ,  $P = 0.345$  ( $P > 0.05$ ). This result indicates that the difference is not statistically significant. Although the difference is clear between the means for Level 1 participants and Level 2 participants regarding the recalled semantically clustered words, the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding semantically clustered words is accepted.

*Semantic Clustering (Delayed Tests)*

Table 51 shows mean number of words accurately recalled and standard deviation for first, second, and third Level participants who took the delayed test on the semantic word cluster.

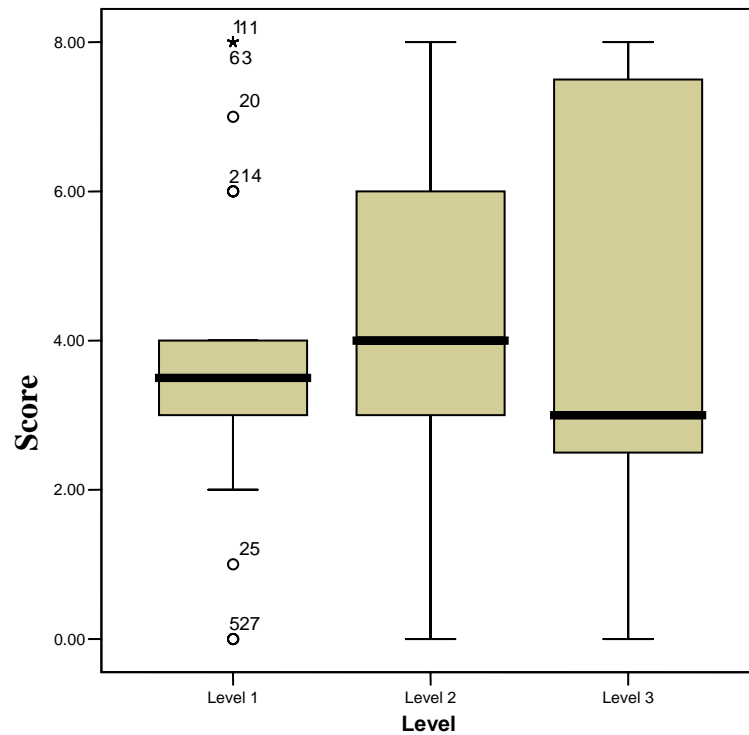
Table 51

*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Semantic list – E-to-A)*

Level	N	Mean	SD
Level 1	30	3.87	2.24
Level 2	30	4.30	2.07
Level 3	20	4.20	2.69

The results in Table 51 are identical to those in Table 49. Level 2 participants recalled more semantically clustered words than Level 3 participants did who, in turn,

recalled more words than Level 1 participants. The difference between the means is small.



*Figure 25*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – semantic list – E-to-A).

The figure above shows that the medians are close. The box for Level 2 is symmetric, while the other two boxes are left skewed toward lower values. Half of the scores for all levels is 3-4 words or higher.

Table 52 compares mean numbers of recalled words for the delayed administration of the test on semantically clustered words, grouped by level and cluster, through an ANOVA.

Table 52

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Semantic list – E-to-A)*

## ANOVA

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.021	2	1.510	.286	.752
Within Groups	406.967	77	5.285		
Total	409.988	79			

Table 52 reports an F of 0.286, which with 2 and 77 degrees of freedom is not statistically significant. The significance value is 0.752 ( $P > 0.05$ ). This is expected since the differences between the means in Table 51 are small. The "Null Hypothesis" of no difference between the three levels in the delayed tests regarding semantically clustered words is accepted.

*Unrelated Clustering (Immediate Tests)*

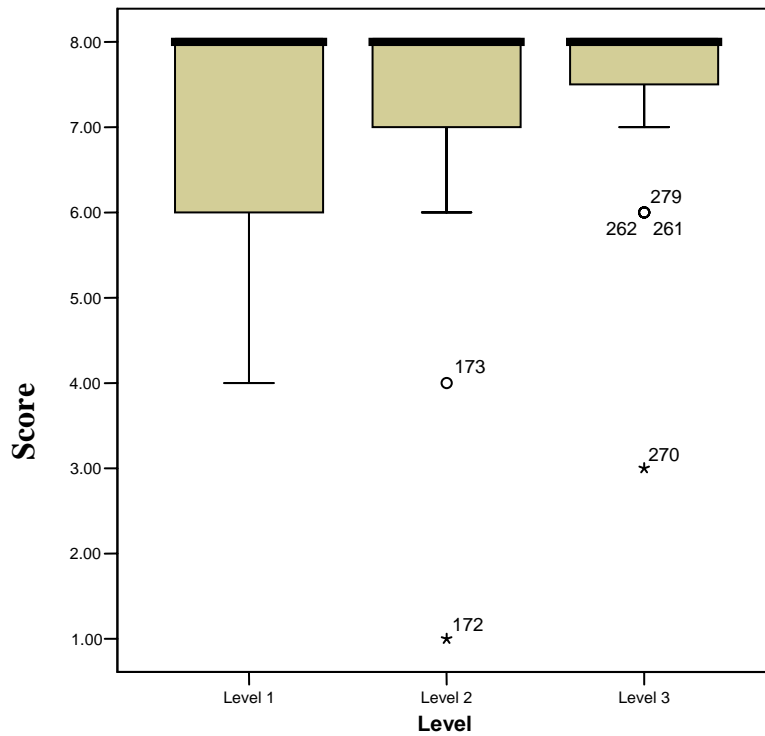
In Table 53, mean number of accurately recalled words and standard deviation are reported for all levels on the immediate test, unrelated word cluster.

Table 53

*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Unrelated list – E-to-A)*

Level	N	Mean	SD
Level 1	30	6.97	1.30
Level 2	30	7.13	1.53
Level 3	20	7.40	1.27

Level 3 participants recalled more unrelated clustered words than did students at other levels. Level 2 participants recalled more unrelated clustered words than those in Level 1. The differences between the means in Table 54 are small.



*Figure 26*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – unrelated list – E-to-A).

Fifty percent of participants from all levels recalled eight words correctly. Half of the participants remembered the full list of unrelated words. All Level 3 participants recalled from seven to eight words. The lowest number of words recalled by Level 1 participants is four. The inter-quartile range of level 3 shows that the variation is small.

Table 54 shows ANOVA results comparing mean scores for the above data, consisting of the number of unrelated words recalled correctly by participants at all levels on the immediate test.

Table 54

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Unrelated list – E-to-A)*

## ANOVA

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.254	2	1.127	.589	.557
Within Groups	147.233	77	1.912		
Total	149.488	79			

With  $F(2, 77) = 0.589$ ,  $P = 0.557$  ( $P > 0.05$ ). This result indicates that the difference is statistically insignificant. Thus, we accept the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding unrelated clustered words.

*Unrelated Clustering (Delayed Tests)*

Table 55 shows mean number of words recalled accurately on the delayed test for the unrelated word cluster, by participant level.

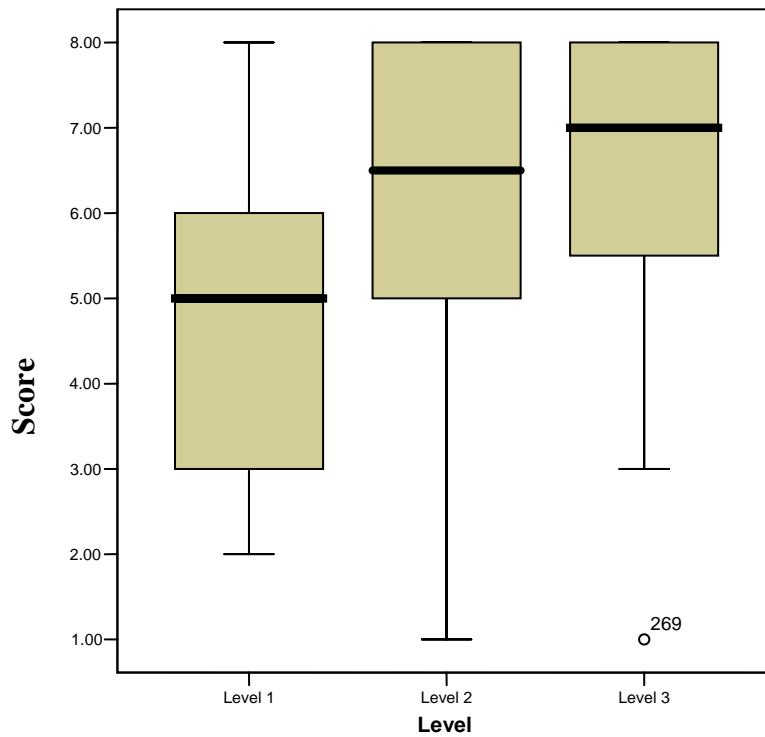
Table 55

*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Unrelated list – E-to-A)*

Level	N	Mean	SD
Level 1	30	5.03	1.96
Level 2	30	6.23	1.83
Level 3	20	6.30	1.92

As Table 55 clearly shows, the mean of recalled unrelated clustered words by Level 2 participants increased from (6.14) on the immediate test to (6.23) on the delayed test. Although the increase is small, this marks the first incidence of this phenomenon. Level 3 participants recalled relatively more unrelated clustered words than other participants did.





*Figure 27*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – unrelated list – E-to-A).

Figure 27 shows that the median of Level 3 is the highest. Of Level 2 and Level 3 participants, 50% recalled all eight words correctly, while 25% of Level 2 participants recalled between one and five words correctly. All boxes are left skewed toward lower values except that of Level 1, which is symmetric.

Table 56 presents ANOVA results for the delayed test on the unrelated cluster, for Levels 1, 2, and 3. Means of number of accurately recalled words are compared, grouped by participant course level.

Table 56

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Unrelated list – E-to-A)*

## ANOVA

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28.267	2	14.133	3.907	.024
Within Groups	278.533	77	3.617		
Total	306.800	79			

With  $F(2, 77) = 3.907$ ,  $P = 0.024$  ( $P < 0.05$ ). Therefore, we can feel comfortable in rejecting the "Null Hypothesis" of no difference between the three levels in the delayed tests regarding unrelated clustered words. Based on Table 56 above,  $E^2$  is 0.09. This means that the three levels of the participants explains about 9 percent of the variation in number of unrelated recalled words. We could say that this relationship is weak but statistically significant.

In Table 57, post hoc tests examine the data described in Table 56 above to explain the significant finding more specifically.

Table 57

*Post Hoc Tests (All Levels – Delayed test – Unrelated list – E-to-A)*

## Multiple Comparisons

Dependent Variable: Delayed  
LSD

(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Level 1	Level 2	-1.20000(*)	.49107	.017	-2.1779	-.2221
	Level 3	-1.26667(*)	.54904	.024	-2.3599	-.1734
Level 2	Level 1	1.20000(*)	.49107	.017	.2221	2.1779
	Level 3	-.06667	.54904	.904	-1.1599	1.0266
Level 3	Level 1	1.26667(*)	.54904	.024	.1734	2.3599
	Level 2	.06667	.54904	.904	-1.0266	1.1599

\* The mean difference is significant at the .05 level.

The difference between the mean of recalled words from the unrelated list by Level 1 participants (5.03) and the mean of recalled words from the unrelated list by Level 2 participants (6.23) is statistically significant ( $P = 0.017$ ) as Table 57 shows. Also, the difference between the mean number of recalled words from the unrelated list by Level 1 participants (5.03) and the mean of recalled words from the unrelated list by Level 3 participants (6.30) is statistically significant ( $P = 0.024$ )

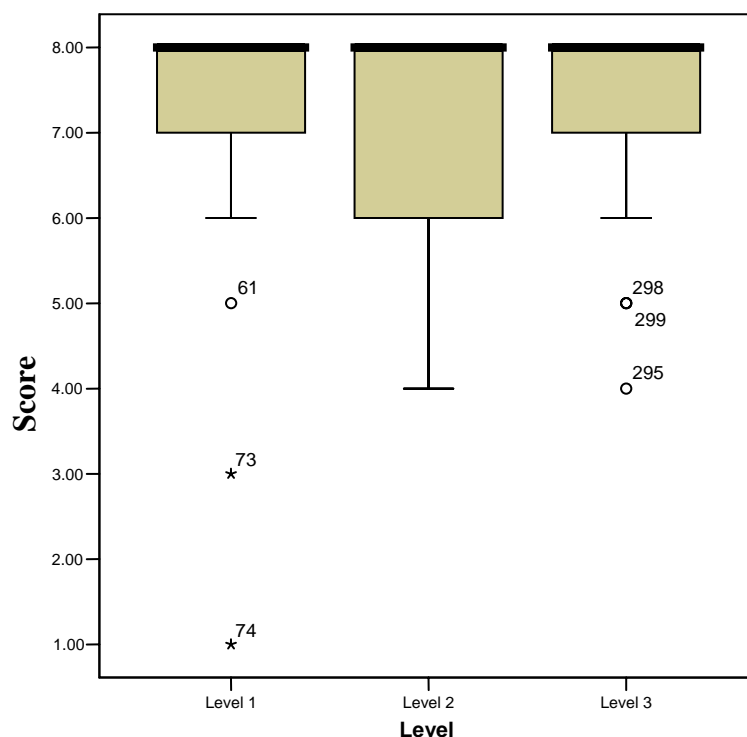
*Thematic Clustering (Immediate Tests)*

Table 58 shows mean number of accurately recalled words and standard deviation for each level on the immediate test of thematically clustered words. Data is grouped and displayed by level.

Table 58  
*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Thematic list – E-to-A)*

Level	N	Mean	SD
Level 1	30	7.03	1.63
Level 2	30	7.13	1.14
Level 3	20	7.25	1.25

Very small differences appear between the means in Table 58. Level 3 participants recalled more thematically clustered words than students in other levels. On the other hand, Level 1 participants recalled fewer thematically clustered words than other student.



*Figure 28*

Box-plot diagram of score in relation to level of participants (all levels – immediate test – thematic list – E-to-A).

Figure 28 shows that the median for number of words accurately recalled in the thematic cluster to be 8 for all levels. Fifty percent of all participants recalled eight words correctly from the thematic list on the immediate test. Variation is small in the inter-quartile ranges for Levels 1 and 3, while it is larger in that for Level 2.

Table 59 shows results of the ANOVA performed to compare means for immediate test results on thematically clustered words.

Table 59

*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Thematic list – E-to-A)*

#### ANOVA

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.567	2	.283	.151	.860
Within Groups	144.183	77	1.873		
Total	144.750	79			

As a result of the tiny differences between the means in Table 58, The ANOVA reports an  $F$  of 0.151, which with 2 and 77 degrees of freedom is not statistically significant ( $P > 0.05$ ). The high  $P$  value in Table 59 indicates that we accept the "Null Hypothesis" of no difference between the three levels in the immediate tests regarding thematically clustered words.

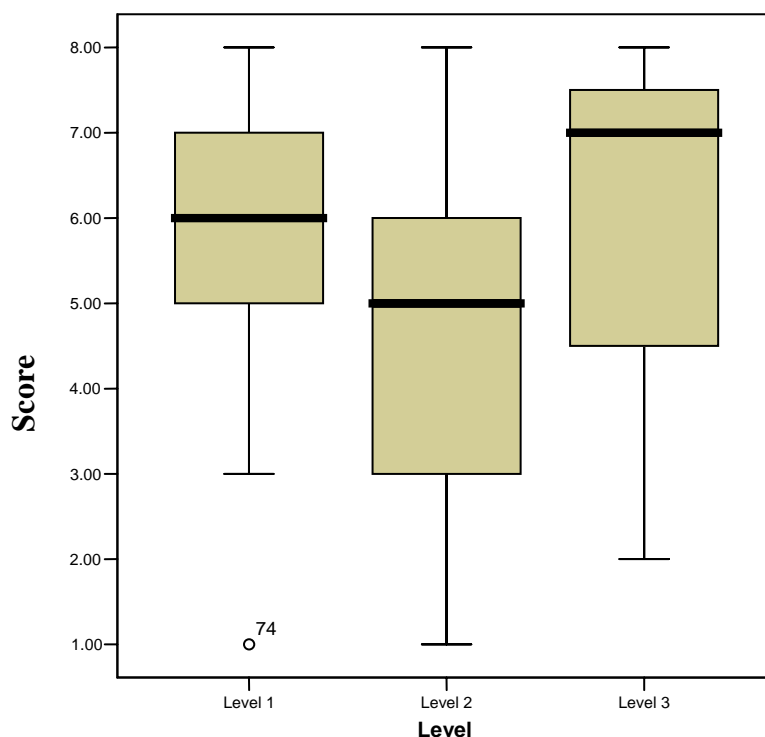
*Thematic Clustering (Delayed Tests)*

Table 60 reports means and standard deviations for the number of accurately recalled words from the thematic cluster, as measured on the delayed test. Results are grouped by level.

Table 60  
*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Thematic list – E-to-A)*

Level	N	Mean	SD
Level 1	30	5.83	1.88
Level 2	30	4.83	1.93
Level 3	20	5.90	2.00

Level 3 participants recalled more thematically clustered words than others in the delayed test as they did in the immediate test. What differs here is that Level 2 participants recalled fewer words than those in Level 1 participants. It seems that there is a big difference between the means for Level 1 participants and Level 2 participants. Whether this difference is significant appears in the following table.



*Figure 29*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – thematic list – E-to-A).

The whiskers for Level 1 and 2 data are more symmetric than the whisker for the Level 3 data in figure 26. The median of Level 3 is the highest while the median of Level 2 is the lowest. 50% of Level 3 participants recalled 7-8 words. The lowest number of recalled words is 0 by Level 2, 2 words for Level 3, and 3 words for Level 1.

Table 61 shows ANOVA results comparing mean scores for the above data, consisting of the number of thematically clustered words recalled correctly by participants at all levels on the delayed test.

Table 61  
*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Thematic list – E-to-A)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.817	2	9.908	2.666	.076
Within Groups	286.133	77	3.716		
Total	305.950	79			

Table 61 reports an F of 2.666, which with 2 and 77 degrees of freedom is not statistically significant ( $P > 0.05$ ). The "Null Hypothesis" of no difference between the three levels in the delayed tests regarding thematically clustered words is accepted.

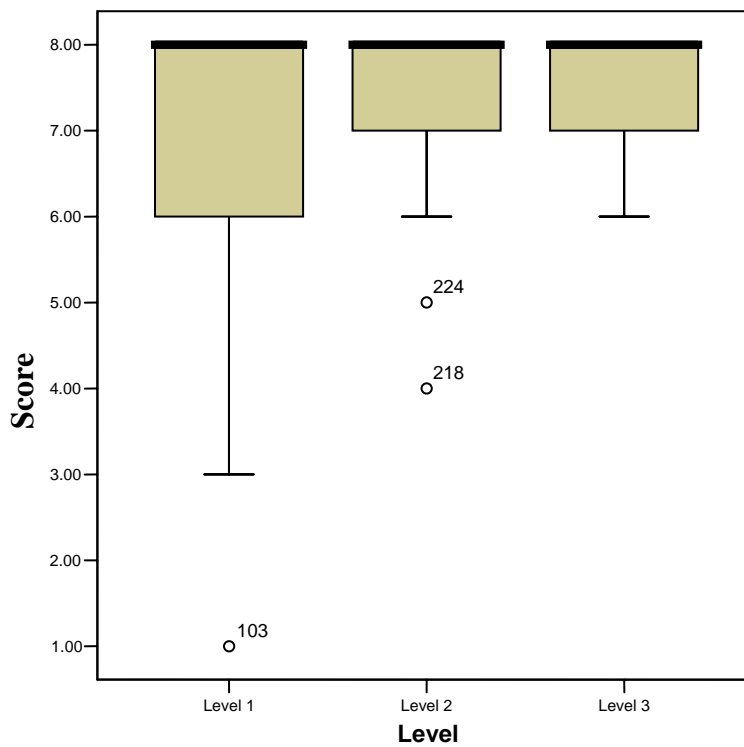
*Context Clustering (Immediate Tests)*

Table 62 reports means and standard deviations for the number of accurately recalled words from the contextualized cluster, as measured on the immediate test. Results are grouped by level.

Table 62  
*Means and Standard Deviation of Recalled Words (All Levels – Immediate test – Context list – E-to-A)*

Level	N	Mean	SD
Level 1	30	6.87	1.85
Level 2	30	7.40	1.10
Level 3	20	7.55	0.76

Table 62 shows that Level 3 participants recalled more contextual clustered words than other participants in the immediate test. Level 1 participants recalled fewer words than other participants. Small differences appear between the means. This might lead to insignificant difference in the ANOVA table.



*Figure 30*  
Box-plot diagram of score in relation to level of participants (all levels – immediate test – context list – E-to-A).

The figure above shows that 50% of all participants recalled 8 words from the context list, while the lowest number of recalled words by Level 2 and Level 3 participants is six words. All the whiskers are left skewed toward lower values.

The ANOVA in Table 63 compares means measuring the number of accurately recalled words from the immediate test on contextualized vocabulary words, grouped by participant level.

Table 63  
*Analysis of Variance of Number of Recalled Words (All Levels – Immediate test – Context list – E-to-A)*

**ANOVA**

Immediate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6.871	2	3.435	1.817	.169
Within Groups	145.617	77	1.891		
Total	152.487	79			



AN ANOVA, reported in Table 63, indicates that the differences among means yield an  $F(2, 77) = 1.817$ ,  $P = 0.169$  ( $P > 0.05$ ). The "Null Hypothesis" of no difference between the three levels in the immediate tests regarding contextually clustered words is accepted.

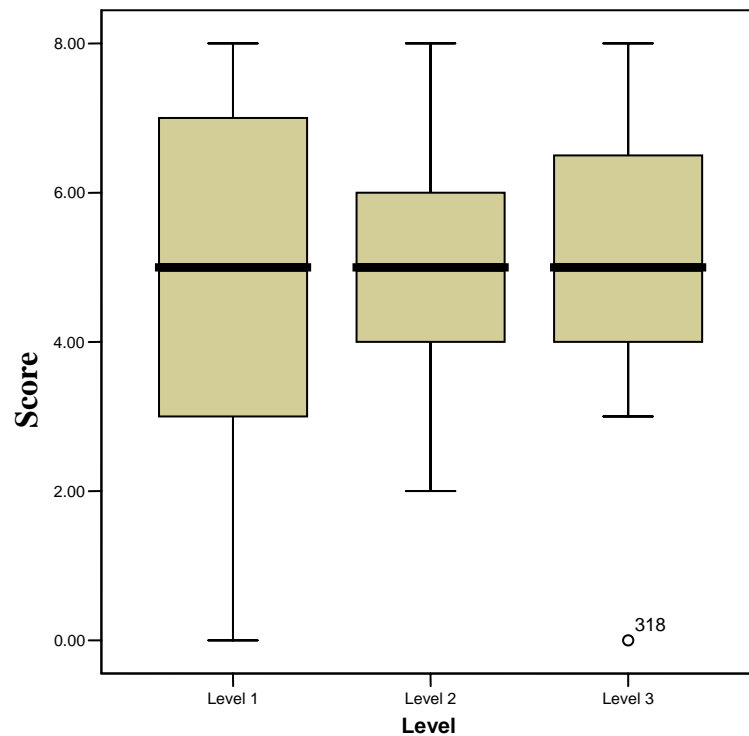
*Context Clustering (Delayed Tests)*

Table 64 displays mean and standard deviations for accurately recalled words in the contextualized cluster group, on the delayed test, with results grouped by level.

Table 64  
*Means and Standard Deviation of Recalled Words (All Levels – Delayed test – Context list – E-to-A)*

Level	N	Mean	SD
Level 1	30	5.00	2.32
Level 2	30	4.90	1.56
Level 3	20	5.15	2.03

Means in Table 64 are similar across levels. However, Level 3 participants recalled more words than others did this time, while Level 2 participants recalled fewer.



*Figure 31*

Box-plot diagram of score in relation to level of participants (all levels – delayed test – context list – E-to-A).

Figure 31 shows similar results. The median is the same for the three levels.

All the whiskers are symmetric. The highest number of recalled words for all three levels is eight. Variation is larger in Level 1 and smaller in Level 2.

The ANOVA in Table 65 shows results for comparison of mean scores for correctly recalled words on all context cluster delayed tests, grouped by level.

Table 65

*Analysis of Variance of Number of Recalled Words (All Levels – Delayed test – Context list – E-to-A)*

**ANOVA**

Delayed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.750	2	.375	.095	.910
Within Groups	305.250	77	3.964		
Total	306.000	79			

With  $F(2, 77) = 0.095$ ,  $P = 0.910$  ( $P > 0.05$ ). This result indicates that the difference is not statistically significant. Thus, we accept the "Null Hypothesis" of no difference between the three levels in the delayed tests regarding contextually clustered words.

### *Interaction Effects*

#### *Immediate Tests*

The descriptive statistics in Table 68 report mean number of words recalled accurately, standard deviation, and number of cases in each group, listed by cluster (condition), level, and direction. These statistics relate to the immediate tests.

Table 66

*Means and Standard Deviation of Recalled Words (Three Levels – Four conditions – Two Translation Directions – Immediate Tests)*

Descriptive Statistics						
Dependent Variable: Immediate						
Condition	Level	Direction	Mean	Std. Deviation	N	
Semantic	Level 1	A2E	4.9333	2.21178	30	
		E2A	5.7667	2.43088	30	
		Total	5.3500	2.34213	60	
	Level 2	A2E	6.0000	2.06782	30	
		E2A	6.5333	1.65536	30	
		Total	6.2667	1.87641	60	
	Level 3	A2E	5.9000	2.04939	20	
		E2A	5.9500	2.11449	20	
		Total	5.9250	2.05548	40	
	Total	A2E	5.5750	2.15110	80	
		E2A	6.1000	2.09036	80	
		Total	5.8375	2.13060	160	
Unrelated	Level 1	A2E	5.0000	2.58644	30	
		E2A	6.9667	1.29943	30	
		Total	5.9833	2.25863	60	
	Level 2	A2E	6.5333	1.83328	30	
		E2A	7.1333	1.52527	30	
		Total	6.8333	1.69912	60	
	Level 3	A2E	6.6500	1.98083	20	
		E2A	7.4000	1.27321	20	
		Total	7.0250	1.68686	40	
	Total	A2E		5.9875	2.28641	80

		E2A	7.1375	1.37559	80
		Total	6.5625	1.96730	160
Thematic	Level 1	A2E	7.0333	1.32570	30
		E2A	7.0333	1.62912	30
		Total	7.0333	1.47254	60
	Level 2	A2E	7.1000	1.09387	30
		E2A	7.1333	1.13664	30
		Total	7.1167	1.10610	60
	Level 3	A2E	7.0500	2.16370	20
		E2A	7.2500	1.25132	20
		Total	7.1500	1.74753	40
	Total	A2E	7.0625	1.48702	80
		E2A	7.1250	1.35362	80
		Total	7.0937	1.41775	160
Context	Level 1	A2E	5.6333	2.35597	30
		E2A	6.8667	1.85199	30
		Total	6.2500	2.19108	60
	Level 2	A2E	5.7333	1.68018	30
		E2A	7.4000	1.10172	30
		Total	6.5667	1.64024	60
	Level 3	A2E	6.5000	1.60591	20
		E2A	7.5500	.75915	20
		Total	7.0250	1.34903	40
	Total	A2E	5.8875	1.95515	80
		E2A	7.2375	1.38932	80
		Total	6.5625	1.82121	160
Total	Level 1	A2E	5.6500	2.30691	120
		E2A	6.6583	1.89868	120
		Total	6.1542	2.16794	240
	Level 2	A2E	6.3417	1.76567	120
		E2A	7.0500	1.39537	120
		Total	6.6958	1.62717	240
	Level 3	A2E	6.5250	1.96794	80
		E2A	7.0375	1.54628	80
		Total	6.7813	1.78277	160
	Total	A2E	6.1281	2.06155	320
		E2A	6.9000	1.64269	320
		Total	6.5141	1.90209	640

Table 67 shows the interaction between the variables; condition, Level, and direction of translation in the immediate tests.

Table 67  
*Three-way ANOVA Model for Immediate Tests*

**Tests of Between-Subjects Effects**

Dependent Variable: Immediate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	315.174(a)	9	35.019	11.049	.000
Intercept	26426.467	1	26426.467	8338.097	.000
Condition	127.755	3	42.585	13.436	.000
Level	50.438	2	25.219	7.957	.000
Direction	95.327	1	95.327	30.078	.000
Condition * Direction	41.655	3	13.885	4.381	.005
Error	1996.699	630	3.169		
Total	29469.000	640			
Corrected Total	2311.873	639			

a R Squared = .136 (Adjusted R Squared = .124)

Table 67 above shows that there is significant interaction between condition and direction of translation in the immediate tests. The P value is 0.005 ( $P < 0.05$ ). No other significant interaction exists between condition and level or level and direction of translation.

Table 68 reports the post hoc tests which compare mean scores by Level for the immediate tests.

Table 68  
*Post Hoc Tests (All Levels – Immediate tests – Two Translation Directions)*

**Multiple Comparisons**

Dependent Variable: Immediate

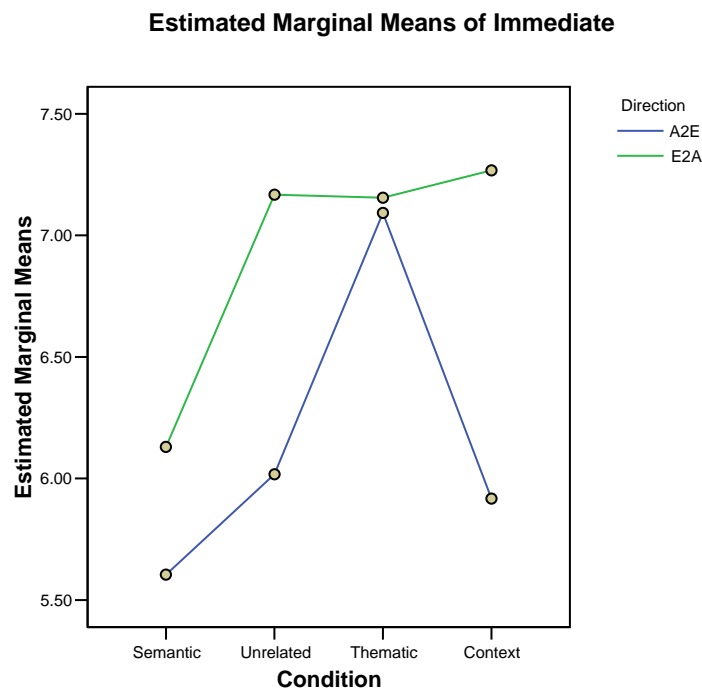
LSD

(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Level 1	Level 2	-.5417(*)	.16252	.001	-.8608	-.2225
	Level 3	-.6271(*)	.18170	.001	-.9839	-.2703
Level 2	Level 1	.5417(*)	.16252	.001	.2225	.8608
	Level 3	-.0854	.18170	.638	-.4422	.2714
Level 3	Level 1	.6271(*)	.18170	.001	.2703	.9839
	Level 2	.0854	.18170	.638	-.2714	.4422

Based on observed means.

\* The mean difference is significant at the .05 level.

Table 68 above compares the three levels based on the number of recalled words from the four lists with the two directions of translation for the immediate tests. The mean of Level 1 is (6.15), that of Level 2 is (6.69), and that of Level 3 is (6.78), as Table 66 shows. The difference between Level 1 and Level 2 results is statistically significant ( $P = 0.001$ ). The difference between Level 1 and Level 3 is also statistically significant ( $P = 0.001$ ). Since the means for Level 2 and Level 3 are close, the difference is not statistically significant ( $P = 0.638$ ).



*Figure 32*  
Interaction plot.

Figure 32 shows that the two lines are not parallel and this means that there is an interaction between Condition and Direction of Translation. Arabic-to-English translation is always lower than English-to-Arabic, except with the Thematic condition in which both directions of translation are almost the same. The lowest means in both directions of translation occurred with the semantic condition.

*Delayed Tests*

The descriptive statistics in Table 71 report mean number of words recalled accurately, standard deviation, and number of cases in each group, listed by cluster (condition), level, and direction. These statistics relate to the delayed tests.

Table 69

*Means and Standard Deviation of Recalled Words (Three Levels – Four conditions – Two Translation Directions – Delayed Tests)*

<b>Descriptive Statistics</b>						
Dependent Variable: Delayed						
Condition	Level	Direction	Mean	Std. Deviation	N	
Semantic	Level 1	A2E	1.5333	1.83328	30	
		E2A	3.8667	2.23966	30	
		Total	2.7000	2.34557	60	
	Level 2	A2E	1.8333	1.72374	30	
		E2A	4.3000	2.07032	30	
		Total	3.0667	2.26144	60	
	Level 3	A2E	2.1000	1.99737	20	
		E2A	4.2000	2.68720	20	
		Total	3.1500	2.56755	40	
	Total	A2E	1.7875	1.82593	80	
		E2A	4.1125	2.27809	80	
		Total	2.9500	2.36537	160	
	Unrelated	Level 1	A2E	1.7667	1.92414	30
			E2A	5.0333	1.95613	30
			Total	3.4000	2.53250	60
Level 2		A2E	2.4000	2.22215	30	
		E2A	6.2333	1.83234	30	
		Total	4.3167	2.79522	60	
Level 3		A2E	3.0000	1.91943	20	
		E2A	6.3000	1.92217	20	
		Total	4.6500	2.52729	40	
Total		A2E	2.3125	2.07208	80	
		E2A	5.8000	1.97067	80	
		Total	4.0563	2.66882	160	
Thematic		Level 1	A2E	2.6667	1.53877	30
			E2A	5.8333	1.87696	30
			Total	4.2500	2.33343	60
	Level 2	A2E	2.5333	1.92503	30	
		E2A	4.8333	1.93129	30	
		Total	3.6833	2.23600	60	
	Level 3	A2E	3.3500	2.47673	20	
		E2A	5.9000	1.99737	20	
		Total	4.6250	2.56892	40	

Context	Total	A2E	2.7875	1.95321	80
		E2A	5.4750	1.96794	80
		Total	4.1312	2.37418	160
	Level 1	A2E	1.9667	1.75152	30
		E2A	5.0000	2.31933	30
		Total	3.4833	2.54779	60
	Level 2	A2E	1.6000	1.61031	30
		E2A	4.9000	1.56139	30
		Total	3.2500	2.28944	60
	Level 3	A2E	2.3000	2.49420	20
		E2A	5.1500	2.03328	20
		Total	3.7250	2.66975	40
Total	Total	A2E	1.9125	1.91062	80
		E2A	5.0000	1.96810	80
		Total	3.4563	2.47718	160
	Level 1	A2E	1.9833	1.79628	120
		E2A	4.9333	2.19523	120
		Total	3.4583	2.48812	240
	Level 2	A2E	2.0917	1.90089	120
		E2A	5.0667	1.96923	120
		Total	3.5792	2.43964	240
	Level 3	A2E	2.6875	2.25351	80
		E2A	5.3875	2.28641	80
		Total	4.0375	2.63715	160
Total	A2E	2.2000	1.97254	320	
	E2A	5.0969	2.13788	320	
	Total	3.6484	2.51502	640	

Table 70 compares condition, level, and direction of translation at the delayed tests.

Table 70  
*Three-way ANOVA model for Delayed tests*

**Tests of Between-Subjects Effects**

Dependent Variable: Delayed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1524.613(a)	6	254.102	63.897	.000
Intercept	8410.671	1	8410.671	2114.959	.000
Condition	147.867	3	49.289	12.394	.000
Level	34.044	2	17.022	4.280	.014
Direction	1342.702	1	1342.702	337.638	.000
Error	2517.285	633	3.977		
Total	12561.000	640			
Corrected Total	4041.898	639			

a R Squared = .377 (Adjusted R Squared = .371)



The table above indicates that there is no interaction at all between any two of the three variables; condition, level, and direction of translation.

The post hoc tests reported in Table 71 compare mean scores by cluster group for the delayed tests.

Table 71

*Post Hoc Tests (All Conditions – Delayed tests – Two Translation Directions)*

**Multiple Comparisons**

Dependent Variable: Delayed  
LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Semantic	Unrelated	-1.1063(*)	.22296	.000	-1.5441	-.6684
	Thematic	-1.1812(*)	.22296	.000	-1.6191	-.7434
	Context	-.5062(*)	.22296	.024	-.9441	-.0684
Unrelated	Semantic	1.1063(*)	.22296	.000	.6684	1.5441
	Thematic	-.0750	.22296	.737	-.5128	.3628
	Context	.6000(*)	.22296	.007	.1622	1.0378
Thematic	Semantic	1.1812(*)	.22296	.000	.7434	1.6191
	Unrelated	.0750	.22296	.737	-.3628	.5128
	Context	.6750(*)	.22296	.003	.2372	1.1128
Context	Semantic	.5062(*)	.22296	.024	.0684	.9441
	Unrelated	-.6000(*)	.22296	.007	-1.0378	-.1622
	Thematic	-.6750(*)	.22296	.003	-1.1128	-.2372

Based on observed means.

\* The mean difference is significant at the .05 level.

Table 71 compares the four conditions based on the number of recalled words by all participants with the two directions of translation at the delayed tests. The means for the semantic, unrelated, thematic, and context lists are (2.95), (4.06), (4.13), and (3.46) respectively. Comparisons show that the difference between the semantic and unrelated lists is statistically significant with a P value of 0.00. The difference between the semantic and thematic lists is also statistically significant and the P value is also 0.00. Moreover, there is a statistically significant difference between the means for the semantic and context lists since the P value is 0.024. As far

as the unrelated list is concerned, there is a statistically significant difference with a P value of 0.007 when we compare it to the context list. Another statistically significant difference exists between the thematic and context lists with a P value of 0.003.

Finally, the only insignificant difference occurs when we compare the thematic and unrelated lists.

Table 72 reports the data described in Table 71 above, in this case for Levels.

Table 72

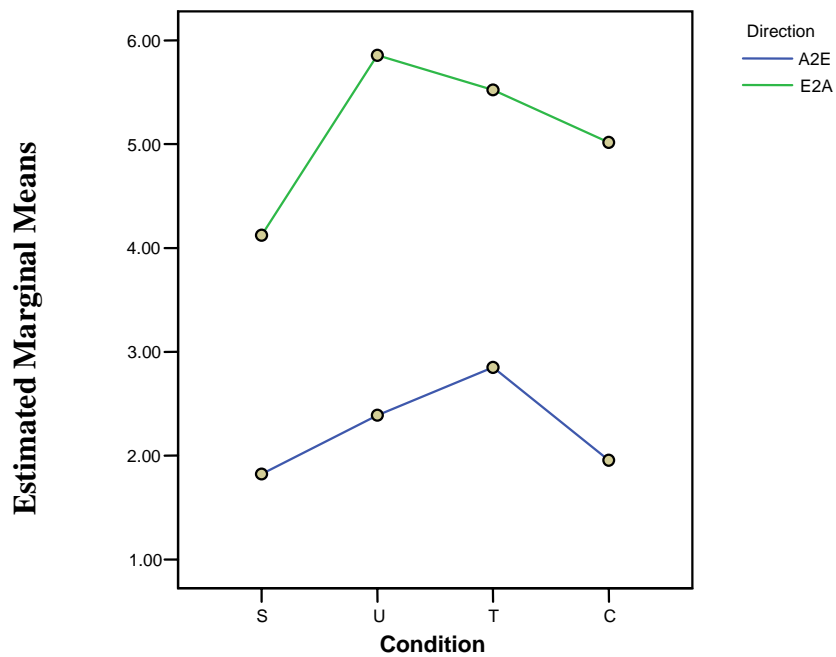
*Post Hoc Tests (All Levels – Delayed tests – Two Translation Directions)*

<b>Multiple Comparisons</b>						
Dependent Variable: Delayed						
LSD						
(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Level 1	Level 2	-.1208	.18204	.507	-.4783	.2366
	Level 3	-.5792(*)	.20353	.005	-.9788	-.1795
Level 2	Level 1	.1208	.18204	.507	-.2366	.4783
	Level 3	-.4583(*)	.20353	.025	-.8580	-.0587
Level 3	Level 1	.5792(*)	.20353	.005	.1795	.9788
	Level 2	.4583(*)	.20353	.025	.0587	.8580

Based on observed means.

\* The mean difference is significant at the .05 level.

From Table 69, the total mean of Level 1 is 3.46, the total mean of Level 2 is 3.58, and the total mean of level 3 is 4.04. The difference between the means for Level 1 and Level 3 is statistically significant since the P value is 0.005. Also, the difference between the means for Level 2 and Level 3 is statistically significant with a P value of 0.025. But there is no statistically significant difference between the means for Level 1 and Level 2.



*Figure 33*  
Interaction plot.

The figure above shows that there is no interaction between condition and directions of translation. Though not fully parallel, the same gap exists between each of the two points of translation direction within each condition. Words from the semantic list were the fewest to be recalled among other words in both directions of translation. Words from the thematic list were the highest in the Arabic-to-English translation, while words from the unrelated list were the highest in the English-to-Arabic translation.

## Participants' Reflection

### *Findings from the Interviews*

As mentioned earlier in chapter 3, students who scored the highest and lowest in the Arabic-to-English tests were questioned briefly. This section reports on their responses to questions focused on their vocabulary learning strategies and experience, as well as reactions to the experimental test they had just completed. For convenience, I will refer to these participants below simply by this criterion. So, the participant who scored highest in the semantic condition in Level I will be referred to as 'Level 1 Highest Semantic,' or simply 'Highest Semantic' in a context where Level 1 is understood.

One of the basic questions I asked is "How do you learn English vocabulary?" In response to this question, the technique of repetition was virtually unanimous, and is claimed by all participants, both in the Highest and Lowest categories. However, a pattern does emerge, in that the more successful, Highest participants more often elaborate on additional strategies.

All four participants from Level 1 claim that their major technique they use to learn English vocabulary is repetition. For instance, Highest Thematic said that he repeats words many times. Then he hides one side of the list and tries to guess the meaning. Once he has done this, he hides the other side and repeats the process again. Thus, he combines repetition with a sort of self-test procedure. Highest Semantic says "I repeat the first two words of the list. Then I revise them a couple of times until I have memorized them well. Then I move to the next two words and apply the same procedure." However, he adds that he uses another secondary technique. He makes connections with words whose pronunciation is similar to the target word. For example, the word "aster" means "زهرة النجمة" and the Arabic word "نجمة" means

"star" which sounds like "aster". This is not unlike the so-called 'key word' strategy recommended by some second language researchers.

In contrast, Lowest Semantic claims that he uses repetition only. Justifications for using repetition as a technique for learning English vocabulary were provided by Lowest Thematic, who feels that repetition is easier and faster than any other technique. Admittedly, this participant, in addition to repetition, claims he connects the first letter of the English word with the Arabic meaning. For example, he connected "T" with "الخرامى" and "C" with "زعفران" and so on.

Repetition is also the main technique reported by Level 2 participants. Highest Semantic and Lowest Thematic in this group both claim that this is the only technique they follow. Lowest Semantic claims that he repeats words many times and writes them down. He says that he does not put words in sentences because sentences need correct grammar and spelling and, therefore, if he were to contextualize the words, his focus would no longer be on the meaning of the individual word. It is interesting here that the participant specifically argues against a strategy that is considered useful for learning, and is in fact at the heart of the considerations that motivated this study.

However, Highest Thematic elaborates considerably beyond the idea of simple repetition: "First of all, I focus on the spelling of each word. Then I focus on the pronunciation, and then the meanings. After that, I repeat each word a couple of times. Also, I use each word in a sentence of my own. If the new words appear in a context, I try to put them in sentences based on the meaning I figure out from the context".

Level 3 participants yield a similar pattern. Highest Semantic claims that he uses the same procedure as his Level 1 colleague. He repeats the first two words in the list and then reviews them. Once he feels he has mastered them, he moves to the

following two words and does the same. In addition, he connects words if pronunciations of the English and Arabic words are similar. Lowest Semantic says that he writes words many times and repeats them as well. Also he focuses on some letter groups in words such as "ea". Lowest Thematic mentions that he pays close attention to spelling, and again depends on repetition as his main learning technique. Highest Thematic, however, mentions a new technique: he uses imagination to help him learn vocabulary. He says, "I use imagination based on the first letter of the English word and its referent. For example, "Dory" which means "boat", I imagine the letter "D" as a boat. Whenever I encounter the word "Dory" I remember the image I formed and get the meaning easily. Also I write English words many times to capture the spelling".

Participants were asked which list was easier to them. Their answers to this question are more mixed, suggesting that it is not easy for them to comment coherently on this subtle difference in the list organization. While a few answers do address the differences, only some of these clearly prefer the thematic presentation for new words.

Level 1 Highest Semantic says that there was no difference between the two lists, while Lowest Semantic says that the unrelated word list was easier to learn than the semantic list. Highest Thematic says that the thematic list was easier than the context list and this was supported by Lowest Thematic who justified his answer as follows, "[reel] and [cast] are not deep. They have simple spelling and pronunciation. Moreover, they are about one single topic". This participant used the phrase "not deep" to mean that words are monosyllabic and easy to pronounce.

Level 2 Highest Semantic says that there is no difference between the unrelated word list and the semantic list; however, I noticed that he recalled more

words from the unrelated list than words from the semantic list. Lowest Semantic says that the semantic list was easier to learn, but he hints at the difficulty of semantic grouping when he says that the words were "confusing" in this list because they are similar. Highest Thematic says that both lists, the thematic list and the context list, were equal and he noticed no difference; however, he does claim that he preferred the context list. Lowest Thematic says that the context list was easier to learn because words of the thematic list looked strange to him.

Level 3 Highest Semantic says that the semantic list was easier to learn, but he gives no particular reason. Lowest Semantic says he found the unrelated list easier to learn because words were not close to each other. Highest Thematic says that the thematic list was easier because the words were in a list form and they were more connected. Lowest Thematic also claims to have found the thematic list easier because words are listed and talked about one topic. This particular contrast suggests that an appreciation for presentation grouping alone does not directly lead to success, but must be accompanied by strategies that reinforce and make use of the thematic relatedness of the items being learned.

Another question that participants answered addressed their preferences more generally, asking which types of vocabulary are difficult for them to learn. The answer to this question, despite potential priming from the previous question, entirely failed to address the presentation grouping of vocabulary items. Instead, all responses to this question aim at the characteristics of individual words (which may be long, unusual, hard to pronounce or spell), rather than grouping. This does not negate the effect of grouping, but suggests that learners may not be consciously aware of the way words are grouped, and may have no idea that grouping might affect their success in learning.

Level 1 Highest Semantic and Lowest Thematic tests say that there is no particular type of vocabulary that they have problem with. Highest Thematic say that long words are difficult to learn. He adds that words with difficult pronunciation are also hard to learn. Lowest Semantic test echoes the problem with long words, when he stats that words with more than one syllable are difficult for him. He says "the more syllables a word has, the more difficult it becomes to learn".

Level 2 participants provided very similar answers. Highest Semantic and Thematic say that there is no particular type that causes them learning problems. In contrast, Lowest Semantic says that he has problems with long words (four syllables and more) and words contain (z, x). Lowest Thematic said that he has problems with words with strange (uncommon) pronunciation and long words of four or more syllables.

Long words are difficult to learn for Level 3 participants, though their concept of 'long' may have adjusted as their English proficiency increased. Highest Semantic said that long words (17 letters!!) and medical terms are difficult to learn. Highest Thematic said that words with three or more syllables are difficult. Since this participant depends on the shape of words to learn them, he claims to have problems learning words with a "strange shape." Long words with three syllables or more, scientific words, and words with many vowels are all problematic to Lowest Semantic, while uncommon words (not used frequently) are problematic to Lowest Thematic.

Participants who had lists in the context condition were asked whether they had read the context. Their responses are quite revealing. Level 1 Highest says, "I read half of the context, because time was not enough" while Lowest said "I have not read it. It is a long context and there was not enough time". Level 2 Highest said "I



read it. Reading the context helps". Recall that this participant puts new words in sentences as a strategy for learning English vocabulary. Lowest said "Yes, but not the whole context". Level 3 participants gave similar answers. Highest score says, "I read some sentences, like two or three". I asked him about the topic but he gave no answer. He said that he did not pay attention to the context, which was confirmed by his inability to say what the contextual story had been about. Lowest said "No. I tried to read it but could not understand it".

This is really a problem. Participants paid either minimal attention or no attention at all, to the context, which in most cases they did not even try to read. Insufficient time does not constitute a good excuse for this, as these participants were given extra time, specifically with the idea that they could then process the story. As a matter of fact, there is evidence that most participants ignored the contextual cues, from the actual test papers, and from my informal observation of them as they learned the words. Since new words were underlined and their meanings were provided, participants had the option of simply turning the new words into an unrelated list. I noticed that most of them wrote the English words with the Arabic equivalents in a list form below the context. Some wrote the list many times, indicating a repetition technique. Therefore, it is best to view the results of groups who had words in context should be basically equivalent to those for unrelated lists. This was an unexpected learning strategy, and it again underlines the importance of simple rote memorization for these students.

When asked about procedures teachers provided them with, participants agreed unanimously that they had never been provided with any techniques. On the contrary, they said that their teachers used to write new words in a list form on the board and ask students to write and memorize them.

### *Participants' Response Patterns*

Around 1280 sheets were analyzed, in an attempt to find mistakes and categorize them according to their type. There were many phonological, semantic, translation, and target language mistakes. The types of mistakes occur in participants' sheets regardless of their level. The quantitative analysis concentrated only on whether participants' answers were right or wrong. The purpose of this section is to examine participants' answers more closely to find patterns in the errors they made.

#### *Level 1 (Phonological Errors)*

Most of the phonological or spelling errors seem to involve vowels; though an occasional form involved inversion (One Level 1 participant wrote the word "donies" instead of "donsie", while another participant interchanged syllable-initial consonant sounds to modify "Lure" to "Rule", which is a real word in English.).

One participant wrote "pampano" instead of "pompano". In another example, a participant raised the vowel so the word "aster" became "astir". On the other hand, the vowel was lowered by another participant so that the word "lily" became "lele".

One participant committed three phonological mistakes, all again involving vowels. The first one is lowering the vowel from "Lure" to "Lore". The second and third ones are deleting vowels: "leister" and "shoal" to "lister" and "shol". Some participants added vowels to the end of words such as "daffodile".

Another participant made two phonological changes as well as the vowel change to the word "pompano". In addition to fronting the vowel "o" to "a", he seems to have inserted a regressive place of assimilation process so that the sound "m" became "n"; finally, he voiced the bilabial consonant, so that "p" became "b". This

last change may have been the result of assimilation, or of interference, as Arabic has no voiceless [p] sound. The final shape of the word became "banbano".

The persistent problems with vowels are interesting, as the students' first language, Arabic, does not mark short vowels; thus, as learners, these students may need training to 'see' and actively process the vocalic shape of words in a language like English.

#### *Level 1 (Semantic Errors)*

Semantic errors refer to errors participants make either by providing wrong meaning or writing the meaning of another word instead. One of the most common and frequent errors occurring in participants' answer sheets is that they confused the two closely related meanings "cast" and "angling". Although these words occur in the thematic clustering list, they might be considered semantically related, as they both refer to similar actions; thus, they may reflect a flaw in the study's design, which is in turn reflected in the participants' problems with these forms. Having these words in a thematic list caused problems to participants who wrote the meaning of one beside the other. Some participants provided one and missed the other. A similar common error occurs in the semantic clustering list in which participants also confused the words "pansy" and "daisy". When asked to write the English word, one participant tried his best and wrote four words, the fourth of which was the correct one. He wrote "dan", "daisy", "dinsy", and "Pansy". The same participant, in his struggle to recall "Daisy", wrote "pansy" and "disy". Another participant wrote "spoeny" instead of "daisy". A third participant wrote "dansy" instead of "daisy". A fourth one wrote "dansy" instead of "pansy" and "pisay" instead of "daisy". I believe that what causes this problem is that the two words end with the same second syllable /zi/, combined with the similarity in their meanings as flower names.

Some students made errors with the word "aster", possibly influenced by the fact that the Arabic equivalent "زهرة النجمة" contains the word "star". Probably with this Arabic form in mind, one participant wrote "star" while another wrote "stair". Another showed phonological interference from the target language when he wrote the familiar word "wheel" instead of the target term "Reel".

#### *Level 1 (Translation Errors)*

This kind of error occurred very frequently whenever the participant could not remember the exact word needed. As a strategy participants followed in these cases, they seemed to go back to their mother tongue to try to find an existing English equivalent word to the word they need to access. Examples are many and are found in many papers. For example, with the context list, a participant wrote "cool" and "sea" instead of "airish" and "bayou" respectively. In the unrelated clustering list, one participant wrote "mixture" instead of "bollix". Another one wrote "soft", "mix", "basket", "boat", and "bug" instead of "doty", "bollix", "hamper", "bayou", and "pismire" respectively. Another participant wrote "softy" instead of "doty". The form "ant" instead of "pismire" occurred in one participant's paper. One participant wrote "extremist" instead of "capsheaf".

Of course, the attempt to find a paraphrase or equivalent form is a well-known communication strategy for second language learners, and its use here is quite natural, though it raises unanswered questions about possible 'blocking' of new forms when the learner already has a closely related word in his vocabulary.

*Level 2 (Phonological and Spelling Errors)*

Level 2 participants also experienced voicing confusion. ". One participant voiced a voiceless sound and the result was "pombano"; note, again, that this error is predictable, since Arabic does not have the voiceless bilabial stop [p]. . However, at level 2, participants actually also reversed the process, and engaged in what might be seen as 'overcorrection', when they produced [p] instead of [b]. Participants wrote "patean" instead of "batean", "pollix" instead of "Bollix", and "poilex" instead of "bollix

Level 2 participants also made a different set of vowel-related errors. For example, one changed the vowel and wrote "pismer" instead of "pismire" and "ires" instead of "iris"; another wrote "capshef" instead of "capsheaf". It is interesting that the vowels continue to pose problems for these learners, though it is hard to see any way in which the level may have influenced the particular vowels that were problematic.

One Level 2 participant tried seven times to write "crocus" correctly: [coreu – coracu – corouce – corecse – cor – coroucs]. He finally settled on "crouces". The Arabic morphophonemic system comes to mind here; note that this participant's first six tries all inserted a vowel between the two initial consonants in "crocus."

One participant confused the doubling of letters as well as changing vowels; the result was "dafidill" instead of "daffodil". The doubling of letters in English spelling is another source of confusion for Arabic speakers. When a consonant is doubled in Arabic, its pronunciation reflects the doubling, as the closure for this consonant is held longer than for single consonants. As this is no longer true for English consonants, learners may generally be confused about the meaning of the double consonants they must learn only for spelling purposes in English.

### *Level 2 (Semantic Errors)*

As in level 1, many students confused items in the semantic list, and "pansy" and "daisy" in particular, regardless of the direction of translation. One participant provided "pansy" correctly but for "daisy" he wrote "dancy" which has the first syllable of "pansy" and the second from "daisy". Another one wrote "dansy" and "pansy", allowing the first syllable of "pansy" to serve for both forms. One participant, violating syllable structure, took the first part from "daisy" and the second part from "pansy" to form "dainsy". Two participants switched the two Arabic equivalents in the immediate test; but in the delayed test, one of them wrote the correct Arabic equivalent for each word. Another participant switched the first syllable of each word, so the results were "paisy" and "dansy".

The same confusion happened at Level 2 as was reported for Level 1, involving the closely related forms "Cast" and "Angling". Some participants kept switching the two words regardless of the translation direction. One participant switched the two words in the delayed test, though he had listed them correctly in the immediate test.

In recalling the Arabic equivalent of "aster", some Level 2 participants wrote "star" instead, a pattern already noted for Level 1 participants in the previous section.

### *Level 2 (Translation Errors)*

As explained above, the incorrect responses I call 'translation errors' involve the participants' falling back on semantically related words already in their vocabulary. These include writing "Hole" and "Bug" instead of "Pothole" and "Pismire". As in the first group, a number of participants wrote "Soft" and "blend" instead of "Doty" and "Bollix". One participant wrote "Hole in the road" and "Boat"

instead of "Pothole" and "Bayou". A third one wrote "cold" and "small river" instead of "airish" and "bayou". Another one wrote "small stones" and "pollution" instead of "dornickets" and "gaummy". The word "Ant" was written by two participants for "Pismire".

Possibly as a quite different kind of target language effect, one participant added the suffix [al] to a word to get "portical". Another one seems to have tried to write the name of a territory instead of "portico". He rendered the word as "portorico".

### *Level 3 (Phonological Errors)*

The sound and letter related errors for Level 3 participants were similar to those found in Level 1. Once again, consonantal errors had to do with voicing. One participant wrote "bunbanian" instead of "pompoan", thus reverting to the familiar voicing of the bilabial stop, which must be voiced in Arabic. In addition to the voicing change, this participant also switched [m] to [n] (reverse assimilation), and then add the suffix [ian] to the end of the word, presumably showing a rare effect from the target language morphology.

Several examples involved the devoicing of target [b], which may again result from overcorrection. One participant, for example, wrote "patean" instead of "batean" while another wrote "payout" instead of "bayou". Another was confused between [d] and [b], which may relate to residual problems with these similar letter shapes. As a result, he wrote "dayou" instead of "bayou".

As before, the vocalic structure of new words continues to cause problems for Level 3 learners. Some participants added vowels, in a pattern reminiscent of the earlier examples, possibly an unconscious strategy for breaking up consonant clusters. For example, one wrote "pansy" as "panasy". Another participant added a vowel and

deleted a final silent e, so that the word "pisamir" was produced instead of "Pismire".

Others wrote "donsie" as "donise" or "donis".

Oddly enough, at this level, some participants changed the position of the vowel in ways that actually produce consonant clusters. For example, one participant wrote "croucs" instead of "crocus". Another produced a cluster that most English speakers would find unpronounceable when he wrote "boillx" instead of "bollix".

As before, consonants are occasionally switched. For example, a participant wrote "satus" instead of "rastus".

### *Level 3 (Semantic Errors)*

Like the learners in Level 1 and 3, some participants at this level made errors involving "pansy" and "daisy". This problem is not restricted to the Arabic-to-English direction only. It affected the English-to Arabic responses as well. One participant wrote the Arabic meaning of "pansy" as the meaning of "daisy" and vice versa.

The two semantically related words within the thematic list also once again confused participants at this level. This type of error occurred when participants were asked to write the Arabic equivalents to the English words. One participant wrote the Arabic meaning of "cast" as the meaning of the word "angling" and vice versa. Two other participants wrote the correct Arabic equivalent for "cast"; but they considered "angling" as a verb, so the words they provided in Arabic for this item were verbs. One participant, when asked to provide the English equivalent for the first term, wrote "casting" instead of "cast"; then, possibly because he had already 'used' the morphological *-ing* form, he did not write anything for "angling". Another one wrote "angling" and "casting". A third one wrote "cast" as the English equivalents to both



Arabic words. Clearly, this confusion in morphological form must be mediated by the similar meanings of the two terms.

As with the other groups, many participants provided real words they already know that just happen to sound like to the correct words. Some wrote "wheel" instead of "reel" and others wrote "layer" instead of "lure". This is interesting, because phonological confusion of this type is commonly believed to occur mainly in the early stages of language learning; here, we see phonological similarity determining responses with quite advanced language learners.

### *Level 3 (Translation Errors)*

As the case with phonological and semantic errors, almost the same translation errors are made by all participants regardless of their level. When they could not remember the exact English words, participants tended to depend on their mental lexicon and look for words with the same meanings. Instead of the word "doty", one wrote "calm" while another wrote "delicate". The third one wrote "Smooth" while two others wrote "soft" and "softy" respectively. One participant wrote "cold" instead of "airish". Another one wrote "shallow" instead of "shoal". A third one wrote "Small Rocks" instead of "dornickets" and this means that he translated the Arabic word literally. A fourth one wrote "fishing" instead of "angling". It is interesting to find such frequent falling back on familiar near-synonyms at this advanced level, as it may suggest resistance to learning a new term when a well-established, semantically related word already exists in the learner's mental lexicon.

### *Discussion*

This part talks about the major findings of tests and questions participants answered. It includes themes about thematic and semantic clusterings, backward and forward translations, and use of context.

#### *Semantic Clustering Vs. Thematic Clustering*

The results of the current study may be surprising to teachers and course designers. Contrary to the idea that grouping related vocabulary items facilitate learning; the results indicated that grouping vocabulary items that share semantic and syntactic characteristics impedes learning and have a negative effect on learning. Participants could recall more words from the list that shared a thematic concept than from the semantic list. These results are not surprising to those who did research on the Interference theory and the distinctiveness hypothesis. The results support what research on Schema theory came up with and support results of similar studies such as Stahl et. al. (1992), Tinkham (1993), and Waring (1997).

The thematic clustering showed better results at the immediate tests than at the delayed test. This could be explained by the forgetting curve. It seems that the amount of forgetting was rapid during the first week so that the means for the recalled words from the four lists were close and the differences were not statistically significant. The reason for this rapid drop-off is that participants did not review the lists between the immediate and delayed tests. But since the thematic clustering has proven effective in the immediate tests, learners can use this method of grouping vocabulary to learn new vocabulary in cases that they need words for a short time or for an immediate purpose. Frequent use of this method in learning English vocabulary might lead to better results over a long time.

### *Backward and Forward Translations*

The findings of English-to-Arabic tests are similar to the findings of Arabic-to-English tests in that grouping vocabulary items according to semantic and syntactic characteristics is a detriment to learning with the two translation directions. Therefore, this method has a negative effect on learning vocabulary items.

Results showed that participants performed better in backward translation (L2-L1) than in forward translation (L1-L2). These results support the results of Prince (1996), Kroll & Curley, (1986), Kroll & Stewart, (1989), and Kroll & Stewart (1994). The reason for this result is provided by Kroll and Stewart (1994) who claim that forward translation takes longer to perform because it requires concept mediation and influence by the presence of semantic context. Backward translation, on the other hand, takes a short time because it requires a lexical mediation and is not influence by the semantic context. The same idea is claimed by Harites and Nelson (2001) who said

L1 initially serves as a lexical intermediary between L2 and conceptual meaning. As a result, lexical links from L2 to L1 are stronger than lexical links from L1 to L2, and conceptual links to L1 are initially stronger than conceptual links to L2 (p. 419).

The negative effect of semantic interference was present in the L2-L1 translation direction too. Words from the semantic list were the least to be recalled. This finding is consistent with results of previous studies such as Altarriba and Mathis (1997), La Heij, Hooglander, Kerling, and van der Velden (1996), and Finkbeiner and Nicol (2003).

### *Use of Context*

I noticed that most of the participants paid minimal or no attention to the context. On the contrary, they turned the new words into an unrelated list. It seems that they resist using the context and believe that the translation condition is superior although learning vocabulary in the context is perceived as desirable (Prince, 1996). When students are faced with a low-effort and a high-effort strategy, they tend to choose the former (Krashen, 1987). Since L2 learners process the context slowly because their L2 networks are not richly developed, then they find translation a rapid way to learn. The best way to overcome this problem is proposed by Prince (1996) who claims, "until such time as an L2 network is sufficiently organized, it may well be automatic, notwithstanding teachers' efforts to use pictures or L2 context to convey meaning" (p. 486).

Participants in this study claimed that they did not have sufficient time. Insufficient time does not constitute a good excuse for this; as these participants were given extra time to process the story. This can be connected with their answer to the question of how teachers used to introduce new vocabulary. Teachers used to write new vocabulary items on the board and ask learners to memorize them. Therefore, participants in this study follow the same technique. They listed the underlined vocabulary items with their meanings and memorized them by repetition. Training learners to use the available contextual cues leads to better learning.

## Summaries of Major Findings

### *Findings from the Quantitative Part*

Participants had to provide the Arabic equivalent to English words of the four lists in two tests; immediately and delayed. Also they had to provide the English

Equivalents to Arabic words of the four lists in other immediate and delayed tests.

The purpose of having two directions of translation is to see which translation direction is preferred by participants. When it comes to Arabic-to-English translation with Level 1 participants, the following are the major findings:

- 1- The participants recalled significantly more words from the thematic list (mean 7.03, SD 1.32) than from all the other lists in the immediate test (mean 5.63, SD 2.35) for the context, (mean 5.00, SD 2.58) for the unrelated list, and (mean 4.93, SD 2.21) for the semantic list.
- 2- The same results appeared in the delayed test. The participants recalled significantly more words from the thematic list (mean 2.66, SD 1.53) than from the context (mean 1.96, SD 1.75), the unrelated list (mean 1.76, SD 1.92), and the semantic list (mean 1.53, SD 1.83).
- 3- Words from the semantic list were the least to be recalled in both tests.
- 4- The participants recalled more words from the context than from the unrelated list in both tests but the difference was small and insignificant statistically.

The followings are the major findings from Level 2 Participants:

- 1- Similar to the result of Level 1 participants, Level 2 participants recalled more words from the thematic list (mean 7.10, SD 1.09) than from the other three lists (mean 6.53, SD 1.83) for the unrelated list, (mean 6.00, SD 2.06) for the semantic list, and (mean 5.73, SD 1.68) for the context list in the immediate test. The difference was significant statistically.
- 2- In the delayed test, the same order appeared: (mean 2.53, SD 1.92) for the thematic list, (mean 2.40, SD 2.22) for the unrelated list, (mean 1.83, SD 1.72)

for the semantic list, and (mean 1.60, SD 1.61) for the context list. The differences between the means are not statistically significant here.

- 3- Words from the context were the fewest to be recalled in both tests but with small insignificant differences.

When it comes to Level 3 Participants, here are the major findings:

- 1- The participants recalled more words from the thematic list (mean 7.05, SD 2.16) than from the unrelated list (mean 6.65, SD 1.98), the context (mean 6.50, SD 1.60), and the semantic list (mean 5.90, SD 2.04) in the immediate test.
- 2- The same order appeared in the delayed test: (mean 3.35, SD 2.47) for the thematic list, (mean 3.00, SD 1.92) for the unrelated list, (mean 2.30, SD 2.49) for the context, and (mean 2.10, SD 1.99) for the semantic list.
- 3- Words from the semantic list were the fewest to be recalled in both tests.
- 4- The participants recalled more words from the context than from the unrelated list in both tests but the difference was small and insignificant statistically.
- 5- The differences between means of recalled words of the lists in the immediate or the delayed tests were not statistically significant.

Studying within conditions results, I found out the following findings regarding participants' recall of the semantic list:

- 1- Level 2 participants, in the immediate test, recalled more semantic listed words (mean 6.00, SD 2.06) than those of Level 3 did (mean 5.90, SD 2.04) who, in turn, recalled more words than Level 1 participants did (mean 4.93, SD 2.21).

- 2- In the delayed test, Level 3 participants were the best recalling words from the semantic list (mean 2.10, SD 1.99). Level 2 participants were the second (mean 1.83, SD 1.72). Level 1 participants recalled the least number of words (mean 1.53, SD 1.83).
- 3- In both tests, the differences between the means are not significant statistically.

When it comes to the unrelated words, these are the major findings:

- 1- Level 3 participants recalled more unrelated words (mean 6.65, SD 1.98) than Level 2 participants (mean 6.53, SD 1.83) who were followed by Level 1 participants (mean 5.00, SD 2.58) in the immediate test. The difference between the means of Level 2 and Level 3 participants on one side and Level 1 participants on the other side is significant statistically.
- 2- The same order occurs in the delayed test. Level 3 participants were the best (mean 3.00, SD 1.91) and were followed by Level 2 participants (mean 2.40, SD 2.22) who recalled more unrelated words than Level 1 participants (mean 1.76, SD 1.92). The differences between the means are not statistically significant.

The followings are the major findings from the tests on recalling thematic words:

- 1- In the immediate test, Level 2 participants recalled the most number of thematic words (mean 7.10, SD 1.09). Level 1 participants recalled the least (mean 7.03, SD 1.32). Level 3 participants came between (mean 7.05, SD 2, 16).

- 2- In the delayed test, Level 3 participants recalled the most number of thematic words (mean 3.35, SD 2.47), and were followed by Level 1 participants (mean 2.66, SD 1.53). Level 2 Participant recalled the least number of words (mean 2.53, SD 1.92).
- 3- The differences between the means at both tests are not statistically significant.

Regarding the contextualized words, here are the major findings:

- 1- In the immediate test, Level 3 participants recalled the most number of thematic words (mean 6.50, SD 1.60). Level 1 participants recalled the least (mean 5.63, SD 2.35). Level 2 participants came between (mean 5.73, SD 1.68).
- 2- In the delayed test, Level 3 participant were the best recalling words from the semantic list (mean 2.30, SD 2.49). Level 1 participants were the second (mean 1.97, SD 1.75). Level 2 participants recalled the least number of words (mean 1.60, SD 1.61).
- 3- The differences between the means at both tests are not statistically significant.

When it comes to English -to- Arabic translation with Level 1 participants, the following are the major findings:

- 1- The participants recalled more words from the thematic list (mean 7.03, SD 1.62) than from all the other lists in the immediate test (mean 6.96, SD 1.29) for the unrelated list, (mean 6.86, SD 1.85) for the context, and (mean 5.76, SD 2.43) for the semantic list.



- 2- The same results appeared in the delayed test. The participants recalled more words from the thematic list (mean 5.83, SD 1.87) than from the unrelated list (mean 5.03, SD 1.95), the context (mean 5.00, SD 2.31), and the semantic list (mean 3.87, SD 2.23).
- 3- The differences were significantly significant between the mean of the semantic list on one hand and the means of the other lists on the other hand in both tests.
- 4- Words from the semantic list were the fewest to be recalled in both tests.
- 5- The participants recalled more words from the unrelated list than from the context in both tests but the difference was small and insignificant statistically.

The followings are the major findings from Level 2 Participants:

- 1- Unlike the result of Level 1 participants, Level 2 participants recalled more words from the context (mean 7.40, SD 1.10) in the immediate test than from the other three lists (mean 7.13, SD 1.13) for the thematic list, (mean 7.13, SD 1.52) for the unrelated list, and (mean 6.53, SD 1.65) for the semantic list in the immediate test. The differences were not significant statistically.
- 2- In the delayed test, a different order appeared: (mean 6.23, SD 1.83) for the unrelated list, (mean 4.90, SD 1.56) for the context list, (mean 4.83, SD 1.93) for the thematic list, and (mean 4.30, SD 2.07) for the semantic list. The difference between the mean of the unrelated list and all the other means is statistically significant.
- 3- Words from the semantic list were the fewest to be recalled in both tests.

When it comes to Level 3 Participants, here are the major findings:

- 1- The participants recalled more words from the context list (mean 7.55, SD 0.75) than from the unrelated list (mean 7.40, SD 1.27), the thematic list (mean 7.25, SD 1.25), and the semantic list (mean 5.95, SD 2.11) in the immediate test.
- 2- The differences between the mean of the semantic list on one hand and each of the other means are statistically significant in the immediate test.
- 3- In the delayed test, Level 3 participants recalled more words from the unrelated list (mean 6.30, SD 1.92) than from the other three lists (mean 5.90, SD 1.99) for the thematic list, (mean 5.15, SD 2.03) for the context, and (mean 4.20, SD 2.68) for the semantic list
- 4- Words from the semantic list were the fewest to be recalled in both tests.
- 5- In the delayed test, the differences between the mean of the semantic list on one hand and the means of the unrelated and thematic lists are statistically significant.
- 6- All participants, regardless of their level, recalled more words when the translation was from English to Arabic.
- 7- The semantic interference effect was present in both translation directions.

#### *Findings from the Participants' Reflection*

- 1- All participants spoken with claim that they use repetition as their technique for learning new vocabulary. It was unanimous. Along with repetition, some Highest participants claim using other techniques such as the Keyword method and the use of sentences.
- 2- The use of repetition as the major technique is not surprising since all participants spoken with claim that teachers in school list new vocabulary

items on the board and ask students to write them down and memorize them.

They claim also that teachers never tried to provide them with different techniques to learn vocabulary. I, personally, agree with this claim since I was a teacher in a high school and know what teachers do.

- 3- When asked about which list was easier to learn, participants provided different answers when it comes to the semantic and unrelated lists. Two out of six questioned participants prefer the unrelated list, two prefer the semantic list, and two have no preference of one over the other. The main reason for preferring the unrelated list is that words were not close to each other. On the other hand, one who prefers the semantic list claims that the list was confusing in the test.
- 4- Most of the questioned participants who studied the thematic list and the context show a clear preference for the thematic presentation for new words. The main reason for this preference is that words are listed and have one topic.
- 5- All questioned participants paid minimal attention to the context. In most cases, they did not even try to read it. Although participants were given extra time to read the context, insufficient time is their excuse for not reading it. Most participants ignored the contextual cues and turned the new words into an unrelated list.
- 6- The results in 3, 4, and 5 above, underline the importance of simple rote memorization for these students. Mixed results in 3 and preference of the thematic list in 4 clearly suggest that there is an appreciation for presenting new vocabulary in groups. They also suggest that learners may not be consciously aware of the way new vocabulary items are grouped, and may have no idea that grouping might affect their success in learning.

*Findings from Analysis of Answers Sheets*

- 1- Regardless of participants' level, most of the phonological and spelling errors seem to involve vowels. Examples of vowel-related errors include deleting, lowering, inserting, and changing vowels.
- 2- There are some consonant-related errors which include voicing and devoicing sounds, applying assimilation processes, doubling letters, and switching consonants occasionally. Many of these errors are predictable since the first language of participants, Arabic, lacks some sounds such as [p] or tends to break consonants clusters since it allows only two consonants cluster in word medial and word final position. Arabic does not mark short vowels in writing.
- 3- Semantic errors included phonological interference from the target language and switching two words. These errors were found in many participants' sheets regardless of their level.
- 4- Participants tended to depend on their mental lexicon to provide equivalent forms instead of the correct ones. This strategy is natural with second language learners. Others had a different kind of the target language effect.

Results Compared with the Research Hypotheses

The first hypothesis, that Saudi students would learn more unrelated words than semantic clusters of new English words, is strongly supported and accepted. Although the differences were not significant in some tests, many tests results showed that participants, regardless of their level, recalled more words from the unrelated list than from the semantic list.

The second hypothesis, that Saudi students would learn more thematic clusters of new English words than semantic clusters or unrelated English words, is also

accepted. Results of tests showed that the words from the thematic list were the most often recalled words while those from the semantic list were the least often recalled. Also words from the unrelated list were recalled more than words from the semantic list.

Participants' reflections did not support the third hypothesis, that Saudi Students would find the semantic related sets the most difficult to learn. As mentioned above, participant's answers to the question on which list was easier to learn were more mixed. Of six participants who were asked this question, two said that the semantic list was easier to learn than the unrelated list, two said that the unrelated list was easier, and two said that there was no difference.

Hypothesis Four, that Saudi Students would find words embedded in a meaningful context the easiest to learn, is not supported. Informal observation and participants' reflection showed that participants either paid minimal or no attention to the context. They wrote the English words with their Arabic equivalents in a list form and studied them out of the context. Therefore, it is impossible to evaluate this hypothesis and hypothesis five which states that the use of context facilitates learning words.

The last hypotheses, that Saudi students with higher levels might be less affected by semantic or thematic clusterings when learning English words, is supported. The differences between the recalled words from the thematic and semantic lists were significant with Level 1 participants in both tests, only significant with Level 2 participants in the immediate test, and not significant at all with Level 3 participants. It seems that grouping vocabulary semantically or thematically loses its effectiveness as participants' level moves higher.

## CHAPTER FIVE

### SUMMARY, IMPLICATIONS, AND SUGGESTIONS FOR FUTURE STUDY

#### Introduction

This chapter starts with a summary of the major findings of this study. It includes reflections on designing the study, limitations of the study, and implications and recommendations for teaching. Then it provides implications for teaching English vocabulary in Saudi school. The chapter ends with some recommendations for further research.

#### Summary

The aim of this research was to compare the effects of semantic and thematic clustering on learning English vocabulary by Saudi students. The study used a quantitative method in addition to asking participants some questions to examine the effects of using semantic and thematic clusterings on learning English vocabulary by Saudi students. In the first part; the quantitative part, data were collected from 160 participants studying in the English Language Department, Umm Al-Qura University; 60 freshmen, 60 sophomores, and 40 juniors. Participants studied four lists of English words representing semantic clustering, unrelated words, thematic clustering, and contextualized words. They were tested twice; immediately after the study phase and a week later. Data from the numbers of recalled words were used to make comparisons demonstrating the effect of each type of clustering on vocabulary learning. In the second part, the informal interview, twelve participants representing the six participants who learned the most number of words and the six who learned the least number of words were questioned individually about their experiences with

this test and with vocabulary learning generally. The interviews asked for the participants' preferred clustering as well as the strategies they use to learn new words.

The results of Arabic-to-English tests show that Level 1 participants recalled significantly more words from the thematic list than from the semantic list immediately and a week after learning. The same result appeared with Level 2 participants except that the difference was not statistically significant in the delayed test. When it comes to Level 3 participants, words from the thematic list were recalled more than words from the semantic list but with no significant differences in both tests.

When it comes to level of participants and type of grouping vocabulary, results show that Level 3 participants recalled most of the words from all the lists while Level 1 participants recalled the least. This order changes with words grouped thematically. Level 2 participants were the best in the immediate test while Level 1 participants recalled more words than Level 2 participants did.

Results of the English-to-Arabic tests show that Level 1 participants recalled more words from the thematic list than from the semantic list. The difference was significant in both tests. With Level 2 participants, thematically grouped words were not the most to be recalled but were recalled more than words from the semantic list in both tests. The difference was significant in the delayed test. With Level 3 participants, thematically grouped words were the most recalled in the delayed test and more recalled than words from the semantic list in the immediate test.

When it comes to level of participants and type of grouping vocabulary, results show that Level 3 participants were the best recalling words from all the lists while Level 1 participants recalled the least words. Differences between levels are not significant. Words grouped semantically were the least to be recalled by all

participants. Although not all of the results were significant, the fact that all tended in the same direction should be noted. Generally, the study has yielded a robust pattern in favor of thematic clustering against semantic clustering.

### Importance of the Study

During the very active decades of the mid-twentieth century, vocabulary building was not a priority for researchers or curriculum designers in the context of language teaching and learning. In fact, vocabulary was ignored and downgraded, while grammatical and phonological structures were given more emphasis because they were considered the first point in the learning process. In the past two decades, however, more emphasis has been placed on vocabulary building and learning. As a result of the growing interest in vocabulary building by researchers, various techniques and strategies have been suggested for learning and teaching the forms of a target language. Researchers started testing and evaluating these techniques in order to reach the best results in the process of language learning and as a result, a growing body of literature now addresses lexical acquisition.

Of course, Thematic and semantic clustering were among those strategies proposed by educational researchers and psychologists. New vocabulary items are typically presented to ESL/EFL students in semantically and thematically related sets in the current ESL (English as a second language) textbooks. Various studies have been done on these techniques during the last two decades such as Tinkham (1993), (1997); La Heij, Hooglander, Kerling, and Van Der Velden (1996); Waring (1997), and Finkbeiner and Nicol (2003). All these studies provide evidence that semantic clustering affects L2 learners negatively.



The current study takes a further step toward studying these two techniques with different participants and with some variation in design as compared with earlier studies. It has shed some light on how Saudi students learn English vocabulary presented in semantic and thematic sets. Although it has been conducted on male learners at an educational institution in one city of Saudi Arabia, it can provide starting point for research on the effect of using these techniques on learning English vocabulary by Saudi students. One important thing to mention is that the results of this study support the claims of researchers discussed in chapter 2, regarding the Interference Theory, the Distinctiveness Hypothesis, and the Schema Theory proposed, and claims about thematic and semantic clusterings.

#### Implications and Recommendations for Teaching

The findings in this study suggest a number of implications that need to be taken into consideration by EFL Course designers, teachers, and writers.

The finding that the differences between the means were significant with Level 1 in both tests, only significant in the immediate test with Level 2, and not significant at all with Level 3 in both tests, suggests that, as participants' level increased, the effect of vocabulary grouping is decreased. Therefore, teachers might consider using thematic clustering especially to introduce new vocabulary to beginners and intermediate level learners more than to advanced learners.

In the current study, it was reported that Level 1 participants recalled the smallest number of words in most of the tests, while Level 3 participants recalled the most number of words. Although the differences were insignificant, this shows that there is a little effect for proficiency level in the study of English vocabulary. Proficiency level might have played a role if the sample of participants and/or the

number of words in lists had been more. Still given the first recommendation above, teachers should emphasize the importance of building vocabulary at earlier stages of learners' progress in L2 learning. They may need to concentrate on lower level learners more than others in building vocabulary.

Results showed that participants performed better in backward translation (L2-L1) than in forward translation (L1-L2). Although this finding supports the findings of other studies, teachers should work hard with their learners to increase their English vocabulary. Since the aim is to learn English, then teachers have to concentrate on L1-L2 translation direction as it forces learners to learn more English vocabulary. Teachers also need to minimize their dependence on L1 during vocabulary instruction. Still, since the L2-L1 direction is much easier, it should be considered as an intermediate step toward mastering of forms in the target language.

It was reported that the negative effect of semantic interference was present in the L2-L1 translation direction too. Words from the semantic list were the least to be recalled. This finding is consistent with results of previous studies. Therefore, teachers might consider avoiding the effect of interference by increasing the differences between the taught items at any time. This can be achieved by introducing related words at different times. If teachers have to introduce related words at the same time, they should inform their learners of the negative effect of learning related items at once and should help learners find explicit strategies to keep the words separate in their minds. Moreover, teachers are encouraged to use different contexts and situations for presenting related items.

One of the findings is that participants in the current study turned the new vocabulary items presented in context into an unrelated list, basically, ignoring the context and using the marginal translation to construct a traditional word list. This is

not surprising since Saudi educational practice depends heavily on memorization. Previous research showed that the use of context and extensive reading had led to vocabulary building (see for instance, Alshamrani, 2003). Moreover, these activities reinforce vocabulary encountered previously. In contrast, learning vocabulary in wordlists might not lead to effective vocabulary building as compared with learning in context. Therefore, teachers might consider using extensive reading to introduce new vocabulary. Teachers might also think of providing some reading comprehension questions to be sure learners read the context in which new words occur. They must also use classroom activities to train learners to use available cues to make correct guesses about the meaning of new words. They might not assign contexts that include uncommon or low frequency words. Teachers also need to include sufficient cues in contexts, as well as presenting unrelated items to minimize the interference effect.

In the interviews, students unanimously agreed that teachers never provide them with techniques to help learning new vocabulary. As a matter of fact, teachers need to play a more active role in guiding students here. First, they need to plan lessons carefully. They need to select the vocabulary items that best suit the learners' needs and then introduce techniques and strategies that learners have found effective in research on this area. Here the teacher has to provide learners with a variety of methods and techniques, so that the learner can choose the methods that they prefer. These techniques, as Prince (1996) claims, should be the object to occasional analysis so that the learning process becomes transparent to the learners who, in turn, have to be able to apply them autonomously. Oxford (1990) suggests a number of memory techniques that can help learning vocabulary and making mental linkages. For creating mental linkages, Oxford suggests grouping, association/elaborating, and

placing new words into a context. For applying images, she suggests using imagery, Semantic mapping, using keywords, and representing sounds in memory.

### Limitations of the Study

As the case with any other research, there are some limitations to this study which do not affect its results. They basically relate to choice of participants and their levels. They are as follows:

- 1- One of the major limitations of this study is that participants were male learners. No female learners were able to participate in this study because females study at a separate campus in Saudi Arabia, and only female professors and teachers are allowed to enter the campus.
- 2- The study used college-level learners as participants. No lower level learners participated, thus, the perspective of the beginning learner is not represented in the results.
- 3- Only students of English Language Department of Umm Al-Qura University participated in the study. There were no participants from other schools.
- 4- The number of words in each list was eight. This low number of words was not sufficient to show clear significant in many cases.
- 5- The study's results may not be immediately applicable to real learning situations, despite the fact that efforts were made to preserve realistic aspects of the learning task (such as the use of real English words.)

### Suggestions for Future Study

Although this study addresses a number of issues regarding various methods of grouping vocabulary items, there are still other issues that need more investigation

to provide more insights into this topic which will help those interested in the field of second/foreign language vocabulary learning. The current study suggests a number of steps that helps to explore more the issue of semantic and thematic clustering.

- 1- This study used eight words in each of the wordlists. Future researchers may increase the number of words in lists as well as to further increase the similarity with real learning contexts to allow for more variation.
- 2- The current study has participants from one school in Saudi Arabia. In order to be able to generalize the findings on Saudi students, more students from different areas of the Kingdom of Saudi Arabia might be involved in similar studies.
- 3- Since this study was conducted on male learners only, further research might be conducted on female Saudi learners by female researchers.
- 4- This study had college-level participants. Further research should be conducted on learners of higher levels. Also, similar research can be conducted with learners at lower levels, such as in high schools or even intermediate school. Since Saudi Arabia has started teaching English language in primary school recently, further research can be done to measure the effect of semantic and thematic groupings on beginners, and ultimately to evaluate the relevant aspects of the syllabi used.
- 5- This study used quantitative tests as well as informal interviews to collect data. Only twelve participants were questioned in this study. Further research can be done using extensive interviews or case studies as a technique for data collection. More learners can be questioned with more deep and varied questions to know their strategies in learning vocabulary and confusing things they encounter when learning English vocabulary.

- 6- Since this study only tested participants twice with one week between, further research might consider testing participants over longer periods of time like three or six months to investigate the effect of different methods of word grouping and to measure the forgetting curve after that.
- 7- Further longitudinal research might be conducted in which participants are provided with various techniques to learn English vocabulary and then they are tested to evaluate the techniques and to find out which ones are more effective to Saudi learners than others.

## References

- Aichison, J. (1987). *Words in the mind*. Oxford: Basil Blackwell.
- Alshamrani, H. (2003). The attitudes and believes of ESL students about extensive reading of authentic texts. Unpublished doctoral dissertation. Indiana University of Pennsylvania, Pennsylvania.
- Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36*, 550-568.
- Baddeley, A.D. (1986). *Working memory*. Oxford: OUP
- Baddeley, A.D. (1990). *Human memory: theory and practice*. Boston: Allyn and Bacon.
- Baddeley, A. D. & Longman, D. J. A. (1978). The influence of length and frequency on training sessions on the rate of learning to type. *Ergonomics, 21*, 627-635.
- Bahrigh, H. P. (1984). Semantic memory content in permastore: 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1-29
- Bahrigh, H. P. & Philps, E. (1987). Retention of Spanish vocabulary over eight year. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 344-349.
- Barsalau, L. W. (1992). Frames, concepts and conceptual fields. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields and contrasts*. (pp. 21-74). Hillsdale, NJ: Erlbaum.
- Bartlett, F. C. (1932). *Remembering*. London: Cambridge University Press
- Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary: all contexts are not created equal. *Elementary School Journal, 83*, 177-181.

- Bransford, J. D., & Johnson M. K. (1972). Contextual prerequisites for understanding: Some investigators of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726
- Brewer, W. F. & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207-230
- Beck, I. L., Perfetti, C. A., & McKeown, M. B. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Beheydt, L. (1987). Vocabulary in foreign language teaching methodology. *Dutch Crossing*, 32, 3-25
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7, 15-32
- Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In Wyer, R. S., & Srull, T. K. (Eds.), *Handbook of social cognition* (vol. 1). (p 119-160). Hillsdale, NJ: Erlbaum.
- Bugelski, B. R., & Cadwallader, T. C. (1956). A reappraisal of the transfer and retroaction surface. *Journal of Experimental Psychology*, 52, 360-366.
- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21, 461-481
- Cassidy, F. G. (1985). *Dictionary of American regional English*. Cambridge, Mass.: Belknap Press of Harvard University Press
- Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching: a guide for language teacher*. New York: Cambridge University Press.
- Costinett, S. (1987). *Spectrum*, 2. Englewood Cliffs, NJ: Prentice Hall.



- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). CA: Sage.
- Crystal, D. (1997). *A dictionary of Linguistics and Phonetics*. 4th edition. Cambridge, MA: Blackwell
- Decarrico, J.S. (2001). Vocabulary Learning and Teaching. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (3rd ed., pp.285-299).
- Digby, C., & Myers, J. (1991). *Making sense of vocabulary*. London: Cassell.
- Dunbar, S. (1992). Developing vocabulary by integrating language and content. *TESL Canada Journal*, 9 (2), 73-79.
- Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 46A, 533-558
- Engelbar, S. M., & Theuerkauf, B. (1999). Defining context within vocabulary acquisition. *Language Teaching Research*, 3(1), 57-69.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Eysenck, M. W. (1979). Depth, elaboration, and distinctiveness. In Cermak, L. S. & Craik, F. I. M. (Eds.), *Levels of processing in human memory* (pp. 89-119). Hillsdale, NJ: Erlbaum.
- Fillmore, C. J. (1985). Semantic fields and semantic frames. *Quaderni di Semantica*, 6(2), 222-254.
- Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: the semantics

- of RISK and its neighbors. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields, and contrasts*. (pp. 75-102). Hillsdale, NJ: Erlbaum.
- Finkbeiner, M., & Nicol, J.L. (2003). Semantic category effects in L2 word learning. *Applied Psycholinguistics* 24 (3), 369-383.
- Franklin, I., & Meyers, C. (1991). *Crossroads, I*. New York: Oxford University Press.
- Freeman, L. C., Romney, A. K. & Freeman. S. C. (1987). Cognitive structure and informant accuracy. *American Anthropologist* 89 (2), 310-325.
- Gass, S. M., & L. Selinker. (2001). *Second language acquisition. An introductory course*. 2nd edition ed. Hillsdale, NJ: Lawrence Erlbaum.
- Gairns, R., & Redman, S. (1986). *Working with words*. Cambridge: Cambridge University Press.
- Gibson, E. J. (1941). Retroactive inhibition as a function of degree of generalization between tasks. *Journal of Experimental Psychology*, 28 (2), 93-115.
- Gipe, J., & Arnold, R. (1979). Teaching vocabulary through familiar associations and contexts. *Journal of Reading Behavior*, 11, 281-285
- Glendon, A. I., McKenna, S. P., Blaylock, S. S., & Hunt, K. (1987). Evaluation mass training in cardiopulmonary resuscitation. *British Medical Journal*, 294, 1182-1183
- Goetz, E. T., Anderson, R. C., & Schallert, D. L. (1981). The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior*, 20, 369-385.
- Grandy, R. E. (1992). Semantic fields, prototypes, and the lexicon. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields, and contrasts*. (pp. 103-122). Hillsdale, NJ: Erlbaum.

- Guba, E. G. & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Harites, C. & Nelson, K. (2001). Bilingual memory: the interaction of language and thought. *Bilingual Research Journal*, 24 (4), 417-438
- Hebb, D. O. (1949). *Organization of behavior*. New York: Wiley
- Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 2, 170-175
- Hippner-Page, T. (2000). Semantic Clustering Versus Thematic Clustering of English Vocabulary Words for Second Language Instruction: Which Method Is More Effective? ED445550
- Hockey, G. R. J., Davies, S. & Gray, M. M. (1972). Forgetting as a function of sleep at different times of day. *Experimental Psychology*, 24, 386-393.
- Hunt, R.R., & Elliott, J.M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, 109, 49-74.
- Hunt, R.R., & Mitchell, D.B. (1982) Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 8, (1), 81-87.
- Jenkins, J. G. & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *American Journal of Psychology*, 35, 605-612
- Johnson, L. M. (1933). Similarity of meaning as a factor in retroactive inhibition. *Journal of General Psychology*, 9, 377-388.
- Judd, E. L. (1978). Vocabulary teaching and TESOL: A need for re-evaluation of existing assumptions. *TESOL Quarterly*, 12, 71-76.

- Kang, S. (1995). The effects of a context-embedded approach to second-language vocabulary learning. *Systems*, 23 (1), 43-55
- Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1, 153-161.
- Kittay, E. F. (1992). Semantic fields and the individuation of content. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields, and contrasts*. (pp. 229-252). Hillsdale, NJ: Erlbaum.
- Kittay, E. F., & Lehrer, A. (1992). Introduction. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields, and contrasts*. Hillsdale, NJ: Erlbaum.
- Knight, S. M. (1994). Dictionary use while reading: the effect on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, 78, 285-299.
- Kroll, J. F. & Curley, J. (1986). Picture naming and bilingual translation. Unpublished manuscript, Mount Holyoke College, South Hadley, MA
- Kroll, J. F. & Stewart, E. (1989). Translating from one language to another: The role of words and concepts in making the connection. Paper presented at the Meeting of the Dutch Psychonomic Society, Noordwijkerhout, The Netherlands.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.
- La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward translation: evidence for concept mediation. *Journal of Memory and Language*, 35, 648-665.
- Lampinen, J., Copeland, S., & Neuschatz, J. (2001). Recollections of things schematic: rooms schemas revisited. *Cognition*, 27, 1211-1222.

- Lavie, N., Briggs, S., Raht, C., & Denman, B. (1991). *In Contact 2*. Glenview, IL: Scott Foresman.
- Li X. (1988). Effects of contextual cues on inferring and remembering meanings. *Applied linguistics*, 9, 402-413
- Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*, Chapter 14. San Francisco: Freeman.
- Lynch, T. (1996). *Communication in the Language Classroom*. Oxford: Oxford University Press.
- Marzano, R. J., & Marzano, J. S. (1988). *A cluster approach to elementary vocabulary instruction*. Newark, DE: International Reading Association.
- McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longman, Green.
- McGeoch, J.A., & McDonald, W.T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology*, 43, 579-588
- McGeoch, J. A., & McGeoch, G. O. (1937). Studies of retroactive inhibition: X. the Influence of similarity of meaning between lists of paired associates. *Journal of Experimental Psychology*, 21, 320-329
- McKenna, S. P. & Glendon, A. I. (1985). Occupational first aid training: Decay in cardiopulmonary resuscitation (CPR) skills. *Journal of Occupational Psychology*, 58, 109-117.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 20(5), 522-535.
- Melton, A. W., & Von Lackum, W. J. (1941). Retroactive and proactive inhibition in

- retention: evidence for a two-factor theory of retroactive inhibition. *American Journal of Psychology*, 54, 157-173.
- Merriam, S. B. & Simpson, E. L. (1995). A guide to research for educators and trainers of adults. (2<sup>nd</sup> ed.). Malabar, FL: Krieger.
- Ministry of Education (2002). English for Saudi Arabia first year intermediate: pupil's book. The General Directorate of Curricula.
- Mitchell, R., & Miles, F. (1998). *Second language learning theories*. London: Arnold.
- Molinsky, S. J., & Bliss, B. (1989). *Side by Side, 1*. Englewood Cliffs, NJ: Prentice Hall.
- Na, L., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16 (1), 33-42
- Nagy, W. E., Anderson, R. C., & Herman, P.A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270.
- Nagy, W. E., Herman, P.A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.
- Nagy, W. E., & Scott, J. A. (1990). Word schemas: expectations about the form and meaning of new words. *Cognition and Instructions*, 7, 105-127.
- Nation, I. S. P. (2000). Learning vocabulary in lexical sets: dangers and guidelines. *TESOL Journal*, 9(2), 6-10
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I. S. P., & Coady, J.(1988). Vocabulary and reading. In Carter, R. & McCarthy, M. (Eds.), *Vocabulary and language teaching*. (pp. 97-110). New York: Longman.

- Nist, S. L., & Olejnik, S. (1995). The role of context and dictionary definitions on varying levels of word knowledge. *Reading Research Quarterly*, 30, 172-193
- Oxford, R. (1990). *Language learning strategies: what every teacher should know*. New York: Newbury House Publishers.
- Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly*, 25, 629-653.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.
- Perkins, K., & Bratten, S. R. (1983). The effect of word frequency and contextual richness on ESL students' word identification abilities. *Journal of Research in Reading*, 6(2), 119-128
- Pollak, G. (1969). Effects of ambiguous conceptual similarity on retroactive interference in verbal memory. *Journal of Experimental Psychology*, 80 (1), 171-174
- Prince, P. (1996). Second language vocabulary learning: the role of context versus translations as a function of proficiency. *Modern Language Journal*, 80, 478-493
- Ramirez, A. (1995). *Creating contexts for second language acquisition*. White Plains, NY: Longman.
- Richards, J. C. (1998). *Interchange, 1*. Cambridge: Cambridge University Press.
- Richards, J.C., & Rodgers, T.S. (2001). *Approaches and methods in language teaching*. New York: Cambridge University Press.
- Richardson-Klavehn, A. and Bjork, R. A. (2002). Long-term memory. Entry from *Encyclopedia of Cognitive Science*, 1096-1105
- Rodgers, T. S. (1969). On measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs. *IRAL*, 7, 327-343

- Rojahan, K. & Pettigrew, T. (1992). Memory for schema relevant information: a meta-analytic resolution. *British Journal of Social Psychology*, 31, 81-109
- Rumelhart, D. (1984). Schemata and the cognitive system. In Wyer, R. S., & Srull, T. K. (Eds.), *Handbook of social cognition* (vol. 1). (p 161-188). Hillsdale, NJ: Erlbaum.
- Schacter, Daniel L. (2001). *The seven sins of memory*. New York: Houghton Mifflin.
- Schmidt, S.R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning Memory and Cognition*, 11, 565- 578.
- Schneider, V. I., Healey, A. F., & Bourne, L. E. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In Healy, A. F. & Bourne, L. E. (Eds.) *Foreign Language Learning: psycholinguistic studies on training and retention*. Mahwah, NJ.: Lawrence Erlbaum.
- Schustack, M. W. & Anderson, J. R. (1979). Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning and verbal Behavior*, 18 (5), 565-584
- Seal, B. (1990). *American vocabulary builder, 1*. White Plains, NY: Longman
- Seal, B.D. (1991). Vocabulary learning and teaching. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language*. (2<sup>nd</sup> ed., pp. 296-311). Boston: Heinle and Heinle.
- Shelfbline, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior*, 22 (1), 71-97
- Solso, R. L. (1995). *Cognitive psychology (4th Ed.)*. Boston: Allyn and Bacon.



- Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology*, *10*, 438-464.
- Sommer, B. & Sommer, R. (1997). A practical guide to behavioral research: tools and techniques. (4<sup>th</sup> ed.). New York: Oxford university press.
- Stahl, S. A., Burdge, J. L., Machuga, M. B., & Stecyk, S. (1992). The effects of semantic grouping on learning word meanings. *Reading Psychology*, *13*, 19-35
- Steffensen, M. S., Joag-dev, C., & Anderson, R. C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, *15*, 10-29
- Stern, H.H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
- Stilwell, C. H. & Markman, A. B. (2003). Schema-driven memory and structural alignment. In *The Proceeding of the 25<sup>th</sup> Annual Meeting of the Cognitive Science Society*. Boston, MA: Lawrence Erlbaum Associates.
- Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: a meta-analysis. *Review of Educational Research*, *69* (3), 261-285
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System* *21* (3), 371-380.
- Tinkham, T. N. (1994). *The effects of semantic and thematic clustering on the learning of second language vocabulary*. Unpublished doctoral dissertation. University of Illinois, Urbana).
- Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second language research*, *13*(2), 138-163
- Tulving, E. (1962). Subject organization in free recall of unrelated words. *Psychological Review*, *69*, 344-354.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

- Underwood, B.J., Ekstrand, B.R., & Keppel, G. (1965). An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, 4, 447-462.
- Wagenaar, W. A. (1986). My memory: a study of autobiographical memory over six years. *Cognitive Psychology*, 18, 225-252
- Waring, R. (1997). The negative effects of learning words in semantic sets: a replication. *System* 25 (2), 261-274.
- Wierzbicka, A. (1992). Semantic primitives and semantic fields. In Lehrer, A., & Kittay, E. F. (Eds.), *Frames, fields, and contrasts*. (pp. 209-227). Hillsdale, NJ: Erlbaum.
- Wilkins, D. A. (1976). *National syllabuses*. London: Oxford University Press.

APPENDICES

APPENDIX A  
SEMANTIC LIST

سوسن	Iris
زنبق	Lily
خزامى	Tulip
الفرجس البري	Daffodil
زهرة الثالوث	Pansy
زهرة الربيع	Daisy
زهرة النجمة	Aster
زعفران	Crocus

APPENDIX B  
ARABIC-TO-ENGLISH TEST  
(SEMANTIC LIST)

## Test 1

Name:

email:

Phone #

Write the English equivalents to the following words:

1. النرجس البري
2. زنيق
3. زهرة النجمة
4. سوسن
5. زهرة الثالوث
6. زعفران
7. زهرة الربيع
8. خزامى

APPENDIX C

ENGLISH -TO- ARABIC TEST  
(SEMANTIC LIST)



## Test 2

Name:

email:

Phone #

Write the Arabic equivalents to the following English words:

1. Pansy
2. Tulip
3. Crocus
4. Daisy
5. Iris
6. Aster
7. Lily
8. Daffodil

APPENDIX D  
UNRELATED LIST

قارب ركاب	Batean
متطرف	Capsheaf
يخاط	Bollix
ناعم	Doty
سبت	Hamper
نملة	Pismire
رواق المنزل	Portico
حفرة في الطريق	Pothole

APPENDIX E  
ARABIC-TO-ENGLISH TEST  
(UNRELATED LIST)

## Test 1

Name:

email:

Phone #

Write the English equivalents to the following words:

1. ناعم
2. رواق المنزل
3. يخلط
4. حفرة في الطريق
5. سبت
6. قارب ركاب
7. نملة
8. متطرف

APPENDIX F  
ENGLISH -TO- ARABIC TEST  
(UNRELATED LIST)

## Test 2

Name:

email:

Phone #

Write the Arabic equivalents to the following English words:

1. Pismire
2. Doty
3. Pothole
4. Bollix
5. Hamper
6. Batean
7. Portico
8. Capsheaf

APPENDIX G  
THEMATIC LIST



رمح	Leister
بكرة	Reel
زورق مسطح	Dory
طعم	Lure
يلقي الصنارة	Cast
ضحل	Shoal
سمك البنبان	Pompano
الصيد بالصنارة	Angling

APPENDIX H  
ARABIC-TO-ENGLISH TEST  
(THEMATIC LIST)

## Test 1

Name:

email:

Phone #

Write the English equivalents to the following words:

1. طعم
2. سمك البنبان
3. بكرة
4. الصيد بالصنارة
5. زورق مسطح
6. يلقي الصنارة
7. رمح
8. ضحل

APPENDIX I  
ENGLISH -TO- ARABIC TEST  
(THEMATIC LIST)

## Test 2

Name:

email:

Phone #

Write the Arabic equivalents to the following English words:

1. Pompano
2. Dory
3. Leister
4. Cast
5. Angling
6. Lure
7. Reel
8. Shoal

APPENDIX J

CONTEXT

*One day Jack felt donsie because he has eaten so much at a party.*

متعب

*His wife and his two kids wanted to spend that day out in the open and to go for a walk. When they left home they realized that it was*

*airish outside since it is the beginning of the winter season so they put jackets on. Jack drove to an area out of town where he saw a*

بارد

*running bayou. The water was moving slowly with a quiet sound.*

نهر صغير

*His kids loved the view of the running water and threw dornickets in the water to make the frogs jump. The water looked gaumy so he did not drink from it fearing of becoming sick of Malaria.*

أحجار صغيرة

ملوث

*Suddenly his wife saw a doty and bright-colored object that*

لين / منقط

*looked strange in the ground. Jack got a rastus, which*

مجرفة

*he always keeps in his car to use on his farm, and started digging.*

*What they found was a very pretty little metal box, with punky*

مهترىء / ردىء

*wooden handle “It must be very old,,: said Jack.*

APPENDIX K  
ARABIC-TO-ENGLISH TEST  
(CONTEXT)



## Test 1

Name:

email:

Phone #

Write the English equivalents to the following words:

1. أحجار صغيرة
2. ناعم
3. مهترىء
4. بارد
5. مجرفة
6. نهر صغير
7. متوعك
8. ملوث

APPENDIX L  
ENGLISH -TO- ARABIC TEST  
(CONTEXT)

## Test 2

Name:

email:

Phone #

Write the Arabic equivalents to the following English words:

1. Airish
2. Punky
3. Gaumy
4. Rastus
5. Bayou
6. Donsie
7. Doty
8. Dornickets

APPENDIX M  
CONSENT FORM

## Informed Consent Form

You are invited to participate in this research study. The following information is provided in order to help you make an informed decision whether or not to participate. If you have any questions please do not hesitate to ask. You are eligible to participate because you are a student at the English Department, Umm Al-Qura University, Makkah, Saudi Arabia.

The purpose of this study is to compare the effects of semantic and thematic clustering on learning of English vocabulary by Saudi students. Participation in this study is not considered a part of any course. Participation or non-participation will not effect the evaluation of your performance in any class. First you will be randomly assigned to either the semantic/unrelated or thematic/contextual groups. Then you will listen to an introduction explaining the purpose of the research and the procedure of the testing. After that you will be given a list of eight English words with their Arabic equivalents to study. In the test, you will write the Arabic equivalents for the English words you studied. They will occur in a different order. Four subjects who learned more words and other four subjects who learned fewer words will be interviewed individually in Arabic and will be asked about nine questions to elicit qualitative data that might improve insight into the quantitative data and analyses.

The information gained from this study may help us to better understand the effectiveness of semantic clustering Vs thematic clustering as two techniques form presenting new vocabulary.

Your participation in this study is voluntary. You are free to decide not to participate in the study or to withdraw at any time without adversely affecting your relationship with the investigators. Your decision will not result in any loss of benefits to which you are otherwise entitled. If you choose to participate, you may withdraw at any time by notifying the researcher. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence and will have no bearing on your academic standing or services you receive from the university. Your response will be considered only in combination with those from other participants. The information obtained in the study may be published in scientific journals or presented at scientific meetings but your identity will be kept strictly confidential.

If you are willing to participate in this study, please sign the statement below and deposit in the designated box by the door. Take the extra unsigned copy with you. If you choose not to participate, deposit the unsigned copies in the designated box by the door.

Researcher:

Mr. Sameer Al-Jabri (qrhl@iup.edu)  
Ph.D. student, Department of English  
Indiana University of Pennsylvania  
Indiana, PA 15705 USA  
Phone: 724 – 357 – 2263

Research advisor:

Dr. Jeannine Fontaine ([jfontain@iup.edu](mailto:jfontain@iup.edu))  
Associate Professor, Department of English  
Indiana University of Pennsylvania  
Indiana, PA 15705 USA  
Phone: 724 – 357 - 2263

This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (Phone: 724-357-7730).

Informed Consent Form (continued)

Voluntary Consent Form:

I have read and understand the information on the form and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed Consent Form to keep in my possession.

Name:

Signature:

Date:

Phone number or location where you can be reached:

Best days and times to reach you

Email:

I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.

Researcher's Signature

Date: