

**THE EFFECTS OF TEST AUTHENTICITY AND TEST DELIVERY MEDIUMS ON THE
COMPUTER-BASED READING COMPREHENSION ABILITY AND THE TEST
TAKERS' ATTITUDES TOWARDS THE TEST**

Miss Supalak Nakhornsri

**A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in English as an International Language**

(Interdisciplinary Program)

Graduate School

Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University

ผลกระทบของแบบทดสอบที่มีความสมจริงและวิธีการจัดสอบต่อความสามารถในการอ่านเพื่อความเข้าใจ
ภาษาอังกฤษด้วยคอมพิวเตอร์ และทัศนคติของผู้เข้าสอบต่อแบบทดสอบ



นางสาวสุภาลักษณ์ นครศรี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

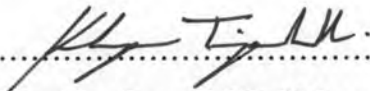
Thesis Title THE EFFECTS OF TEST AUTHENTICITY AND TEST DELIVERY
MEDIUMS ON THE COMPUTER-BASED READING
COMPREHENSION ABILITY AND THE TEST TAKERS'
ATTITUDES TOWARDS THE TEST

By Miss Supalak Nakhornsri

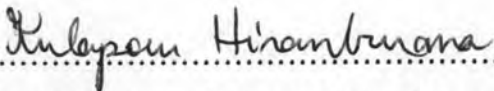
Field of Study English as an International Language


Thesis Advisor Professor Kanchana Prapphal, Ph.D.


Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment
of the Requirements for the Doctoral Degree



..... Dean of the Graduate School
(Assistant Professor M.R. Kalaya Tingsabadh, Ph.D.)

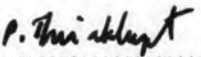
THESIS COMMITTEE


..... Chairman
(Assistant Professor Kulaporn Hiranburana, Ph.D.)


..... Thesis Advisor
(Professor Kanchana Prapphal, Ph.D.)


..... Member
(Assistant Professor Somsak Boonsathorn, Ph.D.)


..... Member
(Assistant Professor Charun Sanrach, Ph.D.)


..... Member
(Assistant Professor Pavinee Thirakhupt, Ph.D.)

สุภาลักษณ์ นครศรี: ผลกระทบของแบบทดสอบที่มีความสมจริงและวิธีการจัดสอบต่อความสามารถในการอ่านเพื่อความเข้าใจภาษาอังกฤษด้วยคอมพิวเตอร์ และทัศนคติของผู้เข้าสอบต่อแบบทดสอบ (THE EFFECTS OF TEST AUTHENTICITY AND TEST DELIVERY MEDIUMS ON THE COMPUTER-BASED READING COMPREHENSION ABILITY AND THE TEST TAKERS' ATTITUDES TOWARDS THE TEST.) อ. ที่ปรึกษา: ศ.ดร. กาญจนา ปรามพาล, 299 หน้า.

การวิจัยนี้มีจุดมุ่งหมายเพื่อสร้างแบบสอบวัดความสามารถในการอ่านเพื่อทดสอบนักศึกษาชั้นปีที่ 1 สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ โดยกำหนดขอบเขตการใช้ภาษาที่เรียน (Target Language Use domain) เพื่อนำไปสร้างแบบสอบวัดความสามารถในการอ่านที่มีความสมจริงและไม่สมจริง ที่มีการจัดสอบโดยใช้ด้วยกระดาษและคอมพิวเตอร์ นอกจากนี้มีการศึกษาทัศนคติของผู้เข้าสอบที่มีแบบสอบต่างรูปแบบดังกล่าว การวิจัยนี้มีจุดประสงค์หลักดังนี้ (1) เพื่อศึกษาผลกระทบของแบบสอบที่มีความสมจริงและวิธีการจัดสอบที่มีต่อความสามารถในการอ่าน และ (2) เพื่อศึกษาทัศนคติของผู้เข้าสอบที่มีต่อแบบสอบ

เครื่องมือในการวิจัยคือ แบบสอบวัดความสามารถในการอ่านแบบสมจริง จัดสอบแบบกระดาษ (ACON), แบบสอบวัดความสามารถในการอ่านแบบไม่สมจริง จัดสอบแบบกระดาษ (ICON), แบบสอบวัดความสามารถในการอ่านแบบสมจริง จัดสอบด้วยคอมพิวเตอร์ (ACOM) และ แบบสอบวัดความสามารถในการอ่านแบบไม่สมจริง จัดสอบด้วยคอมพิวเตอร์ (ICOM) นักศึกษาถูกสัมภาษณ์หลังจากการทำแบบสอบเพื่อศึกษาทัศนคติที่มีต่อแบบสอบ ในการวิเคราะห์ผู้วิจัยใช้การวิเคราะห์ความแปรปรวนเพื่อทดสอบผลกระทบหลักและปฏิสัมพันธ์ระหว่างสองตัวแปร ใช้พาร์เชียล เอต้า สแควร์ (Partial Eta Square) เพื่อวิเคราะห์ขนาดของผลกระทบ ในการวิเคราะห์ทัศนคติของผู้เข้าสอบใช้การวิเคราะห์เนื้อหาและไค สแควร์

ผลการวิจัยพบว่า (1) มีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ระหว่างคะแนนเฉลี่ยจากแบบสอบที่มีความสมจริง (ACOM และ ACON) และคะแนนเฉลี่ยที่ได้รับจากแบบทดสอบที่ไม่สมจริง (ICOM และ ICON) ความสมจริงมีผลกระทบในขนาดปานกลาง (2) ไม่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ระหว่างคะแนนเฉลี่ยจากแบบสอบที่จัดสอบด้วยคอมพิวเตอร์ (ACOM และ ICOM) และคะแนนเฉลี่ยจากแบบสอบที่จัดสอบด้วยกระดาษ (ACON และ ICON) และวิธีการจัดสอบมีผลกระทบในขนาดเล็ก (3) มีปฏิสัมพันธ์ระหว่างความสมจริงของแบบสอบและวิธีการจัดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 นักศึกษาส่วนใหญ่มีทัศนคติเชิงบวกต่อแบบสอบที่ตนทำ อย่างไรก็ตามนักศึกษาแสดงทัศนคติเชิงลบในบางประเด็น

สาขาวิชา ภาษาอังกฤษเป็นภาษานานาชาติ
ปีการศึกษา 2550

ลายมือชื่อนิติศ..... สุภาลักษณ์ นครศรี
ลายมือชื่ออาจารย์ที่ปรึกษา.....

#4789691120 : MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
 KEY WORD: TEST AUTHENTICITY/ TEST DELIVERY MEDIUMS/ COMPUTER-
 BASED READING COMPREHENSION ABILITY/ ATTITUDES TOWARDS THE TESTS
 SUPALAK NAKHORNSRI: THE EFFECTS OF TEST AUTHENTICITY AND TEST
 DELIVERY MEDIUMS ON THE COMPUTER-BASED READING
 COMPREHENSION ABILITY AND THE TEST TAKERS' ATTITUDES TOWARDS
 THE TEST. THESIS ADVISOR: PROF. KANCHANA PRAPPHAL, Ph.D., 299pp.

This study proposed, designed and created Reading Comprehension Tests for assessing first year students' reading proficiency at King Mongkut's Institute of Technology North Bangkok. The objectives of this study were 1) to investigate the effects of test authenticity and test delivery mediums on test takers' reading proficiency and 2) to explore the attitudes of the test takers towards test authenticity and test delivery mediums. The Target Language Use domain was considered to develop authentic and inauthentic reading tests administered via paper-based and computer-based (computer-adaptive online) tests. Students' attitudes towards these different versions of the tests were also investigated.

The main research instruments included the Authentic English Reading Comprehension Conventional Paper-and-Pencil Test (ACON), the Inauthentic English Reading Comprehension Conventional Paper-and-Pencil Test (ICON), the Authentic English Reading Comprehension Computer-Adaptive Online Test (ACOM), and the Inauthentic English Reading Comprehension Computer-Adaptive Online Test (ACOM). A retrospective interview was conducted to investigate students' attitudes towards those tests. The data obtained were analyzed by two-way analysis of variance (2*2 ANOVA) to identify the main effects and the interaction effects. Partial Eta square reported by SPSS program was used to measure the effect sizes of test authenticity and test delivery mediums. Content analysis and chi square (χ^2) tests were applied to find out whether there was any significant difference in the proportions of subjects' attitudes towards test authenticity and test delivery mediums in each group.

Findings reveal that (1) there was significant difference between the mean score obtained from the test considered authentic (ACOM and ICOM) and those obtained from the test considered inauthentic (ICOM and ICON) at the 0.05 level ($p < .05$, $F = 49.164$). Partial Eta Squared value (0.142) shows the medium effect size of the test authenticity on test takers' reading ability score (2) there was no significant difference between the mean score obtained from the test delivered by means of computer (ACOM and ICOM) and that obtained from the test delivered by means of paper (ACON and ICON) at the 0.05 level ($p > .05$, $F = 2.556$). The effect size value of test delivery mediums is 0.009. This means that the effect size of the mediums on test takers' reading ability scores was small. (3) The interaction effect between Mediums and Authenticity value obtained from the 2*2 ANOVA was significant at the level of 0.05 ($p \leq .05$, $F = 14.474$). Regarding the test takers' attitudes towards the test authenticity and test delivery mediums, it can be concluded that test takers tend to have positive attitudes towards the tests they took. However, there were some issues on which they reported negative attitudes.

Field of study English as an International Language

Student's signature *Supalok Nakhorn Sri*

Academic year 2007

Advisor's signature *Kanchana Prapphal*

ACKNOWLEDGEMENTS

This dissertation could not have been completed without Professor Dr. Kanchana Prapphal who not only served as my thesis advisor but also encouraged me throughout my research. She guided me through the dissertation process, never accepting less than my best efforts.

I would like to express my deep gratitude to the entire dissertation committee, Assistant Professor Dr. Kulaporn Hiranburana, Assistant Professor Dr. Somsak Boonsathorn, Assistant Professor Dr. Charun Sanrach and Assistant Professor Dr. Pavinee Thirakhupt, for devoting their time to critically read my work and providing valuable advice.

In this regard, I am also indebted to Associate Professor Dr. Sripen Srestasathien and Associate Professor Dr. Ngamthip Wimolkasem, who, as reviewers of the research instruments, instigated my revisions and improvements of the tests.

I am particularly grateful to Associate Professor Dr. John Read for his thoughtful and creative comments on my work.

I would like to acknowledge the debt I owe to all my colleagues at the faculty of Applied Arts, King Mongkut's Institute of Technology North Bangkok. Their friendship and professional collaboration has meant a great deal to me, particularly Assistant Professor Bubpha Setobol for her generous assistance in the qualitative analysis of this study. I am grateful to Mrs. Kaythi Myo Than for the care with which she reviewed this dissertation paper. I am also very thankful to a number of students at King Monkut's Institute of Technology North Bangkok who sacrificed hours to participate in this study. Importantly, this study would not be possible without great assistance from a brilliant computer programmer, Ajarn Thanyarat Nomponkrang.

My special gratitude belongs to Associate Professor Dr. Sumitra Angwattanakul, my Master's Degree thesis advisor who has always given me great advice and support.

Above all, I wish to thank my beloved parents, Mr. Boontham and Mrs. Ubon Nakhornsri, my greatest sources of inspiration. Without them, I would never have achieved this goal. I owe an unrepayable debt of gratitude for their support in all forms, material as well as emotional. They are indeed "the wind beneath my wings".

CONTENTS

	Page
Abstract (Thai)	iv
Abstract (English)	v
Acknowledgements	vi
Contents	vii
List of Tables	x
List of Figures	xiii
Chapter I Introduction	1
1.1 Background of the study	1
1.2 Research questions	10
1.3 Objectives of the study	10
1.4 Statement of hypotheses	10
1.5 Scope of the study	11
1.6 Assumptions	12
1.7 Definition of terms	12
1.8 Significance of the study	13
Chapter II Literature Review	15
2.1 Introduction	15
2.2 Reading assessment	15
2.3 Test authenticity	25
2.4 Mediums of test delivery	38
2.5 Computer-adaptive testing	58
2.6 Attitudes towards test authenticity and test delivery medium ..	75
Conclusion	79
Chapter III Research Methodology	83
3.1 Introduction	83
3.2 Research design	83
3.3 Population and samples	84

	3.4 Stages of research	87
Chapter IV	Results and Findings	117
	4.1 Introduction	117
	4.2 The results and findings	117
	Part I: the descriptive statistic (arithmetic means Standard deviations) and Analysis of Variance (ANOVA) of the midterm and final scores of the four groups	118
	Part II: the testing of ANOVA assumptions	120
	Part III: the results of the study	122
	Part IV: the qualitative aspects of the test takers' attitudes towards test authenticity and test delivery mediums	127
	4.3 Summary of the results	169
Chapter V	Conclusions and Recommendations	170
	5.1 Introduction	170
	5.2 Research summary and summary of the findings	170
	5.3 Conclusions and implication of the study	174
	5.4 Recommendations for future research	182
	References	184
	Appendices	193
	Appendix 1: Open-ended Survey Questionnaire: English Reading in Real Life	194
	Appendix 2: Questionnaire: English Reading in Real Life	195
	Appendix 3: The most frequently selected reading topics answered By 5 groups	199
	Appendix 4: The Frequency of Purposes, Materials and Settings of Reading Topic, Entertainment, selected by 5 groups...	201

	Page
Appendix 5: The Frequency of Purposes, Materials and Settings of Reading Topic, Jobs, selected by 5 groups	207
Appendix 6: The Frequency of Purposes, Materials and Settings of Reading Topic, Technology, selected by 5 groups	213
Appendix 7: Target Language Use Domain for Authentic Reading Tests	219
Appendix 8: Evaluation Form of Test Authenticity	235
Appendix 9: Table of Specification of Reading Tests	236
Appendix 10: Passage Selecting Form	241
Appendix 11: The flowchart of the first pilot	243
Appendix 12: The flowchart of the second pilot	246
Appendix 13: The flowchart of the third pilot	249
Appendix 14: The flowchart of the fourth pilot	252
Appendix 15: Factor analysis	255
Appendix 16: Authentic Reading Comprehension Test (ACON) ...	258
Appendix 17: Inauthentic Reading Comprehension Test (ICON) ...	271
Appendix 18: Test Characteristic Curve	287
Appendix 19: An encoding scheme	288
Appendix 20: Summed up score for calculating the inter-coder reliability	292
Biography	299

List of Tables

Table	Page
2.1 The Comparison between Paper-based Tests and Computer-based Tests	45
3.1 Four Groups of the Subjects Assigned to Take Four Forms of Reading Tests	86
3.2 The Degree of Test Authenticity in ACON	98
3.3 The Degree of Test Authenticity in ICON	100
4.1 Mean and Standard Deviation of English Language Ability of Group One, Two, Three and Four	119
4.2 One-way ANOVA of English Language Ability of the Four Experimental Groups	119
4.3 Descriptive Statistics of the Data	120
4.4 Tests of Normality	121
4.5 Test of Homogeneity of Variance	121
4.6 Tests of Between-Subjects Effects	122
4.7 Pair wise Comparisons between the Authentic and Inauthentic Tests	123
4.8 Pair wise Comparisons between Tests Delivered by Computer and by Paper	125
4.9 Correlations between two coders	127
4.10 Chi-Square Tests of the Test Takers' Opinions on the Opportunity to Demonstrate Strengths and Weaknesses in Reading Ability	129
4.11 Frequency and Percentage of the Test Takers' Opinions on the Opportunity to Demonstrate Strengths and Weaknesses in Reading Ability	129
4.12 Chi-Square Tests of the Test Takers' Opinions on the Perception of Test Difficulty	130
4.13 Frequency and Percentage of the Test Takers' Opinions on the Perception of Test Difficulty	131
4.14 Chi-Square Tests of the Test Takers' Perceptions of Test Fairness: The Appropriateness of the Questions	132
4.15 Frequency and Percentage of the Test Takers' Perceptions of Test Fairness: The Appropriateness of the Questions	133

4.16	Chi-Square Tests of the Test Takers' Perceptions of Test Fairness: the Effect of the Ways to Answer the Tests	134
4.17	Frequency and Percentage of the Test Takers' Perceptions of Test Fairness: the Effect of the Ways to Answer the Tests	134
4.18	Chi-Square Tests of the Test Takers' Perceptions of Test Fairness: the Situations in the Tests	135
4.19	Frequency and Percentage of the Test Takers' Perceptions of Test Fairness: the Situations in the Tests	136
4.20	Chi-Square Tests of the Test Takers' Perceptions of Nervousness While Taking the Tests	138
4.21	Frequency and Percentage of the Test Takers' Perceptions of Nervousness While Taking the Test	138
4.22	Frequency and Percentage of the Test Takers' Perceptions of the Clarity of the Test Directions	139
4.23	Chi-Square Tests of the Test Takers' Opinions on the Accuracy of the Test to Elicit Their True Ability in Reading	140
4.24	Frequency and Percentage of the Test Takers' Opinions on the Accuracy of the Test to Elicit Their True Ability in Reading	141
4.25	Chi-Square Tests of the Test Takers' Opinions on the Accuracy of the Test to Elicit Their True Ability in Reading: the Relevance of the Test Tasks to Their Real Life Reading Activities	142
4.26	Frequency and Percentage of the Test Takers' Opinions on the Accuracy of the Test to Elicit Their True Ability in Reading: the Relevance of the Test Tasks to Their Real Life Reading Activities	143
4.27	Chi-Square Tests of the Test Takers' Opinions on the Test Administration: The Perceptions on the Test Administration They Took	144
4.28	Frequency and Percentage of the Test Takers' Opinions on the Test Administration: The Perceptions on the Test Administration They Took ...	145
4.29	Chi-Square Tests of the Test Takers' Opinions on the Test Administration: Compared to Other Versions of the Tests	146

4.30	Frequency and Percentage of the Test Takers' Opinions on the Test Administration: Compared to Other Versions of the Tests	147
4.31	Chi-Square Tests of the Test Takers' Opinions on the Time Length	148
4.32	Frequency and Percentage of the Test Takers' Opinions on the Time Length Time Length	149
4.33	Chi-Square Tests of the Test Takers' Opinions on the Facilities	151
4.34	Frequency and Percentage of the Test Takers' Opinions on the Facilities	151
4.35	Chi-Square Tests of the Test Takers' Opinions on the Test Contents	152
4.36	Frequency and Percentage of the Test Takers' Opinions on the Test Contents	153
4.37	Chi-Square Tests of the Test Takers' Opinions on the Scoring	155
4.38	Frequency and Percentage of the Test Takers' Opinions on the Scoring	156
4.39	Chi-Square Tests of the Test Takers' Opinions on Interactiveness	158
4.40	Frequency and Percentage of the Test Takers' Opinions on Interactiveness	159
4.41	Chi-Square Tests of the Test Takers' Opinions on the Test Authenticity ...	161
4.42	Frequency and Percentage of the Test Takers' Opinions on the Test Authenticity	162
4.43	Chi-Square Tests of the Test Takers Opinions on Their Familiarity to the Tests	164
4.44	Frequency and Percentage of the Test Takers' Opinions on Their Familiarity to the Tests	164
4.45	Chi-Square Tests of the Test Takers' Opinions on Perseverance	165
4.46	Frequency and Percentage of the Test Takers' Opinions on Perseverance ..	166
4.47	Chi-Square Tests of the Test Takers' Attitudes towards the Tests	167
4.48	Frequency and Percentage of the Test Takers' Attitudes towards the Tests	167
4.49	Investigated Topics Reported as Significant Different among the Four Groups	168

List of Figures

Figure	Page
1.1 Test and criterion	4
2.1 Authenticity	26
2.2 The continuum of traditional assessment and authentic assessment	31
2.3 A flowchart describing an adaptive test	63
2.4 Constant step size pyramidal models	67
3.1 Four blocked groups	85
3.2 The flowchart of a content-based CAT	95
3.3 The system of the computer-adaptive test	101
3.4 The ways to get access to the system	102
3.5 The signing in page	102
3.6 ICOM test takers' scores	103
3.7 ACOM test takers' scores	104
3.8 ACOM homepage	105
3.9 The introduction of the program	106
3.10 Sign in page	106
3.11 The page presenting the test instruction (ACOM)	107
3.12 The page showing an example of the test	108
3.13 Authentic English reading test item	108
3.14 The page presenting the test taker's test result	109
3.15 The page presenting the test instructions (ICOM)	110
3.16 Inauthentic English reading test item	110
3.17 The page presenting the test taker's test result	111
4.1 Plot of the interaction effect between mediums and test authenticity	126
4.2 Difference frequencies among the four groups in terms of the test takers' perceptions of test fairness: the situations in the tests	137
4.3 Difference frequencies among the four groups in terms of the test takers' opinions on the time length	150
4.4 Difference frequencies among the four groups in terms of the test takers' Opinions on the test contents	154

4.5	Difference frequencies among the four groups in terms of the test takers' opinions on the scoring	157
4.6	Difference frequencies among the four groups in terms of the test takers' opinions on interactiveness	160
4.7	Difference frequencies among the four groups in terms of the test takers' Opinions on the test authenticity	163