

CHAPTER 6

GENERAL DISCUSSION

Introduction

Studies 1, 2 and 3, which used the introductory sequences of two films *Airplane* and *The Graduate*, had one general research question.

What are the effects of captioning upon learners' comprehension and acquisition? For the three studies, this main question can be separated into two parts. First, how and why were the effects of captioning on comprehension and acquisition different for the various post-tests? Specifically the tests were: for Stage 1, the *Multiple-Choice Comprehension Tests*; and for Stage 2, the *Depth of Knowledge Tests (Story Partial-Cloze, Synonym, and Non-Synonym)*. Second, in comparing the introductory sequences of *Airplane* and *The Graduate*, what was the effect of genre between *Airplane* and *The Graduate*? In the definitions section in Chapter 1, the introduction, it was stated that in this dissertation, genre is defined as types of video, particularly with film, that are perceived as being different by the viewer. This means genre is has a meaning equivalent to different kinds of film and therefore is wider than the standard definitions found in either linguistics (Hoey, 1991; Swales, 1990) or film theory (Nichols, 1976).

The three studies showed that the effects of captioning are clearly not consistent across different multimedia texts, and that there is considerable variation in the effect of the presence of captioning, depending on the kind of test, or measures, that are used.

Before discussing this question at length, the next section will summarize the results of the research questions relating specifically to each of the three studies.

Study 1

Study 1 asked three questions.

1. In what way are the introductory sequences from *Airplane* and *The Graduate* different? (See Measurement of the Input in the literature review, Section 4)
2. In the introductory sequences, what do learners perceive as being interesting or important?
3. What tests can be developed to measure comprehension?

The introductory sequences were different in terms of the number of scenes: thirty-two for *Airplane*, compared with seven for *The Graduate*. The number and patterns of ideas, or propositions, as identified by the participants, also indicated that learners perceived the introductory sequences as being different. Despite the considerably larger number of scenes in *Airplane*, the participants perceived the discourse of *The Graduate* as being more complex. The *Multiple-Choice Test*, as a way of measuring comprehension, was found to be both reliable and valid. However, as will be discussed later, it should not be the only test used in any assessment of the effect of captioning upon comprehension.

Study 2

Study 2 asked three questions.

1. What tests can be developed that measure, not only learners' breadth of vocabulary knowledge of language, but also their depth of vocabulary knowledge. (See Broadening and Deepening Knowledge in the literature review, section 2; and Learners' Knowledge in the literature review, Section 4)
2. How reliable and valid are such tests?
3. Which tests should be used?

The *Depth of Knowledge Tests* (*Story Partial-Cloze*, *Synonym*, *Non-Synonym*, and *Definitions*) showed that it is possible to construct a battery of tests, which can measure, not only that they know or do not know a language item, in this case vocabulary, but also to what extent learners know it. This distinction has been described throughout this dissertation as learners' breadth vs. depth of knowledge. The Cronbach alpha reliabilities of the tests were good, ranging from $\alpha = .66$ to $\alpha = .91$, apart from the *Definitions Test*, $\alpha = .60$. It could be that translation of the tests in which the options became shorter reduced the content and increased student guessing, while at the same time not offering a true *Non-Synonym Test*. Furthermore, for both introductory sequences, *Airplane* and *The Graduate*, the participants' mean scores for the *Definitions Tests* were not significantly different from the mean scores of the *Non-Synonym Tests*. Therefore, the *Definitions Tests* were not used in Study 3.

Study 3

Study 3 asked three questions.

1. What is the effect of captioning upon learners' comprehension?
2. What is the effect of captioning upon learners' breadth and depth of vocabulary knowledge?
3. What is the effect of captioning upon learners' production?

Concerning Question 1, the effect of captioning upon learners' comprehension, *The Graduate Multiple-Choice Comprehension Test* showed that the presence of captioning significantly increased learners' comprehension. For *Airplane*, captioning had no significant effect. However, one key consideration is the role of comprehension in the results of the *Story Partial-Cloze Tests*, which registered significantly higher scores on both *Airplane* and *The Graduate* for the closed captioned viewing condition. Cloze-tests based on written texts have a stand-alone context. However, the *Story Partial-Cloze Tests* in this dissertation are not based on a written text, but on the introductory sequences from *Airplane* and *The Graduate*. The linguistic input, spoken or captioned, is closely related to the timing and presence of visual input. The results of the *Story Partial-Cloze Tests* showed that presence of captioning has a significant effect for both films, not only upon production, but also upon comprehension, because the presence of captioning had an overall effect on the comprehension of the context and story.

Concerning Question 2, the effect of captioning upon learners' breadth and depth of knowledge, for *The Graduate*, captioning had a significant effect upon learners' scores

for the *Synonym Test*, but not for the *Non-Synonym Test*. For *Airplane*, captioning did not have a significant effect upon learners' scores for either test. However, if production, as was argued in the literature review, is an indication of a greater depth of knowledge, by inference it can be argued that captioning did have a significant effect upon learners' breadth and depth of knowledge, as was shown by the *Story Partial-Cloze Test* results.

Concerning Question 3, the effect of captioning upon learners' production, for both *Airplane* and *The Graduate*, viewing the captioned version of the introductory sequence had a highly significant effect upon production.

Studies 1, 2, and 3

The three studies showed that the effects of captioning are clearly not consistent across different multimedia texts, and that there is considerable variation in the effect of the presence of captioning, with respect to the kind of tests, or measures, that are used. Therefore, to answer the main research question, what are the effects of captioning upon learners' comprehension and acquisition, two questions have to be answered, as was stated at the beginning of this chapter.

1. How, and why, were the effects of captioning different for the various post-tests?
2. In comparing the introductory sequences of *Airplane* and *The Graduate*, what was the effect of genre between *Airplane* and *The Graduate*?

Effects of Captioning: A Comparison of the Tests

Significance Levels

This dissertation has argued in the literature review that it is important to incorporate a battery of tests. Therefore, by doing so in Study 3, several statistical analyses were necessary. There were two ANOVAs for *Airplane* and *The Graduate Multiple-Choice Comprehension Tests*. For the *Depth of Knowledge Tests*, there was also one MANOVA for *Airplane*, and one MANCOVA for *The Graduate*. As stated in Chapter 5, statisticians would apply the Bonferroni test and reset the significance level of $p = .05$ at $p = .0125$.

An increase in the number of tests results in a loss of power, or in a lower p value.

O'Brien (1983) summarizes this well:

It seems ironic that when investigators publish their separate findings in the medical literature, per-comparison error rates are routinely accepted. However, when one investigator takes on the entire job himself, the same approach may no longer be deemed valid. Rather, he is required to achieve a considerably higher level of significance with each comparison, virtually as a penalty for undertaking such an extensive effort. (p. 788, cited in Keppel, 1991, p. 168)

This issue is further compounded by the repeated target words in the three *Depth of Knowledge Tests*. Conservative statisticians would require that the p value be set lower than $p = .0125$.

Multiple-Choice Comprehension Tests

A summary of the descriptive statistics, reliabilities, and Study 3 one-way ANOVAs are shown in Tables 6.1 and 6.2. The modifications made after Study 1 resulted in improved scores. Despite the fact that in Study 3 both tests had two questions

less than the original Study 1 test, the mean scores of the classes which viewed the captioned conditions increased (*Airplane*, Class A, 18.40 vs. Class C, 20.74; *The Graduate*, Class B, 15.16 vs. Class D, 16.13). Furthermore, the modifications also improved the Cronbach alpha reliabilities of both tests ($\alpha = .30$ vs. $.54$, $.35$ vs. $.69$).

Table 6.1. Summary of *Airplane* Multiple-Choice Comprehension Test Descriptive Statistics, One-Way ANOVA p Value (Classes C & F), and Cronbach alpha Reliabilities

Class	Study	Condition	Questions	Mean	SD	p	α
A ($n = 39$)	1	Captioned	29	18.40	3.20	-	.30
C ($n = 39$)	3	Captioned	27	20.74	3.26	.071	.54
F ($n = 47$)	3	Normal	27	19.49	3.07		

Table 6.2 Summary of *The Graduate* Multiple-Choice Comprehension Test Descriptive Statistics, One-Way ANOVA p Value (Classes D & E), and Cronbach alpha Reliabilities

Class	Study	Condition	Questions	Mean	SD	p	α
B ($n = 54$)	1	Captioned	25	15.16	2.91	-	.35
D ($n = 46$)	3	Captioned	23	16.13	3.12	.000	.69
E ($n = 34$)	3	Normal	23	12.82	2.41		

However, despite the modifications, and the similar procedures used in the constructions of the tests, the presence of captioning had a significant effect for *The Graduate*, but not for *Airplane*.

Depth of Knowledge Post-Tests

A summary of the descriptive statistics and reliabilities for Study 2 (pilot study) and Study 3 (main study) are shown in Tables 6.3 and 6.4. The Study 3 captioning vs. normal means are also displayed in plot form in Figures 6.1 and 6.2, after the respective

Table 6.3. Summary of *Airplane* Descriptive Statistics for *Depth of Knowledge Tests*, Study 2 (pilot) and Study 3 (main study): 35 Target Words (V = Viewing Condition: Captioned, CC vs. Normal, N)

	Study	Class	Test	V	Mean	SD	α
Airplane	2	A	Story Partial-Cloze	CC	20.96	4.33	.62
Pilot	2	A	Synonym	CC	28.61	4.85	.89
(n = 43)	2	A	Non-Synonym	CC	31.90	3.44	.85
Airplane	3	E	Story Partial-Cloze	CC	21.82	4.78	.89
(n = 47)	3	E	Synonym	CC	30.41	3.54	.79
	3	E	Non-Synonym	CC	31.66	4.74	.91
Airplane	3	D	Story Partial-Cloze	N	13.04	4.13	.89
(n = 29)	3	D	Synonym	N	29.48	4.11	.79
	3	D	Non-Synonym	N	32.21	2.91	.91

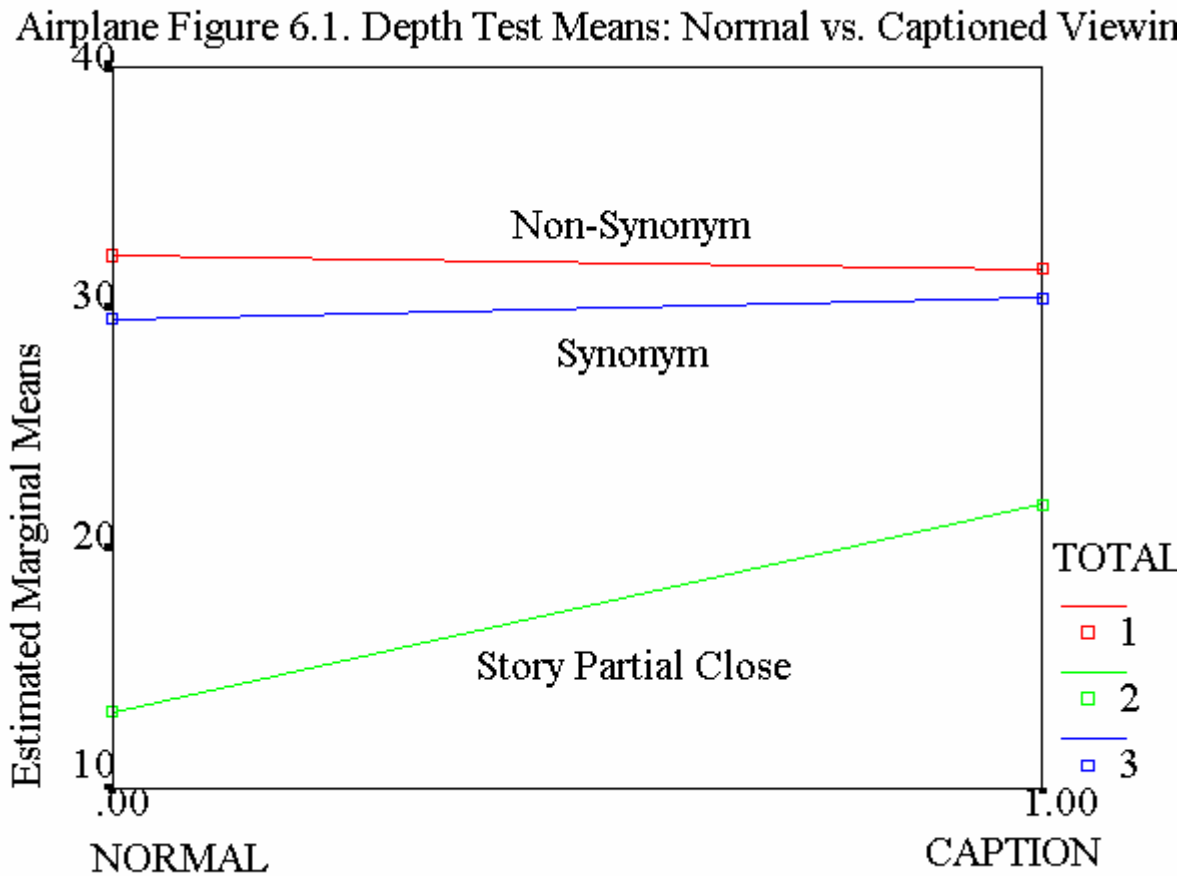
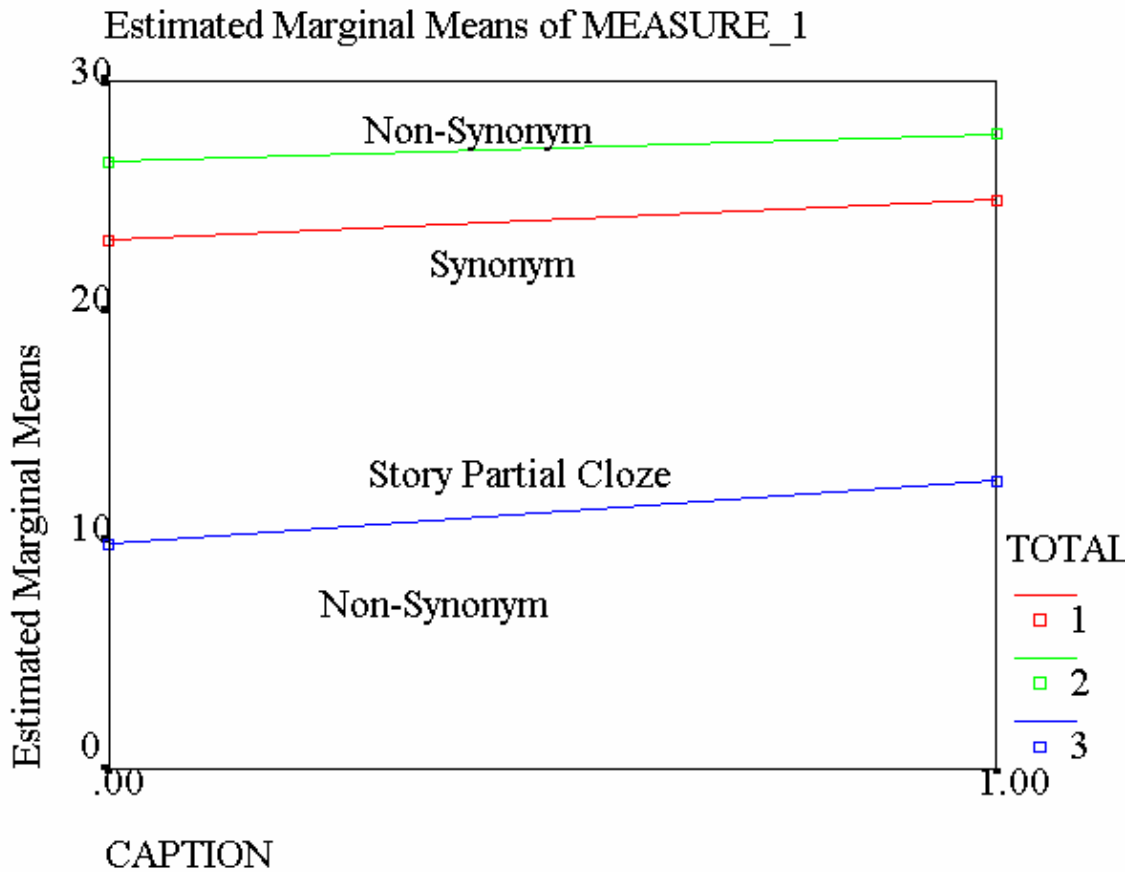


Table 6.4. Summary of *The Graduate* Descriptive Statistics for *Depth of Knowledge Tests*, Study 2 (pilot) and Study 3 (main study): of 30 Target Words (V = Viewing Condition: Captioned, CC vs. Normal, N)

	Study	Class	Test	V	Mean	SD	α
Graduate	2	B	Story Partial-Cloze	CC	16.60	4.33	.62
Pilot	2	B	Synonym	CC	25.46	2.84	.65
(n =47)	2	B	Non-Synonym	CC	27.80	2.13	.67
Graduate	3	C	Story Partial-Cloze	CC	12.56	6.08	.88
(n =38)	3	C	Synonym	CC	25.39	2.55	.69
	3	C	Non-Synonym	CC	27.94	1.76	.66
Graduate	3	F	Story Partial-Cloze	N	9.13	4.15	.88
(n =40)	3	F	Synonym	N	22.50	3.35	.69
	3	F	Non-Synonym	N	26.20	2.72	.66



tables. As the target words for all the tests for *Airplane* and *The Graduate* were based on the *Story Partial-Cloze Tests*, the only change made in the *Story Partial-Cloze Tests* was to increase the number of letters provided in four-letter target words from one to two. Between Studies 2 and 3, no major modifications had been made to the *Synonym* or *Non-Synonym Tests*. However, there were two different means scores: for the *Airplane Synonym Test*, Class A 28.61 vs. Class E 30.41; and for *The Graduate Story Partial-Cloze Test*, Class B. 16.60 vs. 12.56.

The major differences between Studies 2 and 3 were the considerably increased reliabilities of the *Story Partial-Cloze Tests* (*Airplane*, $\alpha = .62$ vs. $.89$; *The Graduate*, $\alpha = .62$ vs. $.88$), and of the decreased reliability of the *Airplane Synonym Test* ($\alpha = .89$, vs. $.79$).

The results of the Study 3 *Airplane* MANOVA found that the effect of captioning was significant for the *Story Partial-Cloze Test*, $p = .001$, but not for the *Synonym* or *Non-Synonym Tests*, $p = .110$, $.941$, respectively. The results of *The Graduate* MANCOVA found that the effect of captioning was significant for the *Synonym Test*, $p = .001$, and might be for *Non-Synonym Tests*, $p = .009$. The effect of captioning was not significant for the *Story Partial-Cloze Test*, $p = .013$.

A Comparison of the Tests

Probably the most important general point to be made is that, if only the results of the *Multiple-Choice Comprehension Test*, and/or the *Synonym Test* and *Non-Synonym Test* had been used, the conclusion would have been that, for *Airplane*, the presence of

captioning had had no effect. However, the significant effect for captioning, $p = .001$, found for the *Airplane Story Partial-Cloze Test*, showed that possible conclusion to be false.

One puzzling difference has to do with why the reliabilities of the *Depth of Knowledge Tests* are consistently higher than those of the *Multiple-Choice Comprehension Tests*. The *Multiple-Choice Comprehension Tests* selected questions that identified the main events in the introductory sequences of *Airplane* and *The Graduate*. Similarly, the *Depth of Knowledge Tests* selected target words considered to be important, or interesting. Yet, the better of the two *Multiple-Choice Comprehension Test* reliabilities: that of *Airplane* is $\alpha = .69$. This only equals the weakest of the *Depth of Knowledge Test* reliabilities: those of *The Graduate Synonym Test* and *Non-Synonym Tests*, $\alpha = .69$ and $.67$. It is possible that the amount of detail contained in the introductory sequence is too great for the viewers' to grasp in the viewing of introductory sequences that were only rerun once. Therefore, in the *Multiple-Choice Comprehension Test* individual differences in perception and in ability may have played a greater role, than in target words questions, such as those of the *Depth of Knowledge Tests*.

Of all the tests, only the *Airplane* and *The Graduate Story Partial-Cloze Tests* had very high reliabilities for both tests. There are a number of possible explanations. First, the rich story context enabled participants to draw on various parts, which they may have partially comprehended, whereas the *Multiple-Choice Comprehension Test* questions were separate items. Second, the context and the required production reduced the

guessing that occurred in all the other tests. Third, as the test is more difficult, there was a wider distribution of participants' scores.

This third point also may explain why the *Non-Synonym Tests* results were not significant. The negatively skewed distributions indicated that the tests were too easy for most of the learners. However, as stated in the sections in Chapter 4 and 5 on assumptions, ANOVAs and MANOVAs are fairly robust to the assumptions of normality. This skewedness does not contravene rules of usage but does result in a possible significant effect being missed. Therefore, groups with less vocabulary knowledge should have results with a more normal distribution, and may also reveal significant differences for the effect of captioning.

Three main conclusions can be drawn from the discussion in this section which has compared the tests. First, it is difficult to construct a range of tests in which there will be neither a positive nor a negative skew present in one of the tests: for example, the *Story Partial-Cloze Test* given to a group with a lower proficiency, or less vocabulary knowledge. Second, in the literature review, mention was made of the emphasis in education in Japan on increasing the amount of vocabulary recognized (breadth), rather than improving knowledge and usage of words already known (depth). The wide range of results may be an indication of one effect of the policy. Third, to accurately assess the effect of captioning, a number of different tests must be given. Differences between the two introductory sequences, or genre, are discussed in the next section.

Effect of Genre

The previous section has shown that the occurrence of significant differences, and of test reliabilities, was not consistent across the two introductory sequences of *Airplane* and *The Graduate*. To understand why results are not consistent, it is best to first summarize the Study 1 findings on the nature of the sequences and the students' *Idea Units*. Then, their possible effects upon the construction of the *Airplane* and *The Graduate Multiple-Choice Comprehension Tests* and the *Depth of Knowledge Tests* will be discussed.

The Introductory Sequences

In comparing the introductory sequences of *Airplane* and *The Graduate*, there were four findings. First, the number of scenes was different: thirty-two for *Airplane*, compared with seven for *The Graduate*. Second, the number of words in the dialogues was similar: 1210 vs. 1153. Third, for the captioned versions of the dialogue, less modification of the original dialogue was necessary for *Airplane* than for *The Graduate*: 1164 vs. 878. Fourth, the *VocabProfile* analysis of *Airplane*, and *The Graduate*, introductory sequences indicated that the vocabulary content of *Airplane* was more difficult than that of *The Graduate*.

Effect of Modification. More modification was required for *The Graduate* than for *Airplane*. Theoretically, the larger the differences between the dialogue and the captioning become, the more negative effects should be found, such as slower verbal

comprehension, or a focus only on the non-verbal input. This is because, where there are differences, learners can only pay attention to the visual/verbal (captioning) or to the aural/verbal (dialogue) other, but not to both. However, more significant differences in the *Multiple-Choice Tests* and *Depth of Knowledge Tests* were found for *The Graduate*, than were for *Airplane*. Therefore, for the proficiency level, or vocabulary knowledge level of the participants in these studies, any effects resulting from modifications were less than effects resulting from the presence of the captioning itself, or of the nature of the discourse. The main reason for the lack of effect of modification is that key content vocabulary is not removed. Most frequently modified are syntax, peripheral noun phrases, and repetitions.

VocabProfile Analysis. Analyzing the percentages, *The Graduate* had a higher amount of vocabulary from the *1000 Word List*; therefore, it was easier for students to understand. Therefore, it is likely that some parts of *Airplane* were beyond the comprehension of most students. However, what cannot be known from this analysis is what parts of both introductory sequences were perceived as important, or interesting. Furthermore, the analysis does not allow a measure of how much these were influenced by the larger number of scenes and accompanying visual input.

Idea Units

There were two main findings. First, despite a large difference in the number of scenes, the number of students' *Idea Units* (important or interesting facts in the

introductory sequences) was almost the same: fifteen vs. sixteen. Second, multidimensional scaling revealed that the students considered the discourse of *The Graduate* to be more complex than *Airplane*. Whether it is more complex when measured by more traditional methods is a separate question. The learners' perception, and the relationship between the students' *Idea Units* and scenes and tests is examined in the next two sections. In these cross-analyses there are two considerations. The first consideration is the effect of proficiency upon the *Idea Units* that students regard as interesting or important. It is possible that students regarded scenes, or parts of scenes which they did not understand, as being less important or interesting. The second consideration is the extent to which test questions should be linked to the *Idea Units*, if at all.

Cross-analysis of Idea Units and scenes. The introductory sequence of *Airplane* had thirty-two scenes, but students identified only sixteen ideas. A cross-analysis of scenes and *Idea Units* reveals two patterns of scenes that have no associated *Idea Units*. First, there were short scenes, which are fillers, but have no dialogue. These scenes reinforced, or introduced, a theme: for example, Scene 8, the ambulance by the plane, which reinforces the mini-story of the little girl who needed an operation; or Scene 15, Ted's first sight of the plane with accompanying fear-instilling music, as an introduction to the next scene of Ted's flashback to a disastrous bombing mission. Second, there were longer scenes which had no apparent importance or impact. These were: Scene 12, the cockpit - captain and crew (175 words); Scene 17, the cabin - two passengers talking in

jive (149 words); Scene 26, the cockpit - captain and crew (61 words); Scene 29, cockpit and cabin - captain talking to passengers (48 words).

The introductory sequence of *The Graduate* had seven scenes, but students identified sixteen ideas. Students recorded *Idea Units* for all of the scenes. The longest scene, Scene 3, in the living room, also had the most *Idea Units*: four. Two scenes had only one *Idea Unit*: Scene 2, outside Mrs. Robinson's house, and Scene 5, on the stairs and upstairs landing. Both of these scenes had a single theme. It was Mrs. Robinson's ploys to seduce Ben: first, by getting him into the house, and second, by getting him into a bedroom. The last two scenes were short (37 & 30 seconds), but together they contained five *Idea Units*: approximately one-third of the *Idea Units*. Specifically, the scenes were Scene 6, in Elaine's room, where Mrs. Robinson appeared naked, and Scene 7, in the living room, when Mr. Robinson returned home unexpectedly. This was because of the climactic nature of the discourse which concluded the introductory sequence of *The Graduate*.

Cross-analysis of Idea Units and Test Questions. For both *Airplane* and *The Graduate*, the test questions for both the *Multiple-Choice Comprehension Tests* and the *Depth of Knowledge Tests* were based on the discourse. However, they were not linked to, or based on, the students' *Idea Units* discussed in the previous section. Even if it had been argued that they should have been, this would have not been possible with Study 1. In Study 1, both classes, Class A which viewed *Airplane*, and Class B which viewed *The Graduate*, wrote down their *Idea Units* immediately after completing the *Multiple-Choice*

Comprehension Test questions. In fact, most questions *did* parallel the *Idea Units*. The ones which *did not* are mentioned below.

Concerning *Airplane*, for the *Multiple-Choice Comprehension Test* questions, there were no questions relating to the last three *Idea Units* (13-15), particularly the brief ten-second encounter in the airplane cabin between Ted and Elaine. Conversely, there were four questions on Scenes 21 through Scene 25 of the young sweetheart couple saying goodbye before their impending separation. Maybe, because this is a simple theme, it only produced one *Idea Unit*. For the *Depth of Knowledge Tests* target words, the words *cruising*, *feet*, and *temperature* were the only words which do not appear in any of the *Idea Units*. These are all from Scene 29, the captain addressing the passengers about the flight. The test results showed that they knew the words *feet* and *temperature*. The conclusion is that they did not regard this scene as important.

Concerning *The Graduate*, for the *Multiple-Choice Comprehension Test* questions, only Question 25 related to the last five *Idea Units*. Six of the target words from the *Depth of Knowledge Tests* related to the same five *Idea Units*.

Idea Units Summary. The first consideration was that of student proficiency. The relationship between *Idea Units*, and scenes and test questions, is easier to identify in *Airplane*, because of the larger number of scenes. In *Airplane*, the scenes which did not have *Idea Units*, the cockpit dialogues and the scene with jive English, did contain more slang and technical English. However, they also were not central to the tenuous central theme of the relationship between Ted and Elaine. For *The Graduate*, because each

scene is longer, it is more difficult to ascertain at what points the dialogue is not being understood at all, or is being misunderstood. The central theme of the discourse, Mrs. Robinson's attempted seduction of Ben, is more easily understood. The contextualization of this is considerably reinforced by non-verbal visual input. For both *Airplane* and *The Graduate*, one way to clarify this issue is to run the *VocabProfile* analysis on each scene. A second way is to create different levels of difficulty of *Multiple-Choice Comprehension Test* questions, and of the target words themselves.

The second consideration was the linking of the *Idea Units* to the test questions. For this, it is probably true to say that more notice should be taken of the students' perspective in the construction of tests. Irrespective of proficiency, a closer linking of *Idea Units* to test questions would probably result in a more significant effect for the presence of captioning.

However, concerning the comments and recommendations made about the considerations, there are two dangers. The first danger is that the main aim of this research has been to examine the overall effect of captioning upon comprehension and acquisition. This was why, for both films, as far as possible, the *Multiple-Choice Comprehension* and *Depth of Knowledge Tests* were based on the general stories, rather than specific scenes, or sections. The second danger is that the more specific and detailed test questions become, the more likely it is that something is missed simply because it is not considered important or interesting. Even L1 native speakers do not take in all the detail contained in such rich introductory sequences. When watching the beginning of

The Shining, this researcher, not being American, missed the reference to the Donner Party, an event famous in American history, but not in British or European history.

Effect of Genre: Summary. Between *Airplane* and *The Graduate*, the introductory sequences and the resulting *Idea Unit* patterns are clearly different. The term *marked* in linguistics refers to something which is less frequently found than the more commonly *unmarked*: for example the occurrence in languages of the consonant *v* compared with that of *b*, or the marked distinction found in English in describing the past, between the simple past, *I went*, and the present perfect, *I have gone*, a distinction not always present in other languages. It could similarly be argued that *Airplane* is more marked in terms of its discourse than is *The Graduate* (Strict film theory would argue that *The Graduate* does not have a genre). With such marked genres as *Airplane*, more attention needs to be given to the nature of introductory sequence and student *Idea Units*, than has to be given to *The Graduate*.

Conclusion to General Discussion

The measurements of the effects of captioning have already been discussed at length in this chapter. However, to summarize, the results indicated the importance of using a number of different tests in any one study. The four analyses used in Study 3 meant that the significance level had to be reset at $p = .0125$ at which point effects might be significant. Without these tests, particularly the *Story Partial-Cloze Test*, the conclusion would have been drawn for *Airplane* that the effect of captioning was not

significant. The focus would have then shifted to differences in the genre between the two introductory sequences of *Airplane* and *The Graduate*.

The development of the tests was time-consuming, and the administering of them highly problematic. Despite this, they were found to be reliable, especially the *Story Partial-Cloze Tests*. The *Depth of Knowledge Tests* have shown that it is possible to construct a framework of tests that can examine in two dimensions, breadth and depth, the effect of captioning upon comprehension and acquisition.

Table 6.5. *Airplane Depth of Knowledge Tests* MANOVA Univariate F Tests: Within-Participants. *Story Partial-Cloze, Synonym, and Non-Synonym Tests*; Between-Participants, Caption vs. Normal Viewing Condition

	Test	Hyp SS	Df	Hyp MS	F	p
DEPTH	Non-Syn vs. Synonym	267.45	1	267.45	13.46	.000
	Non-Syn vs. Partial Cloze	14059.68	1	14059.68	335.23	.000
	Syn vs. Partial-Cloze	10448.88	1	10448.88	231.34	.000
DEPTH	Non-Syn vs. Synonym	37.22	1	37.22	1.87	.176
CAPTION	Non-Syn vs. Partial Cloze	5351.47	1	5351.47	44.80	.000
	Syn vs. Partial-Cloze	1001.96	1	1001.96	22.18	.000
CAPTION	Normal vs. Caption	688.91	1	688.91	34.73	.000

Table 6.6. *The Graduate Depth of Knowledge Tests* MANCOVA Univariate F Tests: Within-Participants. *Story Partial-Cloze, Synonym, and Non-Synonym Tests*; Between-Participants, Caption vs. Normal Viewing Condition

	Test	Hyp SS	Df	Hyp MS	F	p
DEPTH	Non-Syn vs. Synonym	6.24	1	6.24	1.18	.280
	Non-Syn vs. Partial Cloze	491.83	1	491.83	22.26	.000
	Syn vs. Partial-Cloze	388.30	1	388.30	17.25	.000
DEPTH	Non-Syn vs. Synonym	9.81	1	9.81	1.87	.176
CAPTION	Non-Syn vs. Partial Cloze	47.01	1	47.01	2.13	.149
	Syn vs. Partial-Cloze	13.900	1	13.900	.62	.434
CAPTION	Normal vs. Caption	66.38	1	66.38	13.80	.000

Table 6.1. Summary of *Airplane* Multiple-Choice Comprehension Test Descriptive Statistics, One-Way ANOVA *p* Value (Classes C & F), and Reliabilities

Table 6.2. Summary of *The Graduate* Multiple-Choice Comprehension Test Descriptive Statistics, One-Way ANOVA *p* Value (Classes C & F), and Reliabilities

Table 6.3. Summary of *Airplane* Descriptive Statistics for *Depth of Knowledge Tests*, Study 2 (pilot) and Study 3 (main study): 35 Target Words (V = Viewing Condition: Captioned, CC vs. Normal, N)

Table 6.4. Summary of *The Graduate* Descriptive Statistics for *Depth of Knowledge Tests*, Study 2 (pilot) and Study 3 (main study): 35 Target Words (V = Viewing Condition: Captioned, CC vs. Normal, N)

Figure 6.1 Plot of *Airplane* Descriptive Statistics

Figure 6.1 Plot of *The Graduate* Descriptive Statistics

CHAPTER 6	187
GENERAL DISCUSSION	187
Introduction.....	187
<i>Study 1</i>	188
<i>Study 2</i>	189
<i>Study 3</i>	190
Studies 1, 2, and 3	191
<i>Effects of Captioning: A Comparison of the Tests</i>	192
Significance Levels.....	192
Multiple-Choice Comprehension Tests	192
Questions.....	193
Questions.....	193
Depth of Knowledge Post-Tests	193
A Comparison of the Tests	196
<i>Effect of Genre</i>	199
The Introductory Sequences	199
Idea Units	200
Conclusion to General Discussion.....	205