Chapter 3

Results and Analysis

3.1 Scoring and Analysis

3.1.1 Recall Scores

After the experimental stage was completed, the recall protocols were scored according to the number of idea units recalled (Carrell, 1985; Lee, 1986; VanPatten, 1990; & Bouden, Greenslade, & Sanz, 1999). The experimental text contained 47 semantic and syntactic idea units. Each subject's score was computed according to the raw number of idea units contained in the written recall (see Appendix F). A drawback of the original VanPatten (1990) and Bouden, Greenslade, & Sanz (1999) studies was that recall protocols were scored based on the researcher's subjective opinion that an idea unit found in the free writing recall was similar to one found in the experimental text, with no other way to verify that it was actually the same idea unit. In order to remedy this problem, a point system was developed to more rigorously score each individual idea unit. The selection of an individual idea unit within the free writing recall was based on the following criteria:

a). The similarity of the idea unit written in the free writing recall to that of one of the 47 idea units found in the experimental text (VanPatten, 1990; & Bouden, Greenslade, & Sanz, 1999).

b). The number of semantic heads, which are content words found within an idea unit that the idea unit's meaning is built around, and syntactic heads, which are syntactic categories found in an idea unit that the idea unit's phrase structure is built around, contained within a particular idea unit.

Example 1.

36) (countries) that opened their markets to global commerce

a) que abrieron algo. (Similar to Idea Unit 36, 33.3% of syntactic or semantic heads identified, 0 points)

b) que abrieron sus mercados. (Similar to Idea Unit 36, 66.6% of syntactic or semantic heads identified, 1 point)

c) que abrieron sus mercados al comercio. (Similar to Idea Unit 36, 100% of syntactic and semantic heads identified, 2 points)

After an idea unit had been identified, the number of semantic and syntactic heads found in the idea unit was counted. If the idea unit from the free writing recall contained less than fifty percent of the syntactic and semantic heads found in that of the idea unit from the experimental text such as in Example 1a, the subject received no points for that particular idea unit. If the idea unit from the free writing recall contained more than fifty percent but less than one hundred percent of the syntactic and semantic heads found in that of the idea unit from the experimental text such as in Example 1b, the subject received one point for that particular idea unit. If the idea unit from the free writing recall contained one hundred percent of the syntactic and semantic heads found in that of the idea unit from the experimental text such as in Example 1c, the subject received two points for that particular idea unit.

After the data from the eight task groups were scored, the number of recall units was determined, and the mean recall scores for each task group were calculated. This procedure is consistent with VanPatten (1990) and Bouden, Greenslade, & Sanz (1999) studies.

3.1.2 Text Scores

As in VanPatten (1990) and Bouden, Greenslade, & Sanz (1999) mean text scores for each group were determined by calculating the average of the number of target items

marked per task group. The target items were the content lexical item *commerce*, the grammatical item *–ing*, the non-content lexical item *the* marked by the subjects while reading the experimental text for content.

3.2 Results

3.2.1 Recall Scores

Intermediate level mean idea unitrecall scores are displayed in Table 4, and Advanced level mean recall scores are displayed in Table 5.

Table 4. Intermediate Level Mean Recall Scores

Task Group	n	Idea Unit	Std. Dev.
I-NoMarking	13	8.077	4.786
I-Content	14	6.643	4.568
I-BoundMorph	13	4.923	3.947
I-NonContent	12	6.250	3.864

Table 5. Advanced Level Mean Recall Scores

Task Group	n	Idea Unit	Std. Dev.
A-NoMarking	12	13.917	6.708
A-Content	12	14.667	7.011
A-BoundMorph	12	10.500	5.018
A-NonContent	13	11.154	4.652

At the intermediate level, I-NoMarking received the highest recall score, I-Content received the second highest recall score, I-NonContent received the third highest recall score, and I-BoundMorph received the fourth highest recall score. At the advanced level A-Content received the highest recall score, A-NoMarking received the second highest

recall score, A-NonContent received the third highest recall score, and A-BoundMorph received the fourth highest recall score.

For statistical analysis of intermediate recall scores , this study adopted an alpha level of $p<0.05$. At the intermediate level, the results of an ANOVA revealed no significant differences between tasks for the intermediate level recall scores ($F(3,48)=1.171$, $p<0.0001$). This suggests that the variation between tasks was not greater than expected by chance. The results of a post-hoc Tukey's .HSD revealed no significant differences between tasks at the intermediate level ($p<0.05$).

For statistical analysis of advanced recall scores , this study adopted an alpha level of $p<0.05$. At the advanced level, the results of an ANOVA revealed no significant differences between tasks for the advanced level recall scores ($F(3,45)=1.449$, $p<0.0001$). This suggests that the variation between tasks was not greater than expected by chance. The results of a post-hoc Tukey's .HSD revealed no significant differences between tasks at the advanced level ($p<0.05$).

3.2.2 Text Scores

Intermediate level text item-detection scores are displayed in Table 6.

Table 6. Intermediate Level Text Item-Detection Scores

Task Group	n	Item	Std. Dev.
I-Content	14	7.357	3.478
I-BoundMorph	13	8.923	2.813
I-NonContent	12	5.333	2.498*

I-BoundMorph yielded the highest text score, I-Content yielded the second highest text score, and I-NonContent yielded the lowest text score.

For a statistical analysis of intermediate text scores, this study adopted an alpha level of $p<0.05$. The results of an ANOVA reveal significant differences between tasks for intermediate level text scores ($F(2,36)=4.524$, $p>0.0001$). The results of a post-hoc Tukey's .HSD revealed a significant difference between the I-BoundMorph and I-NonContent text scores ($p>.05$) However, a comparison between I-Content and I-BoundMorph and between I-Content and I-NonContent yielded no other significant differences ($p<0.05$).

Table 7. Advanced Level Text Item-Detection Scores

Task Group	n	Item	Std. Dev.
A-Content	12	9.917	0.2887
A-BoundMorph	12	9.917	3.029
A-NonContent	13	8.231	1.964

Advanced level text scores are displayed in Table 7. The advanced level text scores demonstrated a similar pattern to that of the mean intermediate text scores. A-BoundMorph yielded the highest text score, A-Content yielded the second highest text score, and A-NonContent yielded the lowest text score.

For a statistical analysis of advanced text scores, this study adopted an alpha level of $p<0.05$. Unlike the intermediate text scores, the results of an ANOVA revealed no significant differences between advanced level text scores ($F(2,34)=2.750$, $p<0.0001$). This suggests that the variation between means for the text scores greater than expected by

chance. The results of a post-hoc Tukey's .HSD revealed no significant differences between tasks for the text scores at the advanced level ($p<0.05$).

3.3 Scoring and Analysis: Adjusted Recall Scores

3.3.1 Rational for Adjusted Recall Scores

A further drawback of the original VanPatten (1990) and Bouden, Greenslade, & Sanz (1999) studies was that the analyses and the results of these studies were based principally on the recall scores and not on the text scores. Although both studies submitted their text scores to an ANOVA and a Tukey's Test for Honest Significant Difference, they did not address how a lower or higher text score of a particular task group might affect that task group's mean recall scores.

For example, in this study, I-BoundMorph had a lower recall score than I-NonContent. However, I-BoundMorph had a significantly higher text score than I-NonContent. According to the recall scores it appears as though I-NonContent performed better than I-BoundMorph in regards to the number of idea units recalled from the experimental text. But, did I-NonContent acquire higher recall scores because that group marked less target items than I-BoundMorph, or did I-NonContent acquire higher recall scores than I-BoundMorph because the target item that was marked while reading the experimental passage for content caused less of a strain on attentional resources in working memory? To resolve this issue, the recall scores and the text scores must be combined in order to better compare the mean recall scores. So to reliably compare the recall scores between the experimental task groups, the text scores must be balanced and the recall scores must be adjusted according to the differences found between the mean text scores of

each task group. This must be done to account for the differences between the text scores in each experimental group.

3.3.2 Adjusted Recall Scores

The formula for the adjusted recall scores was developed while working with a statistician. Adjusted recall scores were calculated by multiplying the recall score of a particular task group by the text score of that same task group and then dividing that number by the highest average text score at a given level (see Example 2 for details). The recall scores I-NoMarking and A-NonContent were not adjusted because they did not receive the treatment, but were included in the adjusted recall scores in order to make comparisons between the control group and the experimental groups that received the treatment.

Example 2. Formula for Adjusted Recall Scores

Recall Score X Text Score/Highest Average Text Score=Adjusted Recall Score

3.4 Results of Adjusted Recall Scores

3.4.1 Adjusted Recall Scores

Intermediate level mean adjusted recall scores are displayed in Table 8 and Table 4 has been repeated in order to facilitate a comparison between intermediate level mean recall scores and intermediate level adjusted recall scores.

Table 4. Intermediate Level Mean Recall Scores

Task Group	n	Idea Unit	Std. Dev.
I-NoMarking	13	8.077	4.786
I-Content	14	6.643	4.568
I-BoundMorph	13	4.923	3.947
I-NonContent	12	6.250	3.864

Table 8. Intermediate Level Adjusted Recall Scores

Task Group	n	Idea Unit	Std. Dev.
I-NoMarking	13	8.077	4.786
I-Content	14	5.483	3.770
I-BoundMorph	13	4.923	3.947
I-NonContent	12	3.763	2.328*

I-NoMarking received the highest adjusted recall score, I-Content received the second highest adjusted recall score, I-BoundMorph received the third highest adjusted recall score and I-NonContent received the lowest adjusted recall score. The intermediate adjusted recall scores appear to demonstrate a similar pattern to that of the intermediate mean recall scores. I-NoMarking and I-Content yielded the highest recall scores for both the recall scores and the adjusted recall scores. Additionally, I-BoundMorph and I-NonContent yielded the lowest recall and adjusted recall scores. However, in the adjusted recall scores I-BoundMorph yielded the third highest adjusted recall score and I-NonContent yielded the lowest adjusted recall score. In the recall scores, I-NonContent yielded the third highest and I-BoundMorph yielded the lowest at the intermediate level.

For a statistical analysis of intermediate adjusted recall scores, this study adopted an alpha level of $p<0.05$. The results of an ANOVA revealed significant differences between

tasks for intermediate level adjusted recall scores ($F(3,48)=2.864$, $p>0.0001$). The results of a post-hoc Tukey's HSD revealed a significant difference between the I-NoMarking and I-NonContent text scores ($p>.05$) However, there were no significant differences found between any other task groups for the intermediate adjusted recall scores ($p<0.05$).

Table 5. Advanced Level Mean Recall Scores

Task Group	n	Idea Unit	Std. Dev.
A-NoMarking	12	13.917	6.708
A-Content	12	14.667	7.011
A-BoundMorph	12	10.500	5.018
A-NonContent	13	11.154	4.652

Table 9. Advanced Level Adjusted Recall Scores

Task Group	n	Idea Unit	Std. Dev.
A-NoMarking	12	13.917	6.708
A-Content	12	14.667	7.011
A-BoundMorph	12	10.500	5.018
A-NonContent	13	9.255	3.857

Advanced level adjusted recall score are displayed in Table 9 and Table 5 has been repeated in order to facilitate a comparison between advanced level mean recall scores and advanced level adjusted recall scores. A-NoMarking received the highest adjusted recall score, A-Content received the second highest adjusted recall score, A-BoundMorph received the third highest adjusted recall score and A-NonContent received the lowest adjusted recall score. The advanced adjusted recall scores appear to demonstrate a similar pattern to that of the advanced mean recall scores. A-NoMarking and A-Content yielded the highest recall scores for both the recall scores and the adjusted recall scores. Additionally, A-BoundMorph and A-NonContent yielded the lowest recall and adjusted

recall scores. However, in the adjusted recall scores, A-BoundMorph yielded the third highest adjusted recall score and A-NonContent yielded the lowest adjusted recall score. In the recall scores, A-NonContent yielded the third highest and A-BoundMorph yielded the lowest at the intermediate level. These patterns in the adjusted recall scores appear to hold for both proficiency levels.

For statistical analysis of advanced adjusted recall scores , this study adopted an alpha level of $p<0.05$. At the advanced level, the results of an ANOVA reveal no significant differences between tasks for the advanced level adjusted recall scores ($F(3,45)=2.559$, $p<0.0001$). This suggests that the variation between tasks was not greater than expected by chance. The results of a post-hoc Tukey's .HSD revealed no significant differences between tasks for adjusted recall scores at the advanced level ($p<0.05$).