A Statistical Modeling of the Correlation between Island Effects and Working-memory Capacity for L2 Learners

Euhee Kim
Computer Science & Engineering
School of IT Convergence Engineering
Shinhan University
Dongducheon City, Korea
euhkim@shihan.ac.kr

Myung-Kwan Park
Department of English
College of Humanities
Dongguk University
Seoul, Korea
parkmk@dgu.edu

Abstract

The cause of island effects has evoked considerable debate within syntax and other fields of linguistics. The two competing approaches stand out: the grammatical analysis; and the working-memory (WM)-based processing analysis. In this paper we report three experiments designed to test one of the premises of the WM-based processing analysis: that the strength of island effects should vary as a function of individual differences in WM capacity. The results show that island effects present even for L2 learners are more likely attributed to grammatical constraints than to limited processing resources.

1 Introduction

The role of memory in language learning has long received ample attention from researchers in first and second language acquisition (SLA) (Baddeley (1999), Ellis (2001), Juffs (2006)). At an intuitive level, it seems right to reason that individual differences among adult learners in their successful attainment of a second language (L2) are attributable to individual differences in memory capacity. In SLA, researchers have focused on short-term or working rather than long-term memory differences because they think short-term or working-memory (WM) plays a more instrumental role for individual differences in language development. The rationale for this belief is that WM is an on-line capacity for processing and analyzing new information (words, grammatical structures and so on). As a consequence, the bigger the on-line capacity an individual has for new information, the more information will settle into off-line, long-term memory.

In this paper we concentrate on Korean leaners of English (KLEs) to examine the correlation between their individual WM capacity and their knowledge of island constraints on *wh*-dependencies in English. To this end we adopt the methodology that Sprouse, Wagers, and Phillips (SWP) (2012a, b) use for L1 speakers.

2 Hypothesis Testing

The main focus of this paper is to examine the question of whether there is a correlation between KLEs' WM capacity and their knowledge of island constraints on *wh*-dependencies in English. In order to investigate this question, we need (i) a measure of WM capacity, and (ii) a measure of knowledge of *wh*-island constraints. The second measure is often termed a measure of 'island effects', which refer to the relatively low acceptability ratings given to sentences with a *wh*-dependency between a *wh*-phrase and its gap position inside select syntactic environments (cf. Ross (1967), Rizzi (1990), and Chomsky (1995) among many others). Given the foremost interest in the role of such variables as GAP-POSITION (i.e. where a gap is) and STRUCTURE (i.e.

29th Pacific Asia Conference on Language, Information and Computation pages 422 - 430
Shanghai, China, October 30 - November 1, 2015

whether island structure is involved or not) in the instantiation of island effects, we want to bring forth the following two hypotheses.

Table 1: Proposed hypotheses

(i) KLEs recognize the island effects of GAP-POSITION and STRUCTURE for each island type.
(ii) KLEs' recognition of the island strength for each island type correlates with their WM capacity.

3 Materials and Methods

To investigate the correlation between KLEs' perception of the strength of island effects and their WM capacity, we employed the participants and tasks described below.

3.1 Participants

Forty KLEs participated in this experiment for 10,000 Korean Won. The experiment was carried out during a single visit to the lab during which the participants completed the reading span task, the n-back task, and the acceptability-rating task (in that order).

3.2 The Acceptability-rating Task

The materials we used were adopted from SWP (2012a, b). They contained four island types: *Whether*, Complex NP, Subject, and Adjunct islands. For each type of island, gap/extraction site and structural environment were manipulated in a 2×2 factorial design. For example, the *Whether* island type/condition has the four levels/subtypes of the following kind:

(1) a. Non-island/Matrix
Who __ thinks that John bought a car?
b. Non-island/Embedded
What do you think that John bought __?
c. Island/Matrix
Who __ wonders whether John bought a car?
d. Island/Embedded
What do you wonder whether John bought __?

The 2×2 factorial design of each island effect as in (1) controls for the two syntactic properties of island-violating sentences: (i) they contain a long-distance *wh*-dependency, and (ii) they contain an island structure. By converting these two properties into the two main factors such as GAP-POSITION and STRUCTURE, each with two levels (for the first factor: Matrix and Embedded; for the second factor: Non-island and Island), SWP (2012a, b) defined island effects as a superadditive interaction effects that exist between two factors. Recall that the island effects are understood as the effects on acceptability of processing both long-distance *wh*-dependency and island structure contained in a single sentence like (1d) above (see Fodor (1983), Stowe (1986), Kluender (1998, 2004), and more recently Hofmeister & Sag (2010) for the studies on L1 processing of *wh*-dependencies; Juffs & Harrington (1995; 1996), White & Juffs (1998), Williams et al (2001), and Juffs (2005) for the studies on their L2 processing). In other words, the combined effects of the two factors are greater (i.e. superadditive) than the linear sum of the individual factors; that is, ((1a) - (1b)) + ((1a) - (1c)) < ((1a) - (1d)).

The acceptability-rating task using the materials was administered as a paper survey. The surveys were one hundred and twenty-eight token sentences long (8 token sentences for each level of an island type). The task was a 4-point scale acceptability-rating one where 1 represents ‘least acceptable’ and 4 represents ‘most acceptable’. The 4-point scale acceptability-rating task thus employs a continuous scale (the positive number line) for acceptability ratings (cf. Bard, Robertson, & Sorace (1996)). Participants were under no time constraints during the survey.

3.3 The Reading Span Task

The reading span (RS) task which was originally developed by Conway et al. (2005) was designed to assess participants’ WM capacity and was run using E-prime (Psychology software tools Inc.). In the version of the RS task we used, participants were tested on sets of sentences ranging from two to five sentences per set. There were three trials for each set size, totaling forty-two sentences for the entire task (3×(2+3+4+5)=42). Each item was composed of a complete sentence followed by a question mark and then a capital alphabet letter.

Participants read each sentence aloud, paused at the question mark, and answered ‘yes’ or ‘no,’ depending on the semantic plausibility of the sentence. After the answer, they were to read the capital letter aloud also. By pressing the space bar, they proceeded to the next item. After they reached the last sentence in a set, they were to see three question marks ('???') on the screen. They stopped at this point and wrote down each of the letters in the order in which they had appeared in the set. A sample set of three items is shown in (2).

(2) a. No matter how much we talk to him, he is never going to change.? J
b. The prosecutor's dish was lost because it was not based on fact.? M
c. Every now and then I catch myself swimming blankly at the wall.? F ???

The correct responses to the semantic plausibility questions are 'yes, no, no,' and one point was given for every letter correctly written in the correct order on the answer sheet (J, M, F).

3.4 The N-back Task

To get a more reliable measure of WM capacity, the version of n-back (NB) task developed by Ragland et al. (2002) was administered on top of the RS task. In this task, participants were shown a sequence of visual stimuli and they had to respond each time the current stimulus was identical to the one presented n positions back in the sequence. The stimulus material consisted of 20 different consonants in English. The upper case consonants were all shown in white and presented centrally on a black background for 500 ms each, followed by a 2000 ms interstimulus interval. Participants were required to press a pre-defined key ("ENTER") for targets, and their response window lasted from the onset of the stimulus until the presentation of the next stimulus (2500 ms); no response was required for non-targets. Participants were tested on 0-, 1-, 2- and 3-back levels in a pseudo-randomized order, with each level presented for 3 blocks, resulting in a total of 12 blocks. A block consisted of 15 + n stimuli and contained 5 targets and 10 + n non-targets each. The dependent measure was the proportion of hits minus false alarms averaged over all n-back levels.

In short, the results of data in our experiments are reported in Table 2.

Table 2: The descriptive statistics of the experimental data

	READING SPAN	N-BACK	ACCEPTABILITY
Min	.4800	3.083	1.00
1st Qu.	.5700	3.917	2.00
Median	.6400	4.167	3.00
Mean	.6645	4.171	2.89
3rd Qu.	.7450	4.500	4.00
Max.	.9300	4.917	4.00

4 Experiments and Results

4.1 The Syntactic Island Effects

In this section we report the formal acceptability-rating experiment that was used to quantitatively measure the target state for L2 learners' knowledge of island constraints on *wh*-dependencies in English. The acceptability ratings from each participant were z-score transformed. The z-score transformation was intended to eliminate the influence of scale bias on the size of the differences-in-differences (DD) scores (which are used to measure the strength of island effects) and therefore validate its comparison with the measure of WM capacity, which is the main focus in this paper.

The means and standard deviations for each condition (i.e. each of the island types) are presented in Table 3.

Table 3: The means and standard deviations for each condition (N = 40)

		Adjunct	Complex NP	Subject	Whether
Embedded	Island	-.61(.89)	-.70(.81)	-.86(.88)	-.85(.82)
	Non-island	-.72(.88)	-.27(.92)	-.46(.90)	.08(.92)
Matrix	Island	.30(.84)	.65(.64)	.39(.81)	.64(.61)
	Non-island	.51(.67)	.62(.66)	.52(.77)	.74(.56)

To test the first hypothesis (i) of Table 1, the question we examine with this set of data is whether the island effects for each condition are statistically present in the acceptability RATING. To answer this question, we constructed the linear mixed-effects regression models with GP (i.e. GAP-POSITION) and ST (i.e. STRUCTURE) as two fixed factors and with PA (i.e. participants) and ITEM (i.e. items) included as two random factors.

We assumed that fixed effects vary for all participants and items for each island type. In other words, we accounted for by-participant and by-item variations in overall acceptability ratings. So, what we need was random slope models, where participants and items have different intercepts, and where they also have different slopes for the fixed effects of the two factors.

4.1.1 The Interaction Plots for Each Island Type

We now turn to plots of the interaction (GP : ST; island effects) for each island type. The four panels in Figure 1 plotted the acceptability ratings for the four island types. Note that a superadditive effect is reflected statistically as an interaction, since the response to each level of one factor depends upon the level of the other. While linear additivity is visually identified by parallel lines, superadditivity is visually identified by nonparallel ones.

In the "cross-over" graph of the Adjunct island type in Figure 1, we see that the Island/Embedded group does better than the NonIsland/Embedded group. It is evident that ((1a) - (1b)) + ((1a) - (1c)) > ((1a) - (1d)). There is thus no superadditive interaction effect with Adjunct *wh*-dependencies in English when tested for KLEs.

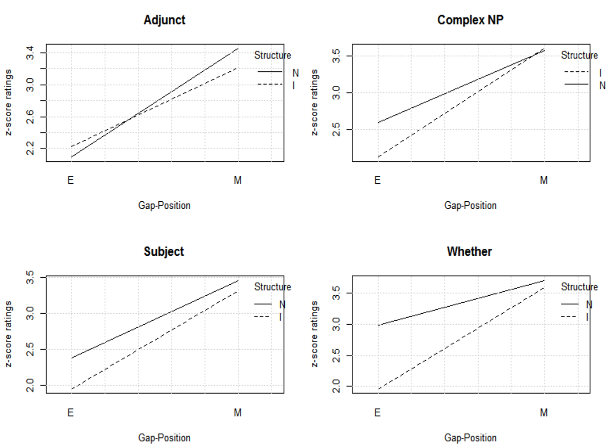

Figure 1: The interaction plots

In the "almost paralleling" graph of the Subject island type we see that ((1a) - (1b)) + ((1a) - (1c)) ≅((1a) - (1d)). So we cannot spot a superadditive interaction in the graph, either.

In the "almost intersecting at the level Matrix" graphs of the Complex NP and *Whether* island types, by contrast, we can spot superadditive interaction effects -- whenever there are no parallel lines there is an superadditive interaction present; in other words, ((1a) - (1b)) + ((1a) - (1c)) < ((1a) - (1d)). All in all, based on KLEs' acceptabiity ratings for the four island types, the graphs show that the island effects on acceptability are present for Complex NP and *Whether* islands, and absent for Adjunct and Subject islands.

4.1.2 The Selection of the Best Fit Regression Model on Island Types

To select a better fit regression model among simulated models, we used the lmerTest package for the statistical programming language R to perform a linear mixed effects analysis of the relationship between overall acceptability ratings and island effects.

What we need was a random slope model, where participants and items are allowed to have both different intercepts and slopes for the fixed effects. As fixed effects, we entered GAP-POSITION and STRUCTURE with an interaction term into the model. As random effects, we had intercepts for participants and items as well as by-participant and by-item random slopes for the fixed effects. The *p*-values were obtained by the likelihood ratio tests of the full model with the effects in question against the model without the effects in question.

With the 2x2 full factorial models for the island types, we constructed linear mixed-effects regressions, but, for lack of space, we won't describe them. Here's what we selected as the best fit model for the four island types:

Formula$_1$*:Rating~*
GP+ST+GP:ST+(1+GP+ST+GP:ST |ITEM)+
(1+GP+ST+GP:ST | PA).

4.1.3 The 2x2 Factorial Design Analysis

Using the lmer() method implemented in the lmerTest package, we estimated all *p*-values via the formula$_1$. Table 4 reports the *p*-values for main effects and the interaction effects of the formula$_1$.

The *p*-values for the coefficients of the interaction factor (GP_M : ST_N)[1] for the Adjunct and Subject island types are greater than the significance level (i.e. $p > 0.05$). Crucially, there

[1] In the description here and below, E and M refer to Embedded and Matrix (as two levels of GAP-POSITION), and I and N to Island and Non-island (as two levels of STRUCTURE), respectively.

are no significant interaction effects of GAP-POSITION and STRUCTURE for the Adjunct and Subject island types. Besides, the *p*-values for the coefficient of the interaction effects for the Complex NP and *Whether* island types are less than the significance level (p = .01581*; p =1.07e-07***). This experiment showed statistically significant interaction effects for the Complex NP and *Whether* island types.

Table 4: The fitted linear mixed-effects regression for the formula$_1$

Fixed Effects:

Type	effects	Estimate	SE	df	t-value	p-value
Adjunct	GP_M	.9149	.1540	36.74	5.940	7.76e-07***
	ST_N	-.1144	.1515	34.94	-.755	.455
	GP_M : ST_N	.3317	.2094	32.70	1.584	.123
Complex NP	GP_M	1.3609	.1369	44.11	9.941	7.80e-13***
	ST_N	.4289	.1310	40.11	3.273	.00219**
	GP_M : ST_N	-.4546	.1798	37.54	-2.528	.01581*
Subject	GP_M	1.2580	.1645	43.58	7.648	1.38e-09***
	ST_N	.3946	.1518	35.35	2.599	.0135*
	GP_M : ST_N	-.2688	.2119	34.04	-1.268	2133
Whether	GP_M	1.4981	.1231	44.78	12.167	8.88e-16***
	ST_N	.9406	.1234	44.77	7.620	1.28e-09***
	GP_M : ST_N	-.8405	.1310	41.37	-6.414	1.07e-07***

As predicted in the plots of interactions for each island type in Figure 1, the 2x2 factorial design analysis with a linear mixed-effects regression model reveals that KLEs recognize the island effects of GAP-POSITION and STRUCTURE for both Complex NP and *Whether* island types.

4.1.4 Pairwise Comparisons of Main Factors

However, because the interaction effects are present in the island STRUCTURE within the embedded GAP-POSITION, it is possible that the embedded island condition is driving these main effects. Therefore we performed the two pairwise comparisons on the embedded GAP-POSITION condition and the non-island STRUCTURE condition to test for each independent effect of STRUCTURE and GAP-POSITION.

Below, Table 5 shows the coefficients of linear mixed-effects regression models of the pairwise comparisons on STRUCTURE at the two island/embedded and non-island/embedded conditions for each island type when the lmer() method applied to the linear mixed-effects regression model with ST random slope:

$$Formula_2: Rating \sim ST + (1+ST|ITEM) + (1+ST|PA)$$

Likewise, Table 5 shows the *p*-values of the pairwise comparisons on the GAP-POSITION at the two matrix/non-island and embedded/non-island conditions when the lmer() method applied to the model with GP random slope:

$$Formula_3: Rating \sim GP + (1+GP|ITEM) + (1+GP|PA)$$

Table 5: The pairwise comparisons: STRUCTURE and GAP-POSITION

Pairwise Comparison

Condition	Factor	Type	Estimate	SE	df	t-value	p-value
GP==E (formula$_2$)	ST_N	Adjunct	.1144	.1779	16.515	.643	.529
		Complex NP	-.4289	.1413	18.909	-3.035	.0068**
		Subject	-.3946	.1445	17.938	-2.731	.0137*
		Whether	-.9406	.1246	36.330	-7.548	.0000***
ST==N (formula$_3$)	GP_M	Adjunct	-1.2465	.1293	31.69	-9.637	.0000***
		Complex NP	-.9063	.1526	18.468	-5.940	.0000***
		Subject	-.9892	.1436	21.313	-6.888	.0000***
		Whether	-.6575	.1094	30.872	-6.007	.0000**

As the above table indicates, the pairwise comparison on GAP-POSITION for each island type with embedded/non-island and matrix/non-island conditions shows that it reaches a statistical significance for each island type (p < .005). As expected, the length cost of gap position was isolated from the structure of non-island condition.

4.2 The Strength of Island Effects and Working-Memory Capacity

Now that we have seen that for L2 learners, island effects are robust in both Complex NP and *Whether* island types, the question is whether their awareness of the effects is attributed to constraints on the amount of WM capacity that any language user can have. This question gains more significance, as one account of *wh*-islands predicts that there is inverse relationship across language users between the strength of island effects and WM capacity (see Hofmeister and Sag (2010) among many others). We indeed tested this prediction for L2 learners.

We measured the strength of island effects by adopting the idea of a differences-in-differences (DD) score (Maxwell & Delaney (2003); SWP (2012a, b)). Intuitively, the DD score measures how much greater the effects of an island structure are in a long-distance dependency sentence than in a sentence with a local dependency. As it is calculated for each individual tested by using the acceptability-rating experiment, it serves as a measure of the superadditive component of the interaction for each individual and for each island type. Thus the score is thought of as the strength of island effects for that individual. More concretely, the DD score is calculated for a two-way interaction as follows. First, calculate the difference (D1) between the scores for two of the four levels. More specifically, we define D1 as the difference between the Non-island/Embedded and the Island/Embedded levels. Second, calculate the difference (D2) between the scores for the other two levels. For our purposes, D2 is the difference between the Non-island/Matrix and the Island/Matrix levels. Finally, calculate the difference between these two difference scores (i.e. D1 and D2) to produce a DD score.

We constructed a set of three linear regressions for each island type using DD scores and the WM capacity (i.e. reading span (RS) and n-back (NB) scores, which will be reported in the next subsections), as follows:

$Formula_4 : DD \sim RS$
$Formula_5 : DD \sim NB$
$Formula_6 : DD \sim RS + NB$

The first set of linear regressions was run on the set of all DD scores for each island type. The second set of linear regressions was run on only the DD scores that were greater than or equal to zero for each island type. The logic behind the second analysis is that DD scores below 0 are indicative of a sub-additive interaction. No theory predicts the existence of sub-additive interactions, which raises questions about how to interpret participants who produce sub-additive island effects. One possibility is that DD scores below 0 may reflect a type of noise that we may not want to influence the linear regression. If they are indeed noise, then eliminating these scores from the analysis should increase the likelihood of finding a significant correlation in the data. On the other hand, it is possible that these DD scores represent participants who truly do not perceive a classic superadditive island effect. In this case, including these scores should increase the likelihood of finding a significant correlation in the data. We report both analyses for these two possibilities

4.2.1 The Reading Span Task

Table 6 reports the results of the simple linear regressions: line-of-best-fit (intercept and slope), goodness-of-fit (R^2), and significance of the slope (t-statistic and p-value).

Table 6: $Formula_4$ for all DDs (DD≥0) (N = 40)

scores	Type	line-of-best-fit		goodness-of-fit	significance test	
		intercept	Slope	R^2	*t*-statistic	*p*-value
all DDs (DDs >=0)	Adjunct	-.4969 (.6607)	-.1495 (-.0024)	-.0022 (-.0294)	-.909 (-.014)	.3661 (.989)
	Complex NP	.4444 (.9500)	.2001 (.0747)	.0129 (-.0097)	1.439 (.632)	.1539 (.53)
	Subject	.4378 (.9524)	.0487 (-.1133)	-.0104 (-.0016)	.335 (-.947)	.7385 (.347)
	Whether	.8095 (1.1065)	.0475 (-.0418)	-.0115 (-.0135)	.333 (-.324)	.74 (.747)

The results in Table 6 concern the two sets of all DD scores and non-negative DD scores (i.e. values in parentheses) for each island type. On the first set of all DDs, three out of four slopes of the line-of-best-fit have positive slopes, but the slope for Adjunct island type has a negative slope. On the other hand, after removing negative DDs scores from the first set DDs, we see that the line-of-best-fit has three negative slopes and one positive slope for Complex NP island type.

The goodness of fit of the line-of-best-fit captured 0-2% of the variance in the data set, which is explained by the line for the four island types, as all the R^2 statistic absolute values were between 0 and 0.02.

Even after removing the potentially noisy DD scores, the four regressions for non-negative DD scores returned the lines with slopes that were not significantly different from 0 at the significance level ($p > 0.05$), thereby failing to reject the null hypothesis. In short, the results above indicate that there is no correlation between the all DD scores and the RS scores.

Figure 2 plots the relationship between the two sets of DD scores for each island type and the RS scores. The solid line represents the line of best fit for all of the DD scores. The dashed line represents the line of best fit when DD scores below zero are removed from the analysis. As predicted in Table 6, the solid and dashed lines for each island type behave like horizontal lines.

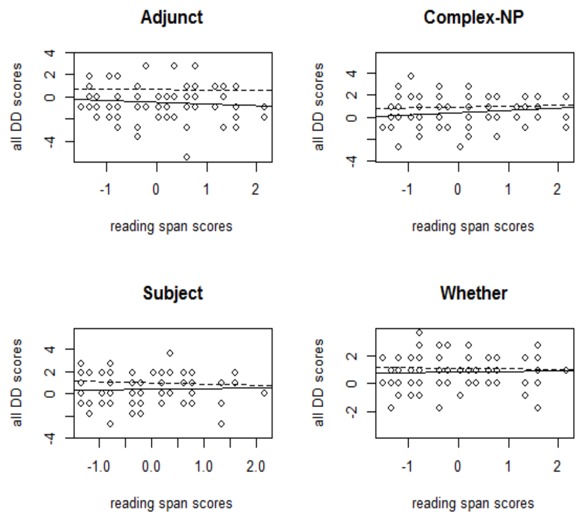

Figure 2: Plots for all DDs & RSs (N = 40)

4.2.2 The N-back Task

Table 7 shows that in the first set of all DDs, two out of four slopes of the line-of-best-fit have positive slopes, but the slopes for Complex NP and Subject island type are negative. On the other hand, after removing noisy scores from the first set DDs, we see that the line-of-best-fit has two negative slopes for Subject and *Whether* island type.

Table 7 shows that three of the four linear regressions of the set of all DD scores for Adjunct, Subject and *Whether* island types on the NB yielded R^2 statistic values that were approximately at 0, and the one for the Complex NP island type did so at .0167. Even after removing the noisy scores from the complete set of all DD scores, three island types such as Adjunct, Complex NP, and Subject have approximately zero R^2 statistic values, and *Whether* island type has it at -.0153. Because the goodness-of-fit of the lines was so extremely low, these results were not particularly meaningful for all DD scores.

The linear regression for four island types each returned the line-of-best-fit with a slope that was not significantly different from 0 at the significance level ($p > 0.1$) at the two sets of DD scores, thereby failing to reject the null hypothesis.

Table 7: Formula$_5$ for all DDs(DDs≥0) (N = 40)

scores	Islands	line-of-best-fit		goodness-of-fit	significance test	
		intercept	slope	R^2	*t*-statistic	*p*-value
all DDs (DDs >=0)	Adjunct	-.0938 (8797)	.1576 (.1173)	.0038 (-.0035)	1.138 (925)	.259 (.36)
	Complex NP	.3342 (1.0282)	-.2391 (.1106)	.0167 (-.0051)	-1.549 (848)	.1252 (.4)
	Subject	.1916 (1.2111)	-.1685 (-.1277)	-.0008 (-.0070)	-.962 (-.794)	.339 (.431)
	Whether	.7530 (1.1485)	.1534 (-.0094)	.0011 (-.0153)	1.044 (-.067)	.299 (.947)

Figure 3 plots the correlation between the set of DD scores for each island type and the NB scores. Each solid line and dashed line for each island type represents the line-of-best-fit with the intercept and slope. As predicted in Table 7, the solid line and dashed line for each island type behave like horizontal lines. Based on Figure 3, we can make a conclusion that there is no correlation between the NB scores and the DD scores for each island type.

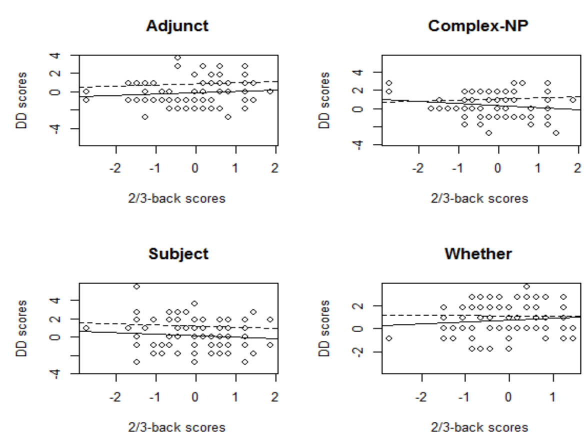

Figure 3: Plots for all DDs & NBs (N = 40)

4.2.3 Combining both RS and NB Scores

As a final analysis, we ran the multiple linear regression model for each island type, namely the $formula_6$, for the combined scores from both RS and NB tasks to ascertain if combining both scores of WM affects their relationship with the strength of island effects.

As Table 8 shows, even when doing the multiple regression analysis for the combined scores from both RS and NB tasks, there is no evidence of a significant correlation between WM and island effects. The four adjusted R^2 values of the regressions for all island types are at 0. After removing the noisy DD scores, the four adjusted R^2 values were improved and greater than 0. Although the regressions for all island types had the adjusted R^2 values that were slightly higher close to zero, their *p*-values for slope of NB and RS scores are not statistically significant ($p>0.05$), thus do not explain variation of the DD scores. Note that the *p*-value for slope of RS scores at the Complex NP type is statistically significant ($p<0.05$) after removing the noisy set of DD scores.

We draw the same conclusion as we did before, confirming that there is no correlation between WM scores and the DD scores for each island type even after combining the scores of both RS and NB.

Table 8: $Formula_6$ for all DDs(DDs≥0) (N = 40)

scores	islands	line-of-best-fit			goodness-of- fit	significance test	
		Intercept	slope (NB)	slope (RS)	R^2	*p*-value(NB)	*p*-value(RS)
ll DDs (DDs>=0)	Adjunct	-.2556 (.6316)	-.2533 (-.3411)	-.1262 (-.1385)	.0064 (.0804)	.212 (.0544)	.545 (.4448)
	Complex NP	.5264 (.5437)	-.1333 (-.3205)	.2028 (.3132)	.0032 (.0643)	.3388 (.0668)	.1788 (.0493*)
	Subject	.1745 (.7791)	.1285 (-.0513)	-.2544 (-.0325)	.0055 (-.0368)	.479 (.747)	.132 (.816)
	Whether	.9955 (1.5277)	-.2302 (-.0216)	.167 (.0539)	-.0035 (-.0300)	.238 (.89)	.328 (.715)

5 Discussion and Conclusion

In the previous literature on island effects in English and other languages there have been two diverging analyses for them: (i) the grammatical theory; (ii) the WM or processing resource capacity-based theory. The former grammatical theory predicts that the statistical GAP-POSITION : STRUCTURE interaction should not correlate with WM capacity measures, whereas the latter WM-based processing theory predicts that the interaction should correlate with such measures.

In this paper we reported three experiments that were designed to test for a correlation between the strength of the interaction and WM capacity. We used the acceptability-judgment task for the response scales, and two different types of WM measures (reading span and n-back), but found no evidence of a correlation between the statistical interaction and WM capacity. In fact, though Korean learners of English registered the GAP-POSITION : STRUCTURE interaction for the Complex NP and *Whether* islands, we didn't find evidence of their correlation with WM scores, refuting the main thesis of the WM-based processing theory. But this lack of the evidence is what is predicted by the grammatical theory of island effects. In short, the results of the experiments in this paper render strong support for a grammatical theory of island effects because we find no evidence of their correlation with WM or processing resource capacity.

Acknowledgement

We are grateful to the three anonymous reviewers for their comments and suggestions on the earlier version of the paper. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF- 2015S1A5A2A01010233).

References

Agresti, A. 2012. An introduction to Categorical Data Analysis (2nd ed.). Wiley & Sons, Inc, Hoboken, NJ.

Baayena, R. H., D. J. Davidsonb & D. M. Bates. 2008. Mixed-effects Modeling with Crossed Random Effects for Subjects and Items. Journal of Memory and Language, 59(4), 390-412.

Bard, E. G., D. Robertson & A. Sorace. 1996. Magnitude Estimation of Linguistic Acceptability. Language, 72, 32-68.

Chomsky, N. 1995. The Minimalist Program. MIT Press, Cambridge, MA.

Conway, A. R. A., M. J. Kane., M. F. Bunting., D. Z. Hambrick., O. Wilhelm., and R. W. Engle. 2005. Working Memory Span Tasks: A Methodological Review and User's Guide. Psychonomic Bulletin & Review, 12(5), 769-786.

Fodor, J. D. 1983. Phrase Structure Parsing and Island Constraints. Linguistics and Philosophy, 6, 163-223.

Hofmeister, P. and I. A. Sag. 2010. Cognitive Constraints and Island Effects. Language, 86, 366-415.

Juffs, A. 2005. The Influence of First Language on the Processing of Wh-movement in English as a Second Language. Second Language Research, 21(2), 121-151.

Juffs, A. 2006. Working-memory, Second Language Acquisition and Low-educated Second Language and Literacy Learners. LOT Occasional Papers: Netherlands Graduate School of Linguistics, 89-104.

Juffs, A. & M. Harrington. 1995. Parsing Effects in Second Language Sentence Processing: Subject and Object Asymmetries in Wh-extraction. Studies in Second Language Acquisition, 17, 483-516.

Juffs, A. & M. Harrington. 1996. Garden Path Sentences and Error Data in Second Language Processing Research. Language Learning, 46(2), 283-326.

Kluender, R. 1998. On the Distinction between Strong and Weak Islands: A Processing Perspective. In Culicover, P. and L. McNally (eds), Syntax and Semantics, vol. 29: The Limits of Syntax, 241-79. Academic Press, New York.

Kluender, R. 2004. Are Subject Islands Subject to a Processing Account? In West Coast Conference on Formal Linguistics (WCCFL) 23, 101-125.

Ragland, J.D., Turetsky, B.I., Gur, R.C, Gunning-Dixon, F., Turner, T, Schroeder, L., Chan, R., & Gur, R.E. 2002. Working Memory for Complex Figures: An fMRI Comparison of Letter and Fractal n-Back Tasks, Neuropsychology, 16, 370-379.

Rizzi, L. 1990. Relativized Minimality. MIT Press, Cambridge, MA.

Ross, J. 1967. Constraints on Variables in Syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Sprouse, J., M. Wagers & C. Phillips. 2012a. A Test of the Relation between Working-memory Capacity and Syntactic Island Effects. Language, 88, 82-123.

Sprouse, J., M. Wagers & C. Phillips. 2012b. Working-memory Capacity and Island Effects: A Reminder of the Issues and the Facts. Language 88, 401-407.

Stowe, L. A. 1986. Parsing Wh-constructions: Evidence for On-line Gap Location. Language and Cognitive Processes, 227-45.

White, L. and A. Juffs. 1998. Constraints on Wh-movement in Two Different Contexts of Non-native Language Acquisition: Competence and Processing. In Flynn, S., G. Martohardjono & W. O'Neil (eds), The Generative Study of Second Language Acquisition, 111-129. Erlbaum, Hillsdale, NJ.

Williams, J. N., P. Mobius & C. Kim. 2001. Native and Non-native Processing of English Wh-questions: Parsing strategies and Plausibility Constraints. Applied Psycholinguistics, 22, 509-540.