**Results of Power Analyses**

**Untimed Auditory Grammaticality Judgment Task (U-AGJT)**

Currently, there are no studies of L2 learning under incidental conditions that used a multivariate analysis of variance (MANOVA). This may be due to the fact that the studies often fall short of the sample size required to conduct such higher-order analytical procedure. Hence, power analyses were only conducted for univariate ANOVAs. As the original multivariate analysis in the study produced significant results for U-AGJT with all subsequent follow-up univariate counterparts showing paralleling results, this was considered as a negligible issue.

Despite the fact that there is an abundance of experimental studies on L2 learning under incidental conditions, Hamrick (2014) was the only one that adopted a trained control group (as was the case in the present study). Participants were 27 native speakers of English assigned to either the experimental (*n* = 14) or the control group (*n* = 13). During the training phase, they were exposed to 96 sentences of a semi-artificial language specifically devised for the study (see Hamrick, 2013 for materials). Learning was assessed immediately after the training by means of a recognition memory test, in which participants had to indicate whether they had heard each sentence before using a rating scale ranging from “1 (*I have seen this sentence before and I am very sure*) to 6 (*I have not seen this sentence before and I am very sure*) [italics in original]” (p. 5). The dependent variable was a recognition difference score computed by subtracting his or her mean rating for old items from that for new items. A univariate ANOVA with Group as a between-subjects factor showed a significant main effect of Group, *F*(1, 25) = 6.41, *p* = .01, *ηp2* = .21.1 Based on the partial eta-squared, the sample size required was computed using G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007). In the present study (*N* = 49), a series of univariate ANOVAs was conducted with Group as a between-subjects factor, following a significant multivariate effect of Group in MANOVA. The power analysis based on the effect size from Hamrick (2014) and the research design of the present study showed that each ANOVA required 51 participants. As the study recruited 49 participants (*n*experimental = 28 and *n*control = 21), which just missed the mark by two people, the bias in the ANOVAs caused by the small sample size of the study should be minimal. Note that the same can be said for the delayed posttest as well, because the same analytical procedure was applied for the immediate and the delated posttest.

**Word Monitoring Task (WMT)**

 Unlike U-AGJT, for which it was possible to come up with the only one sample size required, power analyses for WMT data can be far more complex. First, there is currently no study of learning under incidental conditions that conducted multivariate analyses. Thus, only univariate ANOVAs can be the subject of analysis here. Second, Godfroid (2016) is the only study that adopted WMT in the context of learning under incidental conditions. However, the dependent variable of the study was based on grammaticality sensitivity index (Ungrammatical RTs – Grammatical RTs) which is of different nature than the one in the current study, which compared raw RTs between grammatical and ungrammatical sentences as a within-subjects factor in ANOVAs. Hence, there was no study in the literature from which we can directly retrieve an appropriate effect size. As an alternative, we used an effect size reported in Leung and Williams (2012), which investigated learning under incidental conditions using raw RT measures but not with WMT. Their analysis involved only one within-subjects factor, Block, that consisted of two levels: Control and Violation (see below for more details). Note, however, that they did not recruit any control group. Third, G\*Power requires a correlation between within-subject measures (Grammatical RTs and Ungrammatical RTs in the present study) for computing an ideal sample size. It was expected that the correlation would vary across the two groups (Group), two testing sessions (Time), and six construction types (Construction). Hence, a separate analysis was carried out for each combination of Group x Test x Construction (*k* = 24).

 In Leung and Williams (2012), 33 participants went through an incidental learning condition in which they were simultaneously presented with two types of stimuli (i) two pictures of an object (e.g., a bull and a telephone) and (ii) an audio recording that denoted either of the pictures (e.g., *gi* bull). The target of the training was four artificial article determiners, *gi*, *ro*, *ul*, and *ne*, adopted from Williams (2005). The participants were explicitly instructed before the training exposure that the articles encoded distance of the object from the speaker (i.e., *gi* and *ul* for near objects and *ro* and *ne* for far objects) and that their task was to decide, as soon as they heard the recording, whether it referred to which one of the two pictures. Throughout the training phase, the presentation of the stimuli was structured such that knowing which articles encoded near or far objects would facilitate participants’ performance. Unbeknownst to the participants, however, these articles also referred to the animacy of the noun they modified (i.e., *gi* and *ro* for animate objects and *ul* and *ne* for inanimate objects).

The participants were first trained with 88 animate and inanimate objects used in different combinations of distance and animacy, amounting to 272 trials in total. 72 testing trials with novel objects followed thereafter, presented in three conditions: Transition (*k* = 8), Control (*k* = 32) and Violation (*k* = 32). In these conditions, the two pictures had objects that matched in distance but differed in animacy, which made it impossible for the participants to use distance as the basis for their decision-making. Of particular interest to the current power analyses, however, is the difference between Control and Violation block; while in Control block the recording (e.g., *gi* bull) agreed with the object (e.g., a bull near the speaker) in terms of animacy, it did not in Violation block (e.g., *ul* bull for a bull near the speaker). RTs of the response were used as a dependent variable of the study, and it was expected that the participants would slow down in Violation block, compared to Control block.

 A post-experimental interview revealed that 20 out of the 33 participants became aware of the animacy rule, with the remaining 13 participants subject to further analyses.2 The researchers conducted a univariate ANOVA with Block (Control vs. Violation) as a sole within-subjects factor, so as to test whether the unaware participants significantly slowed down on Violation block. The main effect of Block demonstrated a significant effect in the results, *F*(1, 19) = 6.66, *p* < .05, *ηp2* = .26, which confirmed their original hypothesis. Using this effect size estimate reported in Leung and Williams (2012), the sample size required for the present study was computed (see Table 1). As a result, the number of participants recruited in the present study was found to be sufficient to capture the effect of implicit learning on WMT, had there been any. We thus conclude that the non-significant results obtained in the current study (except for the case-missing violation at the delayed posttest) were due to the true null-effect of incidental exposure on developing implicit knowledge rather than the fact that the study lacked the sufficient sample size to reach significance. Furthermore, 95% CIs of mean raw RTs reported in Table S19 (immediate) and Table S21 (delayed) in Appendix H reinforce this conclusion because they overlap with each other to a significant degree across grammatical and ungrammatical item types for all construction types. It must be the case, therefore, that the participants did not develop robust implicit knowledge (as measured by WMT) to begin with, and the non-significant results on WMT just reflect this result.3

**Note**.

1. Hamrick (2014) in fact conducted a factorial ANOVA with Group as a between-subjects factor and Source as a within-subjects factor. Source had two levels, early or late, in each of which 18 first or 18 late items were presented from the exposure phase, respectively. As a main effect of Source and the interaction of Group and Source were not significant, these were dismissed from the current power analysis, *F*(1, 25) = 0.89, *p* = .35 and *F*(1, 25) = 0.22, *p* = .64 (effect sizes not reported), respectively.

2. They in fact conducted an analysis for both aware and unaware participants, but the latter pertains to our interests, as WMT was employed as a measure of implicit knowledge.

3. As reiterated in the main manuscript, this is not to deny that developing implicit knowledge from incidental exposure is impossible for adult learners in the long run. Note, however, that almost all of the previous studies only afforded as many exposure items as the present study at best, and in many cases, fewer.

**Table 1.** *The sample size required for WMT in the present study*.

|  |  |  |
| --- | --- | --- |
|  | Immediate | Delayed |
|  | Experimental | Control | Experimental | Control |
| OSV | 10 (.62) | 12 (.48) | 8 (.71) | 12 (.50) |
| OSIV | 7 (.77) | 11 (.56) | 13 (.42) | 11 (.55) |
| OSSVV | 5 (.89) | 7 (.75) | 21 (- .01) | 9 (.64) |
| OSSIVV | 8 (.69) | 9 (.67) | 10 (.62) | 11 (.57) |
| Case Mis | 10 (.63) | 12 (.47) | 13 (.42) | 8 (.70) |
| Case Mix | 6 (.84) | 13 (.43) | 10 (.58) | 10 (.62) |

*Note*. Experimental (*n* = 28) and Control (*n* = 21). Inside parentheses are correlations between the two within-subjects measures (i.e., raw RTs on grammatical and ungrammatical items).

**References**

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical

power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge

development. *Studies in Second Language Acquisition*, *38*(2), 177–215.

Hamrick, P. (2013). *Development of conscious knowledge during early incidental learning of L2*

*syntax*. Unpublished Ph.D. Dissertation. Georgetown University, Washington, D.C.

Hamrick, P. (2014). Recognition memory for novel syntactic structures. *Canadian Journal of*

*Experimental Psychology*, *68*(1), 2–7.

Leung, J. H., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-

meaning connections. *Language Learning*, *62*(2), 634–662.