**Supplementary Materials C: Effect Size Calculations**

**Phase I**

Cohen’s *ds*was estimated using:

                                          (1)

As Cohen’s *d*effect sizes tend to overestimate the population parameter with small samples, Hedges (1981) *g* correction was employed with the following equation (Lakens, 2013):

(2)

Recent L2 meta analyses (e.g., Lin & Lin, 2019; Bryfonski & McKay, 2019) have used between-subject *d*s/*g*s calculations for within-subject comparisons, which is somewhat justified because ds and *d*av generally produce near identical values (Lakens, 2013). However, we sought to employ a dedicated within-subject d metric to provide a model for future L2 researchers. Although Cohen’s *d*z, *dr*m, and *d*av are commonly employed by researchers for within-group comparisons (Lakens, 2013), *d*z and *dr*m were problematic for the present study. The formulae for these two effect sizes require the correlation between the repeated measures of the dependent variable (Cohen, 1988), but none the L2 IVA studies in the sample provided these values, and only three provided the relevant data to calculate the correlations (Huang, Kim, & Christianson, 2019; van den Broek, Takashima, Segers, & Verhoeven, 2018), while one provided enough information to estimate *d*z (Kaplan-Rakowski & Loranc-Paszylk, 2019). Furthermore, *d*z overestimates the true effect by reducing the standard error through the process by which the correlation is placed into its equation (Dunlap, Cortina, Vaslow, & Burke, 1996). While *dr*m corrects for this overestimation, Lakens noted that under certain conditions it is too conservative and recommended the use of *d*av instead. Thus, there were both practical and conceptual justifications for the use of *d*av to estimate within-subject effects, and the effect size was calculated with the following equation:

(3)

Hedges’ *gav*was calculated from the resulting *dav* effect sizes utilizing the same adjustment as for the *ds* effect sizes (2).

**Phase II**

The following formulae, which are transformations of those published by Brysbaert (2019), were utilized to calculate *ηp2* from the *g*s acquired in Phase I for (4) between- and (5) within-subject effects. The transformations were necessary because Phase II required the estimation of *ηp2* as opposed to *d*.

(4)

(5)

Although the case for utilizing *g*av in within-group meta-analytic inquiries is strong, within-group power calculations require the *d*z/*g*z effect size (Lakens, 2013). However, this metric requires the correlation coefficient (*r*) between the repeated measurements. Of the L2 studies featuring W-CB comparisons, only three provided enough information to estimate the required *r*coefficients. For each of the three studies, the observed correlations were averaged, resulting in three coefficients: *r* = .31 (Huang et al., 2019); *r* = .50 (Kaplan-Rakowski & Loranc-Paszylk, 2019); *r*= .68 (van den Broek et al., 2018; see Supplementary Materials E). These three *r*coefficients were utilized in equations to convert the median W-CB *g*av effect size (*g*av = 0.25) into three *g*zvalues with equation (6) from Brysbaert (2019):

(6)