**Table 2.** Independent Welch’s *t*-tests comparing the hierarchical and distinction conditions at the post intervention time point. Adjusted *p* values using Holm corrections.

|  |  | Distinction | | | Hierarchy | | |  | 95% CI | |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Outcome | *M* | *SD* | *n* | *M* | *SD* | *n* | Hedges’ *g* | Lower | Upper | *p* | *p*adj |
| Foody et al. (2013) | Discomfort | 24.76 | 16.97 | 18 | 15.43 | 19.09 | 18 | 0.51 | -0.16 | 1.17 | .131 | .392 |
| Foody et al. (2013) | Anxiety | 17.97 | 23.86 | 18 | 17.14 | 23.86 | 18 | 0.03 | -0.62 | 0.69 | .918 | .918 |
| Foody et al. (2013) | Stress | 21.71 | 21.21 | 18 | 12.18 | 19.09 | 18 | 0.46 | -0.20 | 1.12 | .166 | .392 |
| Foody et al. (2013) | Pooled\* | 21.48 | 20.88 | 18 | 14.92 | 20.80 | 18 | 0.31 | -0.35 | 0.96 | .352 | - |
| Foody et al. (2015) | Discomfort | 19.86 | 24.98 | 12 | 14.17 | 17.50 | 12 | 0.25 | -0.55 | 1.06 | .526 | >.999 |
| Foody et al. (2015) | Anxiety | 25.23 | 25.91 | 12 | 15.33 | 16.37 | 12 | 0.44 | -0.37 | 1.25 | .278 | .830 |
| Foody et al. (2015) | Stress | 17.23 | 17.79 | 12 | 13.92 | 17.82 | 12 | 0.18 | -0.62 | 0.98 | .654 | >.999 |
| Foody et al. (2015) | Pooled\* | 20.77 | 23.18 | 12 | 14.47 | 17.24 | 12 | 0.30 | -0.51 | 1.10 | .459 | - |
| *Notes:*  \* Pooled outcome measure were created for these reanalyses and were not reported in the original articles. | | | | | | | | | | | | |

## Assessment of whether intervals were likely to be 95% CIs or SEMs

Given that several of the previously reported reanalyses provide different results depending on whether we assume the intervals reported in Foody et al. (2013) were 95% CIs or SEMs, and the authors of Foody et al. (2013) are unable or unwilling to answer this question, it seems important to gather relevant information to aid our abduction here.

One useful source of information is what the authors’ of Foody et al. (2013) report in their other publications. Publications that came from the Barnes-Holmes research group around the time as Foody’s articles were published employed SEMs as error bars in their plots (e.g., Hughes & Barnes-Holmes, 2011). This includes one of my own studies which, I should note, also failed to clarify this in text (Hussey & Barnes-Holmes, 2012). My personal recollection is that we were generally trained by them to plot SEMs.

Another source of information is to compare the known standard deviations reported in previously published studies with the standard deviations extracted from Foody’s work under both assumptions of what the intervals represent (Foody et al., 2012; Foody et al., 2013; Foody, 2013). Given that the interventions involve an unknown degree of inter-individual heterogeneity, it is safer to compare the standard deviations of the baseline and post induction time points between studies.

There is good reason to believe that Foody et al. were aware that some form of pairwise comparisons between groups or timepoints beyond the RM-ANOVAs would be informative: because Foody, in her unpublished direct replication of their 2013 study (Foody, 2013, experiment 10) reported such pairwise comparisons between timepoints for each condition. – and claimed she found significant results where she did not (e.g., p.180).

Foody et al. (2013) do present mean change scores between time points for each of the visual analogue scales: “distinction resulted in a very small increase in discomfort (+.76), while hierarchy resulted in a decrease (-7.57)”; “Anxiety subsequently decreased for both conditions, although the larger change was recorded for the hierarchical intervention (distinction: -.03; hierarchy: -3.86)”; and “distinction resulted in an increase in stress (+4.71), while hierarchy reduced stress (-8.82).” (pp. 381-382). However, inferences about the population effect cannot be made on the basis of the sample means alone.