**Berkeley Replicating UVA Preregistration Study**

We worked with Luth Research to recruit 1,500 survey respondents. To maximize opportunity to find decline in our effects, per the Operations Manual, we collected these respondents in two rounds of 750, although, due to time constraints in communicating with the survey company, more than 1,500 participants were received. No data analyses took place until after the completion of both data collection for this study and data analysis for the other unblinded studies in this wave of the project.

Per the analysis plan developed by the originating school (UVA), we present three different analysis approaches. Per the assignment detailed in the Operations Manual, we first analyzed and report the first set 750 respondents, then the second. We conclude here with an analysis of the full set.

*First 750*

In the first half of data analysis, we received 824 compete responses. For the primary analysis, we conducted a 2 (Time: Pre-anagram vs. Post-anagram; within-subjects) x 2 (Condition: Pre-Registration vs. Hindsight; between subjects) mixed measures ANOVA on participants’ verbal ratings. We found a significant main effect of time, such that participants rated their verbal ability much higher before attempting the anagrams (M = 8.12, SD = 1.87) than after (M = 6.02, SD = 2.47), *F*(1, 822) = 637.16, *p* < .001. There was no significant main effect of condition, *F*(1, 822) = 0.70, *p* = .404. Contrary to predictions, there was also no interaction effect, *F*(1, 822) = 1.33, *p* = .249. Although there was no difference between conditions before attempting the anagrams, as predicted (Mprereg = 8.12, SDprereg = 1.87; Mhindsight = 8.13, SDhindsight = 1.87), *t*(822) = 0.12, *p* = .919, that null effect persisted into the ratings made after attempting the anagrams, contrary to predictions (Mprereg = 5.96, SDprereg = 2.55; Mhindsight = 6.16, SDhindsight = 2.38), *t*(822) = 1.17, *p* = .242. There was not differential dropout between conditions, *X*2(1, N = 1190) = 0.21, *p* = .645.

The secondary analysis of employing a difference score (post- minus pre-anagram) also failed to show conditional differences, *t*(822) = 1.15, *p* = .249, *d* = .08.

*Second 750*

In the second half of data analysis, we received 999 compete responses. For the primary analysis, we conducted a 2 (Time: Pre-anagram vs. Post-anagram; within-subjects) x 2 (Condition: Pre-Registration vs. Hindsight; between subjects) mixed measures ANOVA on participants’ verbal ratings. We found a significant main effect of time, such that participants rated their verbal ability much higher before attempting the anagrams (M = 8.03, SD = 1.34) than after (M = 6.06, SD = 2.45), *F*(1, 997) = 791.63, *p* < .001. There was no significant main effect of condition, *F*(1, 997) = 0.12, *p* = .732. Contrary to predictions, there was also no interaction effect, *F*(1, 997) = 2.41, *p* = .121. Although there was no difference between conditions before attempting the anagrams, as predicted (Mprereg = 8.06, SDprereg = 1.79; Mhindsight = 7.99, SDhindsight = 1.88), *t*(997) = 0.59, *p* = .556, that null effect persisted into the ratings made after attempting the anagrams, contrary to predictions (Mprereg = 5.98, SDprereg = 2.54; Mhindsight = 6.13, SDhindsight = 2.36), *t*(997) = 0.96, *p* = .336. There was not differential dropout between conditions, *X*2(1, N = 1402) = 1.24, *p* = .267.

The secondary analysis of employing a difference score (post- minus pre-anagram) also failed to show conditional differences, *t*(997) = 1.55, *p* = .121, *d* = .09.

*Full Set*

In total, then, we received 1,823 complete responses. For the primary analysis, we conducted a 2 (Time: Pre-anagram vs. Post-anagram; within-subjects) x 2 (Condition: Pre-Registration vs. Hindsight; between subjects) mixed measures ANOVA on participants’ verbal ratings. We found a significant main effect of time, such that participants rated their verbal ability much higher before attempting the anagrams (M = 8.07, SD = 1.85) than after (M = 6.06, SD = 2.46), *F*(1, 1821) = 1429.54, *p* < .001. There was no significant main effect of condition, *F*(1, 1821) = 0.63, *p* = .428. Closer to predictions, there was a marginal interaction effect, *F*(1, 1821) = 3.80, *p* = .051. Although there was no difference between conditions before attempting the anagrams, as predicted (Mprereg = 8.09, SDprereg = 1.83; Mhindsight = 7.99, SDhindsight = 1.88), *t*(1821) = 0.40, *p* = .687, that null effect persisted into the ratings made after attempting the anagrams, contrary to predictions (Mprereg = 5.97, SDprereg = 2.54; Mhindsight = 6.14, SDhindsight = 2.37), *t*(1821) = 1.50, *p* = .133. There was not differential dropout between conditions, *X*2(1, N = 2592) = 0.26, *p* = .610.

The secondary analysis of employing a difference score (post- minus pre-anagram) also showed marginal conditional differences, *t*(1821) = 1.95, *p* = .051, *d* = .09.