Pre-Analysis Plan

First set of research questions

In preparation for November 2, 2019 FitW Conference ([www.spelman.edu/fitw](http://www.spelman.edu/fitw))

1. Only consenting students and typical students will be considered for the main analysis (also see “External Validity” further below).
2. Assess attrition and whether there is differential attrition.
3. Establish baseline equivalence by performing difference-in-means tests across T and C for RCT 1 and T-T, T-C, C-T, and C-C for RCT 2:
	1. Student-level
		1. Covariates
			1. Highschool GPA
			2. Median SAT/ACT scores
			3. Pell eligibility
			4. First-generation student
		2. Outcomes
			1. Mean pre-test on metacognitive outcome
	2. Instructors
		1. Pre-exposure to metacognition
		2. Time at Spelman
		3. Has taught ADW before
4. For RCT 1, assess teacher outcomes for evidence of fidelity and then estimate the following equations using intent to treat:
	1. Y=b0+b1\*Treatment+b2\*Cohort2+e, where
		1. Y represents one of the three outcomes listed on the preregistration form; Treatment is a dummy variable that takes on the value 1 for T and 0 for C; Cohort2 is a dummy that takes the value 1 if students are in AY 2017-18 and 0 if in AY 2016-17; and e is clustered at the instructor/classroom level
	2. Y=b0+b1\*Treatment+b2\*Cohort2+b3\*Controls+e, where
		1. Controls represent covariates that are unbalanced across T and C at the student and/or instructor level and all else is as defined previously.
5. For RCT 2, assess peer-tutor outcomes for fidelity (as feasible) and estimate similar equations with Treatment defined to account for four conditions; i.e.,
	1. Treatment1=1 if T-T, 0 otherwise
	2. Treatment2=1 if T-C, 0 otherwise
	3. Treatment3=1 if C-T, 0 otherwise
	4. C-C is omitted.
6. Assess robustness of these findings to:
	1. Difference-in-difference specifications where feasible; e.g., pre versus post of metacognitive outcome across T and C.
	2. Multiple hypothesis test corrections using the wyoung command in Stata (which allows for the Bonferroni-Holm correction).
	3. Bootstrapping of standard errors.
7. Assess ex-post power; in particular, the post-study probability (i.e., updating of priors) following the procedures/steps in Maniadis et al. (AER, 2014).
8. External validity checks
	1. Assess balance across consenting and non-consenting students.
	2. Assess robustness of the above findings to inclusion of atypical students in the analysis.
	3. Assess how comparable the Spelman FitW sample is to the U.S. Department of Education College Scorecard sample (e.g., on SAT/ACT and Pell eligibility) to get a sense of whether resulting effects may be higher or lower elsewhere. In so doing, we will use the discussion by Kowalski (2019) as a guide.