README

Data and Code: ”Exercise Improves Academic Performance”

This folder contains:

1. An anonymized data set in STATA format: “gym_JPE_final_anonymized.dta”.
2. The STATA do-file to recreate the tables and figures in the paper and online appendix: “code_JPE_final_main.do”.
3. Four generated data sets to produce specific figures (datafigure3.xlsx, datafigure4.xlsx, datafigure5.xlsx, gymvisits_perweek.dta).
4. A zip file containing the latest version of the STATA program “mhtexp2-main” that enables multiple hypothesis correction using the Holm-Bonferroni respectively List-Shaikh-Vayalinkal.

To proceed with replication, first save the folder on your computer. Then open STATA and select the folder as your working directory. Open the do-file and run the commands for each table/figure step-by-step. Notice that you likely need to install the following STATA programs before you start:
ssc install “moremata”, “orth_out, “rwolf”, “leebounds”
net install “grc1leg2.pkg”.
To install “mhtexp2” place the mhtexp2.ado and mhtexp2.sthlp files in your stata installation’s "personal ado" folder.

IMPORTANT: To comply with requirements to preserve the anonymity of participants we had to remove the variable “age” (which was used as a control variable in the paper) from the publicly available data set. The regression results *with* control variables on the publicly available dataset can therefore vary slightly from the results reported in the paper. As shown in Table 1 below, however, removing the age variable from the regressions with control variables has no consequences for our main treatment effects reported in Table 3, 4, 6 and 7 of the paper. The point estimates are almost identical, and the significance levels remain unchanged.

Table 1: Main regression results with all controls (as in the paper) and without age (as in the anonymized data).

	Outcome variable									
	Gym		CSP		Lifestyle		Self-control		CSP: hetro	
	w/ age	w/o age	w/ age	w/o age	w/ age	w/o age	w/ age	w/o age	w/ age	w/o age
Treated	0.627***	0.628***	0.146**	0.148**	0.156**	0.156**	0.151**	0.151**	0.505***	0.514***
	(0.0678)	(0.0677)	(0.0664)	(0.0667)	(0.0641)	(0.0641)	(0.0624)	(0.0624)	(0.177)	(0.179)
Above									0.524***	0.528***
									(0.143)	(0.147)
Treated* Above									-0.439**	-0.446**
									(0.191)	(0.193)
Constant	0.715***	0.411***	1.863***	1.170***	-0.825***	-0.773***	3.317***	3.156***	1.659***	0.943***
	(0.250)	(0.115)	(0.322)	(0.164)	(0.243)	(0.108)	(0.249)	(0.101)	(0.332)	(0.190)
N	758	758	758	758	723	723	723	723	758	758

NOTE. —The table reports the main treatment effects from Table 3, 4, 6 and 7 in the paper (w/ age columns), and the same regression results not controlling for age (w/o age columns) Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$