Supplemental Online Materials

Quality of smartphone use measure

Filtering raw data. We filtered out observations from the raw screen on-off timestamps due to low data quality. We converted screen on-off observations to missing values when: 1) two consecutive observations were either 'on on' or 'off off', which indicates a data recording error; 2) observations and periods between observations that covered time where all other available phone sensors were turned off (Bluetooth, location, Wi-Fi); 3) observations that implied that the phone's screen was turned on for more than three hours. This filtering process removed 8 pct. of the screen activity observations.

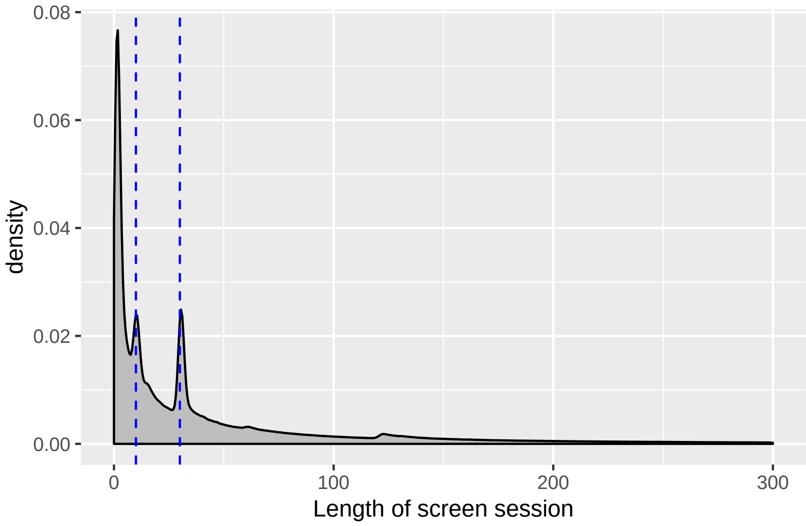

Fig. 1. Kernel density plot of the length of the recorded screen sessions in our study. We cut the x-axis at 300 seconds to make the plot more readable. 10 and 30 seconds are marked by the blue lines.

Active phone engagement. When smartphones are turned on by a touch or an external event, there is a default period of time before the screen turns off again. Thus, our *smartphone use in-class* measure could capture time without active engagement. Figure 1 plots the distribution of the length of our recorded screen sessions. Sessions are defined as periods from the moment that a screen is turned on until it is turned off again. We interpret the two spikes around 10 and 30 seconds as widespread configurations of how long the phone must be left alone until automatic turn-off. The spikes thus show all the times that a screen is quickly turned on and then immediately left alone until automatic turn-off. We therefore defined a long screen session as a screen session that lasts at least 35 seconds, and we define a student's average long smartphone use in-class as the percent of class time spent in long screen sessions averaged over all his/her courses. A linear regression with smartphone use in-class as the response and long smartphone use in-class as the predictor has $R^2 = 0.993$ and thus almost all screentime in-class is explained by long sessions.

Filter student and course observations

Student drop-out. Of the 810 students, only 752 ended up actually receiving a grade for at least one course.

Non-numeric grade data. We further removed observations from courses that only offered pass-fail grades. This procedure reduced data coverage from 752 students and 764 courses to 734 students and 735 courses.

Lacking class attendance. We removed observations, where the student had attended less than 10 hours of class time during the entire course. This removed the number of students from 734 to 674 and the number of courses from 764 to 657. We have confirmed that our results are robust to including these observations and rerunning all our models.

Missing background variables. We also excluded students with missing background variables. Specifically, 130 students lacked registered high school grade average, 109 students lacked registered parent information and 71 students lacked registered gender and age. After performing this procedure, our sample of students dropped to 517 from 674. Since the background variables were not used in some of the fixed effects models, we performed a robustness check of the main analysis by including the students with missing background variables in these models. As shown by Table 1, the resulting estimates were almost identical to our main estimates.

Table 1. Estimated association between smartphone use in-class on student course grades.

	Coefficient, $\hat{\beta}$	**95% confidence interval of coefficient, $\hat{\beta}$**	**Standardized coefficient**	**Adjusted $\mathbf{R}^2$**
(1a)	-0.129	$(-0.158, -0.100)$	-0.208	0.04
(1b)	-	-	-	
(1c)	-	-	-	-
(1d)	-0.043	$(-0.070, -0.016)$	-0.070	0.42
(1e)	-0.022	$(-0.048, 0.004)$	-0.036	0.57

Note: This table is identical to Table 5 from the main text, except that the models were estimated on a sample that includes students with missing background variables. This means that Model (1b) and (1c) cannot be estimated, since their predictors include the background variables. The *Coefficient* column displays the estimate of β from a linear regression defined by the *Specification* column. The standardized coefficient displays the coefficient estimated when variables where scaled to have unit standard deviation.

Ensuring variation within students and courses:. Our fixed effects models requires that every student attended multiple courses and each course were attended by multiple students in our sample. We filtered out students and courses that did not meet this requirement and ended up with sample of 470 students, 401 courses and a total 3,385 observations. The median number of courses per student were 7 and the median number of students per course were 6.

Validating our models

Testing non-linear models. We tested non-linearity of the association between smartphone use in-class and academic performance by re-estimating our models with a regression spline (Durrleman and Simon, 1989) on smartphone use in-class. We used a spline with 3 knots placed at the boundary between each of the quartiles and applied an F-test to compare the model with splines to the simpler model with a linear effect. Our statistical tests did not reject the assumption of linear models as we found $p > .30$ in all specifications. Our tests were performed for the panel model with student and course fixed effects (Model 1e), the pooled model with observed student characteristics (Model 1b) and lastly in the cross-section model (Model 2). Moreover, the linear models had smaller AICs than the non-linear models. We therefore conclude that it is reasonable to use the linear models.

Multicollinearity in fixed effects models. We verified that the student and course intercepts in our fixed effects models did not explain away all the variance of smartphone use in-class. This is essential for being able to accurately estimate an effect of smartphone use in-class on grades, after controlling for fixed student and course characteristics. The highest Variance Inflation Factor for smartphone use in-class was for the model with both student and course fixed effects with a value of 3.9; the second highest Variance Inflation Factor was found in the model with student fixed effects which had a value of 3.0. Thus, all our Variance Inflation Factors were below the recommended threshold of 5-10 (Sheather (2009); Kutner et al. (2004)).

Breaks between classes. DTU has breaks in teaching which normally occur either in the first or last quarter of each hour. We verified that our conclusions are robust to removing all class time scheduled in the first and last quarter of every hour by re-estimating our models with the modified smartphone use in-class measure.

Regression table:

Table 2 displays the estimated regression coefficients with 95% confidence intervals from the pooled model with observed student characteristics (Model 1b) and the cross-sectional model (Model 2).

Table 2. Regression coefficients

	Model 2: Cross-sectional model with student controls			Model 1b: Pooled model with student controls		
Variable	**Coefficient** $\hat{\beta}/\hat{\gamma}$	**CI (95%)**	**Standard. coefficient**	**Coefficient** $\hat{\beta}/\hat{\gamma}$	**CI (95%)**	**Standard. coefficient**
Smartphone use in-class	-0.099	$(-0.142, -0.056)$	-0.181	-0.081	$(-0.110, -0.051)$	-0.129
High school grade average	0.820	$(0.684, 0.956)$	0.495	0.826	$(0.684, 0.968)$	0.382
Age	0.205	$(0.061, 0.349)$	0.111	0.188	$(0.011, 0.366)$	0.076
Male	0.307	$(-0.252, 0.866)$	0.105	0.602	$(0.067, 1.137)$	0.159
Parents' Mean Income	0.011	$(0.000, 0.021)$	0.082	0.006	$(-0.002, 0.015)$	0.036
Parents' Max Years of Education	0.010	$(-0.089, 0.110)$	0.009	0.023	$(-0.070, 0.116)$	0.015
BMI	-0.019	$(-0.094, 0.055)$	-0.020	-0.023	$(-0.092, 0.046)$	-0.018
Smoker	0.004	$(-0.515, 0.523)$	0.001	0.062	$(-0.456, 0.581)$	0.016
Agreeableness	-0.177	$(-0.685, 0.330)$	-0.028	-0.243	$(-0.688, 0.202)$	-0.031
Extraversion	-0.015	$(-0.407, 0.378)$	-0.003	-0.097	$(-0.461, 0.267)$	-0.017
Neuroticism	0.193	$(-0.211, 0.597)$	0.042	0.106	$(-0.255, 0.468)$	0.018
Openness	0.347	$(-0.095, 0.789)$	0.061	0.102	$(-0.301, 0.505)$	0.014
Conscientiousness	0.481	$(0.047, 0.914)$	0.093	0.471	$(0.075, 0.866)$	0.070
Locus of Control	-0.068	$(-0.172, 0.036)$	-0.051	-0.101	$(-0.199, -0.003)$	-0.060

Note: The *Variable* column displays the names of the explanatory variables in the regression models. The three columns under the *Model 2: Cross-sectional model* header display the estimated regression coefficients, 95% confidence intervals and standardized regression coefficients from the cross-sectional model (Model 2). The three columns under the *Model 1b: Pooled model with student controls* header display the estimated regression coefficients, 95% confidence intervals and standardized regression coefficients from the pooled model estimated on panel data with observed students characteristics as controls (Model 1b).

References

Durrleman, S. and Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8(5):551–561.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models.* McGraw-Hill Irwin.

Sheather, S. J. (2009). *A Modern Approach to Regression with R.* Springer.