

Supplementary Material A

Data Collection Procedures for the Three Extant Studies Drawn from for This Study

Formal Setting, ESM: Science in the Moment. Each classroom was observed and videotaped on 10 different occasions (five days each in Fall and Spring semesters), totaling 100 hours. The observational and video data were coded using detailed coding schemes to provide information on features of the learning environment. Observers used brief global rating scales to describe the level of challenge, skill, engagement, and teacher autonomy support. For video data, lessons were coded based on broader categories of instructional practices (e.g., individual student seatwork) and further categorized by specific activities (e.g., review activity). During each day of filming, students participated in the study by responding to questions from pagers which signaled them twice per class period at randomly selected moments. Each student provided up to 20 ESM responses, for a total of 3,963 reports of students' situational states in science classrooms.

Informal Setting, ESM: STEM Interest and Engagement Study. The data collection methods included video and observation (162 hours total) that were linked to ESM, just like the *Science in the Moment* project, with the exception of slightly different periodicity of observations and signals. The coding of ESM signals matched the video data for modeling purposes. Students were signaled four times on each observation day, resulting in as many as 24 ESM responses per student containing self-reports of the three dimensions of engagement, for a total of 2,356 reports of students' situational states.

Formal setting, end-of-class-report: Incremental Mindset and Utility for Science Learning and Engagement. During each observation, a variant of the ESM was administered, collecting self-reported student ratings on the same engagement dimensions as the other projects. The methodology of this project deviated slightly from the other two in several respects. First, while multiple engagement ratings were collected from each student, this occurred only once per

class session (rather than 2-4 times, as in the other studies). Second, all information on classroom context comes from existing observational data (rather than coded video, as in the other studies). While video data were not collected, field notes and observer ratings provide data on features of learning environments. To establish complete comparability with the coding of instructional activity in the other two data sets, some minimal recoding of the activities already recorded in field notes was undertaken. Otherwise, all other contextual features to be examined already exist in usable form in the data set. Video from the other two studies was coded for comparability. Finally, this project involved classroom interventions, whereas the other two did not. Each student in the project provided up to 11 ESM responses for a total of 5,925 reports of students' states in science classrooms.

Supplementary Material B

Intraclass Correlations (ICCs) for Each Source of Variability for Engagement by Dimension

Outcome	Source	ICC [95% Credible Interval]
Affective	Individual	.354 [.339, .369]*
	Classroom	.049 [.032, .068]*
	Situation	.101 [.093, .109]*
Behavioral	Individual	.308 [.294, .322]*
	Classroom	.016 [.007, .027]
	Situation	.049 [.043, .055]*
Cognitive	Individual	.52 [.500, .540]*
	Classroom	.055 [.036, .076]*
	Situation	.036 [.031, .041]*

Note. The asterisk indicates that the credible interval does not overlap with the ROPE and can be considered statistically significantly greater than zero.

Supplementary Material C

Estimated Residual Correlations for the Three Dimensions of Engagement by Source of Variation

Source	Hypothesis	<i>M</i> Difference in Correlation (SE)
Program	Cor(Aff., Beh.) - Cor(Aff., Cog.)	0.006 (0.13), $p = 0.53725$
Program	Cor(Aff., Beh.) - Cor(Beh., Cog.)	0.095 (0.142), $p = 0.23875$
Individual 1	Cor(Aff., Beh.) - Cor(Aff., Cog.)	0.026 (0.019), $p = 0.90525$
Individual 1	Cor(Aff., Beh.) - Cor(Beh., Cog.)	0.146 (0.019), $p < .001$
Situation	Cor(Aff., Beh.) - Cor(Aff., Cog.)	0.439 (0.05), $p < .001$
Situation	Cor(Aff., Beh.) - Cor(Beh., Cog.)	0.089 (0.057), $p = 0.0585$
Overall	Cor(Aff., Beh.) - Cor(Aff., Cog.)	0.066 (0.009), $p < .001$
Overall	Cor(Aff., Beh.) - Cor(Beh., Cog.)	0.139 (0.009), $p < .001$

Note. Aff = affective; Beh = behavioral; Cog = cognitive.

Supplementary Material D

Estimated Intercepts for the Effect of the Dataset Fixed Effects

Dimension	Hypothesis	<i>M</i> Difference in Coefficient (SE)
Affective	IF-ESM - F-ESM	0.632 (0.094), $p < .001$
Affective	IF-ESM - F-DD	0.176 (0.084), $p = 0.016$
Affective	F-ESM - F-DD	0.456 (0.082), $p < .001$
Behavioral	IF-ESM - F-ESM	0.106 (0.072), $p = 0.932$
Behavioral	IF-ESM - F-DD	0.171 (0.062), $p = 0.005$
Behavioral	F-ESM - F-DD	0.276 (0.059), $p < .001$
Cognitive	IF-ESM - F-ESM	0.518 (0.106), $p < .001$
Cognitive	IF-ESM - F-DD	0.107 (0.094), $p = 0.871$
Cognitive	F-ESM - F-DD	0.411 (0.09), $p < .001$

Note. "IF" stands for informal learning environment; "F" stands for formal learning environment; "ESM" stands for experience sampling method; and "DD" stands for daily diary.