Supplementary Materials for "Can a Brief Online Intervention Raise Adolescents' STEM Expectancy and Values? A Four-Arm Randomized Trial of Goal Setting and Role Modeling"

This supplement contains Tables S1 to S20 and Figure S1, which are cited in the main article.

Table S1

Research Questions and Hypotheses

Label	Question, hypothesis, or analysis	Status
Panel A. Research questions		
RQ1	Do goal setting (GS) and role modeling (RM), alone or combined, improve students' overall STEM motivation immediately after the intervention?	Research question
RQ2	Do these effects differ by gender?	Research question
RQ3	Do these effects differ by school location?	Research question
Panel B. Confirmatory hypotheses		
H1	Students in conditions containing GS (GS only, combined) will report more adaptive overall STEM motivation than students in conditions without GS (RM only, active control). The same is predicted for RM (RM only and combined against GS only and active control).	Confirmatory (RQ1)
H2	Students in each intervention condition (GS only, RM only, combined) will report more adaptive overall STEM motivation than students in the active control.	Confirmatory (RQ1)
H3	The benefits of GS and of RM on overall STEM motivation will be larger for female than for male students. Where this moderation is observed, the H2 contrasts will be examined separately for female and male students.	Confirmatory (RQ2); follow-up conditional
Panel C. Exploratory analyses (no directional prediction)		
RQ1	Departure from additivity (RM × GS interaction)	Registered
RQ1	Combined condition against the average of the GS-only and RM-only conditions	Registered
RQ1	All RQ1 effects within biology, chemistry, and physics	Registered
RQ2	RM × GS × gender interaction	Registered
RQ2	Combined condition against the single-component average within each gender	Registered
RQ2	Gender moderation within biology, chemistry, and physics	Registered
RQ3	GS × location and RM × location interactions for overall STEM motivation	Registered
RQ3	Within-location arm contrasts	Added after registration
RQ3	Location moderation within biology, chemistry, and physics	Added after registration
RQ3	RM × GS × location interaction	Added after registration

Note. All models adjust for the baseline outcome, and cost is reverse scored, so higher scores indicate more adaptive motivation. Factorial comparisons average each full activity with its abbreviated version in the combined condition. Labels are this article's; the registration lists these tests without labels, and its third and fourth research questions

concern subject selection, which is reported separately. The registration specified the GS × location and RM × location interactions for overall STEM motivation; the within-location arm contrasts, subject-level location analyses, and RM × GS × location interaction were added afterward. GS = goal setting; RM = role modeling.

Table S2

Chatbot Specifications

Component	Specification
Identity and role	Helpful assistant within a website that educates students about careers in STEM; named “STEMGuide”
Tone	Friendly and helpful
Response length and language	Answers shorter than 40 words; simple language
Prohibited content	No discussion of anything illegal, sexual, or violent; scripted refusal: “My apologies, but that may be inappropriate to discuss”
Audience	All students under 18
Conduct boundaries	Inappropriate language or references beyond the school’s online code of conduct and ICT guidelines not responded to
Scope restriction	Questions beyond the content of the videos not responded to
Inclusivity	Language always inclusive and respectful
Context provided	Student has just watched a collection of videos; answers drawn only from the information contained in the provided videos
Fallback behavior	If no relevant information within the context: “My apologies, I don’t have the ability to answer that”; no made-up answers; never break character

Note. Each condition-specific chatbot used this specification and ran on GPT-4o (OpenAI); Table S3 describes the four chatbot types.

Table S3

Overview of Chatbot Types by Condition

Chatbot type	Purpose	Focus
1. Active-control chatbot	Baseline interactions without specific intervention features	General video/content questions
2. Goal-setting chatbot	Help students connect STEM to personal relevance and future goals	Everyday applications, career exploration
3. Role-model chatbot	Provide career guidance and pathway information	STEM careers, education requirements, skills needed
4. Combined chatbot	Integration of both goal-setting and role-model features	Comprehensive STEM exploration and career planning

Note. All four chatbots ran on GPT-4o (OpenAI) and followed the specification in Table S2.

Table S4

Condition Content and Planned Dose

Component element	Active control	GS only	RM only	Combined
Introductory videos (STEM importance, SMART goals, setting goals)	—	3 videos	—	1 of 3 (SMART goals)
Task-value prompt (open written response on STEM relevance)	—	1 prompt	—	1 prompt
Effort, help-seeking and prior-goals questions	—	3 items	—	3 items
SMART goal (5 elements, menu or own wording)	—	5 elements	—	5 elements (different menu options)
Worked SMART example (shown on the page)	—	1	—	1
Goal review (specific, timeframes, measurable, achievable, effort, realistic)	—	6 ratings	—	Not delivered
Strategy choices (distraction, difficulty)	—	2 items	—	Not delivered
Free-text rewrite of goal with if–then plan	—	6 fields	—	Not delivered
Closing items (obstacles, open to adjusting, confidence)	—	3 items	—	Not delivered
Role-model videos (ATSE, women in STEM, ≈ 3 min each)	—	—	5 videos	3 videos
Post-video career ratings (4 items per video)	—	—	20 responses	12 responses
Most inspiring video (single choice)	—	—	1 item	1 item
Active-control videos (TED-Ed)	4 videos	—	—	—
Post-video response items (e.g., engagement, attention, ease, effort, use of time)	12 items (3 per video)	—	—	—
Condition-specific chatbot (GPT-4o)	Available	Available	Available	Available
SEVT survey (12 items × 4 domains, T1 and T2)	48 × 2	48 × 2	48 × 2	48 × 2
Session length	35 min content	35 min content	35 min content	35 min content

Note. — = not part of that condition; “not delivered” = present in the single-component condition but omitted from the combined condition. The combined condition kept the SMART video and worked example but omitted the STEM-importance video, the third goal-setting video, goal review, strategy choices, the if–then rewrite, and closing items; its goal menus offered more options (67 vs. 41 across the five SMART elements, each including “Other”). With 35 minutes of content per condition, the combined condition received each component at reduced intensity, including three of the five role-model videos. All ATSE role-model videos featured women, and all students saw the same videos. Full materials are on the Open Science Framework (link withheld for anonymized review). ATSE =

Australian Academy of Technological Sciences and Engineering; GS = goal setting; RM = role modeling; SEVT = situated expectancy-value theory; SMART = specific, measurable, achievable, relevant, and time-bound; T1 = baseline; T2 = posttest.

Table S5

The Situated Expectancy-Value Theory (SEVT) Survey

Item wording	Construct	Source
1. I know I can learn the material in [subject].	Expectancy	Kosovich et al. (2015)
2. I believe that I can be successful in [subject].	Expectancy	Kosovich et al. (2015)
3. I am confident that I can understand the material in [subject].	Expectancy	Kosovich et al. (2015)
4. [Subject] is fun for me.	Intrinsic value	Gaspard et al. (2015)
5. I enjoy doing [subject].	Intrinsic value	Gaspard et al. (2015)
6. [Subject] is interesting.	Intrinsic value	Gaspard et al. (2015)
7. I think [subject] is important.	Utility value	Kosovich et al. (2015)
8. I value [subject].	Utility value	Kosovich et al. (2015)
9. I think [subject] is useful.	Utility value	Kosovich et al. (2015)
10. [Subject] requires too much time.	Cost	Kosovich et al. (2015)
11. I have to give up too much to do well in [subject].	Cost	Kosovich et al. (2015)
12. [Subject] is too stressful.	Emotional cost	Flake et al. (2015)

Note. Items were rated on a 7-point Likert-type scale from 1 (*strongly disagree*) to 7 (*strongly agree*), with the intermediate anchors *disagree*, *slightly disagree*, *neither agree nor disagree*, *slightly agree*, and *agree*. In each block, [subject] was replaced by biology, chemistry, physics, or STEM. Cost items (10 to 12) were reverse scored (8 − response) to form the low-cost factor.

Table S6

Internal Consistency and Standardized Factor Loadings for SEVT Constructs by Subject Domain and Wave

	T1 (baseline)				T2 (posttest)			
Construct	*α*	*ω*	*M* loading	Loading range	*α*	*ω*	*M* loading	Loading range
STEM overall								
Expectancy	.939	.939	.915	.905–.920	.950	.954	.934	.921–.950
Intrinsic value	.960	.959	.942	.920–.961	.961	.963	.947	.920–.972
Utility value	.929	.930	.903	.859–.936	.938	.946	.923	.880–.953
Low cost	.840	.841	.799	.786–.810	.889	.891	.855	.826–.880
Biology								
Expectancy	.944	.945	.922	.918–.927	.950	.955	.936	.928–.944
Intrinsic value	.955	.955	.935	.887–.962	.967	.969	.955	.912–.977
Utility value	.928	.930	.903	.872–.921	.951	.956	.937	.916–.948
Low cost	.833	.833	.790	.776–.812	.905	.906	.874	.856–.890
Chemistry								
Expectancy	.942	.942	.919	.906–.930	.953	.956	.938	.925–.950
Intrinsic value	.958	.958	.939	.899–.969	.960	.962	.945	.903–.970
Utility value	.926	.927	.899	.876–.912	.945	.949	.928	.904–.948
Low cost	.838	.838	.796	.784–.816	.903	.906	.873	.861–.890
Physics								
Expectancy	.935	.935	.910	.901–.918	.949	.952	.932	.912–.944
Intrinsic value	.947	.948	.926	.861–.965	.953	.955	.935	.882–.964
Utility value	.914	.916	.885	.843–.917	.935	.942	.918	.862–.948
Low cost	.842	.842	.800	.776–.813	.894	.897	.862	.844–.890

Note. $N = 855$; coefficients use available cases. α = Cronbach's alpha; ω = McDonald's omega. *M* loading is the mean standardized factor loading across the items indexing each construct, and the loading range gives the lowest and highest standardized loading. T1 = baseline; T2 = posttest. Cost items were reverse scored, so the factor is low cost. SEVT = situated expectancy-value theory.

Table S7

Mean Item Intraclass Correlations (ICCs) at Posttest by Construct and Subject Domain

Construct	STEM overall	Biology	Chemistry	Physics
Expectancy	.047	.058	.080	.044
Intrinsic value	.054	.057	.118	.048
Utility value	.055	.048	.058	.040
Low cost	.004	.006	.007	.004

Note. ICCs are for the 48 T2 items, from one-way random-intercept models with school as the grouping factor; baseline ICCs were similar. Values average the school-level ICCs of the three items for each construct in each domain. Across items, ICCs ranged from .000 to .133 (median = .048); clustering was highest for chemistry intrinsic value (about .09 to .13). ICC = intraclass correlation.

Table S8

Measurement Invariance of the SEVT Constructs

Construct	Model	Robust χ^2	*df*	CFI	TLI	RMSEA [90% CI]	ΔCFI	ΔRMSEA
Expectancy	Configural	248.07	176	.996	.994	.022 [.016, .027]	—	—
Expectancy	Metric	262.11	184	.996	.994	.022 [.017, .027]	.000	.000
Expectancy	Scalar	275.16	192	.996	.994	.023 [.017, .028]	.000	.000
Intrinsic value	Configural	304.28	176	.994	.991	.029 [.024, .034]	—	—
Intrinsic value	Metric	309.31	184	.994	.992	.028 [.023, .033]	.000	−.001
Intrinsic value	Scalar	325.09	192	.994	.991	.028 [.024, .033]	.000	.000
Utility value	Configural	265.88	176	.996	.994	.024 [.019, .029]	—	—
Utility value	Metric	272.10	184	.996	.994	.024 [.019, .028]	.000	−.001
Utility value	Scalar	303.90	192	.995	.993	.026 [.021, .031]	−.001	.002
Low cost	Configural	309.72	176	.987	.980	.030 [.025, .034]	—	—
Low cost	Metric	320.62	184	.987	.981	.029 [.025, .034]	.000	.000
Low cost	Scalar	345.62	192	.985	.979	.031 [.026, .035]	−.002	.001

Note. N = 855 (401 female, 454 male); cost models used N = 854 because one male student had no cost data. Each block included the biology, chemistry, physics, and STEM-overall factors at T1 and T2. In Panel A, metric models equated loadings across waves and scalar models also equated intercepts. In Panel B, the gender-configural model kept longitudinal scalar equality within gender, and the gender-metric and gender-scalar models then equated loadings and intercepts across gender. Δ values compare each model with the preceding one (— = no comparison). Invariance was retained when ΔCFI ≥ −.010 and ΔRMSEA ≤ .015; all constructs met these criteria at the scalar level. CFI = comparative fit index; CI = confidence interval; RMSEA = root mean square error of approximation; SEVT = situated expectancy-value theory; TLI = Tucker–Lewis index.

Table S9

Means, Standard Deviations, and Correlations Among the Study Variables for STEM Overall

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. T1 Expectancy	855	4.65	1.39	—								
2. T1 Intrinsic value	855	4.22	1.56	.77**	—							
3. T1 Utility value	855	4.88	1.48	.75**	.75**	—						
4. T1 Low cost	854	3.75	1.22	.16**	.22**	.09**	—					
5. T2 Expectancy	590	4.91	1.34	.76**	.70**	.72**	.26**	—				
6. T2 Intrinsic value	578	4.58	1.46	.68**	.78**	.67**	.31**	.83**	—			
7. T2 Utility value	573	5.16	1.35	.63**	.64**	.76**	.18**	.84**	.78**	—		
8. T2 Low cost	570	3.82	1.24	.24**	.24**	.12**	.61**	.17**	.22**	.10*	—	
9. Female (1 = female)	855	.47	—	−.13**	−.13**	−.03	−.08*	−.10*	−.11**	−.03	−.08*	—
10. Rural school (1 = rural)	855	.56	—	−.11**	−.11**	−.17**	.05	−.07	−.09*	−.08	.05	.03

Note. $N = 855$. Variables 1 to 8 are observed STEM-overall scale scores (mean of three items, 1 to 7); low cost is reverse scored. Correlations use pairwise deletion, so those involving T2 scores rest on 570 to 590 students. For binary variables, *M* is the proportion coded 1; female is coded 1 here, whereas the gender models (Tables S12–S15) code male as 1. Rural school is assigned from the school record. T1 = baseline; T2 = posttest. *$p < .05$. **$p < .01$.

Table S10

Exploratory Incremental Contrast: Combined Condition Versus the Average of the Two Single-Component Conditions

Outcome	Δ	*z*	*p*	95% CI
Expectancy	−0.03	−0.52	.601	[−0.14, 0.08]
Intrinsic value	−0.04	−0.67	.501	[−0.15, 0.07]
Utility value	−0.03	−0.44	.658	[−0.15, 0.10]
Low cost	−0.11	−0.64	.525	[−0.46, 0.23]

Note. $N = 855$. Δ = model-implied standardized mean difference in T1 baseline-SD units; positive values indicate that the combined condition exceeded the average of the GS-only and RM-only conditions. Low cost is reverse scored, so negative values indicate higher perceived cost. Standard errors and confidence intervals are cluster-robust. This contrast was registered as exploratory and is reported uncorrected. CI = confidence interval. GS = goal setting; RM = role modeling.

Table S11

Exploratory Subject-Specific Intervention Effects for Biology, Chemistry, and Physics

Outcome	Effect	β	*p*	95% CI	*q*
Biology: Expectancy	GS main effect	0.12	< .001	[0.07, 0.17]	< .001
	RM main effect	−0.07	.045	[−0.13, −0.002]	.090
	RM × GS	−0.24	.019	[−0.43, −0.04]	—
Biology: Intrinsic value	GS main effect	0.05	.008	[0.01, 0.09]	.019
	RM main effect	−0.10	< .001	[−0.16, −0.05]	< .001
	RM × GS	−0.22	.014	[−0.39, −0.04]	—
Biology: Utility value	GS main effect	0.07	.094	[−0.01, 0.16]	.126
	RM main effect	−0.01	.864	[−0.10, 0.08]	.901
	RM × GS	−0.13	.218	[−0.34, 0.08]	—
Biology: Low cost	GS main effect	−0.13	.003	[−0.22, −0.05]	.017
	RM main effect	−0.02	.642	[−0.10, 0.06]	.770
	RM × GS	0.09	.671	[−0.32, 0.50]	—
Chemistry: Expectancy	GS main effect	0.11	.005	[0.03, 0.18]	.017
	RM main effect	−0.10	.021	[−0.19, −0.02]	.051
	RM × GS	−0.18	.104	[−0.40, 0.04]	—
Chemistry: Intrinsic value	GS main effect	0.05	.078	[−0.01, 0.11]	.117
	RM main effect	−0.12	.002	[−0.19, −0.04]	.010
	RM × GS	−0.08	.453	[−0.28, 0.12]	—
Chemistry: Utility value	GS main effect	0.02	.735	[−0.08, 0.12]	.735
	RM main effect	−0.08	.139	[−0.18, 0.03]	.239
	RM × GS	−0.19	.106	[−0.42, 0.04]	—
Chemistry: Low cost	GS main effect	−0.09	.006	[−0.15, −0.03]	.017
	RM main effect	−0.05	.270	[−0.15, 0.04]	.360
	RM × GS	0.15	.482	[−0.27, 0.57]	—
Physics: Expectancy	GS main effect	0.08	.048	[0.001, 0.16]	.082
	RM main effect	−0.06	.018	[−0.10, −0.01]	.051
	RM × GS	−0.08	.505	[−0.31, 0.15]	—
Physics: Intrinsic value	GS main effect	0.08	.010	[0.02, 0.14]	.020
	RM main effect	−0.07	< .001	[−0.10, −0.04]	< .001
	RM × GS	−0.07	.233	[−0.19, 0.05]	—
Physics: Utility value	GS main effect	0.02	.683	[−0.08, 0.12]	.735
	RM main effect	−0.05	.248	[−0.14, 0.04]	.360
	RM × GS	−0.13	.420	[−0.46, 0.19]	—

Outcome	Effect	β	*p*	95% CI	*q*
Physics: Low cost	GS main effect	−0.03	.482	[−0.12, 0.06]	.579
	RM main effect	−0.01	.901	[−0.15, 0.13]	.901
	RM × GS	0.05	.801	[−0.35, 0.45]	—

Note. $N = 855$. β = factorial coefficient and Δ = planned contrast, each in T1 baseline-SD units. Goal setting and role modeling are effect coded (−0.5 = absent, +0.5 = present), so each main effect averages the full and abbreviated implementations of that component. All analyses are exploratory; the RM × GS rows (Panel A) and combined-versus-average rows (Panel B) are uncorrected (—). Low cost is reverse scored, so negative estimates indicate higher perceived cost. Standard errors and CIs are cluster-robust, with school as the cluster. *q* = Benjamini–Hochberg-adjusted *p* value within the 12-outcome subject-specific family, separately for each effect and contrast. CI = confidence interval; GS = goal setting; RM = role modeling.

Table S12

Gender Moderation of Intervention Effects for Overall STEM Motivation

Outcome	Interaction	β	*p*	*q*	Female factorial effect	Male factorial effect
Expectancy	GS × gender	−0.001	.991	.991	0.11 (.209)	0.11 (.002)
	RM × gender	0.15	.052	.070	−0.10 (.009)	0.05 (.645)
Intrinsic value	GS × gender	−0.12	.141	.514	0.18 (.007)	0.06 (.340)
	RM × gender	0.12	.296	.296	−0.12 (.176)	0.01 (.922)
Utility value	GS × gender	−0.07	.564	.753	0.14 (.209)	0.07 (.340)
	RM × gender	0.15	.047	.070	−0.10 (.015)	0.05 (.710)
Low cost	GS × gender	0.13	.257	.514	−0.22 (.004)	−0.09 (.340)
	RM × gender	0.32	.030	.070	−0.19 (.148)	0.13 (.645)

Note. N = 855. GS and RM were effect coded (−0.5 = absent, +0.5 = present) and gender was coded 0 (female) and 1 (male). β is the male-minus-female difference in each factorial component effect, averaged over the other component; female and male columns give the within-gender factorial effects (*q* in parentheses). Effects are in T1 baseline-SD units; low cost is reverse scored, so negative values indicate higher perceived cost. The uncorrected RM × GS × gender terms are in Table S14. GS = goal setting; RM = role modeling.

Table S13

Exploratory Gender-Specific Arm Contrasts for Overall STEM Motivation

Outcome	Gender	Contrast	Δ	*z*	*p*	*q*
Expectancy	Male	GS only vs AC	0.12	0.91	.362	.661
		RM only vs AC	0.06	0.64	.522	.535
		Combined vs AC	0.16	2.46	.014	.056
		Combined vs average of singles	0.07	0.48	.631	—
		RM × GS	−0.02	−0.08	.933	—
	Female	GS only vs AC	0.23	1.71	.088	.135
		RM only vs AC	0.02	0.26	.798	.984
		Combined vs AC	0.01	0.07	.943	.943
		Combined vs average of singles	−0.12	−2.93	.003	—
		RM × GS	−0.24	−2.09	.037	—
Intrinsic value	Male	GS only vs AC	0.11	1.26	.208	.661
		RM only vs AC	0.06	0.92	.356	.535
		Combined vs AC	0.07	1.04	.298	.397
		Combined vs average of singles	−0.02	−0.27	.785	—
		RM × GS	−0.10	−0.89	.374	—
	Female	GS only vs AC	0.26	2.78	.005	.022
		RM only vs AC	−0.04	−0.40	.689	.984
		Combined vs AC	0.06	0.47	.637	.934
		Combined vs average of singles	−0.05	−0.74	.458	—
		RM × GS	−0.16	−2.04	.041	—
Utility value	Male	GS only vs AC	0.10	0.68	.496	.661
		RM only vs AC	0.08	0.62	.535	.535
		Combined vs AC	0.12	1.15	.250	.397
		Combined vs average of singles	0.03	0.53	.594	—
		RM × GS	−0.06	−0.34	.731	—
	Female	GS only vs AC	0.24	1.64	.101	.135
		RM only vs AC	−0.001	−0.02	.984	.984
		Combined vs AC	0.04	0.38	.701	.934
		Combined vs average of singles	−0.08	−0.79	.432	—
		RM × GS	−0.20	−1.17	.242	—
Low cost	Male	GS only vs AC	−0.09	−0.43	.664	.664
		RM only vs AC	0.13	0.64	.522	.535

Outcome	Gender	Contrast	Δ	*z*	*p*	*q*
		Combined vs AC	0.05	0.47	.640	.640
		Combined vs average of singles	0.03	0.19	.853	—
		RM × GS	0.02	0.04	.965	—
	Female	GS only vs AC	−0.19	−1.13	.259	.259
		RM only vs AC	−0.16	−1.13	.260	.984
		Combined vs AC	−0.41	−2.65	.008	.033
		Combined vs average of singles	−0.23	−1.08	.278	—
		RM × GS	−0.05	−0.16	.877	—

Note. $N = 855$. Δ = arm-versus-active-control or combined-versus-single-component-average contrast, estimated separately by gender; RM × GS rows are interaction terms. Coding is as in Table S12. Estimates are in T1 baseline-SD units; low cost is reverse scored. Benjamini–Hochberg correction was applied within each gender-contrast family; interaction terms and combined-versus-average contrasts are uncorrected (—). These estimates are exploratory; the formal moderation tests are in Tables S12 and S14. AC = active control; GS = goal setting; RM = role modeling.

Table S14

Arm-Specific Gender Differences in Intervention Contrasts

Outcome	Contrast	Estimate	*SE*	*p*	*q*	95% CI
Expectancy	GS only vs AC	−0.11	0.22	.615	.619	[−0.54, 0.32]
	RM only vs AC	0.04	0.14	.767	.767	[−0.23, 0.31]
	RM × GS	0.22	0.31	.481	—	[−0.39, 0.83]
	Combined vs AC	0.15	0.15	.312	.624	[−0.14, 0.44]
	Combined vs average of singles	0.18	0.17	.279	—	[−0.15, 0.52]
Intrinsic value	GS only vs AC	−0.15	0.11	.182	.619	[−0.37, 0.07]
	RM only vs AC	0.09	0.14	.508	.678	[−0.18, 0.37]
	RM × GS	0.06	0.10	.545	—	[−0.14, 0.26]
	Combined vs AC	0.004	0.18	.981	.981	[−0.36, 0.37]
	Combined vs average of singles	0.03	0.09	.710	—	[−0.14, 0.21]
Utility value	GS only vs AC	−0.14	0.14	.338	.619	[−0.41, 0.14]
	RM only vs AC	0.08	0.09	.373	.678	[−0.09, 0.25]
	RM × GS	0.14	0.15	.333	—	[−0.14, 0.42]
	Combined vs AC	0.08	0.17	.631	.841	[−0.25, 0.42]
	Combined vs average of singles	0.11	0.12	.333	—	[−0.11, 0.34]
Low cost	GS only vs AC	0.10	0.20	.619	.619	[−0.29, 0.49]
	RM only vs AC	0.29	0.24	.231	.678	[−0.18, 0.76]
	RM × GS	0.07	0.30	.827	—	[−0.53, 0.66]
	Combined vs AC	0.45	0.22	.034	.138	[0.03, 0.88]
	Combined vs average of singles	0.26	0.16	.108	—	[−0.06, 0.58]

Note. N = 855 (401 female, 454 male). Estimates are male-minus-female differences in each contrast, in T1 baseline-SD units, from the same models as Table S12; negative values indicate a larger contrast for female students. Low cost is reverse scored. Benjamini–Hochberg correction was applied within each contrast type and outcome family; RM × GS and combined-versus-average rows are uncorrected (—). AC = active control; CI = confidence interval; GS = goal setting; RM = role modeling.

Table S15

Gender Moderation of Intervention Effects Within Individual Science Subjects

Outcome	Interaction	β	*p*	*q*	Female factorial effect	Male factorial effect
Biology: Expectancy	GS × gender	−0.11	.245	.420	0.18 (.052)	0.07 (.082)
	RM × gender	0.01	.929	.929	−0.07 (.126)	−0.07 (.627)
Biology: Intrinsic value	GS × gender	−0.08	.227	.420	0.10 (.091)	0.01 (.698)
	RM × gender	−0.04	.697	.846	−0.08 (.206)	−0.12 (.453)
Biology: Utility value	GS × gender	−0.12	.143	.380	0.13 (.052)	0.02 (.917)
	RM × gender	−0.03	.718	.846	0.01 (.920)	−0.02 (.795)
Biology: Low cost	GS × gender	0.16	.328	.437	−0.21 (.082)	−0.05 (.698)
	RM × gender	0.19	.297	.545	−0.11 (.126)	0.07 (.740)
Chemistry: Expectancy	GS × gender	−0.01	.874	.874	0.11 (.168)	0.10 (.055)
	RM × gender	0.12	.101	.302	−0.16 (.042)	−0.05 (.675)
Chemistry: Intrinsic value	GS × gender	−0.13	.018	.214	0.12 (< .001)	−0.01 (.917)
	RM × gender	0.06	.709	.846	−0.14 (.206)	−0.08 (.627)
Chemistry: Utility value	GS × gender	−0.14	.087	.350	0.09 (.169)	−0.05 (.698)
	RM × gender	0.09	.318	.545	−0.12 (.094)	−0.03 (.785)
Chemistry: Low cost	GS × gender	0.16	.158	.380	−0.17 (.052)	−0.01 (.917)
	RM × gender	0.24	.090	.302	−0.17 (.016)	0.07 (.740)
Physics: Expectancy	GS × gender	0.03	.813	.874	0.07 (.486)	0.10 (.055)
	RM × gender	0.18	.052	.302	−0.15 (.021)	0.03 (.740)
Physics: Intrinsic value	GS × gender	−0.08	.290	.435	0.12 (.052)	0.04 (.698)
	RM × gender	−0.03	.775	.846	−0.06 (.328)	−0.09 (.627)
Physics: Utility value	GS × gender	−0.12	.368	.441	0.08 (.486)	−0.04 (.698)
	RM × gender	0.13	.146	.349	−0.12 (.060)	0.01 (.887)
Physics: Low cost	GS × gender	0.20	.039	.232	−0.13 (.069)	0.07 (.698)
	RM × gender	0.36	.095	.302	−0.19 (.126)	0.17 (.627)

Note. N = 855. Coding is as in Table S12. β is the male-minus-female difference in each factorial component effect; female and male columns give the within-gender effects (q in parentheses). Benjamini–Hochberg correction was applied separately within each 12-outcome subject-specific interaction family and each within-gender factorial-effect family. All analyses are exploratory. Estimates are in T1 baseline-SD units; low cost is reverse scored, so negative effects indicate higher perceived cost. GS = goal setting; RM = role modeling.

Table S16

Factorial School-Location Moderation for Overall STEM Motivation

Outcome	Interaction	β	*p*	*q*	Urban factorial effect	Rural factorial effect
Expectancy	GS × location	−0.002	.980	.980	0.11 (< .001)	0.11 (.178)
	RM × location	−0.05	.156	.208	−0.01 (.923)	−0.06 (.092)
Intrinsic value	GS × location	−0.02	.815	.980	0.13 (.005)	0.11 (.178)
	RM × location	−0.11	.061	.121	0.001 (.977)	−0.11 (.063)
Utility value	GS × location	0.23	.035	.140	0.001 (.988)	0.23 (.060)
	RM × location	−0.13	.222	.222	0.02 (.923)	−0.10 (.215)
Low cost	GS × location	0.17	.088	.175	−0.22 (.017)	−0.05 (.196)
	RM × location	−0.35	.010	.039	0.13 (.206)	−0.22 (.092)

Note. $N = 855$. β is the rural-minus-urban difference in the averaged factorial GS or RM effect on the STEM-overall outcomes (GS and RM coded −0.5 = absent, +0.5 = present); urban and rural columns give the within-location factorial effects (*q* in parentheses). Effects are in T1 baseline-SD units; low cost is reverse scored, so negative values indicate higher perceived cost. The uncorrected RM × GS × location terms are in Table S17, and subject-specific results in Table S18. Location is confounded with school (three schools per category). GS = goal setting; RM = role modeling.

Table S17

Location-Specific Arm Contrasts and Interactions for Overall STEM Motivation

Outcome	Location	Contrast	Δ	*z*	*p*	*q*
Expectancy	Urban	GS only vs AC	0.08	2.37	.018	.036
	Urban	RM only vs AC	−0.03	−1.11	.268	.721
	Urban	Combined vs AC	0.11	7.62	< .001	< .001
	Urban	Combined vs average of singles	0.08	1.97	.049	—
	Urban	RM × GS	0.06	0.79	.431	—
	Rural	GS only vs AC	0.28	2.25	.025	.049
	Rural	RM only vs AC	0.11	4.00	< .001	< .001
	Rural	Combined vs AC	0.06	1.06	.289	.385
	Rural	Combined vs average of singles	−0.14	−3.42	< .001	—
	Rural	RM × GS	−0.33	−3.08	.002	—
	Rural − urban	GS only vs AC	0.19	1.52	.127	.255
	Rural − urban	RM only vs AC	0.15	3.59	< .001	.001
	Rural − urban	Combined vs AC	−0.05	−0.93	.355	.355
	Rural − urban	Combined vs average of singles	−0.22	−3.64	< .001	—
	Rural − urban	RM × GS	−0.39	−3.07	.002	—
Intrinsic value	Urban	GS only vs AC	0.15	2.55	.011	.036
	Urban	RM only vs AC	0.01	0.36	.721	.721
	Urban	Combined vs AC	0.14	8.59	< .001	< .001
	Urban	Combined vs average of singles	0.06	2.35	.019	—
	Urban	RM × GS	−0.02	−0.56	.575	—
	Rural	GS only vs AC	0.24	1.92	.054	.073
	Rural	RM only vs AC	0.02	0.66	.512	.512
	Rural	Combined vs AC	0.003	0.05	.957	.957
	Rural	Combined vs average of singles	−0.13	−2.88	.004	—
	Rural	RM × GS	−0.26	−2.32	.020	—
	Rural − urban	GS only vs AC	0.10	0.70	.482	.482
	Rural − urban	RM only vs AC	0.01	0.21	.835	.835
	Rural − urban	Combined vs AC	−0.13	−2.20	.028	.111
	Rural − urban	Combined vs average of singles	−0.19	−4.47	< .001	—
	Rural − urban	RM × GS	−0.24	−2.23	.026	—
Utility value	Urban	GS only vs AC	−0.08	−0.89	.376	.376
	Urban	RM only vs AC	−0.06	−0.39	.695	.721

Outcome	Location	Contrast	Δ	*z*	*p*	*q*
	Urban	Combined vs AC	0.02	1.27	.206	.274
	Urban	Combined vs average of singles	0.09	0.95	.344	—
	Urban	RM × GS	0.16	0.77	.439	—
	Rural	GS only vs AC	0.46	4.19	< .001	< .001
	Rural	RM only vs AC	0.12	1.00	.318	.512
	Rural	Combined vs AC	0.13	1.56	.119	.238
	Rural	Combined vs average of singles	−0.16	−5.72	< .001	—
	Rural	RM × GS	−0.45	−4.78	< .001	—
	Rural − urban	GS only vs AC	0.54	4.04	< .001	< .001
	Rural − urban	RM only vs AC	0.18	0.96	.338	.451
	Rural − urban	Combined vs AC	0.11	1.31	.192	.256
	Rural − urban	Combined vs average of singles	−0.25	−2.75	.006	—
	Rural − urban	RM × GS	−0.61	−2.93	.003	—
Low cost	Urban	GS only vs AC	−0.23	−1.59	.112	.149
	Urban	RM only vs AC	0.12	0.61	.539	.721
	Urban	Combined vs AC	−0.09	−0.90	.368	.368
	Urban	Combined vs average of singles	−0.03	−0.28	.781	—
	Urban	RM × GS	0.03	0.12	.905	—
	Rural	GS only vs AC	0.02	0.06	.956	.956
	Rural	RM only vs AC	−0.15	−0.77	.440	.512
	Rural	Combined vs AC	−0.26	−3.11	.002	.008
	Rural	Combined vs average of singles	−0.20	−0.65	.517	—
	Rural	RM × GS	−0.13	−0.24	.807	—
	Rural − urban	GS only vs AC	0.25	0.76	.445	.482
	Rural − urban	RM only vs AC	−0.26	−0.99	.325	.451
	Rural − urban	Combined vs AC	−0.18	−1.36	.174	.256
	Rural − urban	Combined vs average of singles	−0.17	−0.51	.608	—
	Rural − urban	RM × GS	−0.16	−0.27	.789	—

Note. $N = 855$. Δ = arm-versus-active-control or combined-versus-single-component-average contrast within each set of schools; RM × GS rows are interaction terms, and rural − urban rows are the differences between the two sets. Estimates are in T1 baseline-SD units; low cost is reverse scored. Benjamini–Hochberg correction was applied within each location-contrast family; interaction terms and combined-versus-average contrasts are uncorrected (—). These analyses were added after registration and are exploratory; location is confounded with school (three schools per category). AC = active control; GS = goal setting; RM = role modeling.

Table S18

Factorial School-Location Moderation Within Individual Science Subjects

Outcome	Interaction	β	*p*	*q*	Urban factorial effect	Rural factorial effect
Biology: Expectancy	GS × location	−0.04	.431	.636	0.15 (< .001)	0.10 (.116)
	RM × location	−0.04	.535	.714	−0.05 (.534)	−0.08 (.094)
Biology: Intrinsic value	GS × location	−0.06	.064	.191	0.08 (< .001)	0.02 (.532)
	RM × location	0.07	.129	.299	−0.13 (.007)	−0.07 (.071)
Biology: Utility value	GS × location	0.13	.035	.139	0.02 (.689)	0.15 (.006)
	RM × location	−0.10	.150	.299	0.03 (.619)	−0.07 (.215)
Biology: Low cost	GS × location	0.06	.500	.636	−0.16 (.109)	−0.10 (< .001)
	RM × location	−0.10	.208	.356	0.03 (.563)	−0.07 (.334)
Chemistry: Expectancy	GS × location	−0.14	.004	.026	0.19 (< .001)	0.04 (.421)
	RM × location	−0.14	< .001	.006	−0.03 (.615)	−0.17 (< .001)
Chemistry: Intrinsic value	GS × location	0.004	.950	.950	0.05 (.015)	0.06 (.432)
	RM × location	−0.003	.965	.965	−0.11 (.024)	−0.12 (.094)
Chemistry: Utility value	GS × location	0.18	< .001	< .001	−0.06 (.063)	0.12 (.006)
	RM × location	−0.15	.040	.159	−0.02 (.797)	−0.16 (.008)
Chemistry: Low cost	GS × location	0.08	.215	.403	−0.12 (.075)	−0.04 (< .001)
	RM × location	−0.10	.283	.424	−0.01 (.860)	−0.11 (.215)
Physics: Expectancy	GS × location	−0.01	.884	.950	0.09 (.122)	0.08 (.410)
	RM × location	−0.01	.830	.938	−0.05 (.546)	−0.06 (.015)
Physics: Intrinsic value	GS × location	−0.05	.235	.403	0.10 (.015)	0.05 (.276)
	RM × location	0.04	.084	.253	−0.09 (< .001)	−0.05 (.094)
Physics: Utility value	GS × location	0.13	.087	.209	−0.04 (.260)	0.09 (.312)
	RM × location	0.02	.860	.938	−0.07 (.534)	−0.05 (.410)
Physics: Low cost	GS × location	0.06	.530	.636	−0.05 (.305)	0.02 (.851)
	RM × location	−0.23	< .001	.006	0.10 (.448)	−0.13 (.019)

Note. $N = 855$. β is the rural-minus-urban difference in the averaged factorial GS or RM effect (coded −0.5 = absent, +0.5 = present); urban and rural columns give the within-location factorial effects (*q* in parentheses). Benjamini–Hochberg correction was applied separately within each 12-outcome subject-specific interaction family and each within-location factorial-effect family. These analyses were added after registration and are exploratory; location is confounded with school (three schools per category). Effects are in T1 baseline-SD units; low cost is reverse scored. GS = goal setting; RM = role modeling.

Table S19

Location-Specific Single-Component Contrasts Within Individual Science Subjects

Outcome	Location	Contrast	Δ	*z*	*p*	*q*
Biology: Expectancy	Urban	GS only vs AC	0.28	3.52	< .001	.001
	Urban	RM only vs AC	0.09	1.34	.180	.540
	Rural	GS only vs AC	0.20	2.05	.041	.097
	Rural	RM only vs AC	0.02	0.23	.819	.819
	Rural − urban	GS only vs AC	−0.08	−0.62	.539	.808
	Rural − urban	RM only vs AC	−0.07	−0.80	.421	.632
Biology: Intrinsic value	Urban	GS only vs AC	0.19	5.90	< .001	< .001
	Urban	RM only vs AC	−0.02	−0.34	.731	.855
	Rural	GS only vs AC	0.12	1.31	.190	.284
	Rural	RM only vs AC	0.04	0.60	.552	.672
	Rural − urban	GS only vs AC	−0.07	−0.76	.449	.808
	Rural − urban	RM only vs AC	0.06	0.70	.484	.646
Biology: Utility value	Urban	GS only vs AC	0.01	0.25	.803	.803
	Urban	RM only vs AC	0.02	0.25	.805	.855
	Rural	GS only vs AC	0.30	2.76	.006	.023
	Rural	RM only vs AC	0.08	0.66	.512	.672
	Rural − urban	GS only vs AC	0.29	2.75	.006	.024
	Rural − urban	RM only vs AC	0.06	0.40	.690	.723
Biology: Low cost	Urban	GS only vs AC	−0.14	−1.13	.258	.386
	Urban	RM only vs AC	0.05	0.41	.684	.855
	Rural	GS only vs AC	−0.20	−1.00	.316	.379
	Rural	RM only vs AC	−0.18	−0.85	.396	.672
	Rural − urban	GS only vs AC	−0.07	−0.27	.789	.861
	Rural − urban	RM only vs AC	−0.23	−0.94	.348	.632
Chemistry: Expectancy	Urban	GS only vs AC	0.23	3.51	< .001	.001
	Urban	RM only vs AC	0.02	0.18	.855	.855
	Rural	GS only vs AC	0.18	1.55	.122	.210
	Rural	RM only vs AC	−0.03	−0.33	.740	.807
	Rural − urban	GS only vs AC	−0.06	−0.42	.671	.861
	Rural − urban	RM only vs AC	−0.05	−0.36	.723	.723
Chemistry: Intrinsic value	Urban	GS only vs AC	0.06	0.98	.328	.438
	Urban	RM only vs AC	−0.11	−4.06	< .001	< .001
	Rural	GS only vs AC	0.13	1.16	.248	.331

Outcome	Location	Contrast	Δ	*z*	*p*	*q*
	Rural	RM only vs AC	−0.04	−0.96	.338	.672
	Rural − urban	GS only vs AC	0.08	0.66	.512	.808
	Rural − urban	RM only vs AC	0.07	1.44	.149	.448
Chemistry: Utility value	Urban	GS only vs AC	−0.08	−1.81	.070	.123
	Urban	RM only vs AC	−0.04	−0.43	.667	.855
	Rural	GS only vs AC	0.34	4.74	< .001	< .001
	Rural	RM only vs AC	0.06	0.61	.542	.672
	Rural − urban	GS only vs AC	0.42	6.18	< .001	< .001
	Rural − urban	RM only vs AC	0.10	0.81	.416	.632
Chemistry: Low cost	Urban	GS only vs AC	−0.19	−1.80	.072	.123
	Urban	RM only vs AC	−0.07	−0.62	.534	.855
	Rural	GS only vs AC	−0.13	−0.65	.514	.561
	Rural	RM only vs AC	−0.19	−1.16	.245	.672
	Rural − urban	GS only vs AC	0.06	0.29	.772	.861
	Rural − urban	RM only vs AC	−0.12	−0.61	.542	.651
Physics: Expectancy	Urban	GS only vs AC	0.02	0.26	.793	.803
	Urban	RM only vs AC	−0.12	−1.90	.058	.232
	Rural	GS only vs AC	0.23	1.69	.092	.184
	Rural	RM only vs AC	0.09	1.53	.127	.672
	Rural − urban	GS only vs AC	0.21	1.34	.182	.545
	Rural − urban	RM only vs AC	0.22	2.31	.021	.125
Physics: Intrinsic value	Urban	GS only vs AC	0.11	7.95	< .001	< .001
	Urban	RM only vs AC	−0.08	−5.85	< .001	< .001
	Rural	GS only vs AC	0.12	2.16	.031	.093
	Rural	RM only vs AC	0.02	0.58	.560	.672
	Rural − urban	GS only vs AC	0.01	0.16	.870	.870
	Rural − urban	RM only vs AC	0.10	2.89	.004	.046
Physics: Utility value	Urban	GS only vs AC	−0.10	−0.84	.402	.483
	Urban	RM only vs AC	−0.12	−0.90	.366	.855
	Rural	GS only vs AC	0.29	3.02	.003	.015
	Rural	RM only vs AC	0.15	1.16	.246	.672
	Rural − urban	GS only vs AC	0.39	2.85	.004	.024
	Rural − urban	RM only vs AC	0.27	1.56	.120	.448
Physics: Low cost	Urban	GS only vs AC	−0.10	−2.11	.035	.084
	Urban	RM only vs AC	0.05	0.32	.747	.855
	Rural	GS only vs AC	0.02	0.24	.808	.808

Outcome	Location	Contrast	Δ	*z*	*p*	*q*
	Rural	RM only vs AC	−0.12	−0.94	.346	.672
	Rural − urban	GS only vs AC	0.12	1.18	.239	.574
	Rural − urban	RM only vs AC	−0.17	−0.87	.387	.632

Note. $N = 855$. Δ is the single-component arm-versus-active-control contrast within each set of schools; Rural − urban rows are the difference between them. Estimates are in T1 baseline-SD units. Low cost is reverse scored, so negative values indicate higher perceived cost. Benjamini–Hochberg correction was applied within each subject-specific location-contrast family. These analyses are exploratory, and location is confounded with school, with three schools in each category. AC = active control; GS = goal setting; RM = role modeling. These contrasts were added after registration.

Table S20

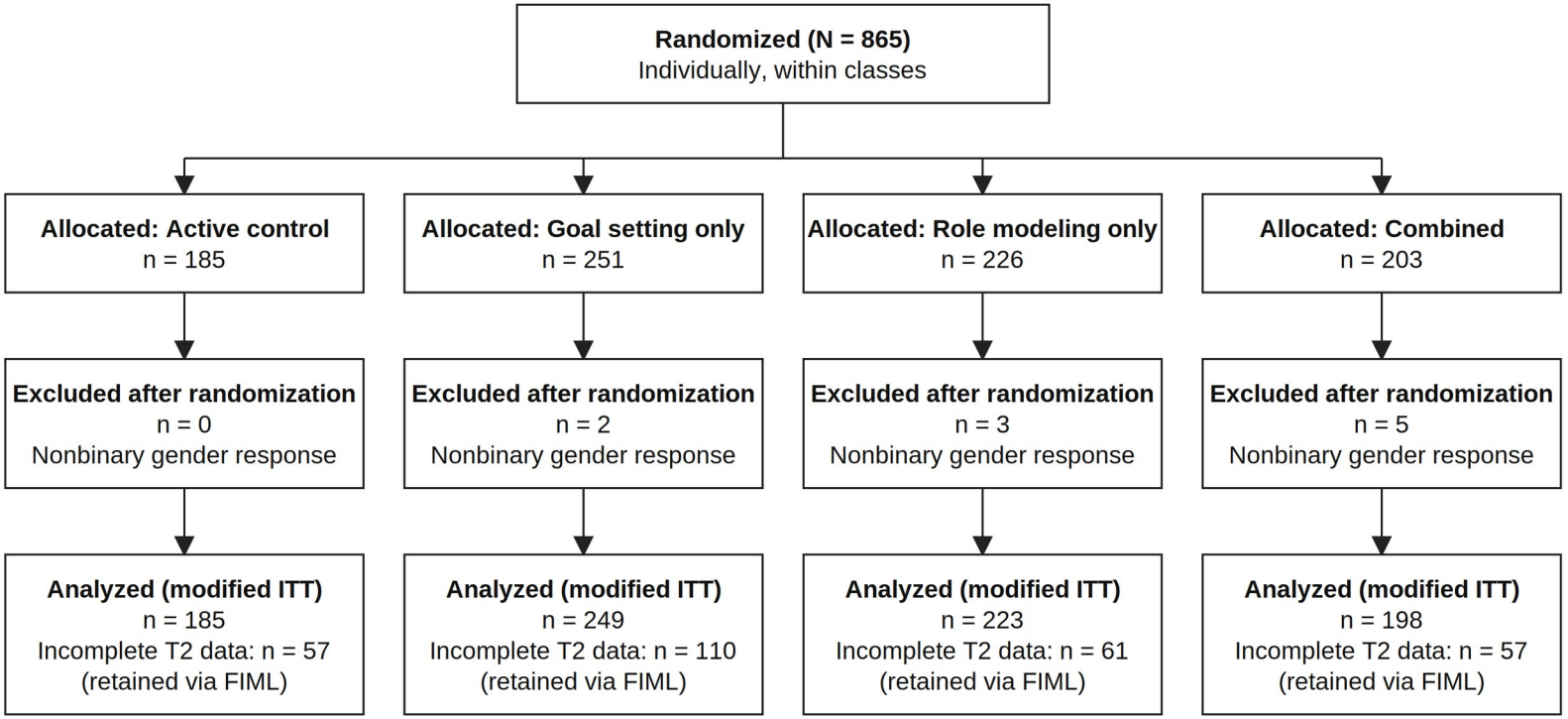
Complete-Case Sensitivity Comparison for STEM-Overall Outcomes

Effect	Outcome	FIML (primary, $N = 855$)	Complete case ($N = 570$)
GS main effect	Expectancy	0.11, $q = .003$	0.11, $q = .005$
	Intrinsic value	0.12, $q = .003$	0.12, $q = .005$
	Utility value	0.10, $q = .190$	0.11, $q = .189$
	Low cost	−0.16, $q = .007$	−0.16, $q = .008$
GS only vs AC	Expectancy	0.17, $q = .034$	0.19, $q = .063$
	Intrinsic value	0.18, $q = .034$	0.20, $q = .063$
Combined vs AC	Low cost	−0.18, $q = .027$	−0.19, $q = .032$

Note. Factorial main effects (β) and planned contrasts (Δ) are in T1 baseline-SD units; low cost is reverse scored, so negative estimates indicate higher perceived cost. q = Benjamini–Hochberg-adjusted p value within the primary STEM-overall family. The preregistered complete-case analysis used the 570 students with complete responses on all 48 SEVT items at T1 and T2. Both columns adjust for school location. AC = active control; FIML = full-information maximum likelihood; GS = goal setting; SEVT = situated expectancy-value theory.

Figure S1

CONSORT Diagram of Participant Flow Through the Trial



Note. Ten students were excluded after randomization because they reported a nonbinary gender identity, which the binary gender covariate could not accommodate (GS only 2, RM only 3, combined 5, active control 0), leaving the analyzed modified intention-to-treat sample (N = 855 of 865). T2 data were counted as incomplete when any of the 48 posttest SEVT items was missing; 570 students (66.7%) had complete T2 data. Retention ranged from 55.8% (GS only) to 72.6% (RM only), $\chi^2(3) = 19.13$, $p < .001$. Of the 285 students with incomplete T2 data, 247 had not finished their assigned intervention content. Students with no T2 data reported lower baseline expectancy, intrinsic value, and utility value (ds = 0.37 to 0.50, all ps < .001) but similar baseline cost. FIML = full-information maximum likelihood; GS = goal setting; ITT = intention-to-treat; RM = role modeling; SEVT = situated expectancy-value theory.

References

Flake, J. K., Barron, K. E., Hulleman, C. S., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Gaspard, H., Dicke, A., Flunger, B., Brisson, B., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence*, *35*(5–6), 790–816. https://doi.org/10.1177/0272431614556890