

Table S1. Mixed-Effects Modeling Results Accounting for Clustering (3-Level Model: Students Nested Within Classes Nested Within Semesters)

Parameter	Estimate (β)	Standard Error (SE)	95% CI (Lower)	95% CI (Upper)	df	t-value	P-value
Fixed Effects							
Intercept	47.12	0.52	46.10	48.14	15.2	90.62	<0.001
Group (Intervention vs. Control)	10.89	0.56	9.79	11.99	18.4	19.45	<0.001
Baseline GPA	1.24	0.31	0.63	1.85	665.0	4.00	<0.001
Prior Clinical Experience (months)	0.18	0.09	0.01	0.35	661.0	2.00	0.046
Random Effects	Variance	SD	ICC				
Level 3: Semester (n=8)	0.92	0.96	0.09				
Level 2: Class/Section (n=24)	1.45	1.20	0.14				
Level 1: Student (n=682)	8.23	2.87	—				

Model Fit Indices:

- AIC = 3421.5
- BIC = 3453.8
- Log-Likelihood = -1704.7
- Marginal R^2 (fixed effects only) = 0.38
- Conditional R^2 (fixed + random effects) = 0.52

Note: Mixed-effects modeling was conducted using the lme4 package in R (version 4.2.1). Degrees of freedom were calculated using the Satterthwaite approximation. The intraclass correlation coefficient (ICC) indicates that 23% of the total variance in teaching quality scores was attributable to clustering at the semester and class levels. After accounting for this clustering,

the intervention effect remained statistically significant ($\beta = 10.89$, $P < 0.001$), consistent with the primary Mann-Whitney U test results.