Table S2 – Appraisal risk of bias criteria defined according to Risk of Bias Assessment tool for Non-randomized Studies (RoBANS) tool.

Domain	Description
Selection of participants	**Selection bias** caused by inadequate selection of participants Selection of participants should be clearly defined to medical undergraduate student population. As high risk, we considered: (i) historical cohorts; (ii) any group composed by student of different curricular years; (iii) different topics or exposure across groups; and (iv) using a restrict group of students without a reasonable and valid explanation.
Confounding variables	**Selection bias** caused by inadequate confirmation and consideration of confounding variable As confounding variables, we considered: (i) group comparison of students from different curricular years (e.g., group 1, first year vs group 2, second year); (ii) unbalanced population in regards to curricular years or sample size across groups; (iii) different methodologies for comparison groups; (iv) cross-over or cross-sectional design without wash-out period (risk of carry-over effect) such as semester effect (e.g., first versus second semester); (v) multicentric study without adjusting for potential related confounders; and (vi) unbalanced time of exposure to class across groups.
Exposure measurement	**Performance bias** caused by inadequate measurement of exposure Academic performance data should be collected with a standardized global test or a topic-focused test. We considered high risk if (i) there was insufficient time of exposure, (ii) measurement of exposure was different between groups in content or time, and/or (iii) when applying non-validated self-reported scales.
Blinding outcome assessment	**Detection bias** caused by inadequate blinding of outcome assessment Automatically assessed with unclear when self-reported measures were used. As per Cochrane recommendations, this type of assessments should always be judged with some concerns regarding blinding outcome assessment. Since RoBANS does not have such category, we classified as "unclear". If tests with subjective answers or subjective clinical evaluations were adopted and the assessor was aware of the intervention/exposure, the study was classified at high risk. In cases that insufficient information was provided, the study was classified as unclear. If only a subset of students was purposefully selected for an assessment, the study was judged at high risk.
Incomplete outcome data	**Attrition bias** caused by inadequate handing of incomplete data outcome Missing data of outcome variables and/or drop-out of >15% or response rate <85%
Selective outcome reporting	**Reporting bias** caused by selective outcome reporting Studies were automatically judged at unclear risk if there was no pre-registered or pre-published protocol. Studies were judged at high risk if: (i) multiple publication was detected (*i.e.*, the same trial, but different outcomes published in different studies); and/or (ii) the authors chose to analyze only a subset of outcomes or a subset of the sample without proper justification; and/or (iii) the reasons for missing data were unreported; and/or (iv) the qualitative reporting was unsubstantiated by the data; and/or (v) the methods provided no information on how the analyzes would be performed; and/or (vi) the methods provided no information on how the outcomes would be assessed; and/or (vii) the methods and the

	results were mutually contradictory; and/or (viii) an ITT analysis was replaced with a per-protocol analysis without proper justification.