

The methodological quality and risk of bias of the observational studies included in this review were assessed using the Newcastle–Ottawa Scale (NOS), following PRISMA 2020 recommendations for transparency in quality appraisal. The NOS evaluates three major domains: Selection, Comparability, and Outcome/Exposure assigning a maximum of nine stars, with a higher number of stars indicating better methodological rigor and lower risk of bias.

1. Selection (Maximum: 4 stars)

This domain evaluates the adequacy and clarity with which each study defines and selects its population. For this review, a star was awarded when studies met the following criteria:

- Clear definition and representativeness of the cohort or case group: Studies explicitly described the medical student population, sampling method, academic year distribution, and target psychosocial variables (stress, depression, anxiety, motivation, burnout, coping, resilience, etc.).
- Appropriate selection or definition of controls (when applicable): In case–control studies, controls were drawn from the same academic population and were free of the exposure under investigation.
- Ascertainment of exposure: Exposure assessment (e.g., stress level, mental health metrics, burnout) used validated instruments such as DASS-21, PSS-10, MBI-SS, PHQ-9, or institutional academic performance records.

2. Comparability (Maximum: 2 stars)

This domain assesses whether studies controlled for confounding factors. Up to two stars were granted when:

- Adjustments were performed for core confounders, including age, sex, socioeconomic status, academic year, and previous academic performance.
- Multivariate analyses or stratified models were used to isolate the effect of psychosocial variables on academic outcomes.

Studies that did not report adjustment methods or failed to consider key confounders received fewer stars in this section.

3. Outcome or Exposure Assessment (Maximum: 3 stars)

Depending on the study design (cohort, cross-sectional, case–control), this domain appraises the reliability of measurement and follow-up.

Stars were awarded when:

- Validated and reliable instruments were used to measure psychosocial factors and academic performance.
- Outcome data were obtained from objective academic records (GPA, exam scores, class rank).
- Statistical methods were appropriate, including reporting of effect measures (OR, β coefficients, p-values, CI 95%).
- For cohort designs, follow-up was complete and clearly reported.

Scoring and Study Classification

Each study could receive a maximum score of 9 stars. Based on total NOS scores, studies were categorized as:

- Low risk of bias: 7–9 stars
- Moderate risk of bias: 5–6 stars
- High risk of bias: ≤ 4 stars

Most studies in this review achieved moderate to high methodological quality, with the majority scoring between 6 and 8 stars, reflecting acceptable internal validity and appropriate control of confounders.

Application to This Review

The NOS was applied independently by two reviewers, and discrepancies were resolved through consensus. This structured approach allowed for consistent evaluation across all included studies and ensured that the synthesis of findings was grounded in methodologically robust evidence.