Canadian Medical Education Journal

Reviews, Theoretical Papers, and Meta-Analyses

Measuring medical student wellbeing longitudinally: a psychometric systematic review of commonly used scales

Henry Li,[1] Youri Kim,[2] Victor Do,[3] Aliya Kassam,[4] Kyle Chankasingh,[2] Melanie Lewis[3]

[1]Department of Emergency Medicine, University of Alberta Faculty of Medicine & Dentistry, Alberta, Canada; [2]University of Alberta Faculty of Medicine & Dentistry, Alberta, Canada; [3]Department of Pediatrics, University of Alberta Faculty of Medicine & Dentistry, Alberta, Canada; [4]Department of Community Health Sciences and Office of Postgraduate Medical Education, Cumming School of Medicine, University of Calgary, Alberta, Canada;

Correspondence to: Henry Li, Department of Emergency Medicine, Faculty of Medicine & Dentistry, University of Alberta, 8303 - 112 Street, 750 University Terrace, Edmonton, Alberta, Canada; e-mail: henry9@ualberta.ca; Twitter: @HenryLiCDN

Published ahead of issue: Nov 25, 2025; CMEJ 2025 Available at https://doi.org/10.36834/cmej.82091

Abstract

Background: Longitudinal measurements of medical student wellbeing are needed to evaluate the impacts of training and potential interventions, but the psychometric evidence underlying commonly used wellbeing scales is unclear, impairing selection decisions. We therefore synthesized the psychometric evidence of the most common scales employed to measure self-reported medical student wellbeing longitudinally.

Methods: We conducted a psychometric systematic review based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines. We searched seven databases and gray literature in March 2023 for psychometric studies in medical students of 53 scales. Two independent reviewers completed screening and data extraction and resolved conflicts via discussion. We assessed study quality and psychometrics using COSMIN methodology and pooled results for internal consistency and test-retest reliability when there were ≥2 studies per scale.

Results: Of 2374 abstracts, we included 133 studies. Over a quarter (26.4%) of study scales lacked psychometric evidence in medical students. Internal consistency was the most studied property (118 studies), while there were no studies on measurement error. There was sufficient evidence of internal consistency for 30 scales and construct validity for 34 scales. However, there were only 1-6 scales with sufficient evidence for each of the remaining properties. Study quality varied widely and only 20 of them reported participant ethno-racial identity.

Conclusions: Many scales commonly used to measure medical student wellbeing longitudinally lack medical student-specific psychometric evidence. Among those that do, few have any evidence beyond internal consistency and construct validity. Future psychometric studies are needed in diverse populations to better inform scale selection.

Résumé

Résumé français à venir.

Introduction

Compared to age-matched college students, medical students start their training with a higher quality of life and lower rates of burnout and depression.[1] However, throughout medical school, these same individuals demonstrate lower quality of life as well as higher rates of burnout, depression, anxiety disorders, and psychological distress.[2–5] In turn, the higher rates of mental health concerns have been linked with unprofessional conduct, reduced altruism and empathy, as well as increased thoughts of suicide and withdrawal from medical school.[4,6–8] Changes in wellbeing can take years to culminate, as can the impact of interventions being studied.[9,10] As a result, there have been multiple calls to address these gaps by measuring wellbeing longitudinally.[11,12]

However, there is a wide breadth of scales used to measure wellbeing-related constructs and a lack of robust, contemporary guidance to help educators, administrators, and researchers select the best scales to measure wellbeing longitudinally.[13] For example, we previously conducted a scoping review of scales used to measure self-reported medical student wellbeing longitudinally and found 140 unique scales across 13 wellbeing-related constructs (most commonly mood/affect, anxiety, and stress).[12]

While prior papers have attempted to provide guidance on how to choose a wellbeing scale, most have only compared a small number of scales and did not use a systematic approach to evaluating the strength of their psychometric evidence or rating the methodological quality of their underlying studies. Dyrbye et al.'s 2018 paper provided recommendations on important characteristics to consider for physician wellbeing scales, such as low respondent burden (i.e. length), low organizational burden (i.e. easy to analyze, cost), and robust psychometric properties (i.e. validity, reliability).[14] However, they only narratively highlighted the strengths and limitations of five burnout measures and two composite well-being measures. Lall et al searched three databases for physician wellbeing scales from 2009-2019 as part of a two-part scoping review. However, they only found 27 unique scales and narratively highlighted the strengths and limitations of only 24 scales based on which ones they felt were most relevant for emergency physicians.[15,16] In the medical student population, Haykal et al conducted a scoping review[17] of wellbeing scales "beyond anxiety and depression" and narratively described six measures of general wellbeing.

We therefore sought to systematically review and compare the psychometric evidence underlying scales used to measure self-reported medical student wellbeing longitudinally. We focused on the most commonly used scales from our scoping review and were guided by Dyrbye et al.'s considerations for scale selection[14] to examine both psychometric evidence and feasibility characteristics.

Methods

We conducted a systematic review based on the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guideline.[18] The COSMIN initiative is an international interdisciplinary team that has developed an array of resources to guide the systematic review and selection of outcome measurement instruments. this includes a taxonomy,[19] search filter,[20] risk of bias checklist,[21] criteria for good measurement properties,[18] and guidance on how to perform a GRADE evaluation of the included evidence.[18] To help clarify the terms used in this paper for a medical education audience, we have compared the definitions used in the COSMIN taxonomy to Messick's validity framework as adopted in the Standards for Educational and Psychological Testing[22] in Table 1. Of note, the COSMIN taxonomy in Table 1 refers to health-related patient-reported outcome (HR-PRO) measures which are more aligned with wellbeing scales rather than assessment of competence measures in medical education that have applied Messick's framework.

We developed a protocol a priori which we registered on Open Science Framework[23] The only changes to the protocol were a reorganization of the included scales for quality of life and general well-being. The results are reported based on the PRISMA-COSMIN checklist (Supplemental Data Table S1)[24] Research ethics approval was not required for this systematic review.

Our team included researchers with broad expertise and perspectives including: a medical student (YK), residents (HL, VD), a medical education researcher with psychometrics expertise (AK), a chief wellness officer (ML), a psychologist with training in psychometrics (RJ), and a health sciences research librarian (JK).

Table 1. Comparing the COSMIN taxonomy and Messick's validity framework

COSMIN Taxonomy			Messick's Validity Evidence Framework	
Domain	Measurement property	Definition	Category	Definition
Reliability	Internal consistency	The degree of the interrelatedness among the items	*Internal structure	The relationships among survey items or sections of a survey, including score consistency/ reliability and subscale structure
	Reliability	The proportion of the total variance in the measurements which is because of "true" differences among patients		
	Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured	Content	The appropriateness of survey content in light of the construct the tool is intended to measure
Validity	Content validity	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured	Relationship to other variables	The associations (positive or negative) between the survey scores and data on other variables
	Construct validity	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured	*Internal structure	The relationships among survey items or sections of a survey, including score consistency/ reliability and subscale structure
	Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured		
	Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument		
	Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"		
Responsiveness		The ability of an HR-PRO instrument to detect change over time in the construct to be measured		
			Response processes	The psychological processes or cognitive operations of survey takers and the "detailed nature of the performance … actually engaged in" while completing the survey
			Consequences of testing	The positive or negative, intended or unintended effects of survey use

**Note internal structure is related to both reliability concepts and structural validity. COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments. HR-PROs = health-related patient-reported outcomes.*

Study scales

Our previous scoping review used Brady *et al's* definition of wellbeing[25] (*quality of life, which includes the absence of ill-being and the presence of positive physical, mental, social, and integrated well-being experienced in connection with activities and environments that allow physicians to develop their full potentials across personal and work-life domains*) and included studies that used a scale to measure medical student wellbeing at two or more timepoints; there were a total of 221 included studies which used 140 unique self-report scales to measure 13 wellbeing-related constructs.[12]

Given the breadth of scales that we found in our previous scoping review,[12] it would be infeasible to examine the psychometric evidence of all of them. From a practical and pragmatic standpoint, understanding the evidence for the most used scales is likely to be most useful to researchers, leaders, and administrators alike. Therefore, from our previous scoping review, we selected the scales that fell into the top quartile of use within each wellbeing construct, ensuring that at least two scales were included per construct. For example, we found 15 total quality of life scales in our previous scoping review. For this systematic review, we included the four quality of life scales that were used in the highest number of studies. Across all 13 constructs, this resulted in a total of 53 scales to compare based on their psychometric and feasibility characteristics. These scales were therefore incorporated into our search strategy, and are hereafter referred to as "study scales" (Supplemental Data, Table S2).

Information sources /searches

Based on the COSMIN search filters[20] and guidance from a research librarian, we conducted a literature search from inception until March 2023 of MEDLINE, EMBASE, Web of Science Core Collection, and Google Scholar (first 200 articles) per Bramer et al.'s suggestions with CINAHL, PsycINFO and ERIC included as specialized databases.[26] Searches were performed with no date, country, or

language restrictions. Supplemental Data, Table S3 includes our full search strategies.

Gray literature sources such as theses, dissertations, and conference abstracts were included via the database searches, such as MedEdPORTAL (via MEDLINE), Conference Proceedings Citation Index (via WOSCC) and abstracts indexed in Embase. We also searched ClinicalTrials.gov as well as ProQuest Dissertations and Theses Global.

Inclusion criteria

Peer-reviewed articles in English were included in the review if they reported original research evaluating a measurement property of a study scale per the COSMIN risk of bias checklist[21] in a medical student sample. Studies where medical students form only part of the study population were included if results specific to the medical student subgroup were described. While the study scales were derived from a scoping review that only included studies measuring wellbeing over multiple timepoints, we recognized that psychometric evidence can also arise from cross-sectional studies and therefore included all psychometric studies regardless of number of timepoints in this follow-up systematic review.

YK and HL participated in the screening phases. After deduplication by Covidence software,[27] title/abstract and full text screening were performed independently by two reviewers. During the full-text screening phase, review article reference lists were hand-searched and relevant articles were extracted and reintroduced to the initial screening stage (Figure 1). Disagreements on inclusion/exclusion at both stages were resolved by consensus between the two reviewers. Inter-rater agreement was not calculated.

Data extraction

Our team collaboratively developed a data charting form (Supplemental Data, Table S4) which included details on the article, sample characteristics, study design, and measurement properties. Data extraction was completed independently by two reviews, with regular meetings to discuss and resolve conflicts. Psychometric properties extracted included details regarding content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, construct validity and responsiveness to change, as per the COSMIN taxonomy[19] (Table 1) and guidelines.[18]

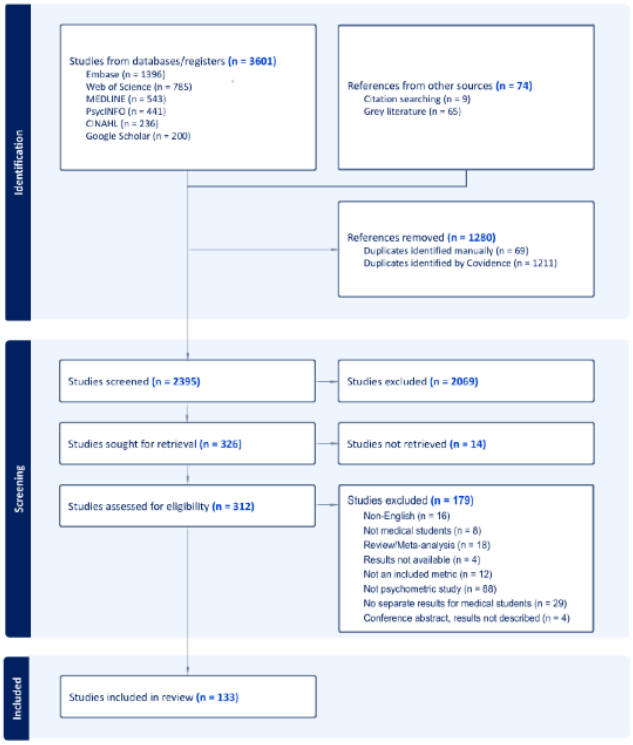

Figure 1. PRISMA Diagram

We also reviewed original derivation studies and distribution webpages to record practical characteristics of each scale such as its purpose, length, completion time, response options, recall period, cost, and other administration logistics. Based on Dyrbye et al.'s guidance on scale selection,[14] we examined feasibility of each scale primarily through the lens of organizational burden (cost) and respondent burden (completion time).[14]

Data analysis

We evaluated the methodological quality of included studies using the COSMIN risk of bias checklist.[21] Only applicable modules of the checklist were applied, depending on the goals of the study in question. For example, studies that only assessed internal consistency were only scored on that domain. As per the COSMIN guideline,[18] we used the "lowest score counts" principle in assessing overall score for a domain, rating it as "very good," "adequate," "doubtful," or "inadequate."

For each study, we then applied the criteria for good measurement properties from the COSMIN guideline,[18] rating property results as sufficient (+), insufficient (-) or indeterminate (?). Structural validity was reported as inconsistent (I) when there were both studies with sufficient evidence as well as studies with insufficient evidence. For studies on construct validity and responsiveness via convergent/divergent validity, we compared the results to pre-defined hypotheses set by the

study team before data analysis (Supplemental Data, Table S5) as recommended by the COSMIN guidelines.[18]

Using R,[28] we quantitatively pooled results on internal consistency and reliability when there were at least two studies that reported the same property for a given scale, using the procedures described by Krieglstein et al.[29] For the remaining properties, we qualitatively summarized the study results.

Lastly, we graded the certainty of the evidence using the modified GRADE approach outlined in the COSMIN guideline (high, moderate, low, very low).[18] This included consideration of risk of bias (methodological quality of the studies), inconsistency (unexplained inconsistency of results), and imprecision (total sample size).

Results

Study characteristics

Out of 2395 total abstracts, 326 passed to full-text screening and 133 studies were included (Table 2). Over a quarter (14/53, 26.4%) of study scales that we sought to compare did not have any studies supporting their validity and reliability in medical students (italicized in Supplemental Data, Table S2). Out of the included studies, 26 (19.5%) were multicentre and the majority (71/133, 53.4%) were conducted in Asia (Table 2). Only 20 studies reported participant race/ethnicity, 12 of which were in Malaysia. Internal consistency was the most studied property (118 studies) whereas there were no studies examining measurement.

Burnout

There were 31 studies examining three burnout scales.[8,30–59] All scales had sufficient evidence of internal consistency and construct validity. The Maslach Burnout Inventory-Student Scale (MBI-SS) was the only scale with sufficient evidence of cross-cultural validity/measurement invariance whereas the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) was the only scale with sufficient evidence of responsiveness. The MBI variants are priced at $2.50 per person whereas the Oldenburg Burnout Inventory (OLBI) is free to use. All scales are available online and require 10-15 minutes for completion.

Anxiety

Twenty-eight papers studied anxiety across six scales,[31,34,35,39,43,60–82] which each had sufficient evidence of internal consistency and construct validity. There was sufficient evidence to support the responsiveness of the Depression and Anxiety Stress Scales-21 (DASS-21) and the cross-cultural validity of the Hospital Anxiety and Depression Scale (HADS). All six scales are available online at no cost, with the General Anxiety Disorder-7 (GAD-7) and HADS being the shortest (five minutes) and the Depression and Anxiety Stress Scales-42 (DASS-42) being the longest (20-30 minutes).

Table 2. Characteristics of included studies

Study Characteristics		Overall (*N* = 133)
Continent	Africa	4 (3.0%)
	Asia	71 (53.4%)
	Europe	20 (15.0%)
	Asia/Europe	4 (3.0%)
	North America	20 (15.0%)
	South America	9 (6.8%)
	Oceania	5 (3.8%)
Decade	1980s	6 (4.5%)
	1990s	2 (1.5%)
	2000s	10 (7.5%)
	2010s	51 (38.3%)
	2020s	16 (12.0%)
Student Type	Missing	48 (36.1%)
	Pre-clinical	31 (23.3%)
	Clinical	12 (9.0%)
	Both	73 (54.9%)
	Missing	17 (12.8%)
Setting	Single Centre	107 (80.5%)
	Multicentre	26 (19.5%)
Sample Size	<100	8 (6.0%)
	100-199	34 (25.6%)
	200-299	27 (20.3%)
	300-399	19 (14.3%)
	400-499	13 (9.8%)
	500+	32 (24.1%)
Property Studied	Content Validity	6 (4.5%)
	Structural Validity	34 (25.6%)
	Internal Consistency	118 (88.7%)
	Cross-cultural validity/Measurement invariance	6 (4.5%)
	Reliability	15 (11.3%)
	Measurement Error	0 (0.0%)
	Criterion Validity	2 (1.5%)
	Construct Validity	106 (79.7%)
	Responsiveness	2 (1.5%)

Characteristics of the included scales are outlined in Appendix A Table 1A,. Psychometric evidence for each scale is summarized in Appendix A, Table 2A with full details in Supplemental Data, Table S6.

Mood & affect

Mood & affect was the most studied construct, encompassing 45 psychometric studies across eight unique metrics,[31,33–35,39,43–46,52,61–65,71–100] three of which concurrently assess anxiety. There was sufficient evidence to support the internal consistency and construct validity of every scale. The Beck Depression Inventory (BDI)-II had sufficient evidence of inter-rater reliability, and the Center for Epidemiologic Studies Depression Scale (CES-D) and Patient Health Questionnaire-9 (PHQ-9) were the only measures with criterion validity out of all study scales.

None of the scales have an associated cost and the HADS takes the shortest amount of time to complete (2-5 minutes).

Stress

There were 28 studies that provided psychometric evidence for seven stress scales. The majority of these studies (61.5%) examined the Perceived Stress Scale-10 (PSS-10),[43,63,70,76,78,81,98,99,101–120] the majority of which (61.5%) studied the Perceived Stress Scale-10 (PSS-10). Only the Perceived Medical School Stress (PMSS) and PSS-10 had sufficient evidence of internal consistency and only the Medical Student Stress Questionnaire-40-Revised (MSSQ-40-R) and PSS-10 had sufficient evidence of inter-rater reliability. The PSS-10 also had sufficient evidence of responsiveness whereas the PMSS had sufficient evidence of cross-cultural validity. All scales are free and available online. The Perceived Stress Scale-4 (PSS-4) has the shortest completion time (5-10 minutes), whereas the Bandura-Rosenthal Metrics for Assessing Stress (BAROMAS) has the longest (30 minutes).

Quality of life

Despite there only being two quality of life scales, they were heavily studied (24 papers).[32,35,45,76,84,87,94,96,115,119,121–134] Both the Satisfaction with Life Scale (SWLS) and the World Health Organization Quality of Life Brief Version (WHOQOL-BREF) had sufficient evidence to support their internal consistency and construct validity, however, only the WHOQOL-BREF had sufficient evidence of structural validity. While the two scales are freely available, the SWLS is significantly shorter (5-10 minutes vs 15-20 minutes for the WHOQOL-BREF).

General well-being

Twelve studies provided psychometric evidence for four scales on general well-being.[59,77,84,110,135,136,137(p12),138–142] The Medical Student Well-Being Index (MSWBI) was unique in being the only included scale overall that had sufficient evidence of content validity. It also had sufficient evidence of structural validity. While the General Health Questionnaire - 12 (GHQ-12) had sufficient evidence of internal consistency, this property was not relevant for the MSWBI given it was designed to measure a variety of different constructs. Both the MSWBI and Symptom Checklist-90-Revised (SCL-90-R) have licensing costs (although the MSWBI is free for research purposes) and also have training available for prospective surveyors. The GHQ-12 is the shortest out of the four scales (five minutes).

Mindfulness

There were five studies examining the two included mindfulness scales.[31,70,76,99,143] While both scales had sufficient evidence of internal consistency and construct validity, only the Mindful Attention Awareness Scale (MAAS) had sufficient evidence of structural validity, inter-rater reliability, and responsiveness. Both scales are freely accessible and available online, with the MAAS being slightly shorter (10-15 minutes) compared to the Five Facets of Mindfulness Questionnaire (FFMQ, 15-20 minutes).

Coping & resilience

Ten studies provided psychometric support for three coping & resilience scales.[66,73,74,82,83,142,144–147] The Brief COPE and Connor-Davidson Resilience Scale (CD-RISC) had sufficient evidence of construct validity. Apart from the CD-RISC, neither of the other two scales had sufficient evidence of structural validity or internal consistency. Among the three scales, the Brief COPE and Ways of Coping Checklist (WCCL) are available at no cost, but the WCCL is slightly longer (15-20 minutes) than the other two (10-15 minutes).

Self-esteem

There was a single study that showed sufficient evidence of internal consistency for the General Self-Efficacy Scale (GSES).[148] The Rosenberg Self-Esteem Scale (RSES) on the other hand had seven studies that when pooled, supported its internal consistency and construct validity.[68,121,127,145,148–150] The GSES, however, is noticeably shorter (four minutes vs 10-15 minutes for the RSES).

Sleep

The Pittsburgh Sleep Quality Index (PSQI, n = 11)[35,51,62,75,106,117,136,151–154] had nearly twice as many studies examining its psychometrics compared to the Epworth Sleep Scale (ESS, n = 6)[35,75,153,155–157] but did not have sufficient evidence supporting its internal consistency. The ESS additionally had evidence of inter-rater reliability whereas the PSQI had evidence of responsiveness. Both scales are freely accessible and available online. The ESS requires 5-10 minutes for completion, while the PSQI requires 10-15 minutes.

Social support & loneliness

There were eleven studies that conducted psychometric analyses on two social support and loneliness scales.[34,68,79,84,86,97,123,139,150,158,159] While the UCLA Loneliness Scale only had sufficient evidence of internal consistency and construct validity, the Multidimensional Scale of Perceived Social Support (MSPSS) additionally had

evidence to support its structural validity, cross-cultural validity/measurement invariance, and inter-rater reliability. Both scales are available online at no cost with an estimated completion time of 10-15 minutes.

Work engagement

The Utrecht Work Engagement Scale (UWES) was the only work-related scale with evidence to support its validity and reliability. Two papers provided support for the internal consistency and construct validity of the UWES.[31,120] The scale is freely accessible online with a completion time of 5-10 minutes.

Discussion

Researchers have used a breadth of scales to measure medical student wellbeing longitudinally, yet the evidence supporting their use in this population varies widely. In this systematic review of 53 commonly used scales, over a quarter had no psychometric evidence in medical students. Furthermore, the availability of evidence depended heavily on the specific psychometric property: most scales had evidence of internal consistency and construct validity via convergence/divergence but very few had data supporting their content validity, cross-cultural validity or responsiveness. Overall, this raises concerns about the ongoing use of these scales and highlights a critical gap that impairs the accurate measurement of medical student wellbeing.

Comparison to prior studies

Lall et al.'s two-part scoping review[15,16] previously found 27 assessment tools used to measure physician burnout, anxiety, depression, resilience, mindfulness, mood, personality, well-being, quality of life, and stress. Haykal et al.'s scoping review[17] also found six measures of general wellbeing used specifically amongst medical students. However, neither of these reviews clearly described how they determined the strength of psychometric evidence (i.e. what is "good" validity) nor considered the methodological quality of their included studies.

The COSMIN methodology has previously been used to examine the psychometrics of specific wellbeing scales without a target population (e.g., DASS-21[160] and CD-RISC[161]). Shoman et al also conducted a systematic review of five burnout measures where they found that the OLBI had the most complete validation.[162]

Our study therefore used this systematic and rigorous approach to select and evaluate the psychometric and feasibility characteristics of 53 scales commonly used to assess medical student wellbeing-related constructs longitudinally. Through directly comparing scales measuring the same wellbeing-related construct, we can provide preliminary recommendations on which scales seem most appropriate for use in this population.

Recommendations for wellbeing scales

Dyrbye et al.[14] previously outlined important considerations for organizations selecting a wellbeing instrument. These include measuring important constructs, low respondent and organizational burden (for example, length and cost, respectively), correlation with other important outcomes, sensitivity to change, and psychometric evidence. Brady et al[163] further highlighted the importance of interpreting psychometric evidence in the context of its proposed use. Indeed, best practices[22] suggest that validity and reliability are not characteristics of a scale itself, but rather of its application in a specific scenario, with a specific population. Unfortunately, many medical student wellbeing scales have never been validated despite extensive use in this population.

In Table 3, we highlight our collective recommendations on the best scales for each construct. We sought to balance the strength of their psychometric evidence, quality of the evidence, and feasibility considerations; readers may come to their own conclusions based on the evidence presented in the results section and their priorities. Furthermore, while psychometric evidence in medical students is ideal (and was the focus of this review), we also acknowledge that evidence from closely related populations (e.g. physicians, other health professional students) could be considered in the totality of evidence for a given scale in different learning environments and contexts.

Strengths and limitations

Our review benefits from a broad search strategy using previously validated search filters for psychometric studies and a comprehensive approach to database and gray literature selection. We focus on scales used for longitudinal measurement given the need to understand wellbeing changes over time including the impact of interventions. However, this means that we are missing instruments that are primarily used in cross-sectional studies. We also excluded non-English papers for feasibility reasons and acknowledge this limits the breadth of our included studies, particularly as it relates to studies on cross-cultural validity. We may also be missing psychometric evidence that has not been published or included in the grey literature that we searched.

Table 3. Recommendations for medical student wellbeing scales

Construct	Metric	Comparative Strengths	Comparative Limitations
Burnout	Oldenburg Burnout Inventory (OLBI)	Free	Only evidence for internal consistency and construct validity
	Maslach Burnout Inventory - Human Services Survey (MBI-HSS)	Good psychometric evidence	Associated cost
	Maslach Burnout Inventory - Student Survey (MBI-SS)	Good psychometric evidence	Associated cost
Anxiety	**General Anxiety Disorder-7 (GAD-7)**	Short, free, moderate psychometric evidence	
Mood & Anxiety	**Hospital Anxiety and Depression Scale (HADS)**	Short, free, evidence for cross-cultural validity, assesses both depression and anxiety	Lower internal consistency
Mood	**Patient Health Questionnaire-9 (PHQ-9)**	Short, free, evidence for criterion validity*	
	Center for Epidemiologic Studies Depression Scale (CES-D)	Free, evidence for criterion validity*	Longer
Stress	**Perceived Stress Scale-10 (PSS-10)**	Short, free, strong psychometric evidence, studied more widely	
Quality of Life	**Satisfaction with Life Scale (SWLS)**	Short, free, moderate psychometric evidence	
	World Health Organization Quality of Life Brief Version (WHOQOL-BREF)	Free, good psychometric evidence	Longer
General Well-being	**Medical Student Well-Being Index (MSWBI)**	Short, free for research purposes, strong psychometric evidence	Associated cost for non-research purposes
	General Health Questionnaire - 12 (GHQ-12)	Short, free	Only evidence for internal consistency and construct validity
Mindfulness	**Mindful Attention Awareness Scale (MAAS)**	Short, free, strong psychometric evidence	
	Five Facets of Mindfulness Questionnaire (FFMQ)	Free	Longer, weaker psychometric evidence
Coping & Resilience	Connor-Davidson Resilience Scale (CD-RISC)	Good psychometric evidence	Associated cost, longer
	Brief COPE	Short, free	Only evidence for construct validity
Self-Esteem	Rosenberg Self-Esteem Scale (RSES)	Free, good psychometric evidence, more validation studies	Longer
	General Self-Efficacy Scale (GSES)	Short, free, good psychometric evidence	Less validation studies
Sleep	**Epworth Sleep Scale (ESS)**	Short, free, good psychometric evidence	
	Pittsburgh Sleep Quality Index (PSQI)	Free	Longer, weaker psychometric evidence
Social Support & Loneliness	**Multidimensional Scale of Perceived Social Support (MSPSS)**	Short, free, strong psychometric evidence	
	UCLA Loneliness Scale	Short, free	Weaker psychometric evidence
Work	**Utrecht Work Engagement Scale (UWES)**	Free, good psychometric evidence	

*Bold refers to the study team's recommended scale based on existing evidence, where a clear best option exists. *Criterion validity for depression scales refers to validity when compared to a diagnostic interview*

Furthermore, we outline evidence for the most used measures specifically to be able to highlight the availability and quality of psychometric evidence for tools that are already being widely used and replicated in the literature. We felt that the results generated from such a review would be more actionable for administrators and researchers who are actively using these measures and may want to understand the underlying evidence and potential alternatives. While we feel this made our review more relevant and feasible, there may be infrequently used tools or newly developed tools that have more robust evidence than the included measures. Lastly, our included constructs are based on our preceding scoping review's search and inclusion strategy. While there are many wellbeing-related constructs included, we acknowledge that some may be missing, depending on the definition of wellbeing that is used.

Future directions

Evidently, future validation studies are needed to fill the observed gaps in psychometric evidence for medical student wellbeing scales. It is also clear that just because scales are being frequently used with medical students, it does not mean that they have psychometric evidence supporting their use in that context. This is particularly important given the increasing diversity of the medical student population and the lack of demographic reporting of the included psychometric studies. Given the breadth of scales in use, it would be inefficient to conduct validation studies on all of them. Therefore, consensus directions are needed to identify the most important constructs to measure (e.g., burnout, depression). Furthermore, within each construct, the scales that have the most potential based on existing psychometric properties in medical students, evidence in other populations, and feasibility

characteristics should be the focus of further validation studies.

Conclusions

Many scales used to measure medical student wellbeing longitudinally have no medical student-specific psychometric evidence. Amongst those that do, few have evidence to support their content validity, cross-cultural validity, and responsiveness. There is a significant gap in context-specific psychometric evidence despite the widespread use of numerous wellbeing scales. Based on the available psychometric evidence and feasibility considerations, we have provided tentative recommendations on which scales are most appropriate to measure medical student wellbeing; however, we acknowledge that significant future research is needed to better inform scale selection.

Acknowledgments: The authors would like to acknowledge Janice Kung for her assistance with the search strategy and Ryan Jacobson for his feedback on the study methods.
Conflicts of interest: Henry Li and Melanie Lewis have received a CMA Healthcare Unburdened Program Grant for unrelated work. Melanie Lewis serves as Chief Wellbeing Officer at the University of Alberta. Victor Do serves as co-chair of the Culture of Academic Medicine Initiative.
Funding: This work received no funding support
Authorship: All authors have each signed their own attestation statement that they meet the requirements of authors. The lead author, in addition to their own attestation as an author, has signed an attestation that all authors listed on this paper meet the requirements for authors.
Edited by: Marco Zaccagnini (senior section editor); Marcel D'Eon (editor-in-chief)

References

1. Brazeau CMLR, Shanafelt T, Durning SJ, et al. Distress among matriculating medical students relative to the general population. *Acad Med*. 2014;89(11):1520-1525. https://doi.org/10.1097/ACM.0000000000000482
2. Maser B, Danilewitz M, Guérin E, Findlay L, Frank E. Medical Student Psychological Distress and Mental Illness Relative to the General Population: A Canadian Cross-Sectional Survey. *Acad Med*. 2019;94(11):1781-1791. https://doi.org/10.1097/ACM.0000000000002958
3. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis. *JAMA*. 2016;316(21):2214. https://doi.org/10.1001/jama.2016.17324
4. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ*. 2016;50(1):132-149. https://doi.org/10.1111/medu.12927
5. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Acad Med*. 2006;81(4):354-373.
6. Dyrbye LN, Harper W, Moutier C, et al. A multi-institutional study exploring the impact of positive mental health on medical students' professionalism in an era of high burnout: *Acad Med*. 2012;87(8):1024-1031. https://doi.org/10.1097/ACM.0b013e31825cfa35
7. Dyrbye LN, Thomas MR, Shanafelt TD, et al. Burnout and serious thoughts of dropping out of medical school: a multi-institutional study. *Acad Med*. 2010;85(1):94-102. https://doi.org/10.1097/ACM.0b013e3181c46aad
8. Brazeau CMLR, Schroeder R, Rovi S, Boyd L. Relationships between medical student burnout, empathy, and professionalism climate. *Acad Med*. 2010;85:S33-S36. https://doi.org/10.1097/ACM.0b013e3181ed4c47
9. Charles ST, Piazza JR, Mogle J, Sliwinski MJ, Almeida DM. The wear and tear of daily stressors on mental health. *Psychol Sci*. 2013;24(5):733-741. https://doi.org/10.1177/0956797612462222
10. Herr R, Barrech A, Riedel N, et al. Long-term effectiveness of stress management at work: effects of the changes in perceived stress reactivity on mental health and sleep problems seven years later. *Int J Environ Res Public Health*. 2018;15(2):255. https://doi.org/10.3390/ijerph15020255
11. Dyrbye LN, Lipscomb W, Thibault G. Redesigning the learning environment to promote learner well-being and professional development. *Acad Med*. 2020;95(5):674-678. https://doi.org/10.1097/ACM.0000000000003094
12. Li H, Upreti T, Do V, et al. Measuring wellbeing: ascoping review of metrics and studies measuring medical student wellbeing across multiple timepoints. *Med Teach*. Published online July 5, 2023:1-20. https://doi.org/10.1080/0142159X.2023.2231625
13. Li H, Do V, Kassam A. Evidence based on correlation with invalid scales? The need to understand medical student-specific validity of wellbeing scales. *Acad Med J Assoc Am Med Coll*. In Press.
14. Dyrbye LN, Meyers D, Ripp Jet al. A pragmatic approach for organizations to measure health care professional well-being. *NAM Perspect*. Published online Oct 1, 2018. https://doi.org/10.31478/201810b
15. Lall MD, Gaeta TJ, Chung AS, et al. Assessment of physician well-being, part two: beyond burnout. *West J Emerg Med*. 2019;20(2):291-304. https://doi.org/10.5811/westjem.2019.1.39666
16. Lall MD, Gaeta TJ, Chung AS, et al. Assessment of physician well-being, part one: burnout and other negative states. *West J Emerg Med*. 2019;20(2):278-290. https://doi.org/10.5811/westjem.2019.1.39665
17. Haykal KA, Pereira L, Power A, Fournier K. Medical student wellness assessment beyond anxiety and depression: a scoping review. *PloS One*. 2022;17(10):e0276894. https://doi.org/10.1371/journal.pone.0276894
18. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. https://doi.org/10.1007/s11136-018-1798-3
19. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006
20. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123. https://doi.org/10.1007/s11136-009-9528-5
21. Mokkink LB, de Vet HCW, Prinsen C a. C, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*.

2018;27(5):1171-1179. https://doi.org/10.1007/s11136-017-1765-4

22. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 2014.
23. Li H. *Review of measurement properties for common metrics used to measure medical student wellbeing over multiple timepoints.* Published online Oct 23, 2023. Available from https://osf.io/s8qfg [Accessed Oct 24, 2025].
24. Elsman EBM, Mokkink LB, Terwee CB, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *J Clin Epidemiol*. 2024;173:111422. https://doi.org/10.1016/j.jclinepi.2024.111422
25. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? a systematic review of its definition and measurement. *Acad Psychiatry*. 2018;42(1):94-108. https://doi.org/10.1007/s40596-017-0781-6
26. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6(1):245. https://doi.org/10.1186/s13643-017-0644-y
27. McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. *Syst Rev*. 2021;10(1):38. https://doi.org/10.1186/s13643-021-01583-y
28. R Core Team. *R: A language and environment for statistical computing.* Published online 2021. https://www.R-project.org/
29. Krieglstein F, Beege M, Rey GD, et al. A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educ Psychol Rev*. 2022;34(4):2485-2541. https://doi.org/10.1007/s10648-022-09683-4
30. Pharasi S, Patra S. Burnout in medical students of a tertiary care Indian medical center: how much protection does resilience confer? *Indian J Psych*. 2020;62(4):407-412. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_681_19
31. Zúñiga D, Torres-Sahli M, Rigotti A, et al. Dispositional mindfulness reduces burnout and promotes flourishing in medical students: a two-wave latent change score model. *Mindfulness*. 2022;13(1):112-122. https://doi.org/10.1007/s12671-021-01774-7
32. Paro HBMS, Silveira PSP, Perotta B, et al. Empathy among medical students: is there a relation with quality of life and burnout? *PloS One*. 2014;9(4):e94133. https://doi.org/10.1371/journal.pone.0094133
33. Lapinski J, Yost M, Sexton P, LaBaere RJ. Factors modifying burnout in osteopathic medical students. *Acad Psychiatry J Am Assoc Dir Psychiatr Resid Train Assoc Acad Psychiatry*. 2016;40(1):55-62. https://doi.org/10.1007/s40596-015-0375-0
34. Sanchez DJ, Strauman TJ, Compton S. Impact of student perceptions of the educational program on burnout in medical school. *Med Sci Educ*. 2019;29(4):1077-1087. https://doi.org/10.1007/s40670-019-00812-3
35. Miguel A de QC, Tempski P, Kobayasi R, Mayer FB, Martins MA. Predictive factors of quality of life among medical students: results from a multicentric study. *BMC Psychol*. 2021;9(1):36. https://doi.org/10.1186/s40359-021-00534-5
36. Shi Y, Gugiu PC, Crowe RP, Way DP. A Rasch analysis validation of the Maslach Burnout Inventory-Student Survey with preclinical medical students. *Teach Learn Med*. 2019;31(2):154-169. https://doi.org/10.1080/10401334.2018.1523010
37. Obregon M, Luo J, Shelton J, Blevins T, MacDowell M. Assessment of burnout in medical students using the Maslach Burnout Inventory-Student Survey: a cross-sectional data analysis. *BMC Med Educ*. 2020;20(1):376. https://doi.org/10.1186/s12909-020-02274-3
38. Periasamy P, Suganthi V, Sukala PM, et al. Burnout among medical students and correlation with academic performance, sleep quality during covid19 pandemic online class in erode district. *Pharmacologyonline*. Published online 2021:962-971.
39. Calcatin S, Sinval J, Neto LL, et al. Burnout and dropout intention in medical students: the protective role of academic engagement - *BMC Med Ed*. 2022 Feb 7;22(1):83. https://doi.org/10.1186/s12909-021-03094-9.
40. Galán F, Sanmartín A, Polo J, Giner L. Burnout risk in medical students in Spain using the Maslach Burnout Inventory-Student Survey. *Int Arch Occup Environ Health*. 2011;84(4):453-459. https://doi.org/10.1007/s00420-011-0623-x
41. Ilic M, Todorovic Z, Jovanovic M, Ilic I. Burnout syndrome among medical students at one university in serbia: validity and reliability of the Maslach Burnout Inventory-Student Survey. *Behav Med Wash DC*. 2017;43(4):323-328. https://doi.org/10.1080/08964289.2016.1170662
42. Rudinskaitė I, Mačiūtė E, Gudžiūnaitė G, Gerulaitytė G. Burnout syndrome amongst medicine students in Lithuania and Germany. *Acta Medica Litu*. 2020;27(2):53-60. https://doi.org/10.15388/Amed.2020.27.2.2
43. Does mental toughness buffer the relationship between perceived stress, depression, burnout, anxiety, and sleep? Available from https://psycnet.apa.org/record/2018-34737-001 [Accessed Fev 4, 2024].
44. Puranitee P, Saetang S, Sumrithe S, Busari JO, van Mook WNKA, Heeneman S. Exploring burnout and depression of Thai medical students: the psychometric properties of the Maslach Burnout Inventory. *Int J Med Educ*. 2019;10:223-229. https://doi.org/10.5116/ijme.5dc6.8228
45. Pagnin D, de Queiroz V. Influence of burnout and sleep difficulties on the quality of life among medical students. *SpringerPlus*. 2015;4:676. https://doi.org/10.1186/s40064-015-1477-6
46. Al-Alawi M, Al-Sinawi H, Al-Qubtan A, et al. Prevalence and determinants of burnout Syndrome and Depression among medical students at Sultan Qaboos University: a cross-sectional analytical study from Oman. *Arch Environ Occup Health*. 2019;74(3):130-139. https://doi.org/10.1080/19338244.2017.1400941
47. Whistle C. Resilience and Burnout in Second- and Third-Year Medical Students. *USF Tampa Grad Theses Diss*. Published online January 26, 2021. Available from https://digitalcommons.usf.edu/etd/8887 [Accessed on Nov 17, 2025].
48. Lee SJ, Choi YJ, Chae H. The effects of personality traits on academic burnout in Korean medical students. *Integr Med Res*. 2017;6(2):207-213. https://doi.org/10.1016/j.imr.2017.03.005
49. Shin HS, Park H, Lee YM. The relationship between medical students' empathy and burnout levels by gender and study years. *Patient Educ Couns*. 2022;105(2):432-439. https://doi.org/10.1016/j.pec.2021.05.036
50. Dyrbye LN, Satele D, West CP. Association of characteristics of the learning environment and US medical student burnout, empathy, and career regret. *JAMA Netw Open*. 2021;4(8):e2119110. https://doi.org/10.1001/jamanetworkopen.2021.19110
51. Shad R, Thawani R, Goel A. Burnout and sleep quality: a cross-sectional questionnaire-based study of medical and non-medical students in India. *Cureus*. 2015;7(10):e361. https://doi.org/10.7759/cureus.361

52. Njim T, Mbanga CM, Tindong M, et al. Burnout as a correlate of depression among medical students in Cameroon: a cross-sectional study. *BMJ Open*. 2019;9(5):e027709. https://doi.org/10.1136/bmjopen-2018-027709
53. Njim T, Makebe H, Toukam L, et al. Burnout syndrome amongst medical students in Cameroon: a cross-sectional analysis of the determinants in preclinical and clinical students. *Psychiatry J*. 2019;2019:4157574. https://doi.org/10.1155/2019/4157574
54. Babenko O, Mosewich A, Abraham J, Lai H. Contributions of psychological needs, self-compassion, leisure-time exercise, and achievement goals to academic engagement and exhaustion in Canadian medical students. *J Educ Eval Health Prof*. 2018;15:2. https://doi.org/10.3352/jeehp.2018.15.2
55. Runyon CR, Paniagua MA, Dyrbye LN. Exploring the validity based on internal structure of the Oldenburg Burnout Inventory - Medical Student (OLBI-MS). *Teach Learn Med*. 2023;35(1):37-51. https://doi.org/10.1080/10401334.2021.2018695
56. Greenmyer JR, Montgomery M, Hosford C, et al. Guilt and burnout in medical students. *Teach Learn Med*. 2022;34(1):69-77. https://doi.org/10.1080/10401334.2021.1891544
57. School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA, et al. Malay Language translation and validation of the Oldenburg Burnout Inventory Measuring Burnout. *Educ Med J*. 2018;10(2):27-40. https://doi.org/10.21315/eimj2018.10.2.4
58. Dahlin M, Joneborg N, Runeson B. Performance-based self-esteem and burnout in a cross-sectional study of medical students. *Med Teach*. 2007;29(1):43-48. https://doi.org/10.1080/01421590601175309
59. Masri R, Kadhum M, Farrell SM, et al. Wellbeing and mental health amongst medical students in Jordan: a descriptive study. *Int Rev Psychiatry Abingdon Engl*. 2019;31(7-8):619-625. https://doi.org/10.1080/09540261.2019.1670402
60. Akhtar M, Herwig BK, Faize FA. Depression and anxiety among International medical students in Germany: the predictive role of coping styles. *JPMA J Pak Med Assoc*. 2019;69(2):230-234.
61. Vasegh S, Mohammadi MR. Religiosity, anxiety, and depression among a sample of Iranian medical students. *Int J Psychiatry Med*. 2007;37(2):213-227. https://doi.org/10.2190/J3V5-L316-0U13-7000
62. Mendes TB, Souza KC de, França CN, et al. Physical activity and symptoms of anxiety and depression among medical students during a pandemic. *Rev Bras Med Esporte*. 2021;27:582-587. https://doi.org/10.1590/1517-8692202127062021_0059
63. Yoon S, Lee Y, Han C, et al. Usefulness of the Patient Health Questionnaire-9 for Korean medical students. *Acad Psychiatry J Am Assoc Dir Psychiatr Resid Train Assoc Acad Psychiatry*. 2014;38(6):661-667. https://doi.org/10.1007/s40596-014-0140-9
64. Bilgel N, Ozçakır A. *Depression and anxiety among medical students: examining scores of the Beck Depression and Anxiety Inventory and the Depression Anxiety and Stress Scale with student characteristics.* 2017. Available from: https://www.academia.edu/31343781/Depression_and_anxiety_among_medical_students_Examining_scores_of_the_beck_depression_and_anxiety_inventory_and_the_depression_anxiety_and_stress_scale_with_student_characteristics. [Accessed on Feb 4, 2024].
65. Ernst J, Jordan KD, Weilenmann S, et al. Burnout, depression and anxiety among Swiss medical students - A network analysis. *J Psychiatr Res*. 2021;143:196-201. https://doi.org/10.1016/j.jpsychires.2021.09.017
66. Peng L, Hu X, Lan L, Xu C, Li M. The moderating role of resilience in the relationship between state and trait anxiety and post-traumatic growth of medical freshmen. *Acta Psychol (Amst).* Oct 2022. https://doi.org/10.1016/j.actpsy.2022.103741
67. Jiang N, Sato T, Hara T, et al. Correlations between trait anxiety, personality and fatigue: study based on the Temperament and Character Inventory. *J Psychosom Res*. 2003;55(6):493-500. https://doi.org/10.1016/s0022-3999(03)00021-7
68. Wongpakaran T, Wongpakaran N, Ruktrakul R. Reliability and validity of the Multidimensional Scale of Perceived Social Support (MSPSS): Thai Version. *Clin Pract Epidemiol Ment Health CP EMH*. 2011;7:161-166. https://doi.org/10.2174/1745017901107010161
69. Hendryx MS, Haviland MG, Shaw DG. Dimensions of alexithymia and their relationships to anxiety and depression. *J Pers Assess*. 1991;56(2):227-237. https://doi.org/10.1207/s15327752jpa5602_4
70. Fino E, Martoni M, Russo PM. Specific mindfulness traits protect against negative effects of trait anxiety on medical student wellbeing during high-pressure periods. *Adv Health Sci Educ Theory Pract*. 2021;26(3):1095-1111. https://doi.org/10.1007/s10459-021-10039-w
71. Concerto C, Rodolico A, La Rosa VL, et al. Flourishing or languishing? predictors of positive mental health in medical students during the COVID-19 pandemic. *Int J Environ Res Public Health*. 2022;19(23):15814. https://doi.org/10.3390/ijerph192315814
72. Jafari P, Nozari F, Ahrari F, Bagheri Z. Measurement invariance of the Depression Anxiety Stress Scales-21 across medical student genders. *Int J Med Educ*. 2017;8:116-122. https://doi.org/10.5116/ijme.58ba.7d8b
73. Ramadianto AS, Kusumadewi I, Agiananda F, Raharjanti NW. Symptoms of depression and anxiety in Indonesian medical students: association with coping strategy and resilience. *BMC Psych*. 2022;22(1):92. https://doi.org/10.1186/s12888-022-03745-1
74. Tee KR, Ismail AS, Ang YH, et al. Prevalence of anxiety and burnout, and coping mechanisms among clinical year medical undergraduate students in Universiti Kebangsaan Malaysia amidst the COVID-19 pandemic. *Int J Environ Res Public Health*. 2022;19(20):13010. https://doi.org/10.3390/ijerph192013010
75. Kumar S, Samal A, Dwivedi S. Sleep quality, daytime sleepiness and mental health in undergraduate medical students: a cross-sectional descriptive study. *J Clin Diagn Res*. Published online 2023. https://doi.org/10.7860/JCDR/2023/60633.17524
76. Phang CK, Mukhtar F, Ibrahim N, Sidik SM. Mindful Attention Awareness Scale (MAAS): factorial validity and psychometric properties in a sample of medical students in Malaysia. *J Ment Health Train Educ Pract.* 2016;11 (5): 305–316. https://doi.org/10.1108/JMHTEP-02-2015-0011
77. Yusoff MSB, Yaacob MJ, Naing NN, Esa AR. Psychometric properties of the Medical Student Well-Being Index among medical students in a Malaysian medical school. *Asian J Psychiatry*. 2013;6(1):60-65. https://doi.org/10.1016/j.ajp.2012.09.001
78. Al-Dubai S, Barua A, Ganasegeran K, Ali Jadoo S, Rampal KG. Concurrent validity of the Malay version of Perceived Stress Scale (PSS-10). *ASEAN J Psychiatry*. 2014;15:8-13.
79. Karaoğlu N, Batı A. Medical students' loneliness, anxiety and depression at the beginning of their studies. *Int J Psychiatry Clin Pract*. 2012;16:33-34.
80. Mumford DB, Tareen IA, Bajwa MA, Bhatti MR, Karim R. The translation and evaluation of an Urdu version of the Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1991;83(2):81-85. https://doi.org/10.1111/j.1600-0447.1991.tb07370.x
81. Voltmer E, Kötter T, Spahn C. Perceived medical school stress and the development of behavior and experience patterns in German medical students. *Med Teach*. 2012;34(10):840-847. https://doi.org/10.3109/0142159X.2012.706339
82. Francis B, Gill JS, Yit Han N, et al. Religious coping, religiosity, depression and anxiety among medical students in a multi-religious

setting. *Int J Environ Res Public Health*. 2019;16(2):259. https://doi.org/10.3390/ijerph16020259
83. Vitaliano PP, Maiuro RD, Russo J, Mitchell ES, Carr JE, Van Citters RL. A biopsychosocial model of medical student distress. *J Behav Med*. 1988;11(4):311-331. https://doi.org/10.1007/BF00844933
84. Swami V, Chamorro-Premuzic T, Sinniah D, et al. General health mediates the relationship between loneliness, life satisfaction and depression. A study with Malaysian medical students. *Soc Psychiatry Psychiatr Epidemiol*. 2007;42(2):161-166. https://doi.org/10.1007/s00127-006-0140-5
85. Shim EJ, Jeon HJ, Kim H, et al. Measuring stress in medical education: validation of the Korean version of the higher education stress inventory with medical students. *BMC Med Educ*. 2016;16(1):302. https://doi.org/10.1186/s12909-016-0824-9
86. Ng CG, Amer Siddiq AN, Aida SA, Zainal NZ, Koh OH. Validation of the Malay version of the Multidimensional Scale of Perceived Social Support (MSPSS-M) among a group of medical students in Faculty of Medicine, University Malaya. *Asian J Psych*. 2010;3(1):3-6. https://doi.org/10.1016/j.ajp.2009.12.001
87. Ghassab-Abdollahi N, Shakouri SK, Aghdam AT, et al. Association of quality of life with physical activity, depression, and demographic characteristics and its predictors among medical students. *J Educ Health Promot*. 2020;9:147. https://doi.org/10.4103/jehp.jehp_91_20
88. Ongel K, Mergen H, Tan S, et al. Psychometric properties of depression in university students of Turkey. *Biomed Res*. 2010;21(3). Available from https://www.alliedacademies.org/abstract/psychometric-properties-of-depression-in-university-students-of-turkey-1316.html. [Accessed Feb 4, 2024].
89. Solibieda A, Rotsaert M, Loas G. Relationship between recent change of anhedonia and suicidal ideation taking into account the severity and the acuteness of suicidal ideation as well as the specific roles of loss of pleasure and loss of interest in people: a study on medical students. *Omega*. 2023;87(4):1280-1292. https://doi.org/10.1177/00302228211037300
90. Gomes-Oliveira MH, Gorenstein C, Lotufo Neto F, Andrade LH, Wang YP. Validation of the Brazilian Portuguese version of the Beck Depression Inventory-II in a community sample. *Rev Bras Psiquiatr Sao Paulo Braz 1999*. 2012;34(4):389-394. https://doi.org/10.1016/j.rbp.2012.03.005
91. Flaherty J, Richman J. Gender differences in the perception and utilization of social support: theoretical perspectives and an empirical test. *Soc Sci Med 1982*. 1989;28(12):1221-1228. https://doi.org/10.1016/0277-9536(89)90340-7
92. Ghubash R, Daradkeh TK, Al Naseri KS, Al Bloushi NB, Al Daheri AM. The performance of the Center for Epidemiologic Study Depression Scale (CES-D) in an Arab female community. *Int J Soc Psychiatry*. 2000;46(4):241-249. https://doi.org/10.1177/002076400004600402
93. Valencia J, Christian F. The relationship between neuroticism, emotion regulation, and academic stress with depression symptoms in Indonesian medical student. *Open Access Maced J Med Sci*. 2022 Feb. 15. https://doi.org/10.3889/oamjms.2022.8322
94. Renn D, Pfaffenberger N, Platter M, et al. International Well-being Index: the Austrian version. *Soc Indic Res*. 2009;90:243-256. https://doi.org/10.1007/s11205-008-9255-2
95. Baptista T, Vargas O, Colmenares R, et al. Positive and Negative Affect Schedule (PANAS): psychometric properties of a Venezuelan Spanish version in medical students. *Investig Clínica*. 2020;61(4):301-315. https://doi.org/10.22209/ic.v61n4a01
96. Asif M, Idrees M, Ghazal S, Ishaq G. Relationship of emotional intelligence and life satisfaction: mediating role of affectivity in medical students. *ASEAN J Psychiatry*. Feb 2022;23:1-8.
97. Karawekpanyawong N, Wongpakaran T, Wongpakaran N, et al. Impact of perceived social support on the relationship between ADHD and depressive symptoms among first year medical students: a structural equation model approach. *Child Basel Switz*. 2021;8(5):401. https://doi.org/10.3390/children8050401
98. Zhao FF, Yang L, Ma JP, Qin ZJ. Path analysis of the association between self-compassion and depressive symptoms among nursing and medical students: a cross-sectional survey. *BMC Nurs*. 2022;21(1):67. https://doi.org/10.1186/s12912-022-00835-z
99. Alzahrani AM, Hakami A, AlHadi A, et al. The interplay between mindfulness, depression, stress and academic performance in medical students: a Saudi perspective. *PloS One*. 2020;15(4):e0231088. https://doi.org/10.1371/journal.pone.0231088
100. Dian CN, Effendy E, Amin MM. The Validation of Indonesian Version of Patient Health Questionnaire-9. *Open Access Maced J Med Sci*. https://doi.org/10.3889/oamjms.2022.9293
101. Rosenthal TL, Rosenthal RH, Edwards NB. Students' self-ratings of stress in medical school: a replication across 20 months. *Behav Res Ther*. 1990;28(2):171-173. https://doi.org/10.1016/0005-7967(90)90030-m
102. Chia T, Oyeniran OI, Oraebosi MI, Dane S. Study of stressors in a cohort of undergraduate medical students: implications for student support. *J Res Med Dent Sci,* 2021, 9 (2): 23-29.
103. Yusoff MSB. A confirmatory factor analysis study on the medical student stressor questionnaire among Malaysian medical students. *Educ Med J*. 2011;3(1). https://doi.org/10.5959/eimj.v3i1.95
104. Yusoff MSB. A multicenter study on validity of the Medical Student Stressor Questionnaire (MSSQ). *Internat Med J*. Mar 1, 2011. Available at https://openurl.ebsco.com/contentitem/gcd:59688789?sid=ebsco:plink:crawler&id=ebsco:gcd:59688789. [Accessed on Feb 4, 2024].
105. Gupta S, Choudhury S, Das M, Mondol A, Pradhan R. Factors causing stress among students of a medical college in Kolkata, India. *Educ Health Abingdon Engl*. 2015;28(1):92-95. https://doi.org/10.4103/1357-6283.161924
106. Itagi ABH, Dharmalingam A, Dipankar S, et al. Influence of stress on quality of sleep and QT interval variables among young adult medical students- a cross-sectional study. *J Clin Diagn Res*. Published online 2021. https://doi.org/10.7860/JCDR/2021/50023.15386
107. Saiful M, Yusoff B, Rahim AF, Jamil M. The development and validity of the medical student stress questionnaire (MSSQ). In: 2010. Available from https://www.semanticscholar.org/paper/THE-DEVELOPMENT-AND-VALIDITY-OF-THE-MEDICAL-STUDENT-Saiful-Yusoff/4c7e3c6a2948bd424adec12e067eb1b9cd4d8f58. [Accessed Feb 4, 2024].
108. Yusoff MSB. The stability of MSSQ to measure stressors among medical students. *Int Med J 1994*. 2013;20:1-3.
109. Dagani J, Buizza C, Ferrari C, Ghilardi A. Psychometric validation and cultural adaptation of the Italian medical student stressor questionnaire. *Curr Psychol*. 2022;41(6):4132-4140. https://doi.org/10.1007/s12144-020-00922-x
110. Jayarajah U, Lakmal K, Athapathu A, Jayawardena AJ, de Silva V. Validating the Medical Students' Stressor Questionnaire (MSSQ) from a Sri Lankan medical faculty. *J Taibah Univ Med Sci*. 2020;15(5):344-350. https://doi.org/10.1016/j.jtumed.2020.08.003
111. Kötter T, Voltmer E. Measurement of specific medical school stress: translation of the "Perceived Medical School Stress Instrument" to

the German language. *GMS Z Med Ausbild*. 2013;30(2):Doc22. https://doi.org/10.3205/zma000865

112. Marchewka W, Loster Z, Marchewka J, Olszewska-Turek K, Kopeć G. Stress associated with undergraduate medical courses: a translation and validation of the Perceived Medical School Stress Instrument into Polish and its adaptation to the Polish environment. *Folia Med Cracov*. 2020;60(2):55-66. https://doi.org/10.24425/fmc.2020.135013
113. Chakraborti A, Ray P, Sanyal D, et al. Assessing perceived stress in medical personnel: in search of an appropriate scale for the Bengali population. *Indian J Psychol Med*. 2013;35(1):29-33. https://doi.org/10.4103/0253-7176.112197
114. Al-Dubai SAR, Alshagga MA, Rampal KG, Sulaiman NA. Factor structure and reliability of the Malay version of the Perceived Stress Scale among Malaysian Medical Students. *Malays J Med Sci MJMS*. 2012;19(3):43-49.
115. Badura-Brzoza K, Dębski P, Główczyński P, Dębska-Janus M, Gorczyca P. Life satisfaction and perceived stress versus health promoting behavior among medical students during the COVID-19 pandemic. *Int J Environ Res Public Health*. 2022;19(11):6706. https://doi.org/10.3390/ijerph19116706
116. Gupta R, Singh N, Kumar R. Longitudinal predictive validity of emotional intelligence on first year medical students perceived stress. *BMC Med Educ*. 2017;17(1):139. https://doi.org/10.1186/s12909-017-0979-z
117. Boyd A, Mealand K, Briggs Early K, Oestreich E. Perceived stress, grit, and self-care behaviors in first-year medical students. *Amer J Lifestyle Med*. 2022;17(6):803-812. https://doi.org/10.1177/15598276221124576
118. Li Z, Yi X, Zhong M, et al. psychological distress, social support, coping style, and perceived stress among medical staff and medical students in the early stages of the COVID-19 epidemic in China. *Front Psychiatry*. 2021;12:664808. https://doi.org/10.3389/fpsyt.2021.664808
119. Shi M, Wang X, Bian Y, Wang L. The mediating role of resilience in the relationship between stress and life satisfaction among Chinese medical students: a cross-sectional study. *BMC Med Educ*. 2015;15:16. https://doi.org/10.1186/s12909-015-0297-2
120. Agarwal G, Mosquera M, Ring M, Victorson D. Work engagement in medical students: an exploratory analysis of the relationship between engagement, burnout, perceived stress, lifestyle factors, and medical student attitudes. *Med Teach*. 2020;42(3):299-305. https://doi.org/10.1080/0142159X.2019.1679746
121. Wang Q, Wu H. Associations between maladaptive perfectionism and life satisfaction among Chinese undergraduate medical students: the mediating role of academic burnout and the moderating role of self-esteem. *Front Psychol*. 2021;12:774622. https://doi.org/10.3389/fpsyg.2021.774622
122. Wang Q, Wang L, Shi M, et al. Empathy, burnout, life satisfaction, correlations and associated socio-demographic factors among Chinese undergraduate medical students: an exploratory cross-sectional study. *BMC Med Educ*. 2019;19(1):341. https://doi.org/10.1186/s12909-019-1788-3
123. Yıldırım M, Ozaslan A. Love of Life Scale: psychometric analysis of a Turkish adaptation and exploration of its relationship with well-being and personality. *Gazi Med J*. 2022;33(2). Available from https://medicaljournal.gazi.edu.tr/index.php/GMJ/article/view/3359. [Accessed Feb 4, 2024].
124. Dev V, Fernando AT, Consedine NS. Self-compassion as a stress moderator: a cross-sectional study of 1700 doctors, nurses, and medical students. *Mindfulness*. 2020;11(5):1170-1181. https://doi.org/10.1007/s12671-020-01325-6
125. Serinolli MI, Novaretti MCZ. A cross-sectional study of sociodemographic factors and their influence on quality of life in medical students at Sao Paulo, Brazil. *PloS One*. 2017;12(7):e0180009. https://doi.org/10.1371/journal.pone.0180009
126. Barreto FS, Carvalho GO de, Silva LCN da, et al. Electronic tools for physical activity and quality of life of medical students during COVID-19 pandemic social isolation. *Mot Rev Educ Física*. Published online Jan 1, 2021. https://doi.org/10.1590/s1980-65742021002121
127. Mohd Salleh Sahimi H, Norzan MH, Nik Jaafar NR, et al. Excessive smartphone use and its correlations with social anxiety and quality of life among medical students in a public university in Malaysia: a cross-sectional study. *Front Psychiatry*. 2022;13:956168. https://doi.org/10.3389/fpsyt.2022.956168
128. Chawla B, Chawla S, Singh H, Jain R, Arora I. Is coronavirus lockdown taking a toll on mental health of medical students? A study using WHOQOL-BREF questionnaire. *J Fam Med Prim Care*. 2020;9(10):5261-5266. https://doi.org/10.4103/jfmpc.jfmpc_715_20
129. Suhail H, Mousa AA, Waleed SM, Hussien YA. Predictors of quality of life (QoL) among Qadisiah medical student, Iraq. *Int J Pharm Res*. 2020;12(04). https://doi.org/10.31838/ijpr/2020.12.04.047
130. Ilić I, Šipetić S, Grujičić J, Mačužić IŽ, Kocić S, Ilić M. Psychometric properties of the World Health Organization's Quality of Life (WHOQOL-BREF) questionnaire in medical students. *Med Kaunas Lith*. 2019;55(12):772. https://doi.org/10.3390/medicina55120772
131. Malibary H, Zagzoog MM, Banjari MA, Bamashmous RO, Omer AR. Quality of Life (QoL) among medical students in Saudi Arabia: a study using the WHOQOL-BREF instrument. *BMC Med Educ*. 2019;19(1):344. https://doi.org/10.1186/s12909-019-1775-8
132. Obad AS, Abdulwali FK, Alaidroos HAH, et al. Relationship between shortage of basic life needs and quality of life of medical students in Yemen: a study utilizing validity and reliability of WHOQOL-BREF questionnaire. *J Fam Med Prim Care*. 2021;10(3):1466-1472. https://doi.org/10.4103/jfmpc.jfmpc_935_20
133. Shareef MA, AlAmodi AA, Al-Khateeb AA, et al. The interplay between academic performance and quality of life among preclinical students. *BMC Med Educ*. 2015;15:193. https://doi.org/10.1186/s12909-015-0476-1
134. Krägeloh CU, Henning MA, Hawken SJ, et al. Validation of the WHOQOL-BREF quality of life questionnaire for use with medical students. *Educ Health Abingdon Engl*. 2011;24(2):545.
135. Mahaur R, Jain P, Jain A. Association of mental health to emotional intelligence in medical undergraduate students: are there gender differences? *Indian J Physiol Pharmacol*. 2017;61:383-391. Available from https://www.researchgate.net/publication/320414609_Association_of_Mental_health_to_Emotional_Intelligence_in_Medical_Undergraduate_Students_Are_there_Gender_Differences [Accessed Feb 4, 2024].
136. Sahraian A, Javadpour A. Sleep disruption and its correlation to psychological distress among medical students. *Shiraz E-Med J*. 2009;11(1):20378. Available from https://brieflands.com/articles/semj-78485 [Accessed Feb 4, 2024].
137. Yusoff MSB, Fuad A, Yaacob MJ. The sensitivity, specificity and reliability of the Malay Version 12-Items General Health Questionnaire (GHQ-12) in detecting distressed medical students. 2009;11:6.
138. Rathner G, Rumpold G. Convergent validity of the eating disorder inventory and the anorexia nervosa inventory for self-rating in an Austrian nonclinical population. *Int J Eat Disord*. 1994;16(4):381-

393. https://doi.org/10.1002/1098-108x(199412)16:4%3C381::aid-eat2260160407%3E3.0.co;2-q
139. Kaur J, Martin G. Non-suicidal self-injury in medical students. *Suicidology*. 2017;8:56-65. Available from https://espace.library.uq.edu.au/view/UQ:681910 [Accessed Feb 4, 2024].
140. Sender R, Salamero M, Vallés A, Valdés M. Psychological variables for identifying susceptibility to mental disorders in medical students at the University of Barcelona. *Med Educ Online*. 2004;9(1):4350. https://doi.org/10.3402/meo.v9i.4350
141. Dyrbye LN, Szydlo DW, Downing SM, Sloan JA, Shanafelt TD. Development and preliminary psychometric properties of a well-being index for medical students. *BMC Med Educ*. 2010;10:8. https://doi.org/10.1186/1472-6920-10-8
142. Peng L, Zhang J, Li M, et al. Negative life events and mental health of Chinese medical students: the effect of resilience, personality and social support. *Psychiatry Res*. 2012;196(1):138-141. https://doi.org/10.1016/j.psychres.2011.12.006
143. van Dijk I, Lucassen PLBJ, van Weel C, Speckens AEM. A cross-sectional examination of psychological distress, positive mental health and their predictors in medical students in their clinical clerkships. *BMC Med Educ*. 2017;17(1):219. https://doi.org/10.1186/s12909-017-1035-8
144. Wu W, Ma X, Liu Y, et al. Empathy alleviates the learning burnout of medical college students through enhancing resilience. *BMC Med Educ*. 2022;22(1):481. https://doi.org/10.1186/s12909-022-03554-w
145. Suh WW, Cho SH, Yoo JY, et al. Relationship between psychological correlates and empathy in medical students: a cross-sectional study. *Psychiatry Investig*. 2019;16(10):766-772. https://doi.org/10.30773/pi.2019.08.31
146. Rezaei AM, Naeim M, Asadi R, et al. The predictive role of emotional intelligence, resilience, and personality traits in addiction potential of students at Arak University of Medical Sciences. *Addict Disord Their Treat*. 2021;20(4):472-478. https://doi.org/10.1097/ADT.0000000000000276
147. Vitaliano PP, Russo J, Carr JE, Maiuro RD, Becker J. The ways of coping checklist: revision and psychometric properties. *Multivar Behav Res*. 1985;20(1):3-26. https://doi.org/10.1207/s15327906mbr2001_1
148. Arima M, Takamiya Y, Furuta A, et al. Factors associated with the mental health status of medical students during the COVID-19 pandemic: a cross-sectional study in Japan. *BMJ Open*. 2020;10(12):e043728. https://doi.org/10.1136/bmjopen-2020-043728
149. Iqbal A, Nawaz H, Sikandar MZ, A M, Shah SIA. A cross-sectional study of happiness, self-esteem and optimism in medical students. *Ann King Edw Med Univ Lahore Pak*. 2019;25:1-4.
150. Chethana K, Nelliyanil M, Anil M. Prevalence of nomophobia and its association with loneliness, self happiness and self esteem among undergraduate medical students of a medical college in Coastal Karnataka. *Indian J Public Health Res Dev*. 2020;11(3):523-529. https://doi.org/10.37506/ijphrd.v11i3.1215
151. Satti MZ, Khan TM, Qurat-Ul-Ain QUA, et al. Association of physical activity and sleep quality with academic performance among fourth-year MBBS students of Rawalpindi Medical University. *Cureus*. 2019;11(7):e5086. https://doi.org/10.7759/cureus.5086
152. James B, Omoaregba J, Igberase O. Prevalence and correlates of poor sleep quality among medical students at a Nigerian university. *Annals Nigerian Med.* 2011. Available from https://go.gale.com/ps/i.do?id=GALE%7CA265947083&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=03313131&p=AONE&sw=w&userGroupName=anon%7E143e4a34&aty=open-web-entry. [Accessed Feb 4, 2024].
153. Anuradha R, Hemachandran S, Patil AB. Sleep quality and daytime sleepiness among medical undergraduate students in Tamil Nadu: a cross-sectional study. *J Clin Diagn Res*. Published online 2022. https://doi.org/10.7860/JCDR/2022/57499.16552
154. Arbabisarjou A, Hashemi SM, Sharif MR, et al. The relationship between sleep quality and social intimacy, and academic burn-out in students of medical sciences. *Glob J Health Sci*. 2015;8(5):231-238. https://doi.org/10.5539/gjhs.v8n5p231
155. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*. 1992;15(4):376-381. https://doi.org/10.1093/sleep/15.4.376
156. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep*. 1994;17(8):703-710. https://doi.org/10.1093/sleep/17.8.703
157. Attal BA, Al-Ammar FK, Bezdan M. Validation of the Arabic version of the Epworth Sleepiness Scale among the Yemeni medical students. *Sleep Disord*. 2020;2020:6760505. https://doi.org/10.1155/2020/6760505
158. Wongpakaran N, Wongpakaran T. A revised Thai Multi-Dimensional Scale of Perceived Social Support. *Span J Psychol*. 2012;15(3):1503-1509. https://doi.org/10.5209/rev_sjop.2012.v15.n3.39434
159. Jordan KD, Foster PS. Medical student empathy: interpersonal distinctions and correlates. *Adv Health Sci Educ Theory Pract*. 2016;21(5):1009-1022. https://doi.org/10.1007/s10459-016-9675-8
160. Lee J, Lee EH, Moon SH. Systematic review of the measurement properties of the Depression Anxiety Stress Scales-21 by applying updated COSMIN methodology. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2019;28(9):2325-2339. https://doi.org/10.1007/s11136-019-02177-x
161. Sharif-Nia H, Sánchez-Teruel D, Sivarajan Froelicher E, et al. Connor-Davidson Resilience Scale: a systematic review psychometrics properties using the COSMIN. *Ann Med Surg 2012*. 2024;86(5):2976-2991. https://doi.org/10.1097/MS9.0000000000001968
162. Shoman Y, Marca SC, Bianchi R, et al. Psychometric properties of burnout measures: a systematic review. *Epidemiol Psychiatr Sci*. 2021;30:e8. https://doi.org/10.1017/S2045796020001134
163. Brady KJS, Kazis LE, Sheldrick RC, Ni P, Trockel MT. Selecting physician well-being measures to assess health system performance and screen for distress: conceptual and methodological considerations. *Curr Probl Pediatr Adolesc Health Care*. 2019;49(12):100662. https://doi.org/10.1016/j.cppeds.2019.100662

Appendix A.

Table A1. Scale characteristics & feasibility considerations

Construct	Metric	# Studies	Year Created	Original Language	# Items and Subdomains	Completion Time	Cost	Administration	Training required?
Burnout	Maslach Burnout Inventory - Human Services Survey (MBI-HSS)	8	1981	English	22 items, 3 subdomains	10-15 minutes	$2.50 per person	Available online	None
	Maslach Burnout Inventory - Student Survey (MBI-SS)	14	2002	English	15 items, 3 subdomains	10-15 minutes	$2.50 per person	Available online	None
	Oldenburg Burnout Inventory (OLBI)	10	2003	English	16 items, 2 subdomains	10-15 minutes	Free	Available online	None
Anxiety	Beck Anxiety Inventory (BAI)	5	1988	English	21 items, 0 subdomains	10-15 minutes	Free	Available online	None
	General Anxiety Disorder-7 (GAD-7)	2	2006	English	7 items, 0 subdomains	5 minutes	Free	Available online	None
	State Trait Anxiety Inventory (STAI)	9	1983	English	20 items, 2 subdomains	10-15 minutes	Free	Available online	None
Mood & Anxiety	Depression and Anxiety Stress Scales-21 (DASS-21)	10	1995	English	21 items, 3 subdomains	10-15 minutes	Free	Available online	None
	Depression and Anxiety Stress Scales-42 (DASS-42)	1	1995	English	42 items, 3 subdomains	20-30 minutes	Free	Available online	None
	Hospital Anxiety and Depression Scale (HADS)	4	1983	English	14 items, 2 subdomains	2-5 minutes	Free	Available online	None
Mood	Beck Depression Inventory (BDI)	8	1961	English	21 items, 0 subdomains	10-15 minutes	Free	Available online	None
	Beck Depression Inventory (BDI)-II	7	1996	English	21 items, 0 subdomains	10-15 minutes	Free	Available online	None
	Center for Epidemiologic Studies Depression Scale (CES-D)	4	1977	English	20 items, 0 subdomains	10-15 minutes	Free	Available online	None
	Patient Health Questionnaire-9 (PHQ-9)	10	1999	English	9 items, 0 subdomains	10-15 minutes	Free	Available online	None
	Positive and Negative Affect Schedule (PANAS)	3	1988	English	20 items, 2 subdomains	10-15 minutes	Free	Available online	None
Stress	Bandura-Rosenthal Metrics for Assessing Stress (BAROMAS)	1	1987	English	85 items, 9 domains	~30 minutes	Free	Available online	None
	Medical Student Stress Questionnaire-20 (MSSQ-20)	1	2011	English	20 items, 6 subdomains	10-15 minutes	Free	Available online	None
	Medical Student Stress Questionnaire-40 (MSSQ-40)	6	2010	English	40 items, 6 subdomains	20-30 minutes	Free	Available online	None
	Medical Student Stress Questionnaire-40-Revised (MSSQ-40-R)	2	2020	Italian	40 items, 5 subdomains	20-30 minutes	Free	Available online	None
	Perceived Medical School Stress (PMSS)	3	1989	English	11 items, 4 subdomains	10-15 minutes	Free	Available online	None
	Perceived Stress Scale-4 (PSS-4)	1	1983	English	4 items, 0 subdomains	5-10 minutes	Free	Available online	None
	Perceived Stress Scale-10 (PSS-10)	14	1983	English	10 items, 0 subdomains	10-15 minutes	Free	Available online	None
Quality of Life	Satisfaction with Life Scale (SWLS)	10	1985	English	5 items, 0 subdomains	5-10 minutes	Free	Available online	None
	World Health Organization Quality of Life Brief Version (WHOQOL-BREF)	14	1995	English	26 items, 4 subdomains	15-20 minutes	Free	Available online (permission from WHO)	None

General Well-being	Medical Student Well-Being Index (MSWBI)	2	2010	English	7 items, 5 subdomains	10-15 minutes	Contact to obtain pricing	Available online (Well-being Index website)	Available online (Well-being Index website)
	General Health Questionnaire - 12 (GHQ-12)	6	1972	English	12 items. 0 subdomains	5 minutes	Free	Available online	None
	General Health Questionnaire - 28 (GHQ-28)	3	1979	English	28 items, 4 subdomains	5-10 minutes	Free	Available online	None
	Symptom Checklist-90-Revised (SCL-90-R)	1	1973	English	90 items, 10 subdomains	10-15 minutes	Free	Available online	Available online (Pearson Assessments)
Mindfulness	Five Facets of Mindfulness Questionnaire (FFMQ)	3	2008	English	39 items, 5 subdomains	15-20 minutes	Free	Available online	None
	Mindful Attention Awareness Scale (MAAS)	2	2005	English	15 items, 0 subdomains	10-15 minutes	Free	Available online	None
Coping & Resilience	Brief COPE	3	1997	English	28 items, 14 subdomains	10-15 minutes	Free	Available online	None
	Connor-Davidson Resilience Scale (CD-RISC)	7	1970	English	25 items, 0 subdomains	10-15 minutes	Contact to obtain pricing	Available online (CD-RISC website)	None
	Ways of Coping Checklist (WCCL)	2	1985	English	66 items, 8 subscales	15-20 minutes	Free	Available online	None
Self-Esteem	General Self-Efficacy Scale (GSES)	1	1995	English	10 items, 0 subdomains	4 minutes	Free	Available online	None
	Rosenberg Self-Esteem Scale (RSES)	7	1965	English	10 items, 0 subdomains	10-15 minutes	Free	Available online	None
Sleep	Epworth Sleep Scale (ESS)	6	1991	English	8 items, 0 subdomains	5-10 minutes	Free	Available online	None
	Pittsburgh Sleep Quality Index (PSQI)	11	1989	English	19 items, 7 subdomains	10-15 minutes	Free	Available online	None
Social Support & Loneliness	Multidimensional Scale of Perceived Social Support (MSPSS)	7	1988	English	12 items, 3 subdomains	10-15 minutes	Free	Available online	None
	UCLA Loneliness Scale	4	1978	English	20 items, 0 subdomains	10-15 minutes	Free	Available online	None
Work	Utrecht Work Engagement Scale (UWES)	2	2002	Dutch	17 items, 3 subdomains	5-10 minutes	Free	Available online	None

Table A2. Summary of psychometric evidence for included scales

Construct	Metric	Content Validity		Structural Validity		Internal Consistency		Cross-cultural validity		Reliability		Criterion Validity		Construct Validity		Responsiveness	
		Q	QE	Q	QE	Q	QE	Q	QE	Q	QE	Q	QE	Q	QE	Q	QE
Burnout	Maslach Burnout Inventory - Human Services Survey (MBI-HSS)					+ (5)	M							+	H	+	H
	Maslach Burnout Inventory - Student Survey (MBI-SS)			I	M	+ (9)	H	+	H	-	M			+	H		
	Oldenburg Burnout Inventory (OLBI)	?	L	I	L	+ (6)	H							+	H		
Anxiety	Beck Anxiety Inventory (BAI)					+ (2)	H							+	H		
	General Anxiety Disorder-7 (GAD-7)					+	H							+	H		
	State Trait Anxiety Inventory (STAI)					+ (2)	H			-	VL			+	H		
Mood & Anxiety	Depression and Anxiety Stress Scales-21 (DASS-21)			-	L	+ (3)	H	-	H					+	H	+	H
	Depression and Anxiety Stress Scales-42 (DASS-42)					+	H							+	H		
	Hospital Anxiety and Depression Scale (HADS)					+	H	+	L					+	H		
Mood	Beck Depression Inventory (BDI)			?	L	+ (4)	H							+	H		
	Beck Depression Inventory (BDI)-II			?	VL	+ (4)	H			+	L			+	H		
	Center for Epidemiologic Studies Depression Scale (CES-D)			?	VL	+ (3)	H			-	L	+	M	+	H		
	Patient Health Questionnaire-9 (PHQ-9)					+ (7)	H			-	M	+	H	+	H		
	Positive and Negative Affect Schedule (PANAS)	-	VL	?	M	+ (2)	H			-	M			+	H		
Stress	Bandura-Rosenthal Metrics for Assessing Stress (BAROMAS)									-	VL						
	Medical Student Stress Questionnaire-20 (MSSQ-20)					-	H										
	Medical Student Stress Questionnaire-40 (MSSQ-40)	?	L	-	H	- (3)	L			-	VL			+	H		
	Medical Student Stress Questionnaire-40-Revised (MSSQ-40-R)	?	L	I	L	- (2)	H			+	M			+	H		
	Perceived Medical School Stress (PMSS)					+ (3)	H	+	M								
	Perceived Stress Scale-4 (PSS-4)													+	H		
	Perceived Stress Scale-10 (PSS-10)	?	M			+ (8)	H	-	VL	+	M			+	H	+	H
Quality of Life	Satisfaction with Life Scale (SWLS)					+ (7)	H							+	H		
	World Health Organization Quality of Life Brief Version (WHOQOL-BREF)			+	L	+ (10)	M			-	M			+	H		
General well-being	Medical Student Well-Being Index (MSWBI)	+	L	+	L									+	H		
	General Health Questionnaire - 12 (GHQ-12)	?	L			+ (3)	H							+	H		
	General Health Questionnaire - 28 (GHQ-28)					+ (2)	M							+	H		
	Symptom Checklist-90-Revised (SCL-90-R)													+	H		
Mindfulness	Five Facets of Mindfulness Questionnaire (FFMQ)					+	H							+	H		
	Mindful Attention Awareness Scale (MAAS)			+	L	+ (2)	H			+	M			+	H	+	H
	Brief COPE													+	H		

Coping & Resilience	Connor-Davidson Resilience Scale (CD-RISC)			+	L	+ (3)	H							+	H		
	Ways of Coping Checklist (WCCL)			?	M	- (2)	M										
Self-Esteem	General Self-Efficacy Scale (GSES)					+	H										
	Rosenberg Self-Esteem Scale (RSES)					+ (4)	H							+	H		
Sleep	Epworth Sleep Scale (ESS)			?	M	+ (4)	H			+	VL			+	H		
	Pittsburgh Sleep Quality Index (PSQI)					- (3)	M							+	H	+	H
Social Support & Loneliness	Multidimensional Scale of Perceived Social Support (MSPSS)			+	H	+ (6)	H	+	H	+ (2)	H			+	H		
	UCLA Loneliness Scale					+ (4)	H							+	H		
Work	Utrecht Work Engagement Scale (UWES)					+ (2)	H							+	H	+	H

Quality of the property (Q) rated as sufficient (+), insufficient (-), inconsistent (I), or indeterminate (?) based on comparison with criteria for good measurement properties. Quality of the evidence (QE) using GRADE rated as high (H), moderate (M), low (L), or very low (VL) based on risk of bias, inconsistency, and imprecision. Where meta-analysis was used for internal consistency and inter-rater reliability, brackets indicate the number of studies contributing to the property.