BMJ Open

Measurement properties of the Mental Health Literacy Scale (MHLS) validation studies: a systematic review protocol

Rouwida ElKhalil ![ORCID],[1] Mohamad AlMekkawi ![ORCID],[2] Matt O'Connor ![ORCID],[3] Moustafa Sherif ![ORCID],[1] Emad Masuadi ![ORCID],[1] Luai A Ahmed ![ORCID],[1] Rami H Al-Rifai ![ORCID],[1] Messaouda Belfakir ![ORCID],[1] Rasha Bayoumi ![ORCID],[4] Iffat Elbarazi ![ORCID][1]

[1]Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, PO.Box:15551, Al Ain, Abu Dhabi, UAE
[2]Nursing Department, Fatima College of Health Sciences, AlAin, UAE
[3]ConnectEd Counselling and Consultancy, Brisbane, Queensland, Australia
[4]School of Psychology, University of Birmingham Dubai, Dubai, UAE

Correspondence to
Dr Iffat Elbarazi;
ielbarazi@uaeu.ac.ae

ABSTRACT

Introduction Mental Health Literacy (MHL) is important for improving mental health and reducing inequities in treatment. The Mental Health Literacy Scale (MHLS) is a valid and reliable assessment tool for MHL. This systematic review will examine and compare the measurement properties of the MHLS in different languages, enabling academics, clinicians and policymakers to make informed judgements regarding its use in assessments.
Methods and analysis The review will adhere to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology for systematic reviews of patient-reported outcome measures and the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis and will be presented following the Preferred Reporting Items for Systematic reviews and Meta-Analysis 2020 checklist. The review will be conducted in four stages, including an initial search confined to PubMed, a search of electronic scientific databases PsycINFO, CINAHL, Scopus, MEDLINE, Embase (Elsevier), PubMed (NLM) and ERIC, an examination of the reference lists of all papers to locate relevant publications and finally contacting the MHLS original author to identify validation studies that the searches will not retrieve. These phases will assist us in locating studies that evaluate the measurement properties of MHLS across various populations, demographics and contexts. The search will focus on articles published in English between May 2015 and December 2023. The methodological quality of the studies will be evaluated using the COSMIN Risk of Bias checklist, and a comprehensive qualitative and quantitative data synthesis will be performed.
Ethics and dissemination Ethics approval is not required. The publication will be in peer-reviewed journals and presented at national and international conferences.
PROSPERO registration number CRD42023430924.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This review evaluates Mental Health Literacy Scale (MHLS) measurement properties across languages, stressing diverse MHL assessments.
⇒ It adheres to the Joanna Briggs Institute Manual and COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology and follows Preferred Reporting Items for Systematic reviews and Meta-Analysis 2020 guidelines.
⇒ Limited by a temporal gap post 2018 due to MHLS development in 2015.
⇒ Limited by exclusion of non-English papers.
⇒ Challenges in meta-analyses are anticipated, given study heterogeneity.

INTRODUCTION

Mental health is an integral part of overall health and well-being. Global rates of mental disorders are significant, with depression alone affecting over 280 million people.[1] Personal health literacy (HL) is defined as 'the degree to which individuals have the ability to find, understand, and use information and services to inform health-related decisions and actions for themselves and others'.[2] Mental HL (MHL), a derivative from and component of HL,[3] is defined as the 'knowledge and beliefs about mental disorders which aid their recognition, management or prevention'.[4] Jorm elaborated on the original definition of MHL to encompass the following: understanding ways to prevent mental illness, recognising early signs and symptoms of mental illness, being aware of various help-seeking choices and treatments, awareness regarding methods of self-help and mental health first aid skills to help and support people who have mental illness.[5] Accordingly, MHL consists of the following attributes: the ability to identify specific disorders, knowledge of how to obtain mental health information, knowledge of risk factors and causes, self-care methods and available professional assistance, and attitudes that encourage recognition and proper seeking of support.[4] Research regarding MHL has covered a wide range of topics, including stigma, help-seeking behaviours and the mental health difficulties experienced by different vulnerable groups.[6] Therefore, MHL plays a crucial role in enhancing

individuals' mental well-being by helping them identify their symptoms, find available resources, and obtain the necessary support.[7 8]

Using validated instruments to assess MHL is vital for developing successful strategies to promote mental health. These instruments can also assist academics and policymakers in identifying knowledge gaps in MHL and designing culturally appropriate solutions tailored to various individual and community needs.[9] Developing an MHL instrument requires having a clear operational definition of the construct.[3 10] Historically, this construct has been evaluated using two approaches, namely the Vignette Approach and the Scale-based Measurements.[11] The Vignette Approach is described as 'stories about individuals and situations which refer to important points in the study of perceptions, beliefs, and attitudes'.[12] This approach has limitations, such as the inability to compare items within the scale, understand the differences between MHL components, and track improvement over time. Scale-based measurements, also called patient-reported outcome measures (PROMs), are 'measurement instruments that patients complete to provide information on aspects of their health status that are relevant to their quality of life, including symptoms, functionality, and physical, mental and social health'.[13]

Following a systematic assessment of MHL instruments in 2014, O'Connor and Casey designed the Mental Health Litercy Scale (MHLS) to address these limitations and to produce a valid and reliable assessment tool for MHL.[11] The rigour with which the MHLS was developed and its subsequent psychometric properties have made it the most reliable and validated instrument for assessing MHL.[14] The scale showed adequate content and structural validity, good test–retest reliability and internal consistency ($\alpha=0.873$).[11] In addition, the MHLS is the only available instrument to measure all aspects of MHL.[15]

The MHLS is a unidimensional measurement scale with 35 items and 6 attributes based on Jorm's six MHL attributes.[4] The scale items were generated using a combination of adaption of existing MHL items, descriptors from the Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR21, national and international data, and the clinical experience of the authors and their clinical panel who advised the item generation. The scale score ranges from 35 to 160, with a higher score implying a higher level of MHL. The scale has the following sections: Recognition of Disorders (8 items measured on a 4-point Likert scale), Knowledge of Risk Factors and Causes (2 items measured on a 4-point Likert scale), Self-Treatment Knowledge (2 items measured on a 4-point Likert scale), Knowledge of Professional Help Available (3 items measured on a 4-point Likert scale), Knowledge of How to Seek Mental Health Information (4 items measured on a 5-point Likert-scale) and Attitudes that Promote Recognition and Appropriate Help-Seeking (16 items measured on a 5-point Likert scale), with items 10, 12, 15 and 20–28 as reverse-scored items.[11]

The scale has been used in various cultural and language contexts, making it a valuable instrument for cross-cultural research studies.[16] Modification and cultural adaptation of research instruments have numerous advantages over creating new ones. It permits comparisons of research outcomes from different cultures, facilitating international scientific collaboration and reducing costs and time.[17 18] According to Arafat *et al*,[17] cross-cultural validation involves translation, adaption, measurement of reliability (repeatability and internal consistency), evaluation of validity (content validity, face validity, construct validity and criterion validity) and responsiveness.

Nevertheless, this study aims to critically examine, summarise and compare the measurement properties of all language versions of the MHLS by systematically examining the methodological quality and findings of the available publications. While the MHLS has been culturally adapted and translated into numerous languages, comprehensive reviews of the adapted versions are lacking, leaving minimal evidence regarding their measurement properties.[16 19] This systematic review is important to researchers aiming to measure MHL in diverse settings as it evaluates and compares the measurement properties of all language versions of the MHLS. The objective is to provide new insights into the measurement properties of the MHLS across different language versions. The findings of this review will be valuable for academics, clinicians and policymakers to enhance their understanding of the MHLS's reliability and validity in various cultural and language contexts. Furthermore, this review will contribute to the theoretical framework surrounding MHLS validation, guide future research initiatives and facilitate collaborations with researchers and publications in the field of MHLS validation.

The objectives of this study are:

1. To summarise the used adaptation/validation processes employed in MHLS validation studies,
2. To assess the methodological quality of the measurement properties of the MHLS across several language versions
3. To compare and synthesise the findings of studies that examined the measurement properties of the MHLS in different language versions, such as its reliability, validity and responsiveness, by qualitatively summarising or quantitatively pooling the results.

METHODS

This systematic review will be conducted between September 2023 and December 2023. This protocol adheres to items outlined under the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) protocol.[20] The proposed systematic review will adhere to the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis (Chapter 12: Systematic Reviews of Measurement Properties)[21] and the COSMIN methodology for systematic reviews of PROMs.[22] The results will be presented according to PRISMA 2020.[23] The systematic
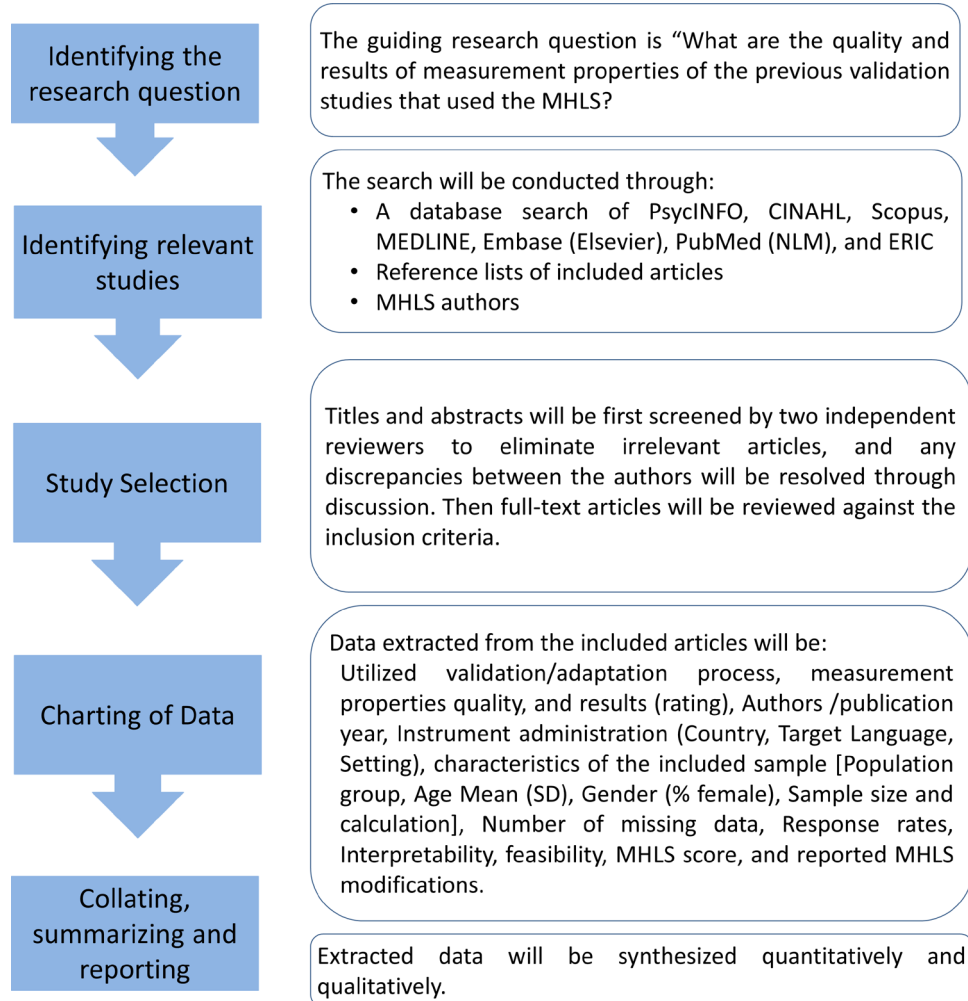
Figure 1 Systematic review methodology summary. MHLS, Mental Health Literacy Scale. SD, Standard Deviation.

review methodology is summarised in figure 1. The study is registered at PROSPERO.

Patient and public involvement
None

Search strategy
The review will begin with forming a research team of individuals with content and methodological competencies.[24] The team will advise on the overarching research question and the entire study protocol, including identifying the search terms and databases. The review will be conducted in four stages per the JBI Standards.[21]

In the first stage, an initial search of the PubMed database will be done using a sensitive search filter[25] to find studies on the measurement properties of MHLS (see online supplemental file 1A). The initial search will follow '*Filter 1: Sensitive search filter for measurement properties*', which guarantees 97.4% sensitive and 4.4% precise results (table 1). In the second stage, we will search the electronic scientific databases PsycINFO, CINAHL, Scopus, MEDLINE, Embase (Elsevier), PubMed (NLM) and ERIC using the final Boolean expression created in the previous phase (see online supplemental file 1B). In

the third stage, the reference lists of all papers included in the second stage will be examined, and more relevant publications will be located and incorporated into this study. In the final stage, the MHLS creators will be contacted to identify validation studies not retrieved in the previous searches.

We have already identified the search filters (see online supplemental file 1A). These were combined with phrases searched for the concept of interest (Mental Health Literacy) 'AND' the measuring instrument of

Table 1 Systematic review search strategy

	Search strategy
#1	Construct Search (Mental Health Literacy)
#2	Instrument Search (Mental Health Literacy Scale)
#3	#1 AND #2 AND Sensitive filter for measurement properties (see online supplemental file 1AI)
#4	#3 NOT exclusion search filter (see online supplemental file 1AII)

Adopted from a study by Terwee et al.[25]
MHL, Mental Health Literacy.

Table 2 Systematic review inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
1. Participants: Any population or age group. Context: All settings in any country. 2. Instrument and Construct: O'Conner and Casey (2015) MHLS to assess the MHL construct. 3. Outcomes: Reliability, validity and responsiveness. 4. Types of sources: Validation studies. 5. Language: English. 6. Date: 2015–2022.	1. Non-English studies. 2. Grey literature (non-peer-reviewed publications or documents of any type). 3. Other MHL measures. 4. Studies that only use the MHLS as an outcome measure.

MHL, Mental Health Literacy; MHLS, Mental Health Literacy Scale.

interest (MHLS). However, no population search was added because there were no population type, age or setting restrictions. These searches were paired with the measurement properties search filter to locate all studies on the MHLS measurement properties that assess MHL in all populations. For a more thorough search, we used the sensitive filter. The exclusion filter was used to eliminate records from the search, such as case studies and animal studies.

Study screening and selection

The screening and selection approach will be summarised using the PRISMA flowchart.[23] Our review question and inclusion criteria are framed using the PICO (Population, Instruments, Construct, Outcomes) method.[21] Eligibility criteria, as shown in table 2, are as follows: (1) *Participants:* The review will consider studies that validate the MHLS in any population (eg, community representation, students, perinatal patients or health professionals) without restricting participants' age group; *Context:* The review will consider all primary research that validated the MHLS in all global settings (ie, as acute care, primary healthcare,

or the community); (2) *Instrument and Construct:* The review will focus solely on O'Connor and Casey MHLS;[11] (3) *Outcomes:* Measurement properties (reliability, validity and responsiveness) of adapted MHLS will be assessed and reported based on the individual study as in table 3[21]; (4) *Types of Sources*: The review will consider primarily published designs empirically validating the MHLS, including translation and cultural adaptation, reliability and validity testing using various statistical analyses.[17] The aim of the included studies should be the evaluation of one or more measurement properties.[22] This review will exclude studies that only use the MHLS as an outcome measure; (5) *Language:* Only English papers published will be eligible for review. Non-English publications will be excluded during the screening phase; (6) *Date:* Since the MHLS was created in 2015, only studies published between 2015 and 2022 will be considered.

The retrieved literature will be imported into Covidence. The publications will be screened in two steps: The title and abstract will be reviewed, and then the full text will be examined. Two reviewers (RE and MA)

Table 3 Systematic review outcomes: measurement properties

Main outcomes	Effect measures
Reliability	Cronbach's alpha coefficients, *or* intra-class correlation coefficients (ICC), *or* weighted or un-weighted Kappa, *or* Standard Error of measurement (SEm), *or* limits of agreement (LoA), *or* smallest detectable change (SDC), *or* concordance correlation coefficients goodness of fit statistics.
Validity	
i. Content validity	Purpose, target population, the comprehensiveness of the instrument, floor or ceiling effects (if available), *and* relevant items for the construct(Content Validity Index(CVI), *or* Index of Item Objective Congruence (IOC)).
ii. Structural validity	Factor analysis and Comparative Fit Index (CFI), *and* Tucker-Lewis Index (TLI), *and* Root Mean Square Error of Approximation (RMSEA), *and* Standardised Root Mean Residuals (SRMR).
iii. Hypothesis testing	Absolute or relative differences or correlations between MHLS with other instruments, *or* Absolute or relative differences or correlations between MHLS with two groups of participants.
iv. Cross-cultural validity	The Differential Item Functioning (DIF).
v. Criterion validity	Correlations, *or* Areas under Receiver Operating Curves (ROC), *or* Sensitivity and Specificity.
Responsiveness	Absolute or relative correlations, *or* Differences of the change scores, *or* The Areas under Receiver Operating Curves (ROC), *or* Sensitivity and specificity.

Adopted from JBI Manual by Aromataris *et al.*[21]

Table 4 Quality criteria for measurement properties

Property	Rating*	Quality criteria
Reliability		
Internal consistency	+	Cronbach alphas ≥0.70.
	?	Cronbach alpha not determined.
	–	Cronbach alphas <0.70.
Reliability	+	ICC/weighted kappa ≥0.70 OR Pearson r ≥0.80.
	?	Neither ICC/weighted kappa nor Pearson r determined.
	–	ICC/weighted kappa 0.70 OR Pearson r 0.80.
Measurement error	+	MIC>SDC OR MIC outside the LOA.
	?	MIC not defined.
	–	MIC≤SDC OR MIC equals or inside LOA.
Validity		
Structural validity	+	**CTT:** CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.08. EFA: Factors should explain at least 60% of the variance. **IRT/Rasch:** No violation of unidimensionality 3: CFI or TLI or comparable measure >0.95 OR RMSEA <0.20 OR Q3's <0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: IRT: $\chi 2$ >0.01 Rasch: infit and outfit mean squares ≥0.5 and ≤1.5 OR Z- standardised values >−2 and <2.
	?	**CTT:** Not all information for '+' reported OR Explained variance not mentioned. **RT/Rasch:** Model fit not reported.
	–	Criteria for '+' not met OR Factors explain <60% of the variance.
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis.
	?	No hypothesis was defined (by the review team).
	–	The result is not in accordance with the hypothesis.
Cross-cultural validity\ measurement invariance	+	No important differences were found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R2 <0.02).
	?	No multiple group factor analysis OR DIF analysis was performed.
	–	Important differences between group factors OR DIF were found.
Criterion validity	+	Correlation with gold standard ≥0.70 OR AUC ≥0.70 X.
	?	Not all information for '+' reported.
	–	Correlation with gold standard <0.70 OR AUC <0.70.
Responsiveness	+	The result is in accordance with the hypothesis 7 OR AUC ≥0.70.
	?	No hypothesis was defined (by the review team).
	–	The result is not in accordance with the hypothesis 7 OR AUC <0.70.
Content validity	+	The Relevance Rating is +, the Comprehensiveness Rating is + and the COMPREHENSIBILITY RATING is +.
	–	The Relevance Rating is –, the Comprehensiveness Rating is – and the Comprehensibility Rating is –.
	±	At least one of the ratings is +, and at least one of the ratings is –.
	?	Two or more of the ratings are rated?

Adapted from studies by Hair *et al,*[32] Prinsen *et al,*[22] and Terwee *et al.*[29]

*+=positive rating, ?=indeterminate rating, −= negative rating, ±= mixed ratings (content validity only).

AUC, Area Under the Curve; CFA, Confirmatory Factor Analysis; CFI, Comparative Fit Index ; CTT, Classical Test Theory ; DIF, Differential Item Functioning; EFA, Exploratory Factor Analysis; ICC, Intraclass Correlation Coefficient; IRT, Item Response Theory; LOA, Limits of Agreement; MIC, Minimal Important Change; RMSEA, Root Mean Square Error of Approximation ; SDC, Smallest Detectable Change; TLI, Tucker–Lewis Index .

will independently examine retrieved abstracts using this review's previously specified eligibility criteria. The author of MHLS will be contacted to identify additional studies, and citations will be searched for additional articles. Covidence will be used to identify and delete the duplicates. The two reviewers will meet at the beginning, midpoint and end of the abstract review process to discuss concerns and uncertainties relating to study selection

Table 5	Definitions of quality levels
Quality level	**Definition**
High	We are very confident that the true measurement property lies close to that of the estimate* of the measurement property.
Moderate	We are moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different.
Low	Our confidence in the measurement property estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property.
Very low	We have very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property.

Adopted from a study by Prinsen *et al.*[22]
*Estimate of the measurement property refers to the pooled or summarised result of the measurement property of a patient-reported outcome measure.

and, if necessary, alter the search approach. Another two researchers (RE and MB) will independently review the full manuscripts. A third reviewer (IE) will make the final judgement when there is disagreement over research inclusion. With IE and MA having been experienced professionals and scholars in the field of public health and RE and MB being doctoral candidates in public health, this group is an optimal team to select and review articles for this study. EM will provide methodological guidance to the research team. The systematic review will document and report the reasons for excluding full-text papers that do not match the inclusion criteria. Finally, reviewed articles will be retained for synthesis.

Data charting
Using the Microsoft Excel 365 spreadsheet template that the reviewers adapted from the COSMIN website,[26] two independent reviewers will perform the data extraction and the methodological quality assessment of full-text articles that meet the inclusion criteria. Before beginning the review, we will conduct calibration exercises, such as piloting the forms on two studies, to ensure consistency among reviewers.[26] The data charting instruments (see online supplemental file 1C) were adapted from the COSMIN methodology for systematic reviews of the user manual (PROMs).[22] Disagreements between the reviewers will be handled through discussion or with the assistance of a third reviewer. We will contact the authors of the study to resolve any uncertainties. The three focus areas, namely, the validation/adaptation process, risk of bias assessment and measurement properties evaluation,

will guide our data 'charting'. We will chart data by publication year, instrument administration (country, target language, setting), included sample characteristics (population group, age mean (SD), gender (% female), sample size and calculation), number of missing data, response rates, interpretability (distribution (skewness and/or kurtosis), percentage of missing items, percentage of missing total scores, floor and ceiling effects), feasibility (completion time, patient's comprehensibility and type and ease of administration), MHLS score, and reported MHLS item modifications.

Assessment of risk of bias
We will determine the quality of the measurement properties by using the COSMIN Risk of Bias (RoB) checklist, which will be filled out to evaluate the methodological quality of each study or the risk of bias in the study's findings. The following nine boxes from the checklist will be used: PROM development, Content validity, Structural validity, Internal consistency, Cross-cultural validity/ Measurement invariance, Reliability, Measurement error, Criterion validity, Hypothesis testing for construct validity and Responsiveness. Only the boxes for the measurement properties reviewed in the article will be evaluated using the RoB, which should be used as a modular tool.[27] Quality rating options for Items under each box are 'very good', 'adequate', 'doubtful', 'inadequate', or 'Not Applicable'. To establish the overall quality of a study, the lowest rating of any standard in the box will be used (ie, 'the worst score counts' principle). For example, if one item in a box is scored as 'inadequate' for a reliability study, the total methodological quality of that reliability research is graded as 'inadequate'. The translation process methodological quality will be determined by using the COSMIN Study Design checklist that provides standards for translating an existing PROM in the box Translation process.[28]

Evaluation of measurement properties
The results of measurement properties will be rated based on the criteria presented in table 4. Ratings will vary from positive (+), negative (−) and indeterminate ratings (?) according to individual study measurement property results.[22] As mentioned, the content validity rating criteria results were based on the COSMIN methodology guidelines for assessing the PROMs User Manual 22 content validity.[29] Specific MHLS hypotheses for 'Hypothesis Testing for Construct Validity' and 'Responsiveness' were developed (online supplemental file 1D).

Data synthesis and levels of evidence
The results will either be quantitatively or qualitatively combined. We will present these pooled or summarised results per measurement property (see online supplemental file 1C), together with a grade for the quality of the evidence (high, moderate, low or very low) and a rating of the pooled or summarised results (+/−/?).

Quantitative pooling of the results

In case of availability of more than two investigations per measurement property and language version, meta-analyses will be conducted, and the findings will be statistically pooled. Calculating weighted averages (depending on the number of participants participating in each research) and 95% CIs will yield pooled estimates of measurement properties. For assessing test–retest reliability, one can calculate weighted mean intraclass correlation coefficients (ICCs) and 95% CIs using a standard generic inverse variance random effects model.[30] ICC values can be combined based on estimates obtained from a Fisher transformation, $z = 0.5 \times \ln((1+ICC)/(1-ICC))$, which has an approximate variance of $(Var(z) = 1/(N-3))$, where N is the sample size.[31] For evaluating construct validity, we will aggregate all correlations between a (PROM) and other PROMs that measure a similar construct. Meanwhile, Cronbach's alpha will be reported as weighted means. To conduct meta-analyses, we will be consulting a statistician.

Qualitative summary of the result

If it is impossible to pool the results statistically, the results of each measurement property will be summed up qualitatively. For example, we will provide the range (lowest and highest) of Cronbach's alpha values found for internal consistency, the percentage of confirmed hypotheses for construct validity or the range of each model fit parameter on a consistently found factor structure in structural validity studies.

Applying measurement properties criteria to the pooled or summarised results

The pooled or summarised result per measurement property per language version of MHLS will again be rated using the same quality standards for good measurement properties (table 4). The overall assessment of the combined or summed outcome may be positive (+), negative (−) or indeterminate rating (?). The ratings will be provided in the summary of findings tables (see online supplemental file 1C).

Using the GRADE approach, which is a systematic approach to rating the certainty of evidence in systematic reviews, the following four factors will be considered when evaluating measurement properties to determine the quality of the evidence in this systematic review (table 5): (1) risk of bias (ie, quality of the studies' methodology), (2) inconsistency (ie, unexplained, inconsistent results across studies), (3) imprecision (ie, the total sample size of the available studies) and (4) indirectness (ie, evidence from different populations than the population of interest in the review).[22]

Data presentation

The data gathered from the included papers will be presented in a tabular format, with the table reporting essential findings relevant to the review topic. The tabulated data will accompany a narrative summary describing how the results relate to the review objective and question.

DISCUSSION

MHL is essential for enhancing mental health and decreasing treatment disparities. It helps healthcare professionals comprehend the educational requirements for mental health among patients and communities. Additionally, it assists individuals in understanding their symptoms, locating relevant resources and receiving appropriate healthcare assistance.[8] Improving and maintaining healthcare provision is a challenge for practitioners and policymakers. Also, patients possess distinct perspectives on healthcare quality; however, their potential for measuring it remains untapped.[13] This systematic review provides a unique insight into the measurement properties of the MHLS in a cross-cultural context. The review uses a rigorous approach to summarise the evidence on MHLS reliability and validity and to assess bias and heterogeneity in the results. It will provide academics, clinicians and policymakers with needed evidence to adopt the MHLS in their research or practice based on its reliability and validity levels and will guide them in selecting the most appropriate version for their specific context. In addition, it will assist in assessing the consistency of results across different populations, settings, and study designs.

Furthermore, the review will provide a robust model and a transparent review of measurement properties using COSMIN guidelines.[21] As such, a notable strength of this review is that it analyses the measurement properties of all language versions of the MHLS, emphasising the importance of researchers measuring MHL in various settings. Additionally, the review will adhere to the JBI Manual for Evidence Synthesis (Chapter 12: Systematic reviews of measurement properties)[21] and the COSMIN methodology for systematic reviews of PROMs user manual[22] and will be reported according to the PRISMA guideline.[23 23] However, this systematic review will be limited by the temporal discrepancy between the MHLS development in 2015 and the available resources for measuring properties' quality evaluation, which existed after 2018. In addition, excluding non-English papers due to logistical constraints could be a limitation. We anticipate that the heterogeneity of the studies will impact the ability to do meta-analyses.

ORCID iDs
Rouwida ElKhalil http://orcid.org/0000-0002-2031-837X
Mohamad AlMekkawi http://orcid.org/0000-0002-2315-7479
Matt O'Connor http://orcid.org/0000-0002-6188-8870
Moustafa Sherif http://orcid.org/0000-0002-8781-5098
Emad Masuadi http://orcid.org/0000-0002-7360-7693
Luai A Ahmed http://orcid.org/0000-0001-5292-8212
Rami H Al-Rifai http://orcid.org/0000-0001-6102-0353
Messaouda Belfakir http://orcid.org/0009-0006-0975-4839
Rasha Bayoumi http://orcid.org/0000-0001-6115-1820
Iffat Elbarazi http://orcid.org/0000-0001-7151-2175

REFERENCES

1 World Health Orgainsation. WHO menu of cost-effective interventions for mental health. 2021. Available: https://www.who.int/publications-detail-redirect/9789240031081 [Accessed 31 Jan 2024].
2 Healthy People 2030. History of health literacy definitions. Available: https://health.gov/healthypeople/priority-areas/health-literacy-healthy-people-2030/history-health-literacy-definitions [Accessed 03 Feb 2024].
3 Kutcher S, Wei Y, Coniglio C. Mental health literacy: past, present, and future. *Can J Psychiatry* 2016;61:154–8.
4 Jorm AF, Korten AE, Jacomb PA, *et al*. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Med J Aust* 1997;166:182–6.
5 Jorm AF. Why we need the concept of 'mental health literacy'. *Health Commun* 2015;30:1166–8.
6 Sweileh WM. Global research activity on mental health literacy. *Middle East Curr Psychiatry* 2021;28.
7 Bennett H, Allitt B, Hanna F. A perspective on mental health literacy and mental health issues among Australian youth: cultural, social, and environmental evidence! *Front Public Health* 2023;11:1065784.
8 Jorm AF. Mental health literacy: empowering the community to take action for better mental health. *Am Psychol* 2012;67:231–43.
9 World Health Organisation. Promoting mental health: concepts, emerging evidence, practice: summary report / a report from the World Health Organization, Department of mental health and substance abuse in collaboration with the Victorian health promotion foundation (Vichealth) and the University of Melbourne, Report no: ISBN 92 4 159159 5-(NLM classification: WM 31.5). Geneva World Health Organization; 2004. 1. Available: https://apps.who.int/iris/bitstream/handle/10665/42940/9241591595.pdf
10 Spiker DA, Hammer JH. Mental health literacy as theory: current challenges and future directions. *J Ment Health* 2019;28:238–42.
11 O'Connor M, Casey L. The mental health literacy scale (MHLS): a new scale-based measure of mental health literacy. *Psychiatry Res* 2015;229:511–6.
12 Hughes R. Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociol Health IIln* 1998;20:381–400.
13 Chow E, Faye L, Sawatzky R, *et al*. Proms background document. Ottawa, Ontario The Canadian Institute for Health Information (CIHI); 2015. 1–38. Available: https://www.cihi.ca/sites/default/files/proms_background_may21_en-web_0.pdf
14 Balula Chaves C, Sequeira C, Carvalho Duarte J, *et al*. Mental health literacy: a systematic review of the measurement instruments. *Revista INFAD de Psicología* 2022;3:181–94.
15 Montagni I, González Caballero JL. Validation of the mental health literacy scale in French university students. *Behav Sci (Basel)* 2022;12:259.
16 Krohne N, Gomboc V, Lavrič M, *et al*. Slovenian validation of the mental health literacy scale (S-MHLS) on the general population: a four-factor model. *Inquiry* 2022;59:469580211047193.
17 Arafat SMY, Chowdhury HR, Qusar MS, *et al*. Cross cultural adaptation & psychometric validation of research instruments: a methodological review. *J Behav Health* 2016;5:129.
18 Huang WY, Wong SH. Cross-cultural validation. In: Michalos AC, ed. *Encyclopedia of quality of life and well-being research*. Dordrecht: Springer Netherlands, 2014: 1369–71. Available: http://link.springer.com/10.1007/978-94-007-0753-5_630
19 Speksnijder CM, Koppenaal T, Knottnerus JA, *et al*. Measurement properties of the Quebec back pain disability scale in patients with nonspecific low back pain. *Phys Ther* 2016;96:1816–31.
20 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
21 Aromataris E, Munn Z. Chapter 12: systematic reviews of measurement properties. In: *JBI manual for evidence synthesis*. JBI, 2020. Available: https://jbi-global-wiki.refined.site/space/MANUAL/4686202/Chapter+12%3A+Systematic+reviews+of+measurement+properties
22 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
23 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
24 Lockwood C, Dos Santos KB, Pap R. Practical guidance for knowledge synthesis: scoping review methods. *Asian Nurs Res (Korean Soc Nurs Sci)* 2019;13:287–94.
25 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
26 Büchter RB, Weise A, Pieper D. Reporting of methods to prepare, pilot and perform data extraction in systematic reviews: analysis of a sample of 152 cochrane and non-cochrane reviews. *BMC Med Res Methodol* 2021;21:240.
27 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
28 Mokkink LB, Prinsen CA, Patrick DL, *et al*. COSMIN study design checklist for patient-reported outcome measurement instruments; 2019. 1–32.
29 Terwee C, Prinsen C, Chiarotto A, *et al*. COSMIN methodology for assessing the content validity of proms user manual. 2018. Available: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf
30 DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials* 2015;45:139–45.
31 Collins NJ, Prinsen CAC, Christensen R, *et al*. Knee injury and osteoarthritis outcome score (KOOS): systematic review and meta-analysis of measurement properties. *Osteoarthr Cartil* 2016;24:1317–29.
32 Hair JF, Black WC, Babin BJ, *et al*. Multivariate data analysis, (always learning). Pearson Education Limited; 2013. Available: https://books.google.ae/books?id=VvXZnQEACAAJ