BMJ Open

Measurement properties of mental health literacy scale (MHLS) validation studies: a systematic review protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-081394
Article Type:	Protocol
Date Submitted by the Author:	31-Oct-2023
Complete List of Authors:	ElKhalil, Rouwida; United Arab Emirates University, Public Health Institute AlMekkawi, Mohamad; Fatima College of Health Sciences, Nursing Department O'Connor, Matt; ConnectEd Counselling and Consultancy Sherif, Moustafa; United Arab Emirates University, Public Health Institute Masuadi, Emad; United Arab Emirates University, Public Health Institute Ahmed, Luai; United Arab Emirates University College of Medicine and Health Sciences, Institute of Public Health Al-Rifai, Rami H.; United Arab Emirates University, Institute of Public Health, College of Medicine and Health Sciences Belfakir, Messaouda; United Arab Emirates University, Public Health Institute Bayoumi, Rasha; University of Birmingham Dubai, School of Psychology Elbarazi, Iffat; United Arab Emirates University
Keywords:	Health Literacy, MENTAL HEALTH, Patient Reported Outcome Measures, Psychometrics, Systematic Review

Measurement properties of mental health literacy scale (MHLS) validation studies: a systematic review protocol

Rouwida ElKhalil [1], Mohamad AlMekkawi [2], Matt O'Connor [3], Moustafa Sherif [1], Emad Masuadi[1], Luai A. Ahmed [1], Rami H. AlRifai [1], Messaouda Belfakir[1], Rasha Bayoumi [4], Iffat Elbarazi [1]*

**Corresponding Author:* Iffat Elbarazi, email ielbarazi@uaeu.ac.ae, P.O. Box 15551, office +97137137643, mobile +971585939520, *Institute of Public Health, College of Medicine and Health Sciences (*CMHS*), United Arab Emirates University, Al Ain, United Arab Emirates.*

[1]Institute of Public Health, United Arab Emirates University, Al Ain, United Arab Emirates.
[2]Nursing Department, Fatima College of Health Sciences, Al Ain, United Arab Emirates.
[3]ConnectEd Counselling and Consultancy, Queensland, Australia.
[4]School of Psychology, Birmingham University, Dubai, United Arab Emirates.

Word count: 3173

Keywords: Cross-Cultural adaptation; Instrument validation; Measurement properties; Psychometric analysis; Translation.

ABSTRACT

Introduction: Mental health literacy MHL is essential for improving mental health and reducing inequities in treatment. Validated and reliable MHL evaluation instruments are essential for accurate data collection and to guide mental health policy and practice. The Mental Health Literacy Scale MHLS was designed to address these limitations and produce a valid and reliable assessment tool for MHL. It has been used in various cultural and language contexts, making it valuable for cross-cultural research studies. This systematic review will examine the measurement properties of the Mental Health Literacy Scale (MHLS) in different languages, enabling academics, clinicians, and policymakers to make informed judgments regarding its use in mental health literacy assessments.

Methods and analysis: The review will adhere to the COSMIN methodology for systematic reviews of Patient Reported Outcome Measures (PROMs) and JBI Manual for Evidence Synthesis and will be presented following PRISMA 2020 Checklist. The review will be conducted in four stages, including an initial search confined to PubMed, a search of electronic scientific databases MEDLINE, Embase, Scopus, and EBSCOhost (PsycINFO, CINAHL, 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.

Ethics and dissemination: Ethics approval is not required. Publication will be done in peer-reviewed journals and at national and international conferences.

PROSPERO registration number: CRD42023430924.

Strengths and Limitations

- This review analyses the measurement properties of all language versions of the MHLS, emphasizing the importance of researchers measuring MHL in various settings.
- This review will adhere to the JBI Manual for Evidence Synthesis (Chapter 12: Systematic reviews of measurement properties) and the COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual and will be reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline (PRISMA 2020).
- This systematic review is limited by the temporal discrepancy between the Mental Health Literacy Scale (MHLS) development in 2015 and the available resources for measurement properties quality evaluation, which existed after 2018.
- We anticipate that the heterogeneity of the studies will impact the ability to do meta-analyses.

INTRODUCTION

Mental health is an integral part of total health and well-being. Millions of individuals worldwide have a mental illness, with depression alone affecting over 280 million people [1]. Mental health literacy (MHL) is the "knowledge and beliefs about mental disorders which aid their recognition, management or prevention"[2]. MHL is critical for improving mental health and reducing inequities in mental health treatment. It assists individuals in recognizing their symptoms, locating resources, and receiving necessary assistance [3].

For accurate data collection and to guide mental health policy and practice, valid and reliable MHL evaluation instruments are essential. Validated instruments assist with collecting more accurate, reliable, and comparable data across contexts and cultures than instruments that have not undergone sufficient

psychometric development. Utilizing validated instruments to evaluate MHL is essential for designing effective strategies to improve mental health outcomes. They assist academics and policymakers in identifying knowledge gaps in MHL and developing culturally appropriate solutions tailored to individual and community needs. In addition, they support the evaluation of mental health interventions to ensure they are evidence-based, influenced by current research findings, and assessed using reliable information [4]. Developing a MHL instrument requires having a clear operational definition of the construct [5,6]. This construct is evaluated using two approaches, namely the Vignette Approach and Scale-based Measurements[7]. The Vignette Approach is " described as stories about individuals and situations which refer to important points in the study of perceptions, beliefs, and attitudes" [8]. 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." [9]. However, the psychometric tests of PROMs have shown significant limitations in measuring MHL [10,11].

Following a systematic assessment of MHL instruments in 2014, O'Connor and Casey designed the Mental Health Literacy Scale (MHLS) to address these limitations and produce a valid and reliable assessment tool for MHL [7]. The MHLS is the only instrument capable of measuring all aspects of mental health literacy [12]. The authors introduced the MHLS as a unidimensional measurement scale with 35 items and six attributes based on Jorm's six mental health literacy attributes [2]. 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 mental health literacy. The scale has the following sections: recognition of disorders (eight items measured on a 4-point Likert scale), knowledge of risk factors and causes (two items measured on a four-point Likert scale), self-treatment knowledge (two items measured on a 4-point Likert scale), knowledge of professional help available (three items measured on a four-point Likert scale), knowledge of how to seek mental health information (four items measured on a five 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 [7].

The MHLS has been used in various cultural and language contexts, making it a valuable instrument for cross-cultural research studies [13]. 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 [14,15]. According to Arafat, Chowdhury, Qusar and Hafez [14], 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.

Aims

this systematic review is significant to researchers aiming to measure MHL in diverse settings as it evaluates and compares the measurement properties of all language versions of the MHLS. 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 [13,16]. Consequently, this review aims to fill this gap by providing 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, enhancing 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 top researchers and publications in the field of MHLS validation. We aim to critically summarize, assess, and compare the measurement properties of all language versions of the MHLS by systematically examining the methodological quality and findings of the available publications. The objectives are to summarize the utilized adaptation /validation processes employed in MHLS validation studies, to assess the methodological quality of studies evaluating the measurement properties of the MHLS across several language versions, and to compare and synthesize 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 summarizing 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 [17]. The proposed systematic review will adhere to the JBI Manual for Evidence Synthesis (Chapter 12: Systematic Reviews of Measurement Properties) [18] and COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) [19]. The results will be presented according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) [20]. The systematic review methodology is summarized in Fig. 1. The study is registered at PROSPERO under the ID number CRD42023430924.

Fig. 1. Systematic review methodology summary

Search strategy

The review will begin with forming a research team of individuals with content and methodological competence [21]. 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, as specified by The Joanna Briggs Institute's Standards [18].

In the first stage, an initial search of the PubMed database will be done using a sensitive search filter to find studies on the measurement properties of MHLS (see Supplementary 3). The initial search will follow [22] '*Filter 1: Sensitive search filter for measurement properties*' because it guarantees 97.4% sensitive and 4.4% precise results (Table 1). In the second stage, we will search the electronic scientific databases MEDLINE, Embase, Scopus, and EBSCOhost (PsycINFO, CINAHL, and ERIC) using the final Boolean expression created in the previous phase (see Supplementary 1). 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 original author will be contacted to identify validation studies not retrieved in the previous searches. The search filters (see Supplementary 3) were combined with phrases searched for the concept of interest (Mental Health Literacy) "AND" the measuring instrument of interest (Mental Health Literacy Scale). 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 mental health literacy 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.

Table 1
Systematic review search strategy. *Adopted form* [22]

	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 Supplementary 3A)
#4	#3 NOT exclusion search filter (See Supplementary 3B)

Study screening and selection

The screening and selection approach will be summarized using the Preferred Reported Items in Systematic Reviews and Meta-analysis (PRISMA) flowchart [23]. Our review question and inclusion criteria are framed using the PICO (Population, Instruments, Construct, Outcomes) method [18].Eligibility criteria, as shown in Table 2, are as follows: (1) *Participants:* The review will consider studies that validate the MHLs in any population (e.g., 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 (i.e., as acute care, primary health care, or the community); (2) *Instrument and Construct:* The review will focus solely on O'Connor and Casey [7] MHLS; 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 [18]; (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 [14]. The aim of the included studies should be the evaluation of one or more measurement properties [19]. 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, then the full text will be examined. Two reviewers 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 and, if necessary, alter the search approach. Another two researchers will independently review the full manuscripts. A third reviewer will make the final judgment when there is disagreement over research inclusion. 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.

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:** Mental Health Literacy Scale (MHLS), O'Conner and Casey 2015 assess the mental health literacy construct. 3. **Outcomes:** Reliability, validity, and responsiveness. 4. **Types of sources:** Validation studies 5. **Language:** English 6. **Date:** 2015 to 2022	1. Non-English studies 2. Grey literature (non-peer-reviewed publications or documents of any type) 3. Other mental health literacy measures 4. Studies that only use the MHLS as an outcome measure.

Table 3
Systematic review outcomes: measurement properties. *Adopted form* [18]

Main Outcomes	Effect Measures
1. 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.
2. 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* Standardized 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.
3. Responsiveness	Absolute or relative correlations, *or* Differences of the change scores, *or* The Areas under Receiver Operating Curves (ROC), *or* Sensitivity and specificity.

Data charting

Two independent reviewers will do data extraction and methodological quality assessment of full-text articles that meet the inclusion criteria using the COSMIN Microsoft Excel 365® spreadsheet template that the reviewers adapted. Before beginning the review, we will conduct calibration exercises to ensure consistency among reviewers.The data charting instruments (See Supplementary 2) were adapted from the COSMIN methodology for systematic reviews of the Patient-Reported Outcome Measures (PROMs) user manual [19]. 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, utilized validation/adaptation process, measurement properties quality, and results (rating), will guide our data "charting." We will also chart data by publication year, Instrument administration (Country, Target Language, Setting), characteristics of the included sample [Age Mean (SD), Gender (% female), Sample size and calculation], MHLS score, and reported MHLS item modifications.

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 [19]. 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 (i.e., "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 [24]. In addition, 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 [19]. The content validity rating criteria results were based on the COSMIN methodology guidelines for assessing the content validity of the PROMs User manual [22].

Data synthesis, levels of evidence, and meta-analyses

The results will either be quantitatively or qualitatively combined. We will present these pooled or summarized results per measurement property (See Supplementary 2), together with a grade for the quality of the evidence (high, moderate, low, or extremely low) and a rating of the pooled or summarized results (+ /- /**?**).

Table 4
Quality criteria for measurement properties. *Adapted from* [19,22,25]

Property	Rating [b]	Quality criteria
Reliability		
Internal Consistency	+	Cronbach alphas ≥.70
	?	Cronbach alpha not determined.
	-	Cronbach alphas <.70
Reliability	+	ICC/weighted kappa ≥.70 OR Pearson r≥.80
	?	Neither ICC/weighted kappa nor Pearson r determined
	-	ICC/weighted kappa .70 OR Pearson r.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
	?	CTT: Not all information for '+' reported Or Explained variance not mentioned
	-	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 hypothesis7 OR AUC ≥ 0.70
	?	No hypothesis was defined (by the review team)
	-	The result is not in accordance with the hypothesis7 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?

[a] MIC=minimal important change, SDC=smallest detectable change, LOA=limits of agreement, ICC=intraclass correlation coefficient, DIF=differential item functioning, AUC=area under the curve.

[b] +=positive rating, ?=indeterminate rating, -= negative rating, ±= mixed ratings *(content validity only)*

Quantitatively pooling the results

When there are 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% confidence intervals will yield pooled estimates of measurement properties. To conduct meta-analyses, we will be consulting a statistician.

Qualitatively summarizing result

If it is impossible to pool the results statistically, the results of each measurement property should 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 summarized results

The pooled or summarized 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 Supplementary 2).

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 (i.e., quality of

the studies' methodology), (2) inconsistency (i.e., unexplained, inconsistent results across studies), (3) imprecision (i.e., the total sample size of the available studies), and (4) indirectness (i.e., evidence from different populations than the population of interest in the review) [19].

Table 5
Definitions of quality levels. *Adopted from* [19]

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.

** Estimate of the measurement property refers to the pooled or summarized result of the measurement property of a PROM.*

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 aids individuals in detecting their symptoms, locating relevant resources, and receiving appropriate assistance[3]. Improving and sustaining healthcare delivery is a challenge for practitioners and policymakers. Patients provide unique insights into healthcare quality, yet they are an underutilized resource for measuring it. This systematic review evaluating and comparing the measurement properties of all language versions of the MHLS will shed new light on the measurement qualities of the MHLS in different language versions. This review will enable academics, clinicians, and policymakers to understand the reliability and validity of the Mental Health Literacy Scale (MHLS) across diverse cultural and linguistic contexts, allowing them to make informed judgments regarding its use in mental health literacy assessments[4]. We believe this systematic review is relevant and will significantly contribute to filling the current research gap.

Author Contributions

The concept and design of the study were conceived by RE, IE, EM, and MA. MS and RE collaborated on developing the search strategy. RE, MA, IE, and MS will finish the literature review. Data extraction will be performed by RE and MB. EM provided statistical expertise. MO provided expert advice on PROM. RA drafted the initial version of this manuscript and will also compose the final systematic review. LA, RA, and RB contributed additional text and revisions and offered further input.MA, MO, and IE reviewed and supervised the work on the manuscript. All authors examined and approved the submitted version.

Funding: The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: None declared.

Patient consent: Not applicable.

Patient and public involvement: Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Acknowledgements: We thank Dr Zufishan Alam for assistance in problem-solving data search issues.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1. WHO. *WHO menu of cost-effective interventions for mental health*. World Health Organization; 2021.
2. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Medical journal of Australia*. 1997;166(4):182-186.
3. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. *American psychologist*. 2012;67(3):231.
4. WHO. *Promoting mental health: Concepts, emerging evidence, practice: Summary report*. World Health Organization; 2004.
5. Kutcher S, Wei Y, Coniglio C. Mental health literacy: Past, present, and future. *The Canadian Journal of Psychiatry*. 2016;61(3):154-158.
6. Spiker DA, Hammer JH. Mental health literacy as theory: current challenges and future directions. *Journal of Mental Health*. 2019;28(3):238-242.
7. O'Connor M, Casey L. The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy. *Psychiatry research*. 2015;229(1-2):511-516.
8. Hughes R. Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of health & illness*. 1998;20(3):381-400.
9. CIHI. PROMs Background Document. presented at: Canadian Institute for Health Information (CIHI); 2015; https://www.cihi.ca/sites/default/files/proms_background_may21_en-web_0.pdf
10. Chaves C, Sequeira C, Duarte JC, Nelas PA, Gonçalves A, Santos EJRd. Mental health literacy: a systematic review of the measurement instruments. 2021;
11. O'Connor M, Casey L, Clough B. Measuring mental health literacy–a review of scale-based measures. *Journal of mental health*. 2014;23(4):197-204.
12. Montagni I, González Caballero JL. Validation of the mental health literacy scale in French university students. *Behavioral Sciences*. 2022;12(8):259.
13. 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: The Journal of Health Care Organization, Provision, and Financing*. 2022;59:00469580211047193.
14. Arafat S, Chowdhury HR, Qusar M, Hafez M. Cross cultural adaptation and psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health*. 2016;5(3):129-136.
15. Huang WY, Wong SH. Cross-Cultural Validation. In: Michalos AC, ed. *Encyclopedia of Quality of Life and Well-Being Research*. Springer Netherlands; 2014:1369-1371.
16. Speksnijder CM, Koppenaal T, Knottnerus JA, Spigt M, Staal JB, Terwee CB. Measurement properties of the quebec back pain disability scale in patients with nonspecific low back pain: systematic review. *Physical Therapy*. 2016;96(11):1816-1831.
17. 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*. Jan 2 2015;350:g7647. doi:10.1136/bmj.g7647
18. Stephenson M, Riitano D, Wilson S, et al. Systematic reviews of measurement properties. *Joanna Briggs Institute Reviewer's Manual The Joanna Briggs Institute [Internet]*. 2020;
19. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*. 2018;27:1147-1157.
20. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*. 2021;88:105906.
21. Lockwood C, Dos Santos KB, Pap R. Practical guidance for knowledge synthesis: Scoping review methods. *Asian nursing research*. 2019;13(5):287-294.

22. Terwee CB, Prinsen C, Chiarotto A, et al. COSMIN methodology for assessing the content validity of PROMs–user manual. *Amsterdam: VU University Medical Center*. 2018;
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*. 2018;169(7):467-473.
24. Mokkink LB, Prinsen C, Patrick DL, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. *Amsterdam, The Netherlands*. 2019:1-32.
25. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. Always Learning. Pearson Education Limited London, UK:; 2013.

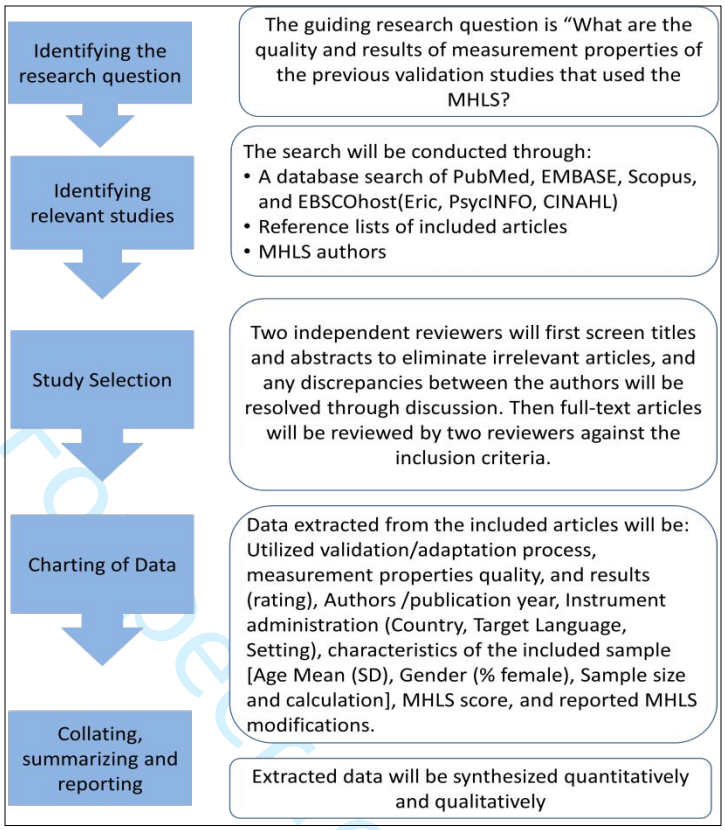

Fig. 1. Systematic review methodology summary

Supplementary 1 Search Strategies

	1.1. Search for PubMed	
	Population Search	This search did not include a population search since there is no restriction on population type, age and settings
#1	*Instrument Search MHLS O'Conner and Casey (2015)*	("mental health literacy scale*") OR (MHLS)
#2	*Construct search mental health literacy*	("mental health"[Title/Abstract] OR "mental health"[Title/Abstract] OR "mental health"[MeSH Terms] OR "mental stabilit*"[Title/Abstract] OR "mental balanc*"[Title/Abstract] OR "mental hygien*"[Title/Abstract] OR "sanit*"[Title/Abstract] OR "psychiatr*"[Title/Abstract] OR "life disrupt*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental disord*"[Title/Abstract] OR "mental wellbeing*"[Title/Abstract] OR "mental well being*"[Title/Abstract] OR "mental condition*"[Title/Abstract] OR "Mental Disorders"[MeSH Terms]) AND ("Health Literacy"[MeSH Terms] OR "health literac*"[Title/Abstract] OR "health education*"[Title/Abstract] OR "health train*"[Title/Abstract] OR "health aware*"[Title/Abstract])
#3	*Filter for measurement properties*	("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ''psychometrics'' [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ''observer variation''[MeSH] OR observer variation[Title/Abstract] OR ''Health Status Indicators''[Mesh] OR ''reproducibility of results''[MeSH] OR reproducib*[Title/Abstract] OR ''discriminant analysis''[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ''internal consistency''[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ''precise values''[Title/Abstract] OR test– retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-

		participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa's[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ''known group''[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ''individual variability''[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ''standard error of measurement''[Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ''ceiling effect''[Title/Abstract] OR ''floor effect''[Title/Abstract] OR ''Item response model''[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ''Differential item functioning''[Title/Abstract] OR DIF[Title/Abstract] OR ''computer adaptive testing''[Title/Abstract] OR ''item bank''[Title/Abstract] OR ''cross-cultural equivalence''[Title/Abstract])
#4	*Exclusion filter*	(''address*''[Publication Type] OR ''biography''[Publication Type] OR ''case reports''[Publication Type] OR ''comment''[Publication Type] OR ''directory''[Publication Type] OR ''editorial''[Publication Type] OR ''festschrift''[Publication Type] OR ''interview''[Publication Type] OR ''lectur*''[Publication Type] OR ''legal case*''[Publication Type] OR ''legislation''[Publication Type] OR ''letter''[Publication Type] OR ''news''[Publication Type] OR ''newspaper article''[Publication Type] OR ''patient education handout''[Publication Type] OR ''popular work*''[Publication Type] OR ''congress*'' [Publication Type] OR ''consensus development conference''[Publication Type] OR ''consensus development conference, nih''[Publication Type] OR ''practice guideline''[Publication Type]) NOT (''animals''[MeSH Terms] NOT ''humans''[MeSH Terms])
#5	*#1 AND #2 AND #3 NOT#4*	((((("mental health"[Title/Abstract] OR "mental health"[Title/Abstract] OR "mental health"[MeSH Terms] OR "mental stabilit*"[Title/Abstract] OR "mental balanc*"[Title/Abstract] OR "mental hygien*"[Title/Abstract] OR "sanit*"[Title/Abstract] OR "psychiatr*"[Title/Abstract] OR "life disrupt*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental disord*"[Title/Abstract] OR "mental wellbeing*"[Title/Abstract] OR "mental well being*"[Title/Abstract] OR "mental condition*"[Title/Abstract] OR "Mental Disorders"[MeSH Terms]) AND ("Health Literacy"[MeSH Terms] OR "health literac*"[Title/Abstract]

		OR "health education*"[Title/Abstract] OR "health train*"[Title/Abstract] OR "health aware*"[Title/Abstract])) AND (("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ‘‘psychometrics’’ [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ‘‘observer variation’’[MeSH] OR observer variation[Title/Abstract] OR ‘‘Health Status Indicators’’[Mesh] OR ‘‘reproducibility of results’’[MeSH] OR reproducib*[Title/Abstract] OR ‘‘discriminant analysis’’[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ‘‘internal consistency’’[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ‘‘precise values’’[Title/Abstract] OR test– retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa’s[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ‘‘known group’’[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ‘‘individual variability’’[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ‘‘standard error of measurement’’[

		Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ‘‘ceiling effect’’[Title/Abstract] OR ‘‘floor effect’’[Title/Abstract] OR ‘‘Item response model’’[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ‘‘Differential item functioning’’[Title/Abstract] OR DIF[Title/Abstract] OR ‘‘computer adaptive testing’’[Title/Abstract] OR ‘‘item bank’’[Title/Abstract] OR ‘‘cross-cultural equivalence’’[Title/Abstract]))) NOT ((‘‘address*’’[Publication Type] OR ‘‘biography’’[Publication Type] OR ‘‘case reports’’[Publication Type] OR ‘‘comment’’[Publication Type] OR ‘‘directory’’[Publication Type] OR ‘‘editorial’’[Publication Type] OR ‘‘festschrift’’[Publication Type] OR ‘‘interview’’[Publication Type] OR ‘‘lectur*’’[Publication Type] OR ‘‘legal case*’’[Publication Type] OR ‘‘legislation’’[Publication Type] OR ‘‘letter’’[Publication Type] OR ‘‘news’’[Publication Type] OR ‘‘newspaper article’’[Publication Type] OR ‘‘patient education handout’’[Publication Type] OR ‘‘popular work*’’[Publication Type] OR ‘‘congress*’’ [Publication Type] OR ‘‘consensus development conference’’[Publication Type] OR ‘‘consensus development conference, nih’’[Publication Type] OR ‘‘practice guideline’’[Publication Type]) NOT (‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) AND (("mental health literacy scale*") OR (MHLS))
	Result in documents Final= 6 Exclusion =14 (Different tool 13-RCT 1)	

**Note. The initial PubMed search was conducted on June 3, 2023, and was limited to English.

Table A2. Final Search Strategy

1.2. Search for Embase
1.3. Search for PsychINFO
1.4. Search for CINAHL
1.5. Search for ERIC
1.6. Search for Medline

Supplementary 2 Data Charting Instruments

2 A. Descriptive Characteristics of The Included Studies

#	Study (Authors/year)	Study Design	Instrument Administration			Population				Methodological Process		MHLS score	Reported MHLS modifications
			Country	Language	Setting	N	Age Mean (SD) yr	Gender (% female)	Selection process	Summary of adaptation process steps	Adaptation process / validation process reported guideline		
1.													
2.													
*Sample size calculation reference provided													

2 B. Results of Studies on Measurement Properties

Study (authors/date)	Country (language) in which the MHLS was evaluated	Structural validity			Internal Consistency			Cross-cultural validity/ measurement invariance			Reliability		
		n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)
Pooled or summary results (overall rating)													

Study (authors/date)	Country (language) in which the MHLS was evaluated	Measurement Error			Criterion validity			Hypothesis testing			Responsiveness		
		n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)
Pooled or summary results (overall rating)													

2 C. Summary of Findings Tables

Structural Validity	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Internal consistency	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Cross-cultural validity\measurement invariance	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Reliability	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Measurement Error	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Hypothesis testing	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Responsiveness	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Supplementary 3 Search Filters

This search filter was adopted from: Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009 Oct;18(8):1115-23. doi: 10.1007/s11136-009-9528-5. Epub 2009 Aug 27. PMID: 19711195; PMCID: PMC2744791.

Sensitive Filter for Measurement Properties

("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ''psychometrics'' [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ''observer variation''[MeSH] OR observer variation[Title/Abstract] OR ''Health Status Indicators''[Mesh] OR ''reproducibility of results''[MeSH] OR reproducib*[Title/Abstract] OR ''discriminant analysis''[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ''internal consistency''[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ''precise values''[Title/Abstract] OR test–retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa's[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ''known group''[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ''individual variability''[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ''standard error of measurement''[Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ''ceiling effect''[Title/Abstract] OR ''floor effect''[Title/Abstract] OR ''Item response model''[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ''Differential item functioning''[Title/Abstract] OR DIF[Title/Abstract] OR ''computer adaptive testing''[Title/Abstract] OR ''item bank''[Title/Abstract] OR ''cross-cultural equivalence''[Title/Abstract])

Exclusion Search Filter

(''address*''[Publication Type] OR ''biography''[Publication Type] OR ''case reports''[Publication Type] OR ''comment''[Publication Type] OR ''directory''[Publication Type] OR ''editorial''[Publication Type] OR ''festschrift''[Publication Type] OR ''interview''[Publication Type] OR ''lectur*''[Publication Type] OR ''legal

case*''[Publication Type] OR ''legislation''[Publication Type] OR ''letter''[Publication Type] OR ''news''[Publication Type] OR ''newspaper article''[Publication Type] OR ''patient education handout''[Publication Type] OR ''popular work*''[Publication Type] OR ''congress*'' [Publication Type] OR ''consensus development conference''[Publication Type] OR ''consensus development conference, nih''[Publication Type] OR ''practice guideline''[Publication Type]) NOT (''animals''[MeSH Terms] NOT ''humans''[MeSH Terms])

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Title: **Measurement properties of mental health literacy scale (MHLS) validation studies: a systematic review protocol**

Section and topic	Item No	Checklist item	Reported on Page number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	Page 1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	NA
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	Page 1 & 4
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	Page 1 &10
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	Page 10
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	NA
Support:			
Sources	5a	Indicate sources of financial or other support for the review	Page 10
Sponsor	5b	Provide name for the review funder and/or sponsor	Page 10
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	Page 10
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	Page 3
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	Page 3
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	Page 5
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	Page 4
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Supplement 1
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	Page5

Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	Page 5
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	Page 6
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	Page 6
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	Page 7
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	Page 6-7
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	Page 8
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	Page 8
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	NA
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	Page 8
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	Page 9
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	Page 9

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-081394.R1
Article Type:	Protocol
Date Submitted by the Author:	14-Feb-2024
Complete List of Authors:	ElKhalil, Rouwida; United Arab Emirates University, Public Health Institute AlMekkawi, Mohamad; Fatima College of Health Sciences, Nursing Department O'Connor, Matt; ConnectEd Counselling and Consultancy Sherif, Moustafa; United Arab Emirates University, Public Health Institute Masuadi, Emad; United Arab Emirates University, Public Health Institute Ahmed, Luai; United Arab Emirates University College of Medicine and Health Sciences, Institute of Public Health Al-Rifai, Rami H.; United Arab Emirates University, Institute of Public Health, College of Medicine and Health Sciences Belfakir, Messaouda; United Arab Emirates University, Public Health Institute Bayoumi, Rasha; University of Birmingham Dubai, School of Psychology Elbarazi, Iffat; United Arab Emirates University
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Public health
Keywords:	MENTAL HEALTH, Patient Reported Outcome Measures, Psychometrics, Systematic Review, Health Literacy

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

Rouwida ElKhalil [1], Mohamad AlMekkawi [2], Matt O'Connor [3], Moustafa Sherif [1], Emad Masuadi [1], Luai A. Ahmed [1], Rami H. AlRifai [1], Messaouda Belfakir [1], Rasha Bayoumi [4], Iffat Elbarazi [1]*

**Corresponding Author:* Iffat Elbarazi, email ielbarazi@uaeu.ac.ae, PO Box 15551, office +97137137643, mobile +971585939520, *Institute of Public Health, College of Medicine and Health Sciences (*CMHS*), United Arab Emirates University, Al Ain, United Arab Emirates.*

[1] Institute of Public Health, United Arab Emirates University, Al Ain, United Arab Emirates.
[2] Nursing Department, Fatima College of Health Sciences, Al Ain, United Arab Emirates.
[3] ConnectEd Counselling and Consultancy, Queensland, Australia.
[4] School of Psychology, Birmingham University, Dubai, United Arab Emirates.

Word count: 3002

Keywords: Cross-Cultural adaptation; Instrument validation; Measurement properties; Psychometric analysis; Translation.

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 judgments regarding its use in assessments.

Methods and analysis: The review will adhere to the COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs) and the JBI Manual for Evidence Synthesis and will be presented following the PRISMA 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 papers 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 (ROB) 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.

Article Summary

Strengths and Limitations of This Study

- This review evaluates MHLS measurement properties across languages, stressing diverse MHL assessments.
- It adheres to the JBI Manual and COSMIN methodology and follows PRISMA 2020 guidelines.
- Limited by a temporal gap post-2018 due to MHLS development in 2015.
- 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 health literacy (MHL), a derivative from and component of Health Literacy (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, recognizing 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 behaviors, 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 a 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 MHLScale (MHLS) to address these limitations and to produce a valid and reliable assessment tool for MHL (11). The rigor 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 (α=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 six 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 (eight items measured on a four-point Likert scale), Knowledge of Risk Factors and Causes (two items measured on a four-point Likert scale), Self-Treatment Knowledge (two items measured on a four-point Likert scale), Knowledge of Professional Help Available (three items measured on a four-point Likert scale), Knowledge of How to Seek Mental Health Information (four items measured on a five Likert-scale), and Attitudes that Promote Recognition and Appropriate Help-Seeking (16 items measured on a five-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, summarize, 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 summarize the utilized adaptation /validation processes employed in MHLS validation studies,
2. To assess the methodological quality of studies
3. To evaluate the measurement properties of the MHLS across several language versions,

4. To compare and synthesize 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 summarizing 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 Patient-Reported Outcome Measures (PROMs) (22). The results will be presented according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020)(23). The systematic review methodology is summarized in Figure 1. The study is registered at PROSPERO under the ID number CRD42023430924.

Insert Figure 1.

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 Supplementary 1). 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 Supplementary 2). 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 creator will be contacted to identify validation studies not retrieved in the previous searches.

We have already identified the search filters (see Supplementary 1). These were combined with phrases searched for the concept of interest (Mental Health Literacy) "AND" the measuring instrument of interest (MHLScale). 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.

Table 1
Systematic review search strategy. *Adopted from Terwee et al.* (25)

	Search Strategy
#1	Construct Search (Mental Health Literacy)
#2	Instrument Search (MHLScale)
#3	#1 AND #2 AND Sensitive filter for measurement properties (See Supplementary 1A)
#4	#3 NOT exclusion search filter (See Supplementary 1B)

Study screening and selection

The screening and selection approach will be summarized using the Preferred Reported Items in Systematic Reviews and Meta-analysis (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 (e.g., 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 (i.e., as acute care, primary health care, 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 ME) 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 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 judgment 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.

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: MHLScale (MHLS), O'Conner and Casey 2015 assess the MHLconstruct. 3. Outcomes: Reliability, validity, and responsiveness. 4. Types of sources: Validation studies 5. Language: English 6. Date: 2015 to 2022	1. Non-English studies 2. Grey literature (non-peer-reviewed publications or documents of any type) 3. Other MHLmeasures 4. Studies that only use the MHLS as an outcome measure.

Table 3
Systematic review outcomes: measurement properties. *Adopted from Stephenson et al.*(21)

Main Outcomes	Effect Measures
1. 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.
2. 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* Standardized 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.
3. Responsiveness	Absolute or relative correlations, *or* Differences of the change scores, *or* The Areas under Receiver Operating Curves (ROC), *or* Sensitivity and specificity.

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 Supplementary 4) 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 (i.e., "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 (Supplementary Appendix 3).

Data synthesis and levels of evidence

The results will either be quantitatively or qualitatively combined. We will present these pooled or summarized results per measurement property (See Supplementary 4C), together with a grade for the quality of the evidence (high, moderate, low, or extremely low) and a rating of the pooled or summarized results (+ /- /**?**).

Table 4

Quality criteria for measurement properties. *Adapted from Hair et al., Prinsen et al., and Terwee et al.* (22,29,30)

Property	Rating [b]	Quality criteria
Reliability		
Internal Consistency	+	Cronbach alphas ≥.70
	?	Cronbach alpha not determined.
	-	Cronbach alphas <.70
Reliability	+	ICC/weighted kappa ≥.70 OR Pearson r≥.80
	?	Neither ICC/weighted kappa nor Pearson r determined
	-	ICC/weighted kappa .70 OR Pearson r.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 unidimensionality3: 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- standardized 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 hypothesis7 OR AUC ≥ 0.70
	?	No hypothesis was defined (by the review team)
	-	The result is not in accordance with the hypothesis7 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?

[a] *MIC=minimal important change, SDC=smallest detectable change, LOA=limits of agreement, ICC=intraclass correlation coefficient, DIF=differential item functioning, AUC=area under the curve.*

[b] *+=positive rating,?=indeterminate rating, -= negative rating, ±= mixed ratings (content validity only)*

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% confidence intervals will yield pooled estimates of measurement properties. For assessing test-retest reliability, one can calculate weighted mean intraclass correlation coefficients (ICCs) and 95% confidence intervals using a standard generic inverse variance random effects model (31). ICC values can be combined based on estimates obtained from a Fisher transformation, z = 0.5 x ln ((1+ICC)/(1-ICC)), which has an approximate variance of (Var(z) = 1/(N-3)), where N is the sample size (32). 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 summarized results

The pooled or summarized 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 Supplementary 4B and 4C).

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 (i.e., quality of the studies' methodology), (2) inconsistency (i.e., unexplained, inconsistent results across studies), (3) imprecision (i.e., the total sample size of the available studies), and (4) indirectness (i.e., evidence from different populations than the population of interest in the review) (22).

Table 5

Definitions of quality levels. *Adopted from Prinsen et al.* (22)

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.

** Estimate of the measurement property refers to the pooled or summarized result of the measurement property of a PROM.*

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 summarize 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, emphasizing 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 Patient-Reported Outcome Measures (PROMs) user manual (22) and will be reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline (PRISMA 2020) (23). However, this systematic review will be limited by the temporal discrepancy between the MHLScale (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.

Contributions

RE, IE, EM, and MA conceived the concept and design of the study. MS and RE collaborated on developing the search strategy. RE, MA, IE, and MS will complete the literature review. RE and MB EM will perform data extraction provided their statistical expertise. MO provided expert advice on the MHLS, methodology plan, and manuscript drafting and editing, but he was not involved in data charting, risk of bias assessment,

or data synthesis. RA drafted the initial version of this manuscript and will also compose the final systematic review. LA, RA, and RB contributed to the additional text and revisions. MA and IE reviewed and supervised the work on the manuscript. All authors examined and approved the submitted version.

Funding: The United Arab Emirates University (UAEU) provided publication funds for this article , and the award/grant number is noted as "Not Applicable."

Competing interests: MO is among the MHLS's authors. The authors do not disclose any additional potential conflicts of interest.

Patient consent: Not required

Provenance and peer review: Not commissioned; externally peer-reviewed

Acknowledgments: The authors would like to acknowledge Dr. Zufishan Alam for assistance in problem-solving data search issues.

Provenance and peer review: Not commissioned; externally peer-reviewed.

ORCID numbers

Ms. Rouwida ElKhalil: 0000-0002-2031-837X

Dr. Mohamad AlMekkawi: 0000-0002-2315-7479

Dr. Matt O'Connor: 0000-0002-6188-8870

Mr. Moustafa Sherif: 0000-0002-8781-5098

Ms. Messaouda Belfakir:0009-0006-0975-4839

Dr. Emad Masaudi: 0000-0002-7360-7693

Dr. Luai A. Ahmed: 0000-0001-5292-8212

Dr. Rami H. AlRifai: 0000-0001-6102-0353

Dr. Rasha Bayoumi : 0000-0001-6115-1820

Dr. Iffat Elbarazi: 0000-0001-7151-2175

Figure. 1. Legend/caption is "Systematic review methodology summary".

REFERENCES

1. World Health Orgainsation. WHO menu of cost-effective interventions for mental health [Internet]. 2021 [cited 2024 Jan 31]. Available from: https://www.who.int/publications-detail-redirect/9789240031081

2. Healthy People 2030. health.gov. [cited 2024 Feb 3]. History of Health Literacy Definitions - Healthy People 2030. Available from: https://health.gov/healthypeople/priority-areas/health-literacy-healthy-people-2030/history-health-literacy-definitions

3. Kutcher S, Wei Y, Coniglio C. Mental Health Literacy: Past, Present, and Future. Can J Psychiatry. 2016 Mar;61(3):154–8.

4. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Medical Journal of Australia. 1997 Feb;166(4):182–6.

5. Jorm AF. Why We Need the Concept of 'Mental Health Literacy'. Health Commun. 2015;30(12):1166–8.

6. Sweileh WM. Global research activity on mental health literacy. Middle East Current Psychiatry. 2021 Sep 2;28(1):43.

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! Frontiers in Public Health [Internet]. 2023 [cited 2024 Feb 3];11. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1065784

8. Jorm AF. Mental health literacy: Empowering the community to take action for better mental health. American Psychologist. 2012 Apr;67(3):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. [Internet]. Geneva: World Health Organization; 2004 p. 1–70. Report No.: ISBN 92 4 159159 5-(NLM classification: WM 31.5). Available from: 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. Journal of Mental Health. 2019 May 4;28(3):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 Sep 30;229(1–2):511–6.

12. Hughes R. Considering the Vignette Technique and its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour. Sociology of Health & Illness. 1998;20(3):381–400.

13. Chow E, Faye L, Sawatzky R, Webster G, Cho MZ. PROMs Background Document [Internet]. Ottawa, Ontario: The Canadian Institute for Health Information (CIHI); 2015 p. 1–38. Available from: https://www.cihi.ca/sites/default/files/proms_background_may21_en-web_0.pdf

14. Chaves C, Sequeira C, Duarte J, Nelas P, Gonçalves A, Santos E. Mental health literacy: a systematic review of the measurement instruments. International Journal of Developmental and Educational Psychology Revista INFAD de psicología. 2022 Jan 14;3:181–94.

15. Montagni I, González Caballero JL. Validation of the Mental Health Literacy Scale in French University Students. Behav Sci (Basel). 2022 Jul 28;12(8):259.

16. Krohne N, Gomboc V, Lavrič M, Podlogar T, Poštuvan V, Šedivy NZ, et al. Slovenian Validation of the Mental Health Literacy Scale (S-MHLS) on the General Population: A Four-Factor Model. Inquiry. 2022 Feb 8;59:00469580211047193.

17. Arafat SMY, Chowdhury HR, Qusar MS, Hafez MA. Cross Cultural Adaptation & Psychometric Validation of Research Instruments: a Methodological Review -. Journal of Behavioral Health. 2016;5(3):129–36.

18. Huang WY, Wong SH. Cross-Cultural Validation. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research [Internet]. Dordrecht: Springer Netherlands; 2014 [cited 2023 May 9]. p. 1369–71. Available from: http://link.springer.com/10.1007/978-94-007-0753-5_630

19. Speksnijder CM, Koppenaal T, Knottnerus JA, Spigt M, Staal JB, Terwee CB. Measurement Properties of the Quebec Back Pain Disability Scale in Patients With Nonspecific Low Back Pain: Systematic Review. Phys Ther. 2016 Nov;96(11):1816–31.

20. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647–g7647.

21. Stephenson M, Riitano D, Wilson S, Leonardi-Bee J, Mabire C, Cooper K, et al. Chapter 12: Systematic reviews of measurement properties. In: JBI Manual for Evidence Synthesis [Internet]. JBI; 2020 [cited 2023 Sep 10]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL/4686202/Chapter+12%3A+Systematic+reviews+of+measurement+properties

22. Prinsen C a. C, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147–57.

23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;n71.

24. Lockwood C, dos Santos KB, Pap R. Practical Guidance for Knowledge Synthesis: Scoping Review Methods. Asian Nursing Research. 2019 Dec 1;13(5):287–94.

25. 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 Oct 1;18(8):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 Medical Research Methodology. 2021 Nov 6;21(1):240.

27. Mokkink LB, de Vet HCW, Prinsen C a. C, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 May;27(5):1171–9.

28. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019;1–32.

29. Terwee C, Prinsen C, Chiarotto A, Vet H, Bouter L, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs User manual. 2018; Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf

30. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis [Internet]. Pearson Education Limited; 2013. (Always learning). Available from: https://books.google.ae/books?id=VvXZnQEACAAJ

31. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials. 2015 Nov;45(Pt A):139–45.

32. Collins NJ, Prinsen C a. C, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthritis and Cartilage. 2016 Aug 1;24(8):1317–29.

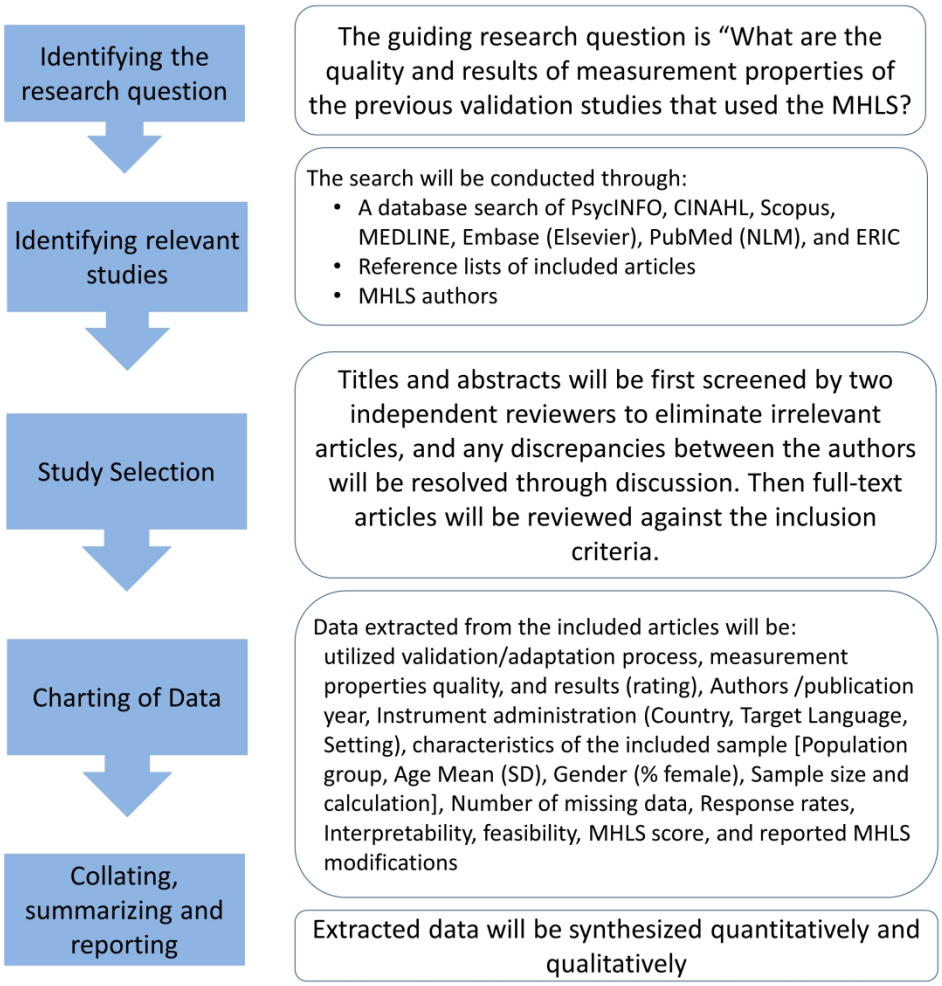

Figure 1- Systematic Review Methodology Summary

276x282mm (300 x 300 DPI)

Supplementary 1 Search Strategies

	1.1. Search for PubMed	
	Population Search	This search did not include a population search since there is no restriction on population type, age and settings
#1	*Instrument Search MHLS O'Conner and Casey (2015)*	("mental health literacy scale*") OR (MHLS)
#2	*Construct search mental health literacy*	("mental health"[Title/Abstract] OR "mental health"[Title/Abstract] OR "mental health"[MeSH Terms] OR "mental stabilit*"[Title/Abstract] OR "mental balanc*"[Title/Abstract] OR "mental hygien*"[Title/Abstract] OR "sanit*"[Title/Abstract] OR "psychiatr*"[Title/Abstract] OR "life disrupt*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental disord*"[Title/Abstract] OR "mental wellbeing*"[Title/Abstract] OR "mental well being*"[Title/Abstract] OR "mental condition*"[Title/Abstract] OR "Mental Disorders"[MeSH Terms]) AND ("Health Literacy"[MeSH Terms] OR "health literac*"[Title/Abstract] OR "health education*"[Title/Abstract] OR "health train*"[Title/Abstract] OR "health aware*"[Title/Abstract])
#3	*Filter for measurement properties*	("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ''psychometrics'' [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ''observer variation''[MeSH] OR observer variation[Title/Abstract] OR ''Health Status Indicators''[Mesh] OR ''reproducibility of results''[MeSH] OR reproducib*[Title/Abstract] OR ''discriminant analysis''[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ''internal consistency''[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ''precise values''[Title/Abstract] OR test– retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-

		participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa's[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ''known group''[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ''individual variability''[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ''standard error of measurement''[Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ''ceiling effect''[Title/Abstract] OR ''floor effect''[Title/Abstract] OR ''Item response model''[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ''Differential item functioning''[Title/Abstract] OR DIF[Title/Abstract] OR ''computer adaptive testing''[Title/Abstract] OR ''item bank''[Title/Abstract] OR ''cross-cultural equivalence''[Title/Abstract])
#4	*Exclusion filter*	(''address*''[Publication Type] OR ''biography''[Publication Type] OR ''case reports''[Publication Type] OR ''comment''[Publication Type] OR ''directory''[Publication Type] OR ''editorial''[Publication Type] OR ''festschrift''[Publication Type] OR ''interview''[Publication Type] OR ''lectur*''[Publication Type] OR ''legal case*''[Publication Type] OR ''legislation''[Publication Type] OR ''letter''[Publication Type] OR ''news''[Publication Type] OR ''newspaper article''[Publication Type] OR ''patient education handout''[Publication Type] OR ''popular work*''[Publication Type] OR ''congress*'' [Publication Type] OR ''consensus development conference''[Publication Type] OR ''consensus development conference, nih''[Publication Type] OR ''practice guideline''[Publication Type]) NOT (''animals''[MeSH Terms] NOT ''humans''[MeSH Terms])
#5	*#1 AND #2 AND #3 NOT#4*	((((("mental health"[Title/Abstract] OR "mental health"[Title/Abstract] OR "mental health"[MeSH Terms] OR "mental stabilit*"[Title/Abstract] OR "mental balanc*"[Title/Abstract] OR "mental hygien*"[Title/Abstract] OR "sanit*"[Title/Abstract] OR "psychiatr*"[Title/Abstract] OR "life disrupt*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental ill*"[Title/Abstract] OR "mental disord*"[Title/Abstract] OR "mental wellbeing*"[Title/Abstract] OR "mental well being*"[Title/Abstract] OR "mental condition*"[Title/Abstract] OR "Mental Disorders"[MeSH Terms]) AND ("Health Literacy"[MeSH Terms] OR "health literac*"[Title/Abstract]

		OR "health education*"[Title/Abstract] OR "health train*"[Title/Abstract] OR "health aware*"[Title/Abstract])) AND (("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ‘‘psychometrics’’ [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ‘‘observer variation’’[MeSH] OR observer variation[Title/Abstract] OR ‘‘Health Status Indicators’’[Mesh] OR ‘‘reproducibility of results’’[MeSH] OR reproducib*[Title/Abstract] OR ‘‘discriminant analysis’’[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ‘‘internal consistency’’[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ‘‘precise values’’[Title/Abstract] OR test– retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa’s[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ‘‘known group’’[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ‘‘individual variability’’[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ‘‘standard error of measurement’’[

		Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ‘‘ceiling effect’’[Title/Abstract] OR ‘‘floor effect’’[Title/Abstract] OR ‘‘Item response model’’[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ‘‘Differential item functioning’’[Title/Abstract] OR DIF[Title/Abstract] OR ‘‘computer adaptive testing’’[Title/Abstract] OR ‘‘item bank’’[Title/Abstract] OR ‘‘cross-cultural equivalence’’[Title/Abstract]))) NOT ((‘‘address*’’[Publication Type] OR ‘‘biography’’[Publication Type] OR ‘‘case reports’’[Publication Type] OR ‘‘comment’’[Publication Type] OR ‘‘directory’’[Publication Type] OR ‘‘editorial’’[Publication Type] OR ‘‘festschrift’’[Publication Type] OR ‘‘interview’’[Publication Type] OR ‘‘lectur*’’[Publication Type] OR ‘‘legal case*’’[Publication Type] OR ‘‘legislation’’[Publication Type] OR ‘‘letter’’[Publication Type] OR ‘‘news’’[Publication Type] OR ‘‘newspaper article’’[Publication Type] OR ‘‘patient education handout’’[Publication Type] OR ‘‘popular work*’’[Publication Type] OR ‘‘congress*’’ [Publication Type] OR ‘‘consensus development conference’’[Publication Type] OR ‘‘consensus development conference, nih’’[Publication Type] OR ‘‘practice guideline’’[Publication Type]) NOT (‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) AND (("mental health literacy scale*") OR (MHLS))
	Result in documents Final= 6 Exclusion =14 (Different tool 13-RCT 1)	

**Note. The initial PubMed search was conducted on June 3, 2023, and was limited to English.

Table A2. Final Search Strategy

1.2. Search for Embase
1.3. Search for PsychINFO
1.4. Search for CINAHL
1.5. Search for ERIC
1.6. Search for Medline

Supplementary 2 Data Charting Instruments

2 A. Descriptive Characteristics of The Included Studies

#	Study (Authors/year)	Study Design	Instrument Administration			Population				Methodological Process		MHLS score	Reported MHLS modifications
			Country	Language	Setting	N	Age Mean (SD) yr	Gender (% female)	Selection process	Summary of adaptation process steps	Adaptation process / validation process reported guideline		
1.													
2.													
*Sample size calculation reference provided													

2 B. Results of Studies on Measurement Properties

Study (authors/date)	Country (language) in which the MHLS was evaluated	Structural validity			Internal Consistency			Cross-cultural validity/ measurement invariance			Reliability		
		n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)
Pooled or summary results (overall rating)													

Study (authors/date)	Country (language) in which the MHLS was evaluated	Measurement Error			Criterion validity			Hypothesis testing			Responsiveness		
		n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)	n	Meth quality	Result (rating)
Pooled or summary results (overall rating)													

2 C. Summary of Findings Tables

Structural Validity	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Internal consistency	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Cross-cultural validity\measurement invariance	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Reliability	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Measurement Error	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Hypothesis testing	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Responsiveness	*Summary or pooled results*	*Overall rating*	*Quality of evidence*
PROM A			
PROM B			
PROM C			

Supplementary 3 Search Filters

This search filter was adopted from: Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009 Oct;18(8):1115-23. doi: 10.1007/s11136-009-9528-5. Epub 2009 Aug 27. PMID: 19711195; PMCID: PMC2744791.

Sensitive Filter for Measurement Properties

("instrumentation" [Subheading]) OR "methods" [Subheading] OR "Validation Stud*"[Publication Type] OR Comparative Study[Publication Type] OR ''psychometrics'' [MeSH] OR psychometr*[Title/Abstract] OR clinimetr*[Text Word] OR clinometr*[Text Word] OR "Outcome Assessment, Health Care"[Mesh] OR outcome assessment[Title/Abstract] OR outcome measure*[Text Word] OR ''observer variation''[MeSH] OR observer variation[Title/Abstract] OR ''Health Status Indicators''[Mesh] OR ''reproducibility of results''[MeSH] OR reproducib*[Title/Abstract] OR ''discriminant analysis''[MeSH] OR reliab*[Title/Abstract] OR unreliab*[Title/Abstract] OR valid*[Title/Abstract] OR coefficient[Title/Abstract] OR homogeneity[Title/Abstract] OR homogeneous[Title/Abstract] OR ''internal consistency''[Title/Abstract] OR (cronbach*[Title/Abstract] AND (alpha[Title/Abstract] OR alphas[Title/Abstract])) OR (item[Title/Abstract] AND (correlation*[Title/Abstract] OR selection*[Title/Abstract] OR reduction*[Title/Abstract])) OR agreement[Title/Abstract] OR precision[Title/Abstract] OR imprecision[Title/Abstract] OR ''precise values''[Title/Abstract] OR test–retest[Title/Abstract] OR (test[Title/Abstract] AND retest[Title/Abstract]) OR (reliab* [Title/Abstract] AND (test[Title/Abstract] OR retest[Title/Abstract])) OR stability[Title/Abstract] OR interrater[Title/Abstract] OR inter-rater[Title/Abstract] OR intrarater[Title/Abstract] OR intra-rater[Title/Abstract] OR intertester[Title/Abstract] OR inter-tester[Title/Abstract] OR intratester[Title/Abstract] OR intra-tester[Title/Abstract] OR interobserver[Title/Abstract] OR inter-observer[Title/Abstract] OR intraobserver[Title/Abstract] OR intraobserver[Title/Abstract] OR intertechnician[Title/Abstract] OR inter-technician[Title/Abstract] OR intratechnician[Title/Abstract] OR intra-technician[Title/Abstract] OR interexaminer[Title/Abstract] OR inter-examiner[Title/Abstract] OR intraexaminer[Title/Abstract] OR intra-examiner[Title/Abstract] OR interassay[Title/Abstract] OR inter-assay[Title/Abstract] OR intraassay[Title/Abstract] OR intra-assay[Title/Abstract] OR interindividual[Title/Abstract] OR inter-individual[Title/Abstract] OR intraindividual[Title/Abstract] OR intra-individual[Title/Abstract] OR interparticipant [Title/Abstract] OR inter-participant[Title/Abstract] OR intraparticipant[Title/Abstract] OR intra-participant[Title/Abstract] OR kappa[Title/Abstract] OR kappa's[Title/Abstract] OR kappas[Title/Abstract] OR repeatab*[Title/Abstract] OR ((replicab*[Title/Abstract] OR repeated[Title/Abstract]) AND (measure[Title/Abstract] OR measures[Title/Abstract] OR findings[Title/Abstract] OR result[Title/Abstract] OR results[Title/Abstract] OR test[Title/Abstract] OR tests[Title/Abstract])) OR generaliza*[Title/Abstract] OR generalisa*[Title/Abstract] OR concordance[Title/Abstract] OR (intraclass[Title/Abstract] AND correlation*[Title/Abstract]) OR discriminative[Title/Abstract] OR ''known group''[Title/Abstract] OR factor analysis[Title/Abstract] OR factor analyses[Title/Abstract] OR dimension*[Title/Abstract] OR subscale*[Title/Abstract] OR (multitrait[Title/Abstract] AND scaling[Title/Abstract] AND (analysis[Title/Abstract] OR analyses[Title/Abstract])) OR item discriminant[Title/Abstract] OR interscale correlation*[Title/Abstract] OR error[Title/Abstract] OR errors[Title/Abstract] OR ''individual variability''[Title/Abstract] OR (variability[Title/Abstract] AND (analysis[Title/Abstract] OR values[Title/Abstract])) OR (uncertainty[Title/Abstract] AND (measurement[Title/Abstract] OR measuring[Title/Abstract])) OR ''standard error of measurement''[Title/Abstract] OR sensitiv*[Title/Abstract] OR responsive*[Title/Abstract] OR ((minimal[Title/Abstract] OR minimally[Title/Abstract] OR clinical[Title/Abstract] OR clinically[Title/Abstract]) AND (important[Title/Abstract] OR significant[Title/Abstract] OR detectable[Title/Abstract]) AND(change[Title/Abstract] OR difference[Title/Abstract])) OR (small*[Title/Abstract] AND (real[Title/Abstract] OR detectable[Title/Abstract]) AND (change[Title/Abstract] OR difference[Title/Abstract]) OR meaningful change [Title/Abstract] OR ''ceiling effect''[Title/Abstract] OR ''floor effect''[Title/Abstract] OR ''Item response model''[Title/Abstract] OR IRT[Title/Abstract] OR Rasch[Title/Abstract] OR ''Differential item functioning''[Title/Abstract] OR DIF[Title/Abstract] OR ''computer adaptive testing''[Title/Abstract] OR ''item bank''[Title/Abstract] OR ''cross-cultural equivalence''[Title/Abstract])

Exclusion Search Filter

(''address*''[Publication Type] OR ''biography''[Publication Type] OR ''case reports''[Publication Type] OR ''comment''[Publication Type] OR ''directory''[Publication Type] OR ''editorial''[Publication Type] OR ''festschrift''[Publication Type] OR ''interview''[Publication Type] OR ''lectur*''[Publication Type] OR ''legal

case*''[Publication Type] OR ''legislation''[Publication Type] OR ''letter''[Publication Type] OR ''news''[Publication Type] OR ''newspaper article''[Publication Type] OR ''patient education handout''[Publication Type] OR ''popular work*''[Publication Type] OR ''congress*'' [Publication Type] OR ''consensus development conference''[Publication Type] OR ''consensus development conference, nih''[Publication Type] OR ''practice guideline''[Publication Type]) NOT (''animals''[MeSH Terms] NOT ''humans''[MeSH Terms])

Supplementary 4-. Hypotheses for construct validity and responsiveness

Specific MHLS hypotheses for 'Hypothesis Testing for Construct Validity' and 'Responsiveness'

Hypothesis Testing for Construct Validity	Hypothesised *a priori* that we would observe: ***Comparison with other outcome measurement instruments (convergent validity)*** 1. In all populations, strong correlations ≥0.5 will be observed between MHLS and comparator instruments (see Table 2) measuring similar constructs [e.g. Mental health literacy questionnaire-short version for adults (MHLq-SVa)] 2. In all populations, medium correlations (≥0.30 and <0.50) will be observed between MHLS and instruments (see Table 2) measuring *related but dissimilar* constructs [e.g. Attitudes Towards Depression (ATD), The Stigmatizing Attitudes-Believability (SAB)] 3. In all populations, ,weak correlations (<0.3) will be observed between MHLS and instruments (see Table 2) measuring a *separate* construct (instruments that do not measure MHL) [e.g. The 12-item General Health Questionnaire (GHQ-12), My lifestyle questionnaire (MLQ)]. ***Comparison between subgroups (Divergent validity)*** MHLS scores should be able to distinguish between the following groups, with statistically significant differences between the groups: • **Alternative Hypothesis 1**:younger vs. older age groups, with older groups having higher MHL. • **Alternative Hypothesis 2**:males vs. females, with females having higher MHL. • **Alternative Hypothesis 3:** those who have direct experience with mental disorders vs. those who do not, with those having direct experience with mental disorders having higher MHL. • **Alternative Hypothesis 4:**those who have indirect experience with mental disorders (family or friends) vs. those who do not, with those having indirect experience with mental disorders having higher MHL.

	• **Alternative Hypothesis 5**:Those with low SES vs. high SES. With those with high SES having higher MHL. • **Alternative Hypothesis 6:**Those with low education levels vs. high education levels, with those with higher education having higher MHL.
Responsiveness	Hypothesized *a priori* that we would observe: ***Before and After Intervention:*** A. ***Null hypothesis:*** In response to an educational intervention, we expect no difference in MHLS scores in the intervention and control groups. B. ***Alternative hypothesis:*** In response to an educational intervention, we expect a statistically significant difference in MHLS scores in the intervention and control groups.

1. O'Connor M, Casey L. The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy. Psychiatry Res. 2015 Sep 30;229(1–2):511–6.

2. 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 Oct 1;18(8):1115–23.

3. Stephenson M, Riitano D, Wilson S, Leonardi-Bee J, Mabire C, Cooper K, et al. Chapter 12: Systematic reviews of measurement properties. In: JBI Manual for Evidence Synthesis [Internet]. JBI; 2020 [cited 2023 Sep 10]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL/4686202/Chapter+12%3A+Systematic+reviews+of+measurement+properties

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Title: **Measurement properties of mental health literacy scale (MHLS) validation studies: a systematic review protocol**

Section and topic	Item No	Checklist item	Reported on Page number
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	Page 1
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	NA
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	In the abstract on page 2 & under methods on page 5
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	Page 1
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	Line 368, page 12
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	NA
Support:			
Sources	5a	Indicate sources of financial or other support for the review	Page 13
Sponsor	5b	Provide name for the review funder and/or sponsor	Page 13
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	Page 13
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	Line 152, page 4
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	Line 162, page 4
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	Line 214, page 6
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	Line 193, page 5
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	Supplement 1

Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	Using Covidence (line 228, page 6) and using COSMIN Excel spreadsheets (Line 250, Page 8)

Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	Page 5
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	Page 6
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	Page 6
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	Page 7
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	Page 6-7
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	Page 8
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	Page 8
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	NA
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	Page 8
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	Page 9
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	Page 9

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.