From Translation to Fair Comparisons: A Practical Framework for Cross-Cultural Adaptation and Measurement Equivalence of Public Health Literacy Instruments

Maria Roumelioti[1], George J. Xydopoulos[2], Georgia Mikrouli[2], Emily Diamantopoulou[2], Eleni Jelastopulu[1]

Abstract

Public health research and education increasingly rely on literacy scales (eHealth, Media/News, and Digital Literacies) to design and evaluate interventions. Instruments are often adapted across languages, yet translation alone is insufficient for valid comparisons unless adapted versions demonstrate measurement equivalence. This narrative article provides an educational, practice-oriented framework that links cross-cultural adaptation workflows with Consensus-based Standards for the selection of Health Measurement Instruments-aligned evidence generation and measurement equivalence concepts, with special attention to objective true/false knowledge scales that are vulnerable to familiarity-related bias. We synthesize established cross-cultural adaptation guidance (preparation, dual forward translation, reconciliation, back-translation, expert review, cognitive debriefing, proofreading, and documentation) with a methodological framework for evidence building and cross-group comparability using multi-group confirmatory factor analysis and differential item functioning. Decision points are illustrated using two contrasting case examples: a self-report eHealth Literacy Scale and a concise 15-item True/False News Literacy Knowledge Scale. Treating cross-cultural adaptation as a validity procedure, and explicitly planning for measurement equivalence, supports fairer group comparisons and stronger evaluation of eHealth and news literacy interventions. The framework emphasizes transparent documentation, student-friendly decision points, and minimum reporting essentials.

Key words: *eHealth literacy; news literacy; public health education; cross-cultural adaptation; measurement invariance*

INTRODUCTION

In the last two decades, accelerated digital transformation has reshaped how individuals access information on health, current affairs and public life [1–3]. eHealth has been defined as an emerging field at the intersection of medical informatics and public health referring to health services and information delivered or enhanced through the internet and related technologies [4]. This shift underscores the growing importance of eHealth literacy in enabling individuals to effectively locate, evaluate and use online health information [1].

Online health information is routinely sought, consumed and distributed via websites, social media platforms, mobile applications and patient portals, while

[1]Department of Public Health, Epidemiology and Quality of Life, School of Medicine, University of Patras, Patras, Greece
[2]Division of Linguistics, Department of Philology, University of Patras, Patras, Greece

Received: 13 Feb 2026; Accepted: 22 May 2026

news on politics, science and health disseminated across increasingly fragmented and interactive media environments [5–7]. In this context, a broader set of competencies-beyond basic reading skills- is required to navigate digital information effectively [1,2]. These competencies are often conceptualized as eHealth literacy in the health domain and news literacy in the media domain [1,5].

With the evolution of digital technologies and online environments, the conceptualization of eHealth literacy has also evolved. Subsequent work has introduced the concepts of eHealth literacy 2.0 and 3.0 to capture interactive, participatory and data-driven aspects of digital health, encompassing social media use, mobile health applications, wearables and algorithmically curated content [2,8]. eHealth literacy was defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem" [2]. This definition is operationalized within the eHealth Literacy Lily Model, which proposes that eHealth literacy is a meta-literacy grounded on six foundational literacies: traditional, health, information, scientific, media and computer literacy [2]. To support measurement in research and practice, the eHealth Literacy Scale (eHEALS) and the revised version, the eHealth Literacy Scale-Revised (eHEALS-R) provide standardized self-report measures of perceived eHealth literacy. In brief, eHEALS assesses individuals' perceived skills and confidence in locating, evaluating, and applying online health information, while eHEALS-R extends this focus to better align with contemporary, interactive digital health contexts and support updated applications in research and practice [9].

In parallel, scholars in communication and media studies have focused on news literacy, a scale that captures individuals' understanding of how news is produced, distributed and consumed, and their ability to critically evaluate news content [5]. Contemporary frameworks conceptualize news literacy in terms of five interrelated domains—the "Five C's": context, creation, content, circulation and consumption—domains that collectively delineate the social, economic and technological conditions under which news is created and accessed [5]. Drawing on this framework, Maksl and colleagues [5] developed a concise, 15-item True/False News Literacy Knowledge Scale, derived from a larger item pool and refined through multiple studies using item response theory (IRT) and differential item functioning (DIF) analyses—approaches used to (a) identify items that most reliably discriminate between individuals with lower versus higher levels of knowledge and (b) assess whether any item performs differently across respondent subgroups despite comparable underlying ability. In broad terms, IRT focuses on how individual items function across levels of the underlying scale, supporting item-level refinement beyond simple total-score approaches. One widely used IRT approach is Rasch modeling, which helps ensure that a set of items forms a coherent scale and that total scores can be interpreted consistently across respondents and groups. DIF analyses, in turn, help evaluate whether items behave equivalently across groups (e.g., demographic subgroups or language versions), which is particularly relevant when instruments are translated and used in new cultural contexts. This combination strengthens the psychometric adequacy and cross-group comparability of the final scale, yielding a psychometrically robust, knowledge-based measure suitable for evaluating news literacy and the impact of educational interventions in diverse populations. In addition to providing a psychometric framework for validation, Maksl et al. [5] underscore item-design choices that facilitate cross-context adaptation (e.g., avoiding overly brand-or ownership-specific content and platform-specific wording) and recommend item-level analyses including IRT/Rasch and DIF to assess comparability after translation. For translated instruments, these item-level checks can help distinguish true differences in the construct from artifacts of wording, familiarity, or culture-specific references.

Both eHealth literacy and news literacy are particularly pertinent in the contemporary information environment, in which health-related misinformation, "fake news" and conspiracy narratives are prevalent. Repeated exposure may increase the perceived veracity of misinformation (the illusory truth effect) even when statements are false [10,11], and familiarity may bias judgments via mere exposure [11,12]. These familiarity-based effects are particularly relevant when instruments rely on objective true/false formats, where recognition can be mistaken for knowledge [11,12]. Individuals with higher eHealth literacy are more likely to navigate online health information effectively, engage in shared decision-making and appropriately use digital health tools, whereas those with higher news literacy are better equipped to distinguish credible news from misleading or deliberately manipulative content. As a result, robust

and culturally transferable measurement instruments are required to support cross-group comparability and evidence generation across cultural and linguistic settings [13–16].

Cross-cultural research has long emphasized that mere linguistic translation is insufficient when adapting instruments for new populations. Differences in language structure, media systems, health care organization and everyday practices may all affect how items are understood and how scales are expressed. Accordingly, several guidelines have been developed to support the translation and cultural adaptation of instruments in health and social research. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Translation and Cultural Adaptation provided a ten-step set of "Principles of Good Practice" for translating patient-reported outcome measures (PROM), measures emphasizing documentation, harmonization across language versions and cognitive debriefing with patients [13]. Alongside, Sousa and Rojjanasrirat proposed a clear seven-step guideline for translation, adaptation and validation of instruments for cross-cultural health care research, emphasizing forward and backward translation, expert committee review, pretesting and psychometric evaluation with particular attention to content validity and psychometric testing in the target context [14]. Complementing these process-focused recommendations, Maksl et al. highlight item-design choices that facilitate cross-context adaptation (e.g., avoiding brand, ownership or platform-specific content) and recommend item-level analyses-such as IRT and DIF- to evaluate comparability after translation [5]. A recent practical guideline for novice researchers reiterates that cross-cultural adaptation must address not only language but also cultural and measurement equivalence, and that documentation of decisions is essential for transparency and replicability [24].

To date, however, there has been limited discussion of how these frameworks can be combined and tailored for literacy-related instruments such as eHEALS-R and news literacy scales, particularly in settings such as Greece, where such measures are still emerging.

While we developed a protocol for the Greek adaptation and planned validation elsewhere, the present article provides a broad, educational framework for public health students and applied researchers. It links cross-cultural adaptation steps to a methodological framework for evidence generation and measurement equivalence, and highlights a key challenge: objective true/false knowledge scales may be biased by familiarity-related effects, representing a threat to validity beyond linguistic translation [10–12]. To appraise methodological quality, we applied the Scale for the Assessment of Narrative Review Articles (SANRA) tool to self-assess key quality domains for narrative reviews and enhance transparency and rigor [17].

MATERIALS AND METHODS

To develop the proposed framework, we conducted a purposive narrative review. We first anchored the framework in widely used guidance for translation and cultural adaptation and for evaluation of measurement properties. We then complemented these sources with targeted searches in PubMed and Scopus using combinations of terms related to cross-cultural adaptation/translation, measurement equivalence (including multigroup confirmatory factor analysis), and item-level approaches (e.g., DIF, IRT), followed by citation tracking to identify additional relevant publications [13–16, 18, 19].

RESULTS OF THE SYNTHESIS

Cross-cultural adaptation as a validity procedure

Cross-cultural adaptation can be understood as a sequence of validity-relevant decisions rather than a purely linguistic task [20]. The central goal is to achieve conceptual and functional equivalence: items should represent the same scale and elicit the same interpretation in the target language and cultural context. Translation and cultural adaptation should be treated as part of construct validation, aiming to preserve semantic, idiomatic, experiential, and conceptual equivalence across languages and contexts [13–15]. Frameworks commonly distinguish multiple forms of equivalence (e.g., conceptual, item, semantic/linguistic, operational, and measurement equivalence), which helps structure committee review and subsequent psychometric testing [20].

The proposed framework is based on a synthesis of three complementary methodological sources. First, the ISPOR Task Force "Principles of Good Practice" offers a more granular ten-step process for translation and cultural adaptation (TCA), i.e., the structured process of translating an instrument and adapting it to the target culture while preserving conceptual equivalence across language versions, supported by transparent documentation. These guidelines emphasize preparation, reconciliation, harmonization across language versions, cognitive debriefing, proofreading, and transparent documentation [13]. Second, Sousa and

Rojjanasrirat provide a clear, user-friendly seven-step guideline that links TCA to content validity assessment and psychometric testing in the target population [14]. Third, because the 15-item News Literacy Knowledge Scale is an objective, knowledge-based true/false scale, we incorporate instrument-specific constraints and validation logic described in Maksl et al. [5], including maintaining coverage of the "Five C's", avoiding ultra-contextual or platform-tied wording, and planning IRT/Rasch and DIF analyses to support comparability.

These sources overlap in core translation steps (forward translation, reconciliation, backward translation, expert review and pretesting), but they differ in emphasis. ISPOR [13] foregrounds standardized operational procedures and reporting, while Sousa and Rojjanasrirat [14] foreground content validity and psychometric evidence in the target context. Maksl et al. [5] add scale-specific requirements that are particularly relevant when adapting objective knowledge items (rather than self-report items), such as ensuring a single defensible correct answer per statement, preserving the intended balance and domain structure, and evaluating item-level behavior (difficulty, fit, DIF) after translation. A comparative overview and implications for the planned pre-test and cross-sectional study are presented in Table 1.

In practice, this third strand provides an instrument-anchored adaptation layer, preserving the five-domain structure and item-level knowledge claims while allowing context-specific adaptations.

By synthesizing overlapping and complementary elements, this approach aims to retain the user-friendly stepwise guidance of major translation frameworks while strengthening documentation and planning for cross-group comparability after translation [13, 14, 16, 18, 19]. Short case examples (eHEALS-R and a 15-item True/False News Literacy Knowledge Scale) highlight how the same logic can apply across different response formats [5, 8, 9, 21]. Maintaining a decision log —an item-level record of translation/adaptation choices and their rationale —strengthens transparency and auditability, supports harmonization across versions, and facilitates downstream psychometric evaluation, consistent with recommendations to systematically document translation and cultural-adaptation decisions throughout the process and in the final report [13–15]. A Greek applied example that illustrates that adaptation extends beyond translation is the cultural adaptation and pilot testing of a Health Literacy Universal Precautions Toolkit for primary care professionals working with older adults. In this case, extensive context-specific modifications were combined with structured training and follow-up implementation, resulting in measurable improvement in communication-related self-efficacy and retained gains over time [22]. A complementary Greek applied example from the digital-caregiving domain is the cultural adaptation and piloting of the WHO iSupport Dementia eLearning platform, where stakeholder feedback during piloting highlighted usability, accessibility and interaction needs (e.g., downloadable content, reduced screen-reading burden, videos/forum support) before wider rollout [23].

A practical workflow for cross-cultural adaptation and validation commonly includes the following elements:

Step 1 – Preparation and permissions

To obtain written permission to translate/adapt the instruments; assemble the translation team and expert committee (including content and measurement specialists); define the target population(s) and mode of administration; prepare a glossary of key terms (e.g., online health resources, algorithms, sponsored content) and a decision log/template to document all translation choices and their rationales [13,14]. When ambiguities arise, consultation with the original instrument developers is recommended; all feedback should be documented in the decision log.

Step 2 – Forward translation

To produce at least two independent forward translations by bilingual translators whose native language is the target language [13,14]. For true/false knowledge items, prioritize renderings that preserve the intended truth value and avoid overly local brand- or platform-specific wording unless essential [5]. WHO guidance additionally recommends that forward translators have the target language as their mother tongue, excellent command of the source language, and familiarity with health/disability concepts; translators should explicitly flag problematic terms/phrases for subsequent linguistic evaluation [24].

Step 3 – Reconciliation

Reconcile the forward translations into a single synthesized version (v1) through structured discussion, documenting alternatives and decisions [13,14]. Verify that each true/false statement remains unambiguous, evidence-consistent, and defensible as true or false in the target language context [5].

Table 1. *Comparison of three complementary methodological sources informing the proposed TCA.*

Topic	ISPOR Task Force (Wild et al.) [13]	Sousa & Rojjanasrirat [14]	Maksl et al. (15-item true/false news literacy knowledge) [5]	Student-oriented public health takeaway
Primary focus	Good-practice workflow for translation and cultural adaptation of patient-reported outcome measures (often used more broadly).	User-friendly stepwise guideline for translation/cultural adaptation followed by psychometric testing in the target population.	Instrument-anchored adaptation/validation constraints for a true/false knowledge test (single correct answer, domain coverage).	Choose a translation workflow, but also plan how evidence will be generated and reported; knowledge tests need extra rules beyond language.
Step structure	Multi-step process: preparation, dual forward translation, reconciliation back-translation, review, cognitive debriefing, finalization, and documentation.	Similar core steps with emphasis on expert committee review, pretesting/cognitive interviews, and subsequent validation.	Not a generic translation framework; emphasizes preserving item meaning and answer key, and evaluating item performance after adaptation.	Use a clear step-by-step flow and produce concrete outputs (versions, decision log, debriefing summary, final report).
Conceptual equivalence	Emphasizes conceptual equivalence through iterative translations, expert review, and documentation.	Highlights 'centering' (symmetrical translation) to optimize conceptual and technical accuracy, not literal word-for-word translation.	Emphasizes that adaptation must not change what counts as correct/incorrect; preserve the construct map (e.g., "Five C's").	Treat adaptation as a validity procedure: preserve meaning first, then test whether items function equivalently across groups.
Roles and actors	Defines project roles (project manager, translators, reconciler, debriefing interviewer, proofreader).	Focuses on translators (informed and uninformed), expert committee, and target-language participants for debriefing.	Requires content experts to verify the answer key and domain mapping; benefits from psychometrics expertise for item-level analyses.	Plan a small but diverse team: translators + domain expert(s) + measurement/psychometrics support, even in student projects.
Cognitive debriefing / pretesting	Core step to evaluate comprehension and cultural relevance in target-language users.	Explicitly recommends pretesting with the target population and using findings to revise items.	Useful but item reuse can create familiarity bias in knowledge tests; consider parallel forms or spacing.	Do debriefing, but avoid training participants on knowledge items; document what changed and why.
Documentation	Stresses transparent reporting and a final translation report.	Encourages clear reporting of each step and psychometric outcomes.	Emphasizes reporting that supports interpretability of item-level results and domain coverage.	Make your work replicable: keep a decision log, version history, and a summary table of changes.
Evidence after translation	Not always detailed; assumes evaluation is needed for intended use.	Explicitly links adaptation to subsequent psychometric validation.	Stresses item-level evaluation and potential DIF/IRT analyses for fairness.	Do not stop at translation: plan structural validity, reliability, and equivalence testing (invariance/DIF) before comparing groups.
Special risks	General translation risks: ambiguity, poor cultural fit, inconsistent terminology.	Same general risks with emphasis on target-user comprehension.	Additional risks: guessing, truth-value drift, and familiarity/illusory truth effects with repeated exposure.	Knowledge tests need safeguards (truth verification, 'I don't know' options if appropriate, avoid repeated exposure, and item diagnostics).

Abbreviations: TCA, translation and cultural adaptation; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; DIF, differential item functioning; IRT, item response theory.

Step 4 – Back-translation

Have an independent translator back-translate v1 into the source language (the instrument's original language), who is a native speaker of the source language and blinded to the original version [13]. Use the back-translation to flag potential conceptual drift, particularly for technical, media-system, and platform-related terms [13,14].

Step 5 – Back-translation review

Compare the back-translation with the source version to identify discrepancies in meaning, tone, and assumed context, and revise v1 accordingly [13]. For the knowledge items, confirm that revisions do not alter the intended truth value, and that each item still supports a single correct answer consistent with the original intent [5].

Step 6 – Harmonization and expert committee review

A multidisciplinary expert committee reviews all items for semantic, idiomatic, experiential, and conceptual equivalence, resolving any remaining issues and confirming cultural fit [13,14]. Additional considerations for true/false knowledge items include: (a) preserving domain coverage, (b) retaining the intended mix of true and false statements, and (c) ensuring the response options (including 'Don't know', if used) function as intended [5]. WHO further describes a structured linguistic evaluation step (bilingual panel, including an editor-in-chief) supported by a standardized linguistic evaluation data sheet to resolve highlighted terms/phrases and document decisions [24]. This step maps closely onto the harmonization/committee phases emphasized in an 8-step adaptation guideline (forward translation, synthesis, back translation, harmonization, pre-testing, field testing, psychometric validation, and analysis of psychometric properties) [20].

Step 7 – Cognitive debriefing / pretesting

Cognitive debriefing is also a key safeguard against cross-cultural bias (method, content, and construct bias) and against systematic response-style differences that can distort comparability (e.g., acquiescence or extreme responding), before proceeding to larger-scale field testing [20]. Conduct cognitive debriefing and, if feasible, a small pilot with members of the target population (ISPOR suggests a small sample, typically 5-8 respondents, for the pre-test) to evaluate comprehension, retrieval, judgment, and response processes, consistent with ISPOR recommendations [13]. In line with Sousa & Rojjanasrirat, consider a pilot of approximately 10–40 participants and structured expert feedback to support content validity (e.g., clarity and relevance) [14]. For true/false knowledge items, verify that respondents interpret each claim as intended and do not encounter ambiguity that could shift the correct answer [5]. For pre-testing, WHO suggests a minimum of 10 respondents per section, with representation across sex and age groups [24].

Step 8 – Review of debriefing results and finalization

Review cognitive debriefing/pilot findings, implement revisions, and re-check conceptual equivalence across items and instructions [13, 14]. Confirm that any changes preserve the intended construct coverage for each instrument and the domain mapping for true/false knowledge items [5].

Step 9 – Proofreading

Proofread the near-final version for language, consistency, formatting, and administration instructions; check numbering, response anchors, and layout for unintended cues, particularly in true/false items [13].

Step 10 – Final report and planned psychometric evaluation

Recommended reporting for the validation phase commonly includes evidence for structural validity, internal consistency, reliability, measurement error, responsiveness, and where relevant measurement invariance/DIF and floor/ceiling effects [24]. Produce the final TCA report and archive all intermediate versions, decisions, and rationales to ensure transparency and reproducibility [13]. Following translation, conduct psychometric evaluation in appropriately powered target language-speaking samples as recommended for cross-cultural instrument use [14]. Planned analyses include reliability and validity testing for each instrument; for the true/false knowledge items, assess item difficulty and fit (e.g., IRT/Rasch where appropriate), differential item functioning across key demographic/educational groups, and convergent/criterion validity via correlations with related constructs [5].

DISCUSSION

Linking adaptation to measurement evidence and equivalence

After an adapted version is finalized, evidence generation should follow a structured plan. COSMIN

methodology frames measurement as a set of properties that must be supported with appropriate analyses and transparent reporting, including content validity, structural validity, internal consistency, reliability, and cross-cultural validity [16,18,19]. Integrating cross-cultural adaptation and COSMIN-based measurement evidence to support the TCA of the instruments, we followed established cross-cultural adaptation procedures—dual forward translation, synthesis, back-translation, expert committee review, and pretesting with cognitive debriefing—to maximize semantic, idiomatic, experiential, and conceptual equivalence between the source and target versions [13,15]. The guideline by Sousa and Rojjanasrirat further links translation decisions to subsequent validation, underscoring that translation is necessary but not sufficient for cross-language comparability [14]. After translation, the adapted versions should be evaluated in line with COSMIN guidance for measurement properties, with particular attention to structural validity, reliability, and cross-cultural validity/measurement invariance [16,18,19]. Cross-cultural validity refers to whether items function equivalently across language versions; this can be examined using multi-group confirmatory factor analysis and/or differential item functioning DIF analyses in IRT/Rasch or regression frameworks, and interpreted using COSMIN criteria for what constitutes important differences [16]. For latent-variable self-report instruments, multi-group confirmatory factor analysis can assess whether the measurement structure holds across groups (e.g., configural, metric, and scalar invariance). For both self-report and knowledge tests, differential item functioning analyses, often within item response theory (IRT)/Rasch frameworks, can identify items that behave differently across groups after controlling for overall trait level. Importantly, adaptation decisions and equivalence testing should inform each other. If cognitive debriefing identifies a term interpreted differently across subgroups, that item becomes a candidate for careful monitoring in subsequent analyses. Conversely, an item that shows substantial differential item functioning (DIF) requires content review and, depending on the purpose, may need rewording, additional clarification, or removal, balanced against domain coverage and interpretability.

In addition to these general equivalence considerations, objective true/false knowledge scales introduce extra adaptation constraints related to truth value, guessing, and familiarity effects.

The special case of true/false knowledge tests: truth value, guessing, and familiarity

Objective true/false knowledge scales impose additional adaptation constraints: each item must preserve its intended truth value in the target context, and correct responses may also occur by guessing [5]. Accordingly, adaptation extends beyond wording to verifying that the underlying factual claim, reference frame, and domain mapping remain stable and meaningful. Knowledge measures are particularly sensitive to context because item statements may embed time-bound, culture-specific, or system-specific assumptions; translation should therefore preserve each item's intended truth value while keeping the statement natural and interpretable in the target language [5].

Beyond translation, familiarity-related effects constitute a distinct threat to validity: repeated exposure can increase processing fluency and perceived accuracy (the illusory truth effect), even when statements are false [10,11], and mere exposure can increase acceptance independently of content quality [12]. To mitigate familiarity-driven bias in cognitive debriefing, piloting, and repeated administrations, we recommend avoiding reuse of the same items between pretesting and the main psychometric study where feasible (or using alternate forms/adequate spacing when repetition is unavoidable), randomizing and counterbalancing item order and condition assignment, and measuring perceived familiarity for analytic adjustment (e.g., covariate or sensitivity checks), to ensure observed differences reflect the intended construct rather than prior exposure [10–12].

A concise reporting checklist: decision points and minimum essentials

For student projects and applied public health studies, it is useful to separate (i) translation outputs, (ii) evidence to be generated, and (iii) documentation for transparent reporting. The aim is to make adaptation decisions traceable and reproducible.

Accordingly, we emphasize documentation of (a) how key terms were negotiated across translators and reviewers, and (b) how the resulting version was checked with the target users to ensure interpretability in the language and cultural context.

Greek applied evidence supports extending TCA reporting beyond linguistic equivalence to include an "implementation layer", documenting context-driven adaptations and whether training-related skills were

sustained in practice [22]. Similarly, the Greek pilot of WHO iSupport Dementia indicates that user-centered adaptation must also address usability and accessibility constraints (e.g., navigation, screen-reading fatigue, and interactive features) to enable successful rollout [23].

The combination of forward translation by both informed and lay translators, back translation and expert committee review offers multiple opportunities to detect conceptual drift and identify areas where simple linguistic translation may not be adequate. Cognitive debriefing with target users is especially important for these instruments, as it can reveal how respondents interpret complex or technical terms, and whether examples drawn from other media systems are meaningful in the language and cultural context.

Minimum reporting essentials:

- Describe permissions, translators' backgrounds, and how reconciliation decisions were made (including a decision log).
- Report cognitive debriefing methods (sample characteristics, interview approach, and the most common issues identified).
- Provide a clear description of the final version and any culturally adapted items (what changed and why).
- Report evidence for structural validity, reliability, and cross-cultural validity (invariance/DIF) before interpreting group differences.

Table 2 summarizes key decision points and common pitfalls, showing how to report these minimum essentials and how to document planned or completed evidence for key measurement properties and cross-group comparability [13–16,18,19].

LIMITATIONS

This narrative review proposes a methodological framework and does not present empirical data; accordingly, the feasibility, acceptability, and measurement performance of the adapted target-language versions should be established in appropriately powered samples. The workflow synthesizes widely used guidance, yet implementation quality may vary across settings and teams, and the framework cannot eliminate all sources of context sensitivity inherent to literacy measures (e.g., time-bound media practices or platform-specific exposure). In addition, the review is not a systematic evidence synthesis and may not capture all relevant methodological discussions; readers should interpret recommendations as a structured, practice-oriented roadmap rather than prescriptive rules.

CONCLUSIONS

By integrating process-oriented translation and cultural adaptation steps with measurement-focused planning for equivalence and item-level diagnostics, the proposed framework provides a transparent and reproducible workflow for adapting literacy-related

Table 2. *Key decision points and common pitfalls for student projects.*

Stage	Minimum deliverable	Common pitfall to avoid
Item content	Truth-value check with justification for the correct answer (evidence/source) and a brief item rationale.	Literal translation that changes the factual claim or makes it time- or culture-bound.
Translation team	Named review group with content and measurement expertise; documented decisions (who reviewed, when, outcomes); Developer feedback log (queries-responses-decisions)	Over-reliance on bilingual translators without construct/measurement expertise; no developer consultation -ambiguous items unsolved.
Pre-testing	Cognitive debriefing report (sample, interview guide, top issues, revisions made + rationale).	No interview guide / no record of changes → pre-test becomes "proofreading" and cannot justify revisions.
Equivalence	Analysis plan + results for measurement invariance (MG-CFA) and/or DIF/IRT with decision rules; or explicit limitation if not feasible.	Reporting between-group comparisons without an equivalence plan (MG-CFA/DIF) or an explicit limitation statement.
Familiarity bias	If reuse is unavoidable: spacing/alternate forms + measure perceived familiarity; run a sensitivity check.	Inflated scores due to prior exposure rather than learning gains.

Note: Table 2 translates the comparative synthesis in Table 1 into minimum deliverables and common pitfalls for student projects.

Abbreviations: MG-CFA, multi-group confirmatory factor analysis; CFA, confirmatory factor analysis; DIF, differential item functioning; IRT, item response theory; TCA, translation and cultural adaptation.

instruments in the target-language context. Applied to eHEALS-R and the 15-item True/False News Literacy Knowledge Scale, the approach emphasizes multidisciplinary collaboration, explicit documentation of translation decisions, and analytic safeguards to strengthen cross-group interpretability. Establishing psychometrically sound target-language versions of these instruments can enable more rigorous evaluation of digital health and media literacy interventions and support meaningful comparisons across populations and settings.

Conflict of interest: *The authors declare no conflicts of interest.*

Declaration of funding sources: *No funding has been received to date.*

Author Contributions: *Maria Roumelioti and Eleni Jelastopulu contributed to the conception and design of the review; MR drafted the article; George J. Xydopoulos, Georgia Mikrouli, and Emily Diamantopoulou contributed to the theoretical framework, with emphasis on translation and the linguistic and cultural adaptation of measurement scales. EJ contributed to the synthesis and interpretation of the findings. All authors substantially revised and approved the final version of the article.*

REFERENCES

1. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res. 2006;8(2):e9.
2. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res. 2011;13(4):e125.
3. Milanti A, Norman C, Chan DNS, So WKW, Skinner H. eHealth Literacy 3.0: Updating the Norman and Skinner 2006 Model. J Med Internet Res 2025;27:e70112.
4. Eysenbach G. What is e-health? J Med Internet Res. 2001;3(2):e20.
5. Maksl A, Boedeker PJ, Vraga EK, Craft S, Tully M, Ashley S. Developing and Validating a 15-Item True/False Measure of News Literacy Knowledge. Commun Res 2024.
6. Fletcher R, Nielsen RK. Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. J Commun. 2017;67(4):476–98.
7. Chung DS. Interactive Features of Online Newspapers: Identifying Patterns and Predicting Use of Engaged Readers. J Comput Mediat Commun. 2008;13(3):658–79.
8. Efthymiou A, Kalaitzaki A, Rovithis M, Petrič G. Validation of the eHealth literacy scales: comparison between the shorter and longer versions. Inform Health Soc Care. 2025;50(1):33–47.
9. Norman CD, Skinner HA. eHEALS: The eHealth literacy scale. J Med Internet Res. 2006;8(4):e27.
10. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol. 1968;9(2Pt2):1-27.
11. Fazio LK, Brashier NM, Payne BK, Marsh EJ. Knowledge does not protect against illusory truth. J Exp Psychol Gen. 2015;144(5):993–1002.
12. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. Nature. 2021;592(7855):590–5.
13. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94–104.
14. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract. 2011;17(2):268–74.
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures: Spine. 2000;25(24):3186–91.
16. Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res. 2024;33(11):2929–39.
17. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4(1):5.
18. 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;27(5):1147–57.
19. 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;27(5):1171–9.
20. Cruchinho P, López-Franco MD, Capelas ML, Almeida S, Bennett PM, Miranda da Silva M, Teixeira G, Nunes E, Lucas P, Gaspar F; Handovers4SafeCare. Translation, Cross-Cultural Adaptation, and Validation of Measurement Instruments: A Practical Guideline for Novice Researchers. J Multidiscip Healthc. 2024;17:2701–28.
21. Efthymiou A, Middleton N, Charalambous A, Papastavrou E. Adapting the eHealth Literacy Scale for Carers of People With Chronic Diseases (eHeals-Carer) in a Sample of Greek and Cypriot Carers of People With Dementia: Reliability and Validation Study. J Med Internet Res. 2019;21(11):e12504.
22. Efthymiou A, Kalaitzaki A, Rovithis M. Cultural adaptation of a health literacy toolkit for healthcare professionals working in the primary care setting with older adults. Healthcare (Basel). 2023;11(5):776.

23. Efthymiou A, Karpathiou N, Dimakopoulou E, Zoi P, Karagianni C, Lavdas M, et al. Cultural Adaptation and Piloting of iSupport Dementia in Greece. Stud Health Technol Inform. 2022;289:184–7.
24. World Health Organization. Process of translation and adaptation of instruments. Geneva: World Health Organization;2016.

Corresponding author:

Maria Roumelioti
Department of Public Health, Epidemiology and Quality of Life, School of Medicine, University of Patras, Patras, Greece,
Tel.: +30 2610969873, +30 6981044775,
E-mail: mroum@upatras.gr