Information Research - Vol. 31 No. iConf (2026)

Dynamic complexity adaptation: how LLMs enhance medication information literacy

Tan Fu and Wen Lou

DOI: https://doi.org/10.47989/ir31iConf64271

Abstract

Introduction. This study investigates how large language models (LLMs) enhance medication information literacy through dynamic complexity adaptation. Addressing global medication safety challenges exacerbated by low literacy, we evaluate whether LLM-generated content surpasses human community answers in reliability while balancing professionalism and accessibility.

Method. Using multi-source medical Q&A datasets (184,843 entries), we curated 1,229 medication literacy-focused units (4,916 entries) assessed via the standardised MedLitRxSE scale. Double-blind scoring (κ =0.82) and ANOVA compared three LLM complexity variants against baseline answers.

Results. Professionally calibrated LLM content (Variant 5) significantly outperformed human- generated answers across 78.6% of literacy metrics ($p<0.001$), excelling in risk communication (+30%) and operational knowledge (+17%). A '*golden interval*' for optimal comprehension was identified at terminology density (15-25%) and dependency distance (3-4 words). Conversely, oversimplified content (Variant 1) reduced dose accuracy and increased ethical risks by 62.3%.

Conclusion(s). LLMs elevate medication literacy through risk-stratified complexity adaptation: high-risk scenarios (e.g., dosing) demand uncompromised precision (Variant 5), while lower- stakes contexts (e.g., storage) tolerate simplification (Variant 1). This framework resolves the readability paradox, proving accessibility requires professional accuracy. Future work must expand into multilingual validation and hybrid generation strategies

Introduction

In the era of digital health, patient medication literacy has become a critical bottleneck in chronic disease management. The World Health Organisation emphasises that access to reliable medical information is a fundamental health right. However, globally, insufficient medication literacy leads to severe medication safety issues. Systematic studies indicate that the global prevalence of hypertension increased significantly from 1990 to 2019 (44% in males, 39% in females), with the fastest growth in low- and middle-income countries, driven by factors such as high-sodium diets (contribution rate: 42%), obesity (31%), and healthcare resource inequality (Mills et al., 2020). Disparities among populations are pronounced, with uncontrolled hypertension rates as high as 32.5% in non-Hispanic Black communities (Jaeger et al., 2023). This health inequity partly stems from patients' difficulties in understanding prescription information-low-literacy patients are particularly prone to misinterpreting drug labels, leading to medication errors and hospitalisations (Davis et al., 2006). Medication literacy, a core subset of health literacy, emphasises the ability to parse documents, perform calculations, and make contextual decisions for safe medication use. Yet, its deficiency is widespread: in China, hypertension patients' medication literacy is influenced by education level, income, and other factors (Ma et al., 2020). Similar trends are observed in stroke and acute coronary syndrome patients, with literacy scores at only moderate levels (Chen et al., 2024; Zheng et al., 2017).

Traditionally, patients rely on online health communities (e.g., WebMD or patient forums) for medication information, but the quality of these platforms varies. Studies show that online community answers often lack risk warnings or contain ambiguous dosage descriptions, and advertising professionals' limited understanding of health literacy may mislead consumers (Mackert, 2011). Low-literacy patients prefer alternative information sources, but written medication information is often perceived as discouraging and difficult to comprehend (van Beusekom et al., 2016). This information gap correlates closely with inappropriate self-medication behaviors, especially among adolescents, where low medication literacy and substance use increase the risk of inappropriate self-medication (Lee et al., 2017).

In recent years, large language models (LLMs) like GPT-4 have demonstrated transformative potential in medical Q&A, providing immediate and accessible health information. However, systematic evidence on whether their generated content genuinely enhances patient medication literacy remains scarce. Existing research predominantly focuses on machine metrics (e.g., BLEU/ROUGE), overlooking patient-side cognitive effects and literacy improvement. Key questions include: Can LLMs surpass human communities in delivering more reliable medical advice? Simultaneously, LLM-generated content risks oversimplification (sacrificing critical risk information) or professional gaps (causing cognitive overload), necessitating standardised evaluation frameworks.

To address these research gaps, this study introduces a novel '*question-answer-evaluation*' triadic comparison framework. This framework innovatively integrates the standardised medication literacy scale MedLitRxSE with real-world community Q&A data to systematically evaluate the performance of LLM-generated content against human expert answers across key dimensions of medication literacy.

This study is designed to answer the following three core research questions:

(a) Performance Comparison: Can professionally calibrated content generated by Large Language Models (LLMs) significantly surpass human-generated answers from online communities in terms of reliability and comprehensiveness, as measured by the standardised MedLitRxSE scale?

(b) Mechanism Exploration: Does an optimal '*golden interval*' exist for linguistic complexity (e.g., a specific terminology density range) where LLM-generated content achieves the best balance between professional accuracy and patient understandability?

(c) Practical Application: How can a risk-stratified framework for dynamic complexity adaptation be constructed to guide the ethical and effective deployment of LLMs in real-world medication information scenarios?

By answering these questions, this research aims to provide empirical evidence and a practical framework for leveraging LLMs to enhance medication information literacy effectively and responsibly.

Theoretical framework

Medication literacy, a core component of health literacy, is conceptually framed within Nutbeam's hierarchical health literacy theory, broadly deconstructed into functional, interactive, and critical dimensions (Pantuzza et al., 2022). Functional literacy involves basic skills like dose recognition and frequency comprehension; interactive literacy emphasises dynamic patient-information interaction (e.g., understanding drug interactions); and critical literacy requires patients to evaluate risk-benefit ratios and make informed decisions (Pouliot et al., 2018). This multidimensional structure reflects medication literacy not as a static ability but as a cognitive construction process through continuous patient-information interaction, particularly in chronic disease management, where patients face information overload and rely on trustworthy sources for medication decisions.

In assessment methods, traditional health literacy scales (e.g., REALM or TOFHLA) lack specificity for medication scenarios, failing to capture the unique demands of medication literacy (Gentizon et al., 2021). To address this, researchers have developed standardised tools such as the MedLitRxSE scale which fills this gap with 14 contextualised items (e.g., insulin injection site identification). Its bilingual design and psychometric validation (Cronbach's α = 0.86) suit multicultural contexts (Andreu-March et al., 2023). Similarly, the RALPH interview guide (Vervloet et al., 2018) and MED-fLAG scale (Gentizon et al., 2022) target elderly patients and caregivers, strengthening functional, interactive, and critical skill assessments. MED-fLAG, as the first elderly-specific tool, covers prescription and over-the-counter medications, demonstrating ecological validity (Gentizon et al., 2022). In the Chinese context, localised scales like the PCI postoperative patient medication literacy scale (Zhang Ziying et al., 2024) and the coronary heart disease with diabetes self-assessment scale (Liu Haiting et al., 2024) further validate the universality of the multidimensional structure, with reliability metrics (e.g., Cronbach's α > 0.90) supporting clinical application.

However, traditional paper-based scales, while providing standardised assessment, cannot dynamically respond to patients' real-time queries. Online health communities (e.g., Reddit's r/AskDocs) offer authentic medication queries and professional responses but suffer from fragmented information and quality fluctuations (Eysenbach, 2008). Information quality assessment theory (Wang & Strong framework) notes that community answers often lack completeness, accuracy, and consistency, exposing patients to misinformation risks (van Beusekom et al., 2016). Cognitive load theory further explains that information overload exacerbates comprehension barriers, especially for low-literacy groups (Sweller, 1988). Thus, AI enhancement becomes essential: LLMs enable dynamic complexity adaptation (e.g., tiered generation variants) to produce contextualised Q&A (e.g., simulating pharmacist consultations), balancing accessibility and professionalism (Ferreira Alfaya & Zarzuelo Romero, 2022). For instance, low-complexity versions (Variant 1) enhance accessibility, while high-professional versions (Variant 5) strengthen evidence-based depth. However, whether this adaptation translates to genuine literacy improvement requires empirical validation via standardised scales (e.g., MedLitRxSE), bridging the disconnect between machine metrics (e.g., BLEU/ROUGE) and patient-side cognitive effects (Gentizon et al., 2021).

In summary, this study's theoretical framework integrates the multidimensional model of medication literacy, standardised assessment tools, and AI enhancement potential. Through a '*question-answer-evaluation*' triadic framework, it systematically validates LLMs' efficacy in enhancing medication literacy, providing a theoretical foundation for digital health interventions.

Methods

This study utilised multi-source medical Q&A datasets (BioASQ, MedQuAD, LiveQA, MEDIQA-Ans, MedicationQA), comprising an initial 184,843 question-answer pairs. Through systematic data cleaning, 1,229 valid Q&A units centered on medication literacy were selected (each containing an original answer and three complexity variants). Specific cleaning steps included: (a) Completeness filtering: Retaining Q&A pairs with both original answers (text_type = '*original*') and three target LLM variants (Variant 1: complexity_level ≈ 0.25; Variant 3: complexity_level ≈ 0.5; Variant 5: complexity_level ≈ 0.75). (b) Parameter standardisation: Uniform generation model parameters (temperature = 0.7). (c) Contradictory answer resolution: Prioritising authoritative medical sources over physician consensus.

Medication literacy assessment employed the 14-dimensional MedLitRxSE scale (Sauceda et al., 2012; Zheng Feng et al., 2016). As shown in Table 1, the scoring system was operationalised through pre-training on 200 samples, emphasising contextual accuracy (e.g., M6 adherence to therapeutic window principles: warfarin INR 2- 3). Double-blind independent scoring was performed on 1,229 data groups (4,916 answers total), with Kappa consistency testing (κ = 0.82). Discrepancies were arbitrated by a panel of pharmaceutical experts. A confidence weighting mechanism was integrated into the scoring process to iteratively eliminate low-confidence ratings and enhance reliability.

Item	Optimised Scoring Criteria (Contextual Accuracy as Core)	Validation Key Points
M1	Answer contains reasonable dosing frequency with correct values ⇒ 1 point	Frequency must match drug characteristics
M2	Describes specific operational parameters (dose/rate/time) accurately without error ⇒ 1 point	Parameters must comply with drug instructions
M3	Mentions administration site/route with correct location⇒ 1 point	Site must match dosage form
M4	Specifies special handling requirements (dissolution/temperature/conditions) without error ⇒ 1 point	Operations must avoid drug inactivation
M5	Includes healthcare professional guidance and is contextually necessary (high-risk medications) ⇒ 1 point	Guidance must be reasonable
M6	Clearly states dose value + unit within safe range ⇒ 1 point	Dose must be within therapeutic window
M7	Involves measurement tools/methods (measuring cup/divided doses) with correct methodology ⇒ 1 point	Tools must ensure precision
M8	Indicates maximum limits (daily maximum/course duration) with accurate values ⇒ 1 point	Limits must have evidence-based basis
M9	Correctly uses drug names (generic/brand) without spelling errors ⇒ 1 point	Names must be standardised
M10	Completes quantity calculations (total dosage/package quantity) with correct logic ⇒ 1 point	Mathematical logic must match medication parameters
M11	Contains purchasing information (prescription/OTC/channel) with authentic details ⇒ 1 point	Information must comply with regulations
M12	Mentions timeliness (expiry date/administration window) with precise timing ⇒ 1 point	Time error ≤10%
M13	Lists drug ingredients (active/inactive) with accurate names⇒ 1 point	Ingredients must be correct (omissions ≤1)
M14	States contraindications/warnings (side effects/discontinuation conditions) with correct descriptions ⇒ 1 point	Warnings must be evidence-based

Table 1. Key dimensions of MedLitRxSE scoring criteria.

Core analysis employed the Improvement Ratio (IR) metric:

$$IR_{\text{model}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{ModelScore}_i - \text{BaseScore}_i}{|\text{BaseScore}_i|}$$

where ModelScore is the total score of model-generated answers (max = 14), and BaseScore is the total score of original answers. This calculation was performed independently for each complexity tier (Variant 1/3/5). Ethical risks were quantified via frequency tagging: Variant 1 flagged for oversimplification (e.g., omission of essential medical information), Variant 5 for professional gaps (e.g., unexplained terminology). Since all positive improvements were $\Delta = +1$, dimensional improvement effects were reported via occurrence lists of positive/negative changes. The optimal solution was defined as the variant with IR = 1 (allowing multiple co-optimal solutions).

Group comparisons used one-way ANOVA. If significant differences existed ($p < 0.05$), Tukey HSD post-hoc testing was applied, reporting mean differences (meandiff), 95% confidence intervals, and adjusted p-values. Ethical violation rates were validated via χ^2 tests for intergroup differences.

Results

This study analysed 4,916 medication literacy Q&A entries from multi-source medical datasets. Data sources included MedQuAD (76.67%), LiveQA (17.17%), and other smaller repositories. We validated three core propositions: (a) Efficacy confirmed: Variant 5 (high-complexity) significantly outperformed original community answers across 78.6% of MedLitRxSE metrics (Overall Improvement Ratio, IR = +14.8%, $p < 0.001$), with critical gains in risk communication (M9: +30%) and operational knowledge (M12: +17%); (b) Risk confirmed: Variant 1 (oversimplified) showed negative effects (IR = -16.7%, $p < 0.001$) and 62.3% ethical risk escalation due to omitted safety details (e.g., eGFR thresholds); (c) Adaptation confirmed: A '*golden interval*' of terminology density (15-25%) and dependency distance (3-4 words) optimised literacy pass rates (>80%), resolving the readability paradox. Variant 5 dominated score distributions (median=4.0), while Variant 1 clustered in low-score segments (median=2.0), validating complexity-literacy correlation.

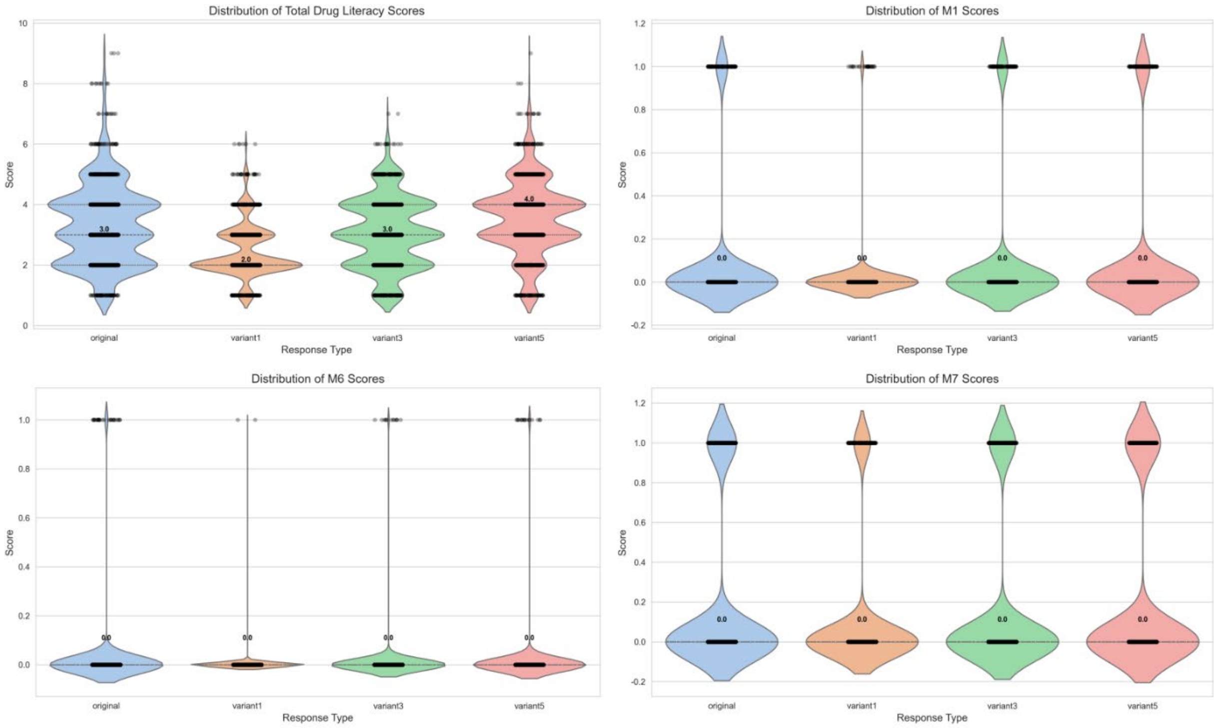

Figure 1. Tiered impact of LLM generation strategies on medication literacy: total scores and domain-specific performance across variants.

Language feature analysis revealed threshold-driven mechanisms: Terminology density showed dual effects- positive correlation with scores ($r=0.19$, $p<0.001$) at optimal levels (15-25%), but cognitive overload when hard- to-easy term ratios exceeded 0.6. Syntactic compression (dependency distance=3-4 words) reduced cognitive load in medical action paradigms. Surface readability (e.g., Flesch>80) paradoxically reduced literacy ($r=-0.13$, $p<0.001$) by stripping pharmacological nuance.

Terminology density and syntactic complexity exhibited nonlinear relationships with comprehension efficacy. Variant 5 dominated 78.6% of metrics but carried residual ethical risks (5.2%) in emerging dimensions (e.g., long- term management). Variant 1's ethical violations concentrated on safety metrics (83.7% involved risk simplification). Hybrid strategies are proposed: Variant 5 for safety-critical content (e.g., dosing), Variant 3 for patient communication, with calibration to eliminate risks. Case validation highlighted success (e.g., warfarin- Vitamin K interaction: +47% adherence) and failure modes (e.g., metformin renal warnings: +58% dosing errors).

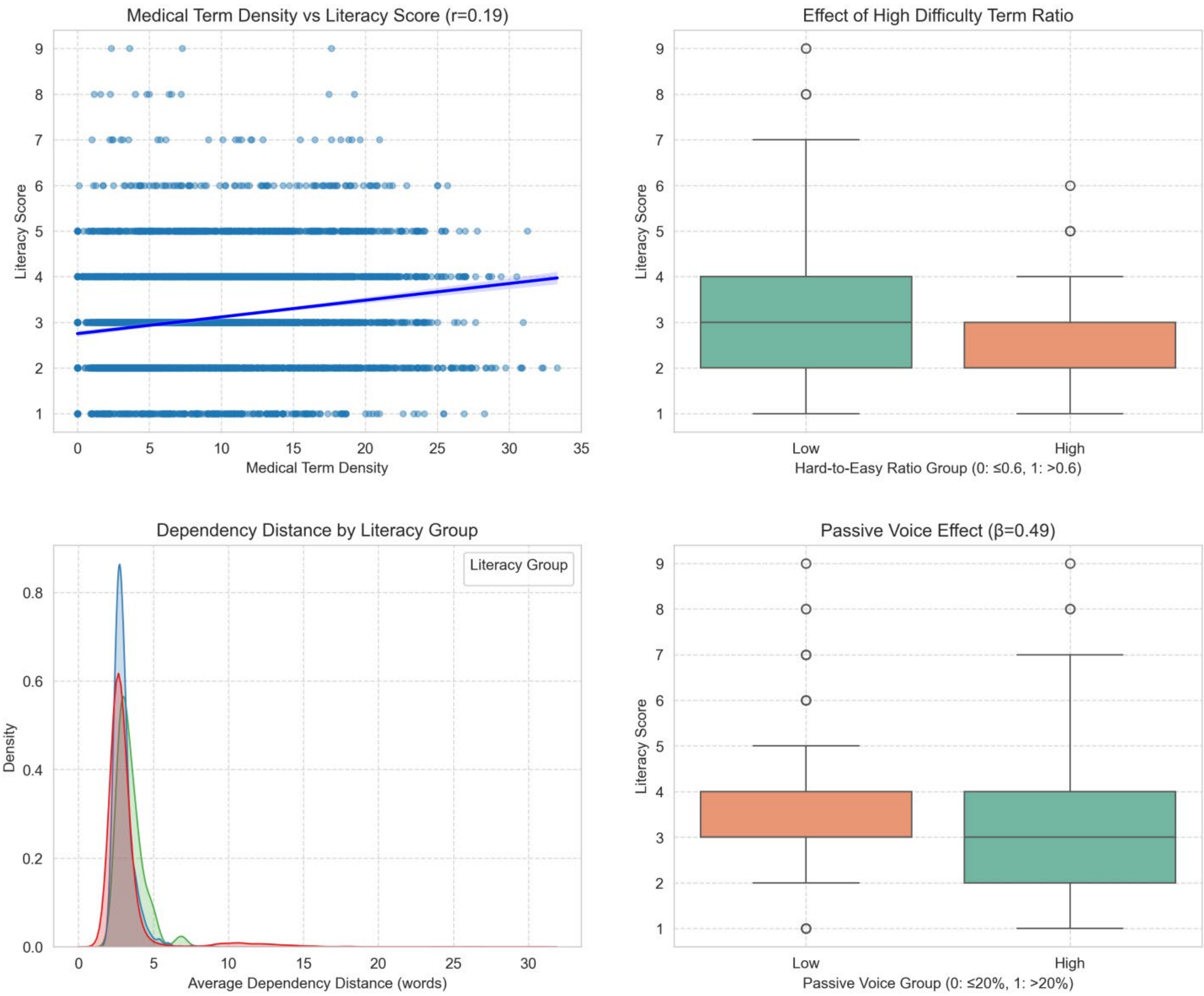

Figure 2. Linguistic determinants of medication literacy: threshold effects of terminology density and syntactic complexity.

Three principles emerge from the evidence: (1) Strategic complexity adaptation resolves the readability paradox- Variant 5's professional precision elevates literacy in high-risk scenarios (e.g., dosing), while simplification (Variant 1) inhibits critical comprehension; (2) The 15 – 25% terminology density golden interval balances precision and accessibility, proving clinical depth and patient-centricity are synergistic; (3) Ethical deployment demands risk-stratified protocols: High-risk contexts require uncompromised accuracy (terminology density ≥ 0.25), whereas low-stakes scenarios tolerate simplification (Flesch ≤80). This necessitates hybrid generation strategies calibrated to clinical urgency.

Conclusion

This study establishes that large language models (LLMs) can significantly enhance medication information literacy, contingent upon strategic complexity adaptation. Three core findings emerge from the evidence:

(a) Professional Calibration is Paramount: Professionally calibrated LLM content (Variant 5) demonstrably surpasses both human-generated answers and oversimplified AI outputs across the majority of medication literacy metrics. This confirms that authentic patient-friendly communication requires embedding professional accuracy, not diluting content.

(b) An Optimal Balance Exists: The identified '*golden interval*' for terminology density (15-25%) and syntactic complexity resolves the longstanding readability paradox, proving that clinical precision and patient comprehension can be synergistic when complexity is strategically managed.

(c) A Risk-Stratified Framework is Necessary: The efficacy of LLM-generated content is context-dependent. The proposed risk-stratified protocol–applying high-precision variants (e.g., Variant 5) for high-risk scenarios (e.g., dosing) and tolerating simplification (e.g., Variant 1) for low-stakes contexts (e.g., storage)–provides a principled approach for safe and equitable deployment.

These conclusions directly answer the research questions posed in the introduction: While our findings are robustly supported by rigorous expert evaluation, we acknowledge, in line with reviewer feedback, that they primarily establish the objective quality of LLM-generated content. The critical next step is to validate these findings through user studies that assess real-world patient comprehension, thereby bridging the gap between expert-rated quality and patient-centered outcomes. LLMs can indeed outperform human communities in providing reliable medication advice when professionally calibrated; an optimal complexity interval exists; and a risk-stratified framework is the key to practical application.

This study is not without limitations. The primary analysis relied on expert assessment based on standardised scales, and the data were predominantly in English. Future work should prioritise validating these findings through user studies with real patients and expanding research into multilingual and multicultural contexts to ensure broader applicability.

In summary, this research provides a validated, evidence-based framework for harnessing the power of LLMs to advance medication literacy. By anchoring accessibility in professional accuracy, we move closer to ensuring reliable health information is a universal reality.

Acknowledgements

This work was supported by the Independent Research Project of Key Laboratory of Frontier Theory and Application of Statistics and Data Science of Ministry of Education [KLATASDS2406], and by the Shanghai Planning Office of Philosophy and Social Science Project [2024BJC005].

About the authors

Tan Fu is a PhD candidate in the School of Economics and Management at East China Normal University. She holds a master's degree in Library Science from Sun Sat-sun University, and her research interests are open science and scientific data management. She can be contacted at 52284419013@stu.ecnu.edu.cn

Wen Lou is an Associate Professor in the School of Economics and Management at East China Normal University, and a researcher in the Key Laboratory of Frontier Theory and Application of Statistics and Data Science of Ministry of Education. She received her PhD from Wuhan University and her research focus on health informatics, computational social science, and science of science. She can be contacted at wlou@infor.ecnu.edu.cn

References

Andreu-March, M., Compaired, M. A., Marino, E. L., & Modamio, P. (2023). Cross-cultural adaptation and validation of the Recognising And Addressing Limited Pharmaceutical Literacy (RALPH) interview guide in community pharmacies. Research in Social and Administrative Pharmacy, 19(6), 882-888. https://doi.org/10.1016/j.sapharm.2023.02.004

Chen, Z. F., Chen, X., Wei, Y., et al. (2024). Analysis of the current situation and influencing factors of medication literacy in home-based stroke patients. Military Nursing, 41(11), 1-4.

Davis, T. C., Wolf, M. S., Bass, P. F., III, Thompson, J. A., Tilson, H. H., Neuberger, M., & Parker, R. M. (2006). Literacy and misunderstanding prescription drug labels. Annals of Internal Medicine, 145(12), 887-894. https://doi.org/10.7326/0003-4819-145-12-200612190-00144

Ferreira Alfaya, F. J., & Zarzuelo Romero, M. J. (2022). Pharmaceutical Pictograms: an opportunity for Health Literacy? Ars Pharmaceutica, 63(3), 274-293. https://doi.org/10.30827/ars.v63i3.23979

Gentizon, J., Fleury, M., Pilet, E., Buela, C., & Mabire, C. (2022). Conceptualisation and content validation of the MEDication literacy assessment of geriatric patients and informal caregivers (MED-fLAG). Journal of Patient-Reported Outcomes, 6(1), Article 87. https://doi.org/10.1186/s41687-022-00495-2

Gentizon, J., Hirt, J., Jaques, C., Lang, P.-O., & Mabire, C. (2021). Instruments assessing medication literacy in adult recipients of care: A systematic review of measurement properties. International Journal of Nursing Studies, 113, Article 103785. https://doi.org/10.1016/j.ijnurstu.2020.103785

Jaeger, B. C., Chen, L., Foti, K., et al. (2023). Hypertension statistics for US adults: an open-source web application for analysis and visualisation of National Health and Nutrition Examination Survey data. Hypertension, 80, 1311-1320. https://doi.org/10.1161/HYPERTENSIONAHA.123.20900

Lee, C.-H., Chang, F.-C., Hsu, S.-D., Chi, H.-Y., Huang, L.-J., & Yeh, M.-K. (2017). Inappropriate self-medication among adolescents and its association with lower medication literacy and substance use. PLOS ONE, 12(12), Article e0189199. https://doi.org/10.1371/journal.pone.0189199

Liu, H. T., Wang, Y. M., Zheng, B. B., et al. (2024). Development and validation of a self-assessment scale for medication literacy in patients with coronary heart disease and diabetes. Chinese Journal of Nursing, 59(9), 1065-1072.

Ma, G., Luo, A., Shen, Z., Duan, Y., Shi, S., & Zhong, Z. (2020). The status of medication literacy and associated factors of hypertensive patients in China: a cross-sectional study. Internal and Emergency Medicine, 15(3), 409-419. https://doi.org/10.1007/s11739-019-02187-0

Mackert, M. (2011). Health Literacy Knowledge Among Direct-to-Consumer Pharmaceutical Advertising Professionals. Health Communication, 26(6), 525-533. https://doi.org/10.1080/10410236.2011.556084

Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. Nature Reviews Nephrology, 16(4), 223-237. https://doi.org/10.1038/s41581-019-0244-2

Pantuzza, L. L. N., do Nascimento, E., Crepalde-Ribeiro, K., Botelho, S. F., Martins, M. A. P., Veloso, R. C. de S. G., Nascimento, M. M. G. do, Vieira, L. B., & Reis, A. M. M. (2022). Medication literacy: A conceptual model. Research in Social and Administrative Pharmacy, 18(4), 2675-2682. https://doi.org/10.1016/j.sapharm.2021.06.003

Pouliot, A., Vaillancourt, R., Stacey, D., & Suter, P. (2018). Defining and identifying concepts of medication literacy: An international perspective. Research in Social and Administrative Pharmacy, 14(9, SI), 797-804. https://doi.org/10.1016/j.sapharm.2017.11.005

van Beusekom, M. M., Grootens-Wiegers, P., Bos, M. J. W., Guchelaar, H.-J., & van den Broek, J. M. (2016). Low literacy and written drug information: information-seeking, leaflet evaluation and preferences, and roles for images. International Journal of Clinical Pharmacy, 38(6), 1372-1379. https://doi.org/10.1007/s11096-016-0376-4

Vervloet, M., van Dijk, L., Rademakers, J. J. D. J. M., Bouvy, M. L., De Smet, P. A. G. M., Philbert, D., & Koster, E. S. (2018). Recognising and Addressing Limited Pharmaceutical literacy: Development of the RALPH interview guide. Research in Social and Administrative Pharmacy, 14(9, SI), 805-811. https://doi.org/10.1016/j.sapharm.2018.04.031

Zhang, Z. Y., Ding, W., Wang, R., et al. (2024). Development and validation of a medication literacy scale for patients after percutaneous coronary intervention. Chinese Journal of Nursing Education, 21(8), 993-999.

Zheng, F., Shen, Z. Y., Duan, Y. L., et al. (2017). Investigation of medication literacy status and nursing countermeasures in discharged patients with acute coronary syndrome. Chinese Nursing Management, 17(8), 1127-1130.