

Chinesization and Validity and Reliability Testing of the Patient Education Materials Assessment Tool for Audiovisual Materials

Keer Huang

School and Hospital of Stomatology, Fujian Medical University

Xiaohua Chen

Fujian Medical University

Yanping Wei

School and Hospital of Stomatology, Fujian Medical University

Jiayuan Zhuang

jiayuanz@fjmu.edu.cn

Fujian Medical University

Article

Keywords: Chinesization of the PEMAT-A/V, understandability, actionability, cultural adjustment

Posted Date: April 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8162141/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 7th, 2026.
See the published version at <https://doi.org/10.1038/s41598-026-68702-4>.

Abstract

Background

The provision of health information that can be easily understood by the general public can help improve health behaviors. However, there is a lack of tools to assess the quality of health education videos in the Chinese context. These tools are essential for the popularization of health education videos. There are various existing tools for evaluating health education video quality in the English context. Among them, the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V) is one of the most commonly used tools. However, this scale has not been adjusted for the Chinese context and verified for use in the Chinese population.

Methods

The Brislin translation model was applied to achieve Chinesization and cultural adjustment of the PEMAT-A/V, resulting in the C-PEMAT-A/V. Videos depicting health information related to periodontitis on TikTok, MicroBlog, and Bilibili were identified using a convenient sampling method. Items in the C-PEMAT-A/V scale were analyzed by the critical ratio method and correlation coefficient method. The content validity of the scale was tested by the expert evaluation method. The criterion validity of the scale was tested using the modified DISCERN. The internal consistency of the scale was tested by Cronbach's α , and the inter-rater reliability was tested by Cohen's Kappa coefficient, the average percentage agreement, and Gwet's AC1.

Results

After literal translation, back translation, and cultural adjustment, the C-PEMAT-A/V scale was reduced from 17 items to 13 items. The item content validity index (I-CVI) of the scale was 0.83-1.00 and the average content validity index (S-CVI) was 0.96. The correlation coefficient of criterion validity was $r = 0.647$ ($p < 0.01$). The Cronbach's α coefficient of the C-PEMAT-A/V scale was 0.703, Cohen's Kappa coefficient was 0.86, the average percentage agreement was 87.7%, and Gwet's AC1 was 0.873.

Conclusions

The C-PEMAT-A/V scale has good reliability and validity. It can be used to evaluate the understandability of health science videos in the Chinese context.

1 Introduction

Health education is an important intervention to prevent disease and improve health¹. The effectiveness of health education is contingent on the understandability of the information for the target audience. With the development of information technology, the internet has become one of the most important information sources in the medical field and an important way by which the public can acquire health information.

It is estimated that around 995 million people in China access online videos, accounting for 94.6% of the total netizens in China². Around 71–75% of people globally acquire health information from the Internet, with this figure reaching around 68% in China^{3,4}. Due to the rapid development of information technology, the public is increasingly acquiring health information from online platforms. Whether online health information is accurate or not may influence the health-related behaviors of the public. Several studies have indicated that online health information is difficult to read and understand, and most people lack the ability to recognize scientific health information online^{5,6}. Therefore, the accurate identification and evaluation of quality online health videos is critical to the scientific popularization of health in the information age.

At present, the common assessment tools for health education videos include the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V), the DISCERN, the Global Quality Score (GQS), the Health On the Net's Foundation Code of Conduct (HONcode), and the Journal of the American Medical Association (JAMA) Benchmark Criteria. These quality assessment tools have various foci^{7–11}. For example, the PEMAT-A/V pays high attention to the understandability and actionability of videos. The DISCERN places emphasis on the integrity and reliability of videos. The GQS considers the fluency and usability of videos. The HONcode assesses the reliability and transparency of videos. The JAMA criteria pay high attention to the reliability of videos. Different from other assessment tools, the PEMAT-A/V has been specifically developed for the assessment of video and audiovisual materials. Its aim is to evaluate the health information quality of videos and audiovisual materials. The other assessment tools cover a range of purposes, from text material development to the evaluation of video quality.

The PEMAT-A/V was compiled by the Agency for Healthcare Research and Quality (AHRQ)⁷ in 2014 to evaluate the understandability and actionability of health information. The Cronbach's α coefficient of the PEMAT-A/V scale is 0.76⁷. The scale contains 17 items that measure two dimensions, understandability and actionability; 13 items measure the understandability of the information, and 4 items measure the actionability of the information. The understandability scale covers five themes, namely, education content, vocabulary selection and style, language organization, layout and design, and visual aid tools. Each item is rated as 1 (agree) or 0 (disagree). Inapplicable items are not scored. The overall score is calculated as follows: total score / the number of all scored items (excluding the inapplicable items) $\times 100\%$. The total score ranges from 0–100%; a higher score indicates that the content of the audiovisual material is easier to understand and operate. Scores higher than 70% indicate that the material has good understandability or actionability⁷.

The PEMAT-A/V has been widely applied to evaluate the quality of various types of videos. Rubel et al¹² used it to evaluate the quality of videos related to nasosinusitis on YouTube. The results showed that there was a lack of high-quality online education videos related to nasosinusitis and most videos could be neither understood nor implemented. Sampige et al¹³ analyzed the quality of ophthalmological health education videos on TikTok and evaluated scores of their understandability and actionability. They found

that ophthalmological health education videos on TikTok had high understandability scores (88.1%) but low actionability scores (50.6%). Overall, 5.4% of videos contained inaccurate information. It has been reported that the PEMAT-A/V can also be used to develop or improve patient health education materials. Mohd et al¹⁴ developed an e-book about bone health and osteoporosis for teenagers, and improved the e-book by using PEMAT-A/V.

The PEMAT-A/V has been translated into Malay, Japanese and Arabic^{15–17}; however, to date, there is no Chinese PEMAT-A/V scale. Hence, it is necessary to Chinesize the PEMAT-A/V scale for use in the Chinese context. The PEMAT-A/V was Chinesized, and its reliability and validity were tested. The aim was to develop an assessment tool for health science video quality in the Chinese context.

2 Methods

2.1 Translation and Cultural Adjustment

2.1.1 Translation

The Brislin translation model^{18,19} was applied to Translation and cultural adjustment of the PEMAT-A/V as follows. ☐ Literal translation: one Master of Science in Nursing (MSN) student and one Doctor of Medicine (MD) who were familiar with the research and development background of measurement tools, and who were native speakers of Chinese, were invited to independently translate the questionnaire from English into Chinese, forming the Chinese versions T1 and T2. T1 and T2 were then compared with the original English version. The translators discussed any disagreements until reaching a consensus, thus forming the Chinese version TI2. ☐ Back translation: a MSN who was a native speaker of Chinese with overseas work experience, but did not have a measurement tool background, was invited to back-translate TI2, forming the back translation version B1. ☐ Combination: a MD with 20 years of overseas learning and work experience was invited to compare the original English questionnaire and B1. When there was disagreement between the two versions, a monolingual translator was engaged to provide three bilingual translators with an explanation of the English item for which there was disagreement. The three bilingual translators then adjusted and revised TI2 after discussion until they reached a consensus with the monolingual translator. As a result, the initial draft of the C-PEMAT-A/V was obtained.

2.1.2 Cultural Adjustment

Six experts who were engaged in health education, including three experts in oral medicine, one expert in nephrology science, one expert in community nursing science, and one expert in maxillofacial surgery, were invited to evaluate and provide comments on the initial draft of the C-PEMAT-A/V²⁰. The experts evaluated the scale according to their professional knowledge and clinical work experience from the perspective of the understandability of the items, idiomatic expression, agreement with the cultural background, and whether the addition or deletion of items was necessary. When reviewing the scale, each item was scored on a four-point Likert scale. A score of 1 indicated “completely unrelated”, 2

indicated “related to some extent but needs great improvement”, 3 indicated “related but needs a small amount of revision”, and 4 indicated “highly related”. This revised scale was the final C-PEMAT-A/V.

2.2 Statistical Analysis

A cross-sectional study was performed to verify the reliability and validity of the developed C-PEMAT-A/V. The Chinese keyword for “牙周病” (“periodontosis”) was used to identify relevant videos on the three most popular social media platforms (TikTok, Weibo, and Bilibili) in China^{21–23}. **Inclusion criteria:** (1) videos containing periodontal health education content; (2) Chinese videos. **Exclusion criteria:** (1) commercial advertisements; (2) duplicate videos; (3) videos with only background music and no narration. According to the Kendall rough estimation method, the sample size required for the study was 5–10 times the number of items in the questionnaire. The original PEMAT-A/V scale contains 17 items. Considering an invalid sample rate of 10%, the desired sample size in this study was 187.

2.2.1 Item Statistics

The critical ratio method and correlation coefficient method were employed in this study for item analysis. For the critical ratio method, the total scores of the questionnaire were placed in descending order. The top 27% were categorized as the high-score group and the bottom 27% were categorized as the low-score group. Differences in the mean scores of each item between the high-score and low-score groups were tested by independent samples t-tests. If the difference was statistically significant ($P < 0.05$), it was concluded that the item is of significance to the questionnaire and could be used to identify the response degrees of different research objects. Otherwise, the item was considered meaningless and was deleted. For the correlation coefficient method, Spearman's correlation coefficients were calculated between the scores for each item and the total questionnaire score. Generally speaking, if the correlation coefficient reaches statistical significance ($p < 0.05$), the item has good homogeneity with the total questionnaire; otherwise, the item should be deleted^{24,25}.

2.2.2 Validity Testing

Content validity was tested by the expert review method. The experts were those chosen during cross-cultural adjustment. All experts were asked to evaluate the correlation between each item and the questionnaire theme using a four-point scoring method. Specifically, a score of 1 reflected “unrelated”, 2 reflected “related to some extent”, 3 reflected “related”, and 4 reflected “highly related”. The content validity of the scale was evaluated using the Item-level Content Validity Index (I-CVI) and the Scale-level Content Validity Index (S-CVI). The criterion validity of the C-PEMAT-A/V scale was analyzed using the modified DISCERN (mDISCERN)²⁶ as the criterion.

2.2.3 Reliability Testing

The internal consistency of the C-PEMAT-A/V was tested by Cronbach's α coefficient. The inter-rater reliability was analyzed to test the external consistency of the C-PEMAT-A/V. Cohen's Kappa coefficient and the average percentage agreement of two reviewers were analyzed. Moreover, Gwet's AC1 was

calculated when we observed a low Kappa value even though the consistency percentage was very high^{7,16,27,28}.

2.3 Data Collection and Analysis

Data were input into Microsoft Excel 2019 and analyzed with SPSS 26.0 software and R version 4.2.2. Statistical significance was determined at $p < 0.05$.

3 Results

3.1 Translation

During scale translation, four translators were invited to discuss and revise the items. A lack of content disagreement was taken to indicate that the item reflected the expression habits of the Chinese population (Table 1).

Table 1
Specific revision results of items

Items	English version	Initial Chinese version	Chinese pre-survey version
Item 3	The material uses common, everyday language.	Medical terms are only applicable to people who are familiar with these terms. Medical terms have to be defined first before use.	Medical terms are only used to help readers be familiar with their meanings. The medical terms shall be defined at use.
Item 5	The material breaks or "chunks" information into short sections.	The material breaks information into chunks or classifies information into sections.	The material decomposes or "chunks" information into short sections.
Item 6	The sections of the material have informative headers.	Each section of the material provides headers with useful information.	Each section of the material has informative headers.
Item 7	The material presents information in a logical sequence.	The material presents information in a logical way.	The material presents information in a logical sequence.
Item 13	The material uses simple tables with short and clear row and column headings.	Rows of material in the table shall be short, simple, and explicit, and accompanied by column headings.	The material uses simple tables with short and clear row and column headings.

3.2 Cultural Adjustment

Three items in the scale were adjusted according to the experts' suggestions. Item 4, "The material uses a positive voice", was not consistent with Chinese language expression. A positive voice is rarely used in Chinese health science videos. Hence, this item was considered inapplicable and was deleted. Item 9, "The material uses visual hints (e.g., arrows, frameworks, symbols, bold font, big font, highlighted font) to attract the attention of readers to the key information", is inapplicable for the evaluation of videos,

according to the introduction of the original author. Hence, this item was deleted. Item 10, “Words on the screen are easy to read”, is inapplicable according to the introduction materials of the original author, specifically if the material has no text or all text is narrated. Chinese health education videos can be divided into two types: videos with and without subtitles. Hence, this item was inapplicable in the Chinese context and was deleted. Finally, 14 items were retained.

3.3 Item Statistics

Critical ratio method

After normality testing, it was concluded that the scores for all items were normally distributed. The independent samples t-test was applied. The results showed significant score differences between the high-score group and the low-score group for all items except for Item 2, Item 13, and Item 17 ($P < 0.05$) (Table 2).

Table 2
C-PEMAT-A/V item scores for the high-score group and low-score group

Item	t value	P value
Item 1	-5.049	<0.01
Item 2	-0.029	>0.05
Item 3	-2.065	<0.05
Item 4	Deleted	Deleted
Item 5	-5.123	<0.01
Item 6	-6.928	<0.01
Item 7	-4.055	<0.01
Item 8	-3.959	<0.01
Item 9	Deleted	Deleted
Item 10	Deleted	Deleted
Item 11	-2.086	<0.05
Item 12	-3.708	<0.01
Item 13	a	a
Item 14	-8.262	<0.01
Item 15	-7.453	<0.01
Item 16	-3.517	<0.01
Item 17	a	a
^a too few videos received ratings on the item		

Correlation coefficient method

There were significant correlations between all items and the total scale score, except for Item 2, Item 13, and Item 17 ($P < 0.05$). The correlation coefficients (r values) ranged from 0.162–0.742. Due to the small sample size, the critical ratio method and correlation coefficient method could not be applied to Item 13 and Item 17. After discussion, the research team decided to retain Item 13 and Item 17 but delete Item 2. However, these items should be analyzed and verified using larger sample sizes in the future (Table 3). All other items were kept and further analyzed.

Table 3
Correlation coefficients between each item and
the total C-PEMAT-A/V score

Item	<i>r value</i>	<i>P value</i>
Item 1	0.382	<0.01
Item 2	-0.001	>0.05
Item 3	0.165	<0.05
Item 4	Deleted	Deleted
Item 5	0.470	<0.01
Item 6	0.429	<0.01
Item 7	0.305	<0.01
Item 8	0.317	<0.01
Item 9	Deleted	Deleted
Item 10	Deleted	Deleted
Item 11	0.162	<0.05
Item 12	0.380	<0.01
Item 13	a	a
Item 14	0.485	<0.01
Item 15	0.524	<0.01
Item 16	0.363	<0.01
Item 17	0.742	<0.05
^a too few videos received ratings on the item		

3.4 Validity Testing

Content validity

the I-CVI ranged from 0.83-1.00 and the S-CVI was 0.96. The item contents were not revised (Table 4).

Table 4
Item-level content validity results for items in C-PEMAT-A/V

Item	Number of experts rated 3 or 4	I-CVI
Item 1	6	1
Item 3	5	0.83
Item 5	6	1
Item 6	6	1
Item 7	6	1
Item 8	5	0.83
Item 11	6	1
Item 12	6	1
Item 13	6	1
Item 14	6	1
Item 15	5	0.83
Item 16	6	1
Item 17	6	1

Criterion validity

the correlation coefficient between the C-PEMAT-A/V scale and mDISCERN was $r = 0.647$ ($p < 0.01$). The correlation coefficient between the understandable scale and mDISCERN was $r = 0.717$ ($p < 0.01$). The correlation coefficient between the actionability scale and mDISCERN was $r = 0.454$ ($p < 0.05$).

3.5 Reliability Testing

The Cronbach's α coefficient of the C-PEMAT-A/V scale was 0.703 and the Cronbach's α coefficients of the different dimensions ranged from 0.617–0.664. The Cohen's Kappa coefficient of the C-PEMAT-A/V scale was 0.86 and the Cohen's Kappa coefficients of the different dimensions ranged from 0.84–0.85. The average percentage agreement was 87.7%, and Gwet's AC1 was 0.873 (Table 5).

Table 5
Item-level reliability results for items in C-PEMAT-A/V

Scales	Cronbach's α	Cohen's Kappa	Average percentage agreement	Gwet's AC1
All scales	0.703	0.86	87.7%	0.873
Understandability scale	0.664	0.85	87.2%	0.866
Actionability scale	0.617	0.84	88.2%	0.875

4 Discussion

4.1 Principal Findings

In this study, the Brislin translation model was applied for literal translation, back translation, cross-cultural adjustment, and pre-survey analysis of the C-PEMAT-A/V scale to obtain maximum congruency of the scale content. One MSN and one MD were invited to undertake literal translation. One MSN with overseas work experience was invited to undertake back-translation. One MD with 20 years of overseas learning and work experience was invited to act as a monolingual translator. In the expert review stage, Item 4, Item 9, and Item 10 were deleted with consideration for the Chinese cultural background and applicability of the scale. After the Chinesization process with scientific and standard translation, back translation, cross-cultural adjustment, and pre-survey analysis, the C-PEMAT-A/V scale maintained good equivalence with the original version.

According to the item analysis in this study, there were no differences in Item 2, Item 13, and Item 17 between the high-score group and the low-score group and the correlation coefficients between these items and the total score were not statistically significant. Item 2, "The material uses common and everyday language", and Item 3, "Medical terms are only used to help readers to be familiar with their meanings. The medical terms shall be defined at use", have similar semantics in the Chinese context. Since the correlation coefficient between Item 3 and the total scale reached statistical significance ($P < 0.05$), while the correlation coefficient for Item 2 was not ideal, Item 2 might be unsatisfactory in distinguishing daily language and medical terms. The critical ratio method and correlation coefficient method could not be used for Item 13 and Item 17 due to the small practical sample size. The results for Item 13 and Item 17 are consistent with the results of the original scale. After discussion, the research team decided to keep Item 13 and Item 17 and delete Item 2. After the cultural adjustment stage, Item 9 and Item 10, which were inapplicable to video evaluations, were also deleted. Overall, the Chinese scale was adjusted due to cultural differences and differences between the Chinese and English contexts, forming the C-PEMAT-A/V, which is suitable for use in the Chinese cultural background.

The Cronbach's α coefficient in this study was 0.703. Specifically, the Cronbach's α coefficient of the understandability scale was 0.664 and that of the actionability scale was 0.617, indicating good internal

consistency of the scale.

The Cohen's Kappa coefficient in this study was 0.86. Specifically, the Cohen's Kappa coefficient of the understandability scale was 0.85 and that of the actionability scale was 0.84. The average percentage agreement was 87.7%. Specifically, the average percentage agreement of the understandability scale was 87.2% and that of the actionability scale was 88.2%. Gwet's AC1 was 0.873. Specifically, Gwet's AC1 of the understandability scale was 0.866 and that of the actionability scale was 0.875. The inter-rater consistency of the scales was significant.

In this study, the I-CVI ranged from 0.83-1.00 and the S-CVI was 0.96. The content validity was satisfactory, and the item contents of the scale were representative.

Although there is no uniform gold standard for the measurement of health science video quality, the mDISCERN, which is widely applied for the evaluation of health information quality globally, was chosen for the correlation analysis based on previous studies^{29–31}. The results showed that the Spearman correlation coefficient between the C-PEMAT-A/V and mDISCERN was 0.647, indicating that the scale has good criterion-related validity.

4.2 Comparison With Previous Studies

The Cronbach's α coefficient of this study was slightly lower than that obtained in the study by Shoemaker⁷ and the Arabic version¹⁷. The Cronbach's alpha coefficient of the original scale is 0.76, and the Cronbach's alpha coefficients of each dimension are 0.74–0.77. The difference might be attributed to the different cultural backgrounds of the populations under study.

The Cohen's Kappa coefficient of this study was higher than that reported by Shoemaker et al⁷. The Cohen's Kappa coefficient of the original scale was 0.50 and the Cronbach's α coefficient of all dimensions ranged from 0.49–0.53. It should be noted that the nine experts who participated in the analysis of the original scale included non-medical staff, such as workers from profit, non-profit, and government institutions. This may have resulted in weak inter-rater consistency.

The average percentage agreement of the original scale⁷ was 80.0%. Specifically, the average percentage agreement of the understandability scale was 79.0% and that of the actionability scale was 83.0%. The average percentage agreement of the Japanese version^[16] was reported to be 85.7%, with a value of 84.3% for the understandability scale and 90.0% for the actionability scale. Gwet's AC1 of the original scale⁷ was 0.71, with a value of 0.71 for the understandability scale and 0.72 for the actionability scale. Gwet's AC1 of the Japanese version¹⁶ was 0.8, with a value of 0.90 for the understandability scale and 0.92 for the actionability scale. The results of the current study indicated that the C-PEMAT-A/V scale is not significantly different from other language versions in terms of consistency.

4.3 Limitations

There are several limitations of this study that should be noted. First, this study evaluated the understandability and actionability of health education videos from the perspective of medical health workers using only the C-PEMAT-A/V scale. However, the perspective of the public was not considered. Future studies should invite non-medical health workers to evaluate the understandability and actionability of health education videos using the C-PEMAT-A/V scale. Secondly, periodontitis videos on different online platforms were chosen as the research objects. The use of a single type of research sample and disease type may limit the generalizability of the findings. Future studies should expand the survey scope and sample size to verify the reliability, validity, and applicability of the C-PEMAT-A/V scale and improve its universality. Third, the criterion validity of the C-PEMAT-A/V was evaluated with reference to the mDISCERN. However, the mDISCERN and C-PEMAT-A/V emphasize different aspects of health information quality. Future studies should further verify the criterion validity of the C-PEMAT-A/V scale using other measures as comparators.

5 Conclusions

After literal translation, back translation, and cultural adjustment of the PEMAT-A/V scale, the C-PEMAT-A/V scale was reduced from 17 items to 13 items. Three items, i.e., “The material uses a positive voice”, “The material uses visual hints (e.g., arrows, frameworks, symbols, bold font, big font, and highlighted font) to attract the attention of readers to the key information”, and “Words on the screen are easy to be read”, were deleted. According to the research results, the C-PEMAT-A/V scale comprises 13 items and 2 dimensions, with good reliability and validity. After rigorous scientific verification, the C-PEMAT-A/V scale can be used as an effective tool to evaluate the understandability and actionability of Chinese health science videos. This tool may be useful in producing and improving health science videos in the future.

Declarations

Ethical Considerations

This study was approved by the Biomedical Research Ethics Committee, Fujian Medical University (2024, Approval No.59). This research adhered to the ethical principles outlined in the Declaration of Helsinki. Additionally, it complied with all relevant research guidelines established by national, local, and institutional governing bodies.

Author contributions

Keer Huang: Writing - Original Draft, Writing - Review & Editing, Validation, Formal analysis. **Xiaohua Chen:** Investigation, Formal analysis, Date curation. **Yanping Wei:** Validation, Methodology. **Jiayuan Zhuang:** Conceptualization, Supervision, Resources, Writing - Review & Editing.

All authors reviewed the manuscript.

Funding

This work was sponsored by Fujian provincial health technology project[2024QNA065].

Consent to publication

Written informed consent was obtained from all participants to publish their information and images in an online open access publication.

Conflict of interest statement

No conflict of interest has been declared by the authors.

Data availability

Data is available on request from the authors.

References

1. Sureda, A. et al. Oral microbiota and Alzheimer's disease: Do all roads lead to Rome?[J]. *Pharmacol. Res.* **151**, 104582 (2020).
2. China Internet Network Information Center. The 50th statistical report on China's development[EB/OL]. [2022-11-18]. <http://www.cnnic.net.cn/n4/2022/0914/c88-10226.html>
3. Zhang, X. et al. How the public uses social media wechat to obtain health information in china: a survey study[J]. *BMC Med. Inf. Decis. Mak.* 2017, **17**(Suppl 2):66 .
4. Bizzi, I. Ghezzi P, Paudyal P. Health information quality of websites on periodontology[J]. *J. Clin. Periodontol.* **44** (3), 308–314 (2017).
5. Suarez-Lledo, V., Alvarez-Galvez, J. & Prevalence of Health Misinformation on Social Media. *Syst. Review[J]* *J. Med. Internet Res.*, **23**(1):e17187. (2021).
6. Yeung, A. W. K. et al. Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature[J]. *J. Med. Internet Res.* **24** (1), e28152 (2022).
7. Shoemaker, S. J. & Wolf, M. S. Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information[J]. *Patient Edu Couns.* **96** (3), 395–403 (2014).
8. Charnock, D. et al. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices[J]. *J. Epidemiol. Community Health.* **53** (2), 105–111 (1999).
9. Bernard, A. et al. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web[J]. *Am. J. Gastroenterol.* 2007, **102**(9):2070–2077 .
10. Al-Silwadi, F. M. et al. Effect of social media in improving knowledge among patients having fixed appliance orthodontic treatment: A single-center randomized controlled trial[J]. *Am. J. Orthod. Dentofac. Orthop.* **148** (2), 231–237 (2015).

11. Silberg, W. M. Assessing,controlling,and assuring the quality of medical information on the Internet:Caveant lector et viewer–Let the reader and viewer beware[J].*JAMA. J. Am. Med. Association* 1997, **277**(15): 1244–1245 .
12. Rubel, K. E. et al. Understandability and actionability of audiovisual patient education materials on sinusitis[J]. *Int. Forum Allergy Rhinol* 2020, **10**(4):564–571 .
13. Sampige, R. et al. Education and misinformation: exploring Ophthalmology content on TikTok[J]. *Ophthalmol. Ther.* **13** (1), 97–112 (2024).
14. Mohd Azmi, N. F. I., Aznul Hisham, M. H. & Jamil, N. A. Development and evaluation of an e-Book for bone health and osteoporosis education in Adolescents[J].*Nutrients*,2023,**15**(8):1899 .
15. Wong, S. T. & Saddki N,Arififin W N.Inter-rater reliability of the Bahasa Malaysia version of patient education materials assessment tool[J]. *Med. J. Malays* ,2019,74,100.
16. Furukawa, E. et al. Translation, cross-cultural adaptation, and validation of the Japanese version of the patient education materials assessment tool (PEMAT)[J]. *Int. J. Environ. Res. Public. Health* 2022, **19**(23):15763 .
17. Alrasheed, M. A. et al. Arabic version of the patient education materials assessment tool (PEMAT): translation and validation[J]. *Saudi Pharm. J.* **33** (3), 15 (2025).
18. Brislin, R. W. Comparative research methodology:cross-cultural studies[J]. *Int. J. Psychol.* **3** (11), 215–229 (1976).
19. Cha, E., Kim, K. H. & Erlen, J. A.Translation of scales in cross-cultural research:issues and techniques[J]. *Adv. Nurs.* 2007, **58**(4):386–395 .
20. Wan Mohamad Darani, W. N. S. et al. Development and validation of the information-motivation-behavioural skills model-based human immunodeficiency virus education kit for adolescents in. *Malaysia[J] Sci. Rep.* **14** (1), 6890 (2024).
21. Tao, Z. Y. Chu G,McGrath C,et al.Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo[J]. *J. Med. Internet Res.* 2020, **22**(6):e19981 .
22. Han, J., Shi, Y. & Ma, H. Assessment of videos related to lung nodules in China[J]. *Front. Surg.* **9**, 1019212 (2022).
23. Liu, H. et al. Assessment of the reliability and quality of breast cancer related videos on TikTok and Bilibili: cross-sectional study in China[J]. *Front. Public. Health.* **11**, 1296386 (2024).
24. Mukaka, M. M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research[J]. *Malawi Med. J.* 2012, **24**(3):69–71 .
25. Rodríguez-García, M. C. et al. Cross-cultural adaptation and validation of the Spanish version of the Cardiopulmonary Resuscitation Motivation Scale (s-CPRMS): a cross sectional study[J].*BMC* 23(1):845. (2024).
26. Singh, A. G., Singh, S. & Singh, P. P. YouTube for Information on Rheumatoid Arthritis -A Wakeup Call? [J].*J Rheumatol*,2012,**39**(5):899–903 .

27. Gwet, K. L. Computing inter-rater reliability and its variance in the presence of high agreement[J]. *Br. J. Math. Stat. Psychol.* **61**, 29–48 (2008).
28. Feinstein, A. R. Cicchetti D V.High agreement but low kappa: I. The problems of two paradoxes[J]. *J. Clin. Epidemiol.* **43**, 543–549 (1990).
29. Demirtas, M. S. Alici N.YouTube as a source of information on infantile colic[J]. *Pediatr. Int.* **65** (1), e15624 (2023).
30. Liu, Z. et al. YouTube/ Bilibili/ TikTok videos as sources of medical information on laryngeal carcinoma: cross-sectional content analysis study[J]. *BMC Public. Health.* **24** (1), 1594 (2024).
31. Chang, M. C. & Park, D. YouTube as a Source of Information on Epidural Steroid Injection[J]. *J. Pain Res.* **14**, 1353–1357 (2021).