ORIGINAL ARTICLE

Cross-cultural Adaptation and Psychometric Testing of the Malay Version of Patient Education Materials Assessment Tool

Siong Ting Wong[1,2], Norkhafizah Saddki[1], Wan Nor Arifin[3]

[1] Dental Public Health Unit, School of Dental Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia
[2] Sibu Divisional Dental Office, Ministry of Health Malaysia, 96000 Sibu, Sarawak, Malaysia.
[3] Biostatistics and Research Methodology Unit, School of Medical Sciences, Health Campus Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

ABSTRACT

Introduction: The Patient Education Materials Assessment Tool (PEMAT) evaluates the understandability and actionability of patient education materials. There are PEMAT-P and PEMAT-A/V for printable and audio-visual materials respectively. This study cross-culturally adapted these versions of PEMAT into Malay and determine their psychometric properties. **Materials and methods**: The study proceeded in four stages: 1) translation and cross-cultural adaptation of PEMAT-P(M) and PEMAT-A/V(M) (four translators, five healthcare workers [HCWs]), followed by cross-sectional studies to 2) determine the internal consistency, inter-rater and test-retest reliabilities (nine HCWs), 3) determine the validity by correlation with the Suitability Assessment of Materials (SAM) (20 HCWs) and 4) determine the validity by correlation with the Consumer Testing Questionnaire (16 adult patients). These stages involved HCWs from different healthcare disciplines. **Results**: Cross-culturally adapted PEMAT-P(M) and PEMAT-A/V(M) were produced. The median Cronbach's alpha for understandability and actionability domains ranged between 0.78 and 0.90. The domain-level inter-rater reliability showed ICCs ranging from 0.72 to 0.81, and test-retest reliability showed ICCs ranging from 0.82 to 0.90. The item-level inter-rater and test-retest reliability showed Gwet's AC1 values above 0.60 for most items. Significant positive correlations were found between all PEMAT-P(M) and PEMAT-A/V(M) scores with the SAM score. Most domain scores were significantly correlated with the consumer testing comprehension score, while only one domain score was significantly correlated with consumer testing numerical score. **Conclusion**: The cross-culturally adapted PEMAT-P(M) and PEMAT-A/V(M) were found to be valid and reliable. PEMAT(M) can be used by raters from diverse backgrounds to evaluate printable and audiovisual patient education materials.
Malaysian Journal of Medicine and Health Sciences (2026) 22(3): 1-11.doi:10.47836/mjmhs.v22i3.1592

Keywords: Malay Version, PEMAT, Patient Education, Reliability, Validity, Health Education Material

Corresponding Author:
Wan Nor Arifin, PhD
Email: wnarifin@usm.my
Tel: +609-7676825

INTRODUCTION

Health educators commonly use patient education materials (PEMs) to improve knowledge, attitudes and skills of patients and the public. PEMs can improve the quality of care provided in clinical practice by facilitating the delivery of relevant health information to patients in busy clinical settings, improving the health literacy of patients, and enabling patients to better manage their own health (1). Therefore, the production of well-designed and easy-to-understand PEMs is essential to ensure the delivery of effective health messages needed to support patient self-care (2). Hence, there is a considerable interest in defining and evaluating the characteristics of PEMs to ensure they deliver high-quality health-related information.

Readability formulas are commonly applied by providing quantitative estimates in reading difficulty level based on the written text. They fail to take into account factors essential for ease of comprehension of written text (3). This prompted development of several tools or instruments in the forms of checklist and questionnaires such as Suitability Assessment of Materials (SAM) (4), DISCERN (5), ClearDoc Index (6), A Plain Language Audit Kit (7), Tool to Evaluate Materials Used in Patient Education (TEMPtEd) (8) and Health Literacy INDEX (9) to assess suitability of PEMs as guidelines for production of low-demand PEMs. Accordingly, these existing instruments or checklists for assessment of PEMs did not report inter-rater reliability (4), were designed for a specific field or purpose (5), or required trained users (9), and none of these checklists and questionnaires

measure actionability of PEMs (10).

The Patient Education Materials Assessment Tool (PEMAT) was developed in English by Shoemaker, Wolf and Brach (11) to measure the ability of patients with diverse backgrounds and levels of health literacy to process and explain the key messages of a PEM (understandability) and to identify appropriate actions that should or can be taken based on the messages (actionability). PEMAT is the first instrument to measure actionability (11). PEMAT comes in two versions: PEMAT-P for assessing printable materials and PEMAT-A/V for assessing audiovisual materials. PEMAT-A/V is the only available instrument that specifically assesses audiovisual materials. PEMAT was developed for use by untrained lay-people and health professionals. PEMAT has been translated into several languages, which include Korean (12), Japanese (13), Turkish (14) and Chinese (15) with acceptable psychometric properties.

The Ministry of Health Malaysia (MOH) is the primary source of PEMs used for health education and promotion activities in Malaysia. Most of these PEMs are in the Malay language, the official language of Malaysia. To the best of our knowledge, the quality of these PEMs in the Malay language has not been adequately examined, primarily due to the lack of suitable evaluation instruments for PEMs in the Malay language. Without such an instrument in the local language for the local setting and context, healthcare workers (HCWs) in Malaysia, who are generally more conversant in the Malay language, will not be able to systematically evaluate PEMs in Malay. These existing PEMs may be questionable in terms of quality or too difficult for patients with low health literacy to comprehend and act upon; and result in compromised quality of health information to the patients (11). Therefore, this study aimed to cross-culturally adapt these versions of PEMAT into the Malay language and determine their psychometric properties, which ultimately resulting in enhanced health literacy, improved patient communication and better health outcome for Malays-speaking individuals in Malaysia specifically, neighbouring South-East Asian and other countries generally. The purpose of the cross-cultural adaptation process is to achieve cross-cultural equivalence between the original version of an instrument or tool and its target culture or population version. This process is followed by psychometric testing to ensure the reliability and validity of the adapted version, thereby achieving acceptable levels of equivalence in terms of its psychometric properties. The Malay version of PEMAT will provide a common and systematic evaluation tool for printable and audiovisual materials, that can be used by untrained raters from diverse backgrounds.

MATERIALS AND METHODS

Study Design, Settings and Participants

This study was a cross-cultural adaptation and validation study to provide the psychometric evidence of validity to the cross-culturally adapted PEMAT. Validity or construct validity in this study was provided by utilizing the unitary validity framework (16, 17), which describes validity in the form of the degree of evidence that comes from content, internal structure, relationship with other variables, response process and consequences.

The study was conducted in four phases. For the study locations, Phase 1 was conducted in the Health Campus of Universiti Sains Malaysia, while Phase 2 to 4 were conducted in Sibu Division, Sarawak, Malaysia. Phase 1 focused on the translation and cross-cultural adaptation of PEMAT into the Malay language, involving four translators and five expert committee members. This phase provided the content and response process evidence of validity. In Phase 2, the reliability of PEMAT-P(M) and PEMAT-A/V(M) was investigated in a cross-sectional study with nine HCWs, providing the internal structure evidence of validity. Phase 3 and 4 provided the evidence of relationships of PEMAT-P(M) and PEMAT-A/V(M) with other related tools. Phase 3 evaluated the validity of PEMAT-P(M) and PEMAT-A/V(M) against the Suitability Assessment of Materials (SAM) in a cross-sectional study, involving 20 HCWs, while Phase 4 evaluated the validity of PEMAT-P(M) and PEMAT-A/V(M) against the Consumer Testing Questionnaire in a cross-sectional study involving 16 adult patients.

All PEMs in this study were selected via the combination of random sampling method and expert committee advice. The expert committee included one dental officer, one family medicine specialist, one medical education specialist, one dental public health specialist and one biostatistician, an expert in research methods. The members of expert committee were native Malay language speakers. HCWs and adult patients were selected via non-probability purposive sampling method. Potential sources of bias in this study were addressed by ensuring that none of the participants were involved in the production of PEMs being evaluated.

Prior to commencement of this study, the corresponding owner of the original PEMAT provided permission via email to perform the cross-cultural adaptation of PEMAT into the Malay language. This study was approved by the Universiti Sains Malaysia Human Research and Ethics Committee (Ref: USM/JEPeM/15050167) and the Ministry of Health Malaysia Medical Research and Ethics Committee (Ref: (5)KKM/NIHSEC/P17-140). This study was registered with the National Medical Research

Register of Ministry of Health Malaysia (NMRR ID: NMRR-16-2452-33670). Written informed consent was obtained from each participant of this study. Anonymity and confidentiality of the participants were ensured.

Phase 1: Cross-cultural adaptation of the PEMAT into Malay

The original English version of PEMAT

PEMAT has two versions: PEMAT-P for assessing printable materials and PEMAT-A/V for assessing audiovisual materials. Each PEMAT (PEMAT-P and PEMAT-A/V) has two domains: understandability and actionability. PEMAT-P consists of 24 items (17 understandability items and seven actionability items) and the PEMAT-A/V has 17 items (13 understandability items and four actionability items) (Table I). PEMAT-P and PEMAT-A/V are self-administered and the PEMAT User's Guide is available (10).

Table I: Items, topics, and domains of Patient Education Materials Assessment Tool (PEMAT).

Domain: Understandability	
Topic: Content	
Item 1	The material makes its purpose completely evident (P and A/V)
Item 2	The material does not include information or content that distracts from its purpose (P)
Topic: Word Choice & Style	
Item 3	The material uses common, everyday language (P and A/V)
Item 4	Medical terms are used only to familiarize audience with the terms. When used, medical terms are defined (P and A/V)
Item 5	The material uses the active voice (P and A/V)
Topic: Use of Numbers	
Item 6	Numbers appearing in the material are clear and easy to understand (P)
Item 7	The material does not expect the user to perform calculations (P)
Topic: Organization	
Item 8	The material breaks or "chunks" information into short sections (P and A/V)
Item 9	The material's sections have informative headers (P and A/V)
Item 10	The material presents information in a logical sequence (P and A/V)
Item 11	The material provides a summary (P and A/V)
Topic: Layout & Design	
Item 12	The material uses visual cues (e.g., arrows, boxes, bullets, bold, larger font, highlighting) to draw attention to key points (P and A/V)
Item 13	Text on the screen is easy to read (A/V)
Item 14	The material allows the user to hear the words clearly (e.g., not too fast, not garbled) (A/V)
Topic: Use of Visual Aids	
Item 15	The material uses visual aids whenever they could make content more easily understood (e.g., illustration of healthy portion size) (P)
Item 16	The material's visual aids reinforce rather than distract from the content (P)
Item 17	The material's visual aids have clear titles or captions (P)
Item 18	The material uses illustrations and photographs that are clear and uncluttered (P and A/V)
Item 19	The material uses simple tables with short and clear row and column headings (P and A/V)
Domain: Understandability	
Domain: Actionability	
Item 20	The material clearly identifies at least one action the user can take (P and A/V)
Item 21	The material addresses the user directly when describing actions (P and A/V)
Item 22	The material breaks down any action into manageable, explicit steps (P and A/V)
Item 23	The material provides a tangible tool (e.g., menu planners, checklists) whenever it could help the user take action (P)
Item 24	The material provides simple instructions or examples of how to perform calculations (P)
Item 25	The material explains how to use the charts, graphs, tables, or diagrams to take actions (P and A/V)
Item 26	The material uses visual aids whenever they could make it easier to act on the instructions (P)

Abbreviations: A/V, PEMAT-A/V; P, PEMAT-P; PEMAT, Patient Education Materials Assessment Tool.

All PEMAT items have response options of "Disagree" and "Agree". Some items also include the response option of "Not Applicable (N/A)". These items are items number 6, 8, 9, 11, 16, 17, 18, 19, and 25 for PEMAT-P, while for PEMAT-A/V these items are items number 8, 9, 11, 12, 13, 14, 18, 19, and 25. Items rated as "Agree" were given 1 point and items rated as "Disagree" were given 0 points. The scores for the understandability and actionability domains are calculated and interpreted separately. Each score is the sum of points for all items in the domain divided by the total possible points (i.e. excluding items rated "N/A") and multiplied by 100 to yield a percentage. A higher score indicates the PEM is more understandable or actionable (10).

Cross-cultural adaptation procedure

The cross-cultural adaptation of the PEMAT-P, PEMAT-A/V, and PEMAT User's Guide into the Malay language proceeded in five stages, following the guidelines proposed by Beaton et al. (18) with slight modifications. The stages implemented were: 1) forward translation, 2) synthesis of forward translations, 3) back translation, 4) expert committee review, and 5) cognitive debriefing.

Forward translation: PEMAT-P and PEMAT-A/V were forward translated from English into Malay by two independent translators, whose mother's tongue is Malay. Both translators were bilingual with distinct backgrounds. The first translator was a dental public health specialist, while the second translator was a naive translator (a language teacher) without medical background. The translation of PEMAT User's Guide involved only the second translator because it required extensive translation work that the first translator could not commit to.

CONTINUE

Synthesis of forward translations: The two forward translations for each PEMAT were reviewed by an expert committee, comprising of a dentist, a family medicine specialist, a medical education specialist, a dental public health specialist and a biostatistician who is an expert in research methods. All the expert committee members were native speakers of Malay and proficient in English.

Back translation: PEMAT-P and PEMAT-A/V in Malay were back translated into English by other two independent translators. The translators were proficient in both English and Malay. Both translators were without medical background. To eliminate bias, they were not informed about the concepts contained in PEMAT and had no knowledge of the English version PEMAT.

Expert committee review: The two back translations were then reviewed by the same committee to compare and evaluate several equivalence aspects (conceptual, item and semantic) between the back translations and original PEMAT.

Cognitive debriefing: Cognitive debriefing sessions were conducted to evaluate the process of responding to the questions in the Malay versions of PEMAT-P and PEMAT-A/V. Five HCWs from Hospital Universiti Sains Malaysia, who were a representative sample of potential users of PEMAT, participated in the debriefing. Each debriefing session took between 1.5 to 2 hours.

At the end of these stages, the final Malay versions of PEMAT-P, PEMAT-A/V, and the user's guide are referred to as PEMAT-P(M), PEMAT-A/V(M), and the PEMAT(M) User's Guide, respectively.

Phase 2: Reliability Evaluation of PEMAT-P(M) and PEMAT-A/V(M)

Sample Size

In this phase, it required evaluation of PEMs by raters, and hence sample sizes were required for both the number of PEMs to be evaluated, and the number of raters to evaluate the PEMs. The following sample sizes were determined using a sample size calculator developed by Arifin (19).

For the internal consistency reliability assessment, the null Cronbach's alpha was set at 0.60 (20), the expected Cronbach's alpha was set at 0.90 (21), with a power of 80% and a significance level of 0.05 (two-tailed), the calculated sample size was 13 PEMs, giving a total of 26 PEMs (13 each for printable and audio-visual PEMs). Nine HCWs from various health care disciplines were recruited as raters.

For the assessment of inter-rater reliability, based on nine observations by nine HCWs, with a power of 80%, a significance level of 0.05 (one-tailed) (22), an acceptable level of intra-class correlation (ICC) of 0.60 and an expected ICC value of 0.85 (23), the calculated sample size was 10 PEMs, giving a total of 20 PEMs (10 each for printable and audio-visual PEMs).

For the assessment of test-retest (intra-rater) reliability, based on two observations by each HCW, with a power of 80%, a level of significance of 0.05 (one-tailed) (22), an acceptable level ICC of 0.60 and an expected ICC of 0.90, the calculated sample size was 12 PEMs, giving a total of 24 PEMs (12 each for printable and audio-visual PEMs). Taking the largest of these sample sizes, a sample size of 26 PEMs (13 for printable PEMs and sample size of 13 for audiovisual PEMs) was taken to cover all statistical analyses to determine the reliability of PEMAT-P(M) and PEMAT-A/V(M).

Sampling Procedure

For the internal consistency reliability analysis, nine MOH HCWs from nine different job categories (medical doctor, medical assistant, medical nurse, dentist, dental therapist, pharmacist, occupational therapist, food technologist, and nutritionist) were invited as raters. For these HCWs, the work responsibilities of each HCW includes patient education. The raters were also involved in the provision of patient education and proficient in the Malay language. Each rater evaluated 13 printable (three booklets, five pamphlets, and five posters) and 13 audiovisual (multimedia/videos) PEMs of various health care topics produced by the MOH. The raters were instructed to read the PEMAT(M) User's Guide, read or view the PEMs, and then assess each PEM using the PEMAT-P(M) or PEMAT-A/V(M). The evaluation was repeated after two weeks to examine the test-retest reliability.

Statistical Analysis

Data were entered and analysed using Microsoft Excel 2016, IBM SPSS Statistics version 22.0 and R version 3.5.1. The internal consistency reliability of the understandability and actionability scales for PEMAT-P(M) and PEMAT-A/V(M) were evaluated using modified Cronbach's alpha (24), owing to the presence of N/A option for some items. A Cronbach's alpha value above 0.60 was considered acceptable (20). It should be noted that, although the reliability of a scale with binary responses is typically determined by Kuder-Richardson Formula 20 (KR20), KR20 is just a special case of Cronbach's alpha (25).

The inter-rater reliability of PEMAT-P(M) and PEMAT-A/V(M) were evaluated at domain and item levels, while the test-retest reliability was evaluated at the domain level. At the domain level, the inter-rater and test-retest reliability of understandability and actionability domains was determined by intra-class correlation (ICC) coefficient. The ICC case for the inter-rater reliability was a two-way random effect model, average measurement and absolute agreement type; while the ICC case for the

test-retest reliability was a two-way mixed effect model, single measurement and consistency type (26). The ICC strength of agreement was defined as poor (<0.40), fair (0.40 to 0.59), good (0.60 to 0.74), or excellent (0.75 to 1.00) (27). At the item level, the inter-rater reliability (agreement) of understandability and actionability items was analyzed by Gwet's AC1 (28, 29). The Gwet's AC1 strength of agreement adopted the interpretation for the kappa coefficient, which was defined as slight (0 to 0.20), fair (0.21 to 0.40), moderate (0.41 to 0.60), substantial (0.61 to 0.80) and almost perfect (0.81 to 1.00) (30). Overall, ICC and Gwet's AC1 values above 0.60 were considered acceptable.

Phase 3: Validity of PEMAT-P(M) and PEMAT-A/V(M) against Suitability Assessment of Materials

The Suitability Assessment of Materials instrument

Developed by Doak, Doak and Root (4), the Suitability Assessment of Materials (SAM) instrument contains 22 items arranged in six domains: content (four items), literacy demand (five items), graphics (five items), layout and type (three items), learning stimulation and motivation (three items), and cultural appropriateness (three items). Each item was rated using a three-point rating scale (0 = not suitable, 1 = adequate, 2 = superior). Items that were irrelevant to the evaluated PEMs were recorded as "not applicable (N/A)". The score for each PEM was calculated by adding the points for all items (excluding "not applicable" items), dividing this number by the highest possible score, and multiplying by 100 to yield a percentage (4).

Sample Size

In Phase 3, the validity of PEMAT(M) with respect to its relationship with other tools was evaluated using the correlation between PEMAT(M) scores and SAM scores. The sample size was determined using the sample size calculator developed by Arifin (19). The expected correlation coefficient (r) was set at 0.6 based on the study by Morony et al. (31). With a power of 80% and a significance level of 0.05, the estimated sample size was 20 PEMs. Hence, 20 printable PEMs and 20 audiovisual PEMs of various health care topics were needed, which gave a total number of 40 PEMs. Accordingly, 20 HCWs were recruited with each of them assessing two PEMs (one printable PEM and one audiovisual PEM).

Sampling Procedure

For the correlation analysis with the SAM instrument, 20 HCWs were invited as raters. The raters were involved in the provision of patient education and bilingual in both Malay and English. Each rater assessed two PEMs (one printed and one audiovisual) using PEMAT-P(M), PEMAT-A/V(M), and the SAM instrument.

Statistical Analysis

Data were entered and analysed using Microsoft Excel 2016 and IBM SPSS Statistics version 22.0. The validity of PEMAT-P(M) and PEMAT-A/V(M) with respect to its relationship with other tools was determined by Spearman's correlation analysis because the data were not normally distributed. PEMAT-P(M) and PEMAT-A/V(M) domain scores were correlated with scores from other instrument with established validity, namely the SAM instrument (4), and the Consumer Testing Questionnaire (10). The correlation coefficients (positive or negative values) were classified as little or none (0.00 to 0.25), low (0.26 to 0.49), moderate (0.50 to 0.69), high (0.70 to 0.89), and very high (0.90 to 1.00) (32).

Phase 4: Validity of PEMAT-P(M) and PEMAT-A/V(M) Against the Consumer Testing Questionnaire

Consumer Testing Questionnaire

The Consumer Testing Questionnaire was developed based on an interview protocol developed by Shoemaker, Wolf and Brach (11) to test the health literacy of patients. Sixteen questionnaires were developed to assess 16 PEMs (eight printed and eight audiovisual). Each questionnaire had four parts: 1) comprehension; 2) numeric; 3) open-ended questions regarding content that could influence understandability or actionability; and 4) verbal description of the contents, including the graphics (only for printable materials).

The comprehension part has two to three questions about the content of the PEM. Each correct answer was given 1 mark. The total mark achieved was divided by the total possible mark, and then multiplied by 100 to yield a percentage comprehension score. The numeric part assessed perceived understandability or actionability of PEM. The respondents were asked to rate the PEM on a scale of 1 to 10, in which "1" indicated the hardest to understand/least actionable and "10" indicated the easiest to understand/most actionable. The consumer testing score was the sum of the comprehension and the numeric scores. The third and fourth parts have open-ended questions to allow respondents to give their opinions and understandings of the PEM.

Sample Size

The validity of PEMAT(M) was further evaluated by correlating PEMAT(M) and with the Consumer Testing Questionnaire scores. The sample size was determined using the sample size calculator developed by Arifin (19). The following parameters were set: an expected correlation coefficient (r) of 0.85 (11), a power of 80% and a significance level of 0.05. The estimated sample size required was eight PEMs. Since there were two versions of PEMAT(M), eight printable PEMs and eight audiovisual PEMs of various health care topics were needed, giving a total of 16 PEMs. Accordingly, 16 adult patients were recruited – one for each PEM.

Sampling Procedure and Statistical Analysis

For the correlation analysis, 16 participants aged 18 and above, proficient in the Malay language and with a high

school education were recruited. Each participant was requested to read or view one of the PEMs, followed by a structured interview using the Consumer Testing Questionnaire. The PEMs were randomly assigned. All PEMs used in this study were either written or broadcasted in the Malay language.

Statistical Analysis

The data handling and statistical analysis procedures were the same as the ones implemented in Phase 3 as described above. The validity of PEMAT-P(M) and PEMAT-A/V(M) in relation to its relationship with the Consumer Testing Questionnaire was determined by Spearman's correlation analysis because the data were not normally distributed.

RESULTS

Cross-cultural adaptation of the PEMAT into PEMAT(M)

The cross-cultural adaptation of the PEMAT-P, PEMAT-A/V, and PEMAT User's Guide into the Malay language proceeded in five stages as planned. During the second stage (synthesis of forward translations), the expert committee resolved discrepancies between the two forward translations and synthesized them into a single Malay translation for each PEMAT. For the PEMAT User's Guide, the committee replaced and made modifications to the graphs, tables, charts, diagrams, and pictures from PEMs developed by the United States National Institutes of Health and the Centers for Disease Control and Prevention with those taken from PEMs developed by the MOH to reflect local context, culture and language. The committee also commissioned a graphic designer to replace an original illustration featuring an American baby with an Asian baby.

Following the back translation stage, expert committee reviewed the back translations and agreed that all PEMAT-P(M) and PEMAT-A/V(M) items were relevant, representative and had adequate coverage of the concepts to be measured. The committee also agreed that domains, subdomains and items of PEMAT(M) were relevant to the Malaysian context, culture and setting. The similarity of meaning for each item between the original and adapted versions were also agreed upon. This led to the establishment of conceptual, item, and semantic equivalences of the PEMAT(M) versions. The committee also agreed that the PEMAT(M) User's Guide was appropriate to the Malaysian context, culture, and setting.

The feedback during the cognitive debriefing was favourable. All respondents commented that the Malay versions of PEMAT-P and PEMAT-A/V were user friendly and easy to apply in assessing PEMs. However, they found that it was necessary to consult the cross-culturally adapted PEMAT User's Guide in Malay during the assessment process. For the last time, the expert committee evaluated all issues, comments and suggestions provided by the respondents, and made necessary revisions. PEMAT-P(M), PEMAT-A/V(M), and the PEMAT(M) User's Guide were finalized at this study phase.

Reliability Evaluation of PEMAT-P(M) and PEMAT-A/V(M)

For the reliability evaluation of PEMAT-P(M) and PEMAT-A/V(M), 13 PEMs each for printable and audiovisual materials were utilized. Nine HCWs were involved in evaluating these PEMs using the Malay version of PEMAT.

Both PEMAT-P(M) and PEMAT-A/V(M) showed good internal consistency. The median (IQR) of the Cronbach's alpha values from nine HCWs were: PEMAT-P(M) understandability, $\alpha = 0.78$ (0.06); PEMAT-P(M) actionability, $\alpha = 0.89$ (0.08); PEMAT-A/V(M) understandability, $\alpha = 0.90$ (0.135); and PEMAT-AV(M) actionability, $\alpha = 0.83$ (0.224).

The domain-level inter-rater reliability of PEMAT-P(M) and PEMAT-A/V(M) domain scores are shown in Table II (highlighted in bold). The ICC values ranged from 0.72 to 0.81, indicating good to excellent inter-rater agreement. The item-level inter-rater reliability for most items showed substantial agreement with Gwet's AC1 values above 0.60 (Table II). The Gwet AC1 values for PEMAT-P(M) were fair (0.21 to 0.40) for Item 15 and Item 26 at 0.26 and 0.39 respectively, while the values were moderate (0.41 to 0.60) for Items 6, 16, 17, 18, 22, and 23. For PEMAT-A/V(M), the Gwet AC1 value was fair for Item 12 at 0.40, whereas the values were moderate for Items 13 and 18 (0.57 and 0.55 respectively). Although these values fell below the general target cutoff value (> 0.60) for reliability in this study, domain-level reliability via ICC was prioritized as it reflects the collective performance of the domain compared to the item-by-item assessment.

Table II: Inter-rater reliability of PEMAT-P(M) and PEMAT-A/V(M) at domain level by intraclass correlation (ICC) and item level by Gwet's AC1 (k = 9 raters, n = 13 patient education materials)

Instrument	PEMAT-P(M)	PEMAT-A/V(M)
	Gwet's AC1 (95% CI)	Gwet's AC1 (95% CI)
	Domain: Understandability	
Domain ICC (95% CI)*	**0.74 (0.57, 0.89)**	**0.81 (0.66, 0.92)**
Item 1	0.68 (0.43, 0.93)	0.73 (0.47, 0.98)
Item 2	0.84 (0.70, 0.98)	–
Item 3	0.85 (0.68, 1.00)	0.83 (0.70, 0.97)
Item 4	0.73 (0.52, 0.94)	0.71 (0.47, 0.96)
Item 5	0.94 (0.84, 1.00)	0.87 (0.76, 0.99)
Item 6	0.52 (0.30, 0.75)	–
Item 7	0.92 (0.79, 1.00)	–
Item 8	0.88 (0.78, 0.98)	0.92 (0.79, 1.00)
Item 9	0.91 (0.80, 1.00)	0.83 (0.63, 1.00)
Item 10	0.79 (0.58, 1.00)	0.71 (0.44, 0.99)
Item 11	0.72 (0.50, 0.94)	0.74 (0.51, 0.97)
Item 12	0.87 (0.76, 0.99)	0.40 (0.04, 0.75)
Item 13	–	0.57 (0.36, 0.78)
Item 14	–	0.76 (0.61, 0.91)
Item 15	0.26 (0.03, 0.49)	–
Item 16	0.56 (0.26, 0.86)	–
Item 17	0.50 (0.26, 0.74)	–
Item 18	0.60 (0.33, 0.88)	0.55 (0.30, 0.80)
Item 19	0.96 (0.91, 1.00)	0.93 (0.84, 1.00)
Median (IQR)	*0.79 (0.31)*	*0.74 (0.21)*
	Domain: Actionability	
Domain ICC (95% CI)*	**0.72 (0.54, 0.88)**	**0.73 (0.56, 0.89)**
Item 20	0.98 (0.93, 1.00)	0.85 (0.66, 1.00)
Item 21	0.88 (0.75, 1.00)	0.65 (0.36, 0.94)
Item 22	0.52 (0.27, 0.77)	0.68 (0.41, 0.94)
Item 23	0.58 (0.33, 0.83)	–
Item 24	0.87 (0.70, 1.00)	–
Item 25	0.90 (0.78, 1.00)	0.91 (0.75, 1.00)
Item 26	0.39 (0.10, 0.69)	–
Median (IQR)	*0.87 (0.38)*	*0.76 (0.24)*

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient ; PEMAT-A/V(M), PEMAT-A/V (Malay version); PEMAT-P(M), PEMAT-P (Malay version); PEMAT, Patient Education Materials Assessment Tool. *ICC case: Two-way random effect model, average measurement, absolute agreement type.

For the test-retest (intra-rater) reliability, PEMAT-P(M) and PEMAT-A/V(M) domain scores showed excellent temporal stability. The median (IQR) of the ICC values from nine HCWs were: PEMAT-P(M) understandability, ICC = 0.90 (0.12); PEMAT-P(M) actionability, ICC = 0.93 (0.07); PEMAT-A/V(M) understandability, ICC = 0.92 (0.06); and PEMAT-AV(M) actionability, ICC = 0.85 (0.20).

Validity of PEMAT-P(M) and PEMAT-A/V(M) Against Other Measures

For the correlation between the scores of the Malay version of PEMAT and SAM, 20 PEMs for printable and audiovisual materials respectively were utilized. These were evaluated by 20 HCWs. For the correlation between the scores of the Malay versions of PEMAT and Consumer Testing Questionnaire (comprehension and numeric parts), eight PEMs for printable and audiovisual materials respectively were utilized. These were evaluated by 16 adult patients. The results are presented in Table III. The understandability and actionability domain scores of PEMAT-P(M) and PEMAT-A/V(M) had significant positive correlations with the SAM scores. The range of correlation coefficients indicated moderate to high correlation.

Table III: Correlation analysis of PEMAT-P(M) and PEMAT-A/V(M) scores with Suitability Assessment of Materials (SAM) (*k* = 20 raters, *n* = 20 patient education materials [PEMs]) and Consumer Testing Questionnaire (*k* = 16 patients, *n* = 8 PEMs) scores

Instrument	Domain	SAM score		Consumer testing comprehension score		Consumer testing numeric score	
		ρ	***p*-value**	ρ	p-value	ρ	p-value
PEMAT-P(M)	Understandability	0.793	<0.001	0.771	0.025	0.490	0.217
	Actionability	0.653	0.002	0.811	0.015	0.563	0.147
PEMAT-A/V(M)	Understandability	0.684	0.001	0.814	0.014	0.798	0.018
	Actionability	0.526	0.017	0.304	0.464	0.578	0.133

Abbreviations: PEM, patient education materials; PEMAT-A/V(M), PEMAT-A/V (Malay version); PEMAT-P(M), PEMAT-P (Malay version); PEMAT, Patient Education Materials Assessment Tool; SAM, . Symbol: ρ, Spearman's rank correlation coefficient.

Both PEMAT-P(M) domains showed significant correlation with the consumer testing comprehension score, while no significant correlation was found with the consumer testing numerical score. The understandability domain of PEMAT-A/V(M) was significantly correlated with both parts of the consumer testing scores, while no significant correlation was found for the actionability domain.

DISCUSSION

This study adequately produced cross-culturally adapted PEMAT-P and PEMAT-A/V in the Malay language and determined their psychometric properties. Apart from adapting PEMAT-P and PEMAT-A/V, it also cross-culturally adapted and produced the User's Guide for PEMAT. The median Cronbach's alpha values for understandability and actionability domains ranged between 0.78 to 0.90. The domain-level inter-rater reliability showed ICC values ranging from 0.72 to 0.81, and the test-retest reliability showed ICC values ranging from 0.82 to 0.90. The item-level inter-rater reliability showed Gwet's AC1 values above 0.60 for most items. Significant positive correlations were found between all PEMAT-P(M) and PEMAT-A/V(M) scores with the SAM score. Most domain scores were significantly correlated with the consumer testing comprehension score, while only one domain score was significantly correlated with consumer testing numerical score.

Kang and Lee (12) translated the PEMAT-A/V into the Korean language. However, the authors did not report the psychometric evaluation of the Korean version. Additionally, although the PEMAT User's Guide was also translated, the content, particularly examples in the User's Guide, were not fully adapted to suit the Korean context and culture (12). In this study, the use of established guidelines (18) to cross-culturally adapt the PEMAT-P and PEMAT-A/V into Malay has helped to ensure that the translated versions are contextually appropriate, consistent and equivalent with the original English versions. The adaptation of the PEMAT User's Guide into Malay can be considered comprehensive as we also replaced all original graphs, tables, charts, diagrams, and pictures with those from the MOH PEMs to ensure relevance of the content to the Malaysian context and culture.

The original PEMAT employed two raters to evaluate 23 printable and 23 audio-visual materials of varied characteristics in English (11). The average Gwet's AC1 for PEMAT-P understandability items, PEMAT-P actionability items, PEMAT-A/V understandability items, and PEMAT-A/V actionability items were 0.76, 0.72, 0.71 and 0.72 respectively. For a study by Vishnevetsky et al. (33), they recruited two raters to evaluate 110 oncology-related materials in English (80 printable and 30 audio-visual). The findings of this study displayed median Gwet's AC1 for each domain for PEMAT version as follows: PEMAT-P understandability items (0.94), PEMAT-P actionability items (0.93), PEMAT-A/V understandability items (0.84), and PEMAT-A/V actionability items (0.97). The findings of this study for the Malay version of PEMAT were not markedly different from the original PEMAT with median Gwet's AC1 for PEMAT-P understandability items of 0.79, PEMAT-P actionability items of 0.87, PEMAT-A/V understandability items of 0.87, and PEMAT-A/V actionability items of 0.76. Furukawa et al. (13) translated PEMAT into Japanese. Their study evaluated 100 printable and 50 audio-visual materials by four raters of diverse backgrounds. The Japanese version of PEMAT also demonstrated strong inter-rater reliability, with an average Gwet's AC1 of 0.83 for the Japanese PEMAT-P and 0.80 for the Japanese PEMAT-A/V, respectively. This finding is comparable to the results of our study.

Similar to the study by Furukawa et al. (13), this study included materials from diverse health topics. The strength of this study compared to the study by Furukawa et al. (13) was that we involved more raters (n = 9) of diverse health backgrounds to evaluate Malay materials. Shan et al. (15) translated PEMAT-P into Chinese. This study involved two health educators as raters of 15 print materials on air pollution and health. The internal consistency of Chinese version of PEMAT-P using Cronbach α was 0.847 and 0.751 for understandability and actionability domains respectively. In addition, Turkish version of PEMAT was produced by Akkoc et al. (14), involving 248 intern nurses for PEMAT-P and 223 intern nurses for PEMAT-A/V evaluating

five printable and five audio-visual health-related materials. This study reported the following findings on internal consistency: PEMAT-P understandability (0.89), PEMAT-P actionability (0.77), PEMAT-A/V understandability (0.85), and PEMAT-A/V actionability (0.93). The current study on Malay version of PEMAT showed almost similar findings for internal consistency with the Chinese version of PEMAT-P; and Turkish version of PEMAT-P and PEMAT-A/V. This study was better with the inclusion of diverse HCWs and health topics.

The Cronbach's alpha values of PEMAT-P(M) and PEMAT-A/V(M) met the recommended internal consistency threshold for research tools (34). Most items in both instruments had good correlation with other items in the same scale (13). Items with low or missing corrected item-total correlations were retained due to their contribution to the overall content validity of the instruments. This is in agreement with Shoemaker, Wolf and Brach (11) who retained items with low item-level reliability values because of a strong theoretical foundation supporting their inclusion.

The results demonstrated that the understandability and actionability domains of both instruments had good to excellent inter-rater reliability although the raters had varying levels of educational attainment and work experience. This may be attributable to the clear instructions and grading of each item provided in the PEMAT(M) User's Guide. The instruments were also extremely reliable when performed at different times by the same rater.

Positive significant correlations of the SAM score with scores of both domains in the PEMAT-P(M) and PEMAT-A/V(M) were observed, and the strength of these correlations ranged from moderate to high, establishing the concurrent validity of PEMAT(M). The highest correlation was between the SAM score and the PEMAT-P(M) understandability score, possibly due to close similarity of the construct being measured. Morony et al. (31) and Avis et al. (35) reported similar findings, with correlations between the PEMAT-P and the SAM scores ranging from 0.62 to 0.78. The measurements of the correlations of the two domains within the PEMAT-P(M) and the PEMAT-A/V(M) with the consumer testing scores provided further evidence of concurrent validity. In general, both instruments captured patients' understandability and actionability about the PEM they read or viewed. The qualitative assessment indicated that certain qualities of the PEMs could have distracted patients from understanding or performing the recommended actions, in agreement with the results of Shoemaker, Wolf and Brach (11).

This study has several limitations. The raters for the reliability testing were purposively sampled and the number was small due to the difficulty in recruiting HCWs who would be willing to assess the PEMs twice. The choice of non-probability, purposive sampling may affect the generalizability of the study findings; however, we attempted to mitigate this issue by recruiting raters from a variety of healthcare disciplines. Future studies may recruit more HCWs for this purpose to ensure more robust reliability results. We also limited the number of PEMs to be assessed by each rater to prevent an excessive burden on the raters. In future research, the burden can be reduced by having a group of raters to rate a larger pool of PEMs, where a rater may rate only a portion of the PEMs. This requires preliminary inter-rater reliability testing of the group of raters to ensure they are interchangeable in rating the PEMs. In addition, both versions of PEMAT(M) should be tested in broader community settings or with different patient populations.

CONCLUSIONS

The results suggest that the PEMAT-P(M) and PEMAT-A/V(M) are culturally appropriate, reliable, and valid for assessment of PEMs in the Malay language. Both instruments can be used by untrained raters from diverse backgrounds. The Malay version of PEMAT can provide a common and systematic evaluation tool for printable and audiovisual materials. By facilitating the production of understandable and actionable PEMs tailored to the linguistic and cultural context of the local Malaysian audience, it ensures effective health communication to the public.

ACKNOWLEDGEMENTS

The authors thank the Agency for Healthcare Research and Quality (AHRQ), Department of Health and Human Resources, United States of America, for giving the permission to cross-culturally adapt PEMAT into the Malay language. The authors also gratefully acknowledge the Director General of Health, Ministry of Health Malaysia, for granting the permission to conduct this study at the Ministry's facilities and for authorizing the publication of this manuscript. The publication of this article was funded by the Research Creativity and Management Office and the School of Medical Sciences, Universiti Sains Malaysia.

REFERENCES

1. Giguère A, Zomahoun HTV, Carmichael PH, Uwizeye CB, Légaré F, Grimshaw JM, et al. Printed educational materials: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2020;8(8):CD004398. doi: 004310.001002/14651858.CD14004398.pub14651854.
2. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int. 2000;15(3):259-267.

3. Pothier L, Day R, Harris C, Pothier DD. Readability statistics of patient information leaflets in a Speech and Language Therapy Department. Int J Lang Commun Disord. 2008;43(6):712-22.
4. Doak CC, Doak LG, Root JH. Teaching patients with low literacy skills. 2nd ed. Philadelphia: J.B. Lippincott; 1996.
5. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53(2):105-11.
6. Moody C, Rose M. Literacy and Health: Defining Links and Developing Partnerships. A Final Report to Population Health. Literacy Partners of Manitoba; 2004.
7. Northwest Territories Literacy Council. A Plain Language Audit Tool [Internet]. 2008. Available from: [https://www.ece.gov.nt.ca/files/T4.05.01_Plain%20Language%20Audit%20Tool.pdf] (https://www.ece.gov.nt.ca/files/T4.05.01_Plain%20Language%20Audit%20Tool.pdf)
8. Clayton LH. TEMPtEd: development and psychometric properties of a tool to evaluate material used in patient education. J Adv Nurs. 2009;65(10):2229-38.
9. Kaphingst KA, Kreuter MW, Casey C, Leme L, Thompson T, Cheng MR, et al. Health Literacy INDEX: development, reliability, and validity of a new tool for evaluating the health literacy demands of health information materials. J Health Commun. 2012;17(sup3):203-21.
10. Shoemaker S, Wolf M, Brach C. The Patient Education Materials Assessment Tool (PEMAT) and User's Guide. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
11. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns. 2014;96(3):395-403. doi: 310.1016/j.pec.2014.1005.1027.
12. Kang SJ, Lee MS. Assessing of the audiovisual patient educational materials on diabetes care with PEMAT. Public Health Nurs. 2019;36(3):379-387. doi: 310.1111/phn.12577.
13. Furukawa E, Okuhara T, Okada H, Shirabe R, Yokota R, Iye R, Kiuchi T. Translation, cross-cultural adaptation, and validation of the Japanese version of the patient education materials assessment tool (PEMAT). Int J Environ Res Public Health. 2022;19(23):15763. doi: 10.3390/ijerph192315763.
14. Akkoз CP, Orgun F. Psychometric testing of the Turkish version of the patient education materials assessment tool. Florence Nightingale J Nurs. 2023;31(3):180. doi: 10.5152/FNJN.2023.22196.
15. Shan Y, Ji M, Dong Z, Xing Z, Wang D, Cao X. The Chinese version of the patient education materials assessment tool for printable materials: Translation, adaptation, and validation study. J Med Internet Res. 2023;25:e39808.
16. AERA, APA, NCME. The standards for educational and psychological testing. Washington DC: American Educational Research Association; 2014.
17. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006;119:166e7–166e16.
18. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. Spine. 2000;25(24):3186-3191.
19. Arifin WN. Sample Size Calculator (Version 2) [Spreadsheet file]. Available from: http://wnarifin.github.io [Accessed 20th January 2017] 2017.
20. Nunnally JC. Psychometric Theory. New York: McGraw-Hill 1967.
21. Bland JM, Altman DG. Statistics Notes: Cronbach's alpha. BMJ. 1997;314(7080):572.
22. Walter S, Eliasziw M, Donner A. Sample Size and Optimal Designs for Reliability Studies. Stat Med. 1998;17(1):101-110.
23. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284-289.
24. Arifin WN. Calculating the Cronbach's alpha coefficients for measurement scales with "not applicable" option. Preprint. 2018. doi: 10.13140/RG.2.2.16955.87843.

25. Anselmi P, Colledani D, Robusto E. A Comparison of Classical and Modern Measures of Internal Consistency. Front Psychol. 2019;10. doi: 10.3389/fpsyg.2019.02714.
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-163. doi: 110.1016/j.jcm.2016.1002.1012.
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284-290. doi: 210.1037/1040-3590.1036.1034.1284.
28. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(1), 29-48.
29. Gwet KL. R functions for calculating agreement coefficients [Online]. Available from: http://agreestat.com/r_functions.html [Accessed 25th March 2018] 2010.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.
31. Morony S, McCaffery KJ, Kirkendall S, Jansen J, Webster AC. Health Literacy Demand of Printed Lifestyle Patient Information Materials Aimed at People With Chronic Kidney Disease: Are Materials Easy to Understand and Act On and Do They Use Meaningful Visual Aids? J Health Commun. 2017;22(2):163-170.
32. Munro BH. Statistical Methods for Health Care Research. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
33. Vishnevetsky J, Walters CB, Tan KS. Interrater reliability of the Patient Education Materials Assessment Tool (PEMAT). Patient Educ Couns. 2018;101(3):490-496. doi: 410.1016/j.pec.2017.1009.1003.
34. DeVellis RF. Scale Development: Theory and Applications. 3rd ed. Thousand Oaks: CA: Sage Publications; 2012.
35. Avis JL, Komarnicki A, Farmer AP, Holt NL, Perez A, Spence N, et al. Tools and resources for preventing childhood obesity in primary care: A method of evaluation and preliminary assessment. Patient Educ Couns. 2016;99(5):769-775. doi: 710.1016/j.pec.2015.1012.1006.