HHS Public Access

Author manuscript

Patient Educ Couns. Author manuscript; available in PMC 2016 October 28.

Published in final edited form as:

Patient Educ Couns. 2014 September ; 96(3): 395–403. doi:10.1016/j.pec.2014.05.027.

Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information

Sarah J. Shoemaker[a,*], **Michael S. Wolf**[b], and **Cindy Brach**[c]

[a]Health Policy, Abt Associates, Inc., Cambridge, USA

[b]Feinberg School of Medicine, Northwestern University, Chicago, USA

[c]Agency for Healthcare Research and Quality (AHRQ), Rockville, USA

Abstract

Objective—To develop a reliable and valid instrument to assess the understandability and actionability of print and audiovisual materials.

Methods—We compiled items from existing instruments/guides that the expert panel assessed for face/content validity. We completed four rounds of reliability testing, and produced evidence of construct validity with consumers and readability assessments.

Results—The experts deemed the PEMAT items face/content valid. Four rounds of reliability testing and refinement were conducted using raters untrained on the PEMAT. Agreement improved across rounds. The final PEMAT showed moderate agreement per Kappa (Average $K = 0.57$) and strong agreement per Gwet's AC1 (Average = 0.74). Internal consistency was strong ($\alpha = 0.71$; Average Item-Total Correlation = 0.62). For construct validation with consumers ($n = 47$), we found significant differences between actionable and poorly-actionable materials in comprehension scores (76% vs. 63%, $p < 0.05$) and ratings (8.9 vs. 7.7, $p < 0.05$). For understandability, there was a significant difference for only one of two topics on consumer numeric scores. For actionability, there were significant positive correlations between PEMAT scores and consumer-testing results, but no relationship for understandability. There were, however, strong, negative correlations between grade-level and both consumer-testing results and PEMAT scores.

Conclusions—The PEMAT demonstrated strong internal consistency, reliability, and evidence of construct validity.

Practice implications—The PEMAT can help professionals judge the quality of materials (available at: http://www.ahrq.gov/pemat).

[*]Corresponding author at: Abt Associates, Inc. Wheeler Street, Cambridge, MA 02138, USA. Tel.: +1 617 349 2472; fax: +1 617 386 7638. sarah_shoemaker@abtassoc.com, sarahjshoemaker@gmail.com (S.J. Shoemaker).

Conflicts of interest

All authors state that there are no other actual or potential conflicts of interest including any financial, personal or other relationships with other people or organizations within 3 years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work.

Keywords

Health literacy; Assessment; Measurement; Instrument development; Patient education; Educational materials; Audiovisual materials; Plain language; Clear communication; Readability

1. Introduction

Health literacy is the capacity to "obtain, process and understand basic health information and services needed to make appropriate health decisions" [1]. Health literacy is well recognized as a challenge for public health, with many adults lacking the requisite skills to engage successfully in their health care. Recent systematic reviews have confirmed that low health literacy is strongly associated with poorer use of health care and subsequent health outcomes, leading to higher use of emergency departments and inpatient beds [2–4].

While the skills of individuals are an important part of health literacy, the field has come to recognize that the demands placed on individuals by the health system and professionals are an important determinant [5–9]. To address health literacy, the U.S. Department of Health and Human Services' *National Action Plan to Improve Health Literacy (National Action Plan)* promotes a multi-sector effort to improve health literacy, including reducing the demands placed on individuals [10]. A key goal of the *National Action Plan* is "to develop and disseminate health and safety information that is accurate, accessible, and actionable" [10]. Studies assessing the readability, suitability or comprehensibility of patient education materials on a myriad of topics are abound, and the evidence is clear that most education materials are too complex for patients with low health literacy.

1.1. Limitations of current patient education materials assessments

Organizations and professionals aspiring to produce low-demand patient education materials have a selection of guides that provide instruction [11–15]. Readability formulas [13,16] are commonly relied upon to assess whether written materials are in fact low-demand. Readability formulas provide quantitative estimates, in the form of a grade level, of the reading difficulty of written information based on word and sentence difficulty. Yet readability formulas ignore several factors that contribute to comprehension [13].

In recognition of the shortcomings of readability formulas, several checklists and instruments that assess the health literacy demand of materials have been developed [17–23], including two newly-developed instruments [24,25]. These assessment tools, however, have not shown inter-rater reliability [17], were developed or tested with a specific topic or aim [18–20], or were tested using only raters trained in the use of the instrument [24,25]. Furthermore, most are applicable only to print material, and none measure whether materials are actionable – an important characteristic of materials called for in the first goal of the *National Action Plan* [10].

1.2. Aim

The aim of this study was to develop a reliable and valid Patient Education Materials Assessment Tool (PEMAT) to be used by untrained lay and health professionals alike to

assess the understandability and actionability of both printable (e.g., printed materials like brochures or pamphlets or materials that can be printed from websites like PDFs) and audiovisual (e.g., video or multi-media presentation with or without narration) patient education materials. We defined understandability and actionability as follows:

> *Understandability*: Patient education materials are *understandable* when consumers of diverse backgrounds and varying levels of health literacy can process and explain key messages.
>
> *Actionability*: Patient education materials are *actionable* when consumers of diverse backgrounds and varying levels of health literacy can identify what they can do based on the information presented.

2. Methods

The PEMAT was iteratively and systematically developed with repeated input from a panel of experts in health literacy; health communications; content creation, including different modalities; patient education; communication; patient engagement, and health information technology. The experts included clinicians, researchers, policymakers, academicians, and staff from nonprofit, for-profit and governmental organizations. (See Acknowledgements for the list of experts.)

Once we defined understandability and actionability in collaboration with the expert panel, the research team: (1) reviewed existing instruments and guides for assessing or developing materials and identified relevant topics and items; (2) assessed the face and content validity using experts; (3) determined the reliability (external and internal consistency); and (4) assessed the construct validity by conducting testing with consumers and comparing understandability results to readability assessments.

2.1. Review existing instruments and guides

We searched for and reviewed existing instruments and guides to assess and develop materials to identify both concepts thought to be related to understandability and actionability and actual items to be considered for inclusion in the PEMAT. We identified these from a literature search of PubMed, online searches of health literacy organizations' websites, resources recommended by the expert panel, and consultation with other experts in the field. Topics (e.g., content, numeracy, quality of visual aids) and associated items were compiled from the initial set of instruments and guides identified. We also discussed challenges and ideas for improvement with individuals developing related instruments [24,25].

2.2. Face and content validity

Our expert panel judged the relevance of the topics and items identified for each scale (understandability and actionability) in the initial item pool to establish face and content validity. There were 36 items in the initial item pool: understandability (7 topics with 28 items), and actionability (8 items). Nine expert panel members indicated whether they thought a patient education material's performance on an item would affect its

understandability or actionability (Yes = 1, Maybe = 2, and No = 3) from which we calculated an average score. We considered dropping items that received an average score ≥1.5. We discussed the results with the expert panel, identified potential gaps, and discussed the relevance of some items for the different types of materials (e.g., multimedia). The results of the initial validation served to develop the item pool for the first version of the PEMAT. Because we made considerable changes to the previously-validated items and developed new ones as a result of the first two rounds of reliability testing (see Section 2.4), we had a subset of the expert panel revalidate the items and provide input on specific items that would be relevant for specific material types (e.g., video, multimedia). The resulting items went through two additional rounds of reliability testing. (See Section 2.3.)

2.3. Reliability

We completed four rounds of reliability testing with a total of 22 different raters. In all rounds, raters were staff at one of the project team organizations (but were not involved in the development of the PEMAT), had research or clinical experience and at least a Bachelor's degree. The rounds varied in terms of the number of raters, material characteristics, and characteristics of the PEMAT (see Table 1). Because of the low reliability found in round 1 (see Section 3.3.1), steps were taken to minimize variation including dropping Spanish materials and changing from the 4-point (strongly disagree = 1, disagree = 2, agree = 3, strongly agree = 4, or not applicable) to the 2-point scale for the remaining rounds (disagree = 1, agree = 2, or not applicable). In rounds 1 and 2 raters were only provided the instrument, whereas in rounds 3 and 4 raters were also given a Users' Guide that was developed because of the low reliability in rounds 1 and 2 No training was provided to raters in any rounds. In all rounds, raters were asked to follow the instrument instructions and rate every material on each item. In round 4 we calibrated the raters to the task by having them each rate four materials, then discussing the disagreement and identifying what was still unclear.

2.3.1. External consistency—Inter-rater reliability (IRR) was used to assess the external consistency of the PEMAT, using percentage agreement and either Fleiss' kappa for more than 2 raters [26] or Cohen's kappa for 2 raters [27]. Agreement was deemed poor (0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.0) [28]. We also calculated Gwet's AC1 [29] – an alternative statistic developed in response to a critique of the kappa statistic when low kappas occur despite high percentage of agreement [30].

2.3.2. Internal consistency—We set Cronbach's coefficient alphas >0.7 and item-total correlations >0.2 as acceptable levels for both understandability and actionability scales.

2.4. Construct validity – consumer testing

To establish construct validity, we conducted consumer testing to determine whether consumers understood materials rated understandable on the PEMAT better than materials rated poorly understandable, and whether consumers know what actions to take based on materials rated actionable on the PEMAT better than materials rated poorly actionable.

2.4.1. Consumers—We conducted interviews with a convenience sample of U.S. adults. Consumers were excluded if they: were under18 years of age; did not speak English; had substantial knowledge of the topics of the materials chosen for testing (i.e., received or cared for someone who had received a colonoscopy; had asthma or used an inhaler; or had participated in research in the past 6 months). We sought diversity in gender, race, ethnicity and education, especially inclusion of adults with a high school diploma or less education. We engaged an outside agency to identify and recruit individuals who met these criteria.

2.4.2. Patient education materials—We used twelve (12) different publicly-available patient education materials in consumer testing, consisting of two modalities (6 printable and 6 audiovisual) and two topics (6 colonoscopy and 6 inhaler/asthma). Two research team members rated the materials using the PEMAT, generating separate understandability and actionability scores. PEMAT scores are calculated by taking the sum of points, dividing by the total possible points (i.e., exclude not applicable items), and multiplying by 100 to get a percentage. For the purposes of selecting patient education materials with varying understandability and actionability, the research team set a threshold of 70% for what should be considered understandable or actionable (i.e., a PEMAT score of 70% or below would be considered poorly understandable or poorly actionable. The set of 12 materials included ones that were both understandable and actionable, understandable but poorly actionable, and poorly understandable materials. Each consumer was randomly assigned one of the six materials on each of the two topics.

2.4.3. Data collection protocols and procedures—Consumer testing protocols included a demographic questionnaire, the English version of the Short Test of Functional Health Literacy in Adults (S-TOFHLA) to determine each consumer's health literacy skill level [31,32], and an interview protocol tailored to each patient education material.

To begin the testing, consumers first read or viewed each randomly assigned material. Then the interviewer asked consumers a set of questions to assess their understanding of each material's content (understandability), and the extent to which they knew what actions to take (actionability). Consumers were not time limited and they could refer back to the material as needed to answer questions. There were 4 types of interview questions in the protocols: (1) comprehension questions; (2) numeric scoring questions (i.e., consumers were asked on a scale from 1 to 10 how easy a material was to understand or act upon); (3) open-ended consumer opinion questions; and (4) questions that asked consumers to describe what each section was saying or visual aid was showing (for printable materials only).

Five experienced interviewers with educational degrees ranging from Bachelor's to Ph.D. conducted the consumer interviews. Interviewers were trained on the protocols and took notes during the interview, which were also recorded.

Consumers provided written informed consent to participate. The study was reviewed by Abt Associates' Institutional Review Board and deemed exempt, and approved by the Office of Management and Budget (OMB#0935-0207).

2.4.4. Analysis—The four types of interview questions in the protocols were analyzed differently. For the comprehension questions, each consumer's percentage correct was calculated for each material and an average comprehension score was calculated for each material. For the rating questions, the rating was recorded and an average was calculated for each material. The results from these two sets of questions were used to compare whether consumers better comprehended or rated higher materials that were understandable versus those that were poorly understandable according to the PEMAT, and the same for actionable versus poorly actionable materials. We compared the results for all materials and within topics and material types using *t*-tests. We also examined whether consumers' demographic characteristics affected their comprehension scores or ratings using ANOVA.

There was not an empirical basis for setting the PEMAT score threshold at 70%. To add precision, we used Pearson's product-moment correlation coefficient to determine whether there was a correlation between the PEMAT scores and the average comprehension score or the average numeric score for the consumer testing materials.

All significance tests were examined at p-values <0.05 and p-values <0.10. All statistical analyses were performed using STATA version 11.2 [33].

The positive and negative comments from the open-ended opinion questions were tabulated and described for each material. Responses to the final type of questions provided insight into whether and how consumers understood what each section was saying or what a visual aid was showing, and what they found difficult or easy. We used recordings of the interviews to supplement interviewer notes.

2.5. Construct validity – additional validation with readability scores

The PEMAT does not assess readability, and it is recommended that a readability assessment be conducted in conjunction with the PEMAT. However, reading difficulty has been consistently associated with poorer comprehension [34] and can be quickly and objectively analyzed. Comparing the PEMAT understandability scores as well as the consumer testing results to well-established readability assessments was an efficient approach to further validate the PEMAT.

To determine reading difficulty of the materials used in Round 4, we calculated an average reading grade level using two readability assessments, the Gunning Fog Index and Lexile analysis [35], employing the approach described in Wolf et al. [36]. Five of the 46 materials were not convertible to the required format for the Lexile analysis, leaving a total of 41 materials that included the 12 materials used in consumer testing. We calculated Pearson product-moment correlation coefficients to determine correlations between the results of the 12 consumer testing materials (i.e., comprehension scores and numeric ratings) and reading difficulty. We also examined correlations between PEMAT understandability scores of the 41 materials and reading difficulty.

3. Results

3.1. Review of existing instruments and guides

From an initially identified set of 41 instruments and guides, 22 had potentially relevant concepts and items for understandability or actionability. From these instruments, 7 topics (i.e., content, word choice and style, numeracy, quality visual aids, organization, layout and design, ease of use for audiovisual materials only) and 64 associated items were initially compiled. The project team reviewed and narrowed the pool of items based on relevance to understandability or actionability and redundancy.

3.2. Face and content validation

The expert panel validated that the understandability scale topics (i.e., content, word choice and style, use of numbers, organization, layout and design, and use of visual aids) were relevant (average score range: 1.0–1.33); the actionability scale did not consist of separate topics. The expert panel also validated that the criteria reflected in the PEMAT items would affect a material's understandability or actionability both in the initial validation and in the revalidation (average score: all <1.5).

3.3. Reliability

3.3.1. Inter-rater reliability (external consistency)—For inter-rater reliability (IRR), the average kappa improved markedly from early rounds to latter rounds. By Round 4, we achieved on average, moderate agreement for both scales and all materials. The kappa range for the understandability items was 0.40–0.84 and for the actionability items was 0.35–0.76 (see Table 2)

For Round 4 IRR, we also calculated the Gwet's AC1 and found strong agreement for both scales and material types. Table 3 presents the IRR results for the items included in the final PEMAT.

3.3.2. Internal consistency—The internal consistency for both scales was strong throughout the various rounds of testing. Table 3 also presents the average Cronbach's alpha and item-total correlations for Round 4 for each scale and by material type for the items included in the final version of the PEMAT.

3.4. Construct validity – consumer testing

Table 4 presents sociodemographic characteristics of the consumers. The consumers were diverse in gender, ethnicity, and race, though the majority were younger than 40 years of age (57.5%) and had at least a high school diploma or equivalent (70.2%). Despite recruiting individuals from groups that typically present with a higher prevalence of limited health literacy, all consumers had adequate functional health literacy skills (mean: 33.8; median: 35).

3.4.1. Consumer testing results—No difference was found between the 6 understandable materials and the 6 poorly understandable materials, as determined with the PEMAT, in terms of the consumer testing comprehension scores. There was, however, a

significant difference in the consumer numeric scores between the 3 understandable and the 3 poorly understandable inhaler/asthma materials (8.6 vs. 6.9, $p = 0.01$, Table 5).

There were significant differences between materials the PEMAT indicated were actionable and the ones the PEMAT indicated were poorly actionable both in the comprehension scores (76% vs. 63%, $p = 0.03$, Table 5) and in the consumer numeric scores (8.9 vs. 7.7, $p = 0.03$) for all materials. There were also significant differences between actionable and poorly actionable materials on the comprehension scores within colonoscopy materials (76% vs. 62%, $p = 0.07$) and on the consumer numeric scores within inhaler/asthma materials (9.1 vs. 7.1, $p = 0.05$). There were significant differences within audiovisual materials on both comprehension scores (78% vs. 63%, $p = 0.05$) and consumer scale ratings (9.3 vs. 7.0, $p = 0.02$).

ANOVA analysis showed that the differences between understandable and poorly understandable materials, and actionable and poorly actionable materials were not dependent on or driven by consumer demographic variables.

3.4.2. Correlation between consumer testing results and PEMAT scores—We examined the correlation between the consumer testing results and PEMAT scores to measure the relationship between them more precisely. No relationship was found between PEMAT results for understandability and consumer testing results for both comprehension scores and consumer numeric scores for understandability (Table 6).

For actionability, there were significant positive correlations between PEMAT results and comprehension scores for all materials ($r = 0.87$, $p < 0.05$) and printable materials ($r = 0.96$, $p < 0.05$), and significant positive correlations between PEMAT results and the consumer numeric score for audiovisual materials ($r = 0.91$, $p < 0.10$).

3.4.3. Qualitative findings from consumer testing—While qualitative findings provided little data to inform the PEMAT scales, they did illustrate instances of the criteria reflected in some of the PEMAT items supporting or detracting from understanding or taking actions. For example, some visual aids without captions were inconsistently or incorrectly identified by consumers; consumers identified jargon, medical terms, and complex terms as difficult to understand; and for some highly understandable materials, the consumers' answers were far more precise and consistent even if the answers were not incorrect for other materials. The qualitative results that supported criteria reflected in specific PEMAT items were indicated in the item-level results below.

3.5. Additional validation with readability assessments

Table 7 presents the consumer testing results and average grade level for the 12 materials used in consumer testing. There was a strong negative correlation between consumer numeric scores and average grade level ($r = -0.72$, 95% CI −0.92 to −0.21). The relationship between comprehension scores and average grade level was negligible ($r = -0.14$, 95% CI −0.66 to 0.47).

3.5.1. Comparison of the PEMAT understandability scores and readability scores of materials—For 41 materials from Round 4 of the IRR we calculated an average reading grade level to examine the correlation with the PEMAT understandability scores. There was a very strong negative correlation between the PEMAT understandability scores and the average grade level for printable materials ($r = -0.74$, 95% CI −0.89 to −0.45), and a strong negative correlation for all materials ($r = -0.61$, 95% CI −0.77 to −0.37) and audiovisual materials ($r = -0.49$, 95% CI −0.77 to −0.07). As PEMAT understandability scores went up, reading difficulty went down.

3.6. Item-level results

Table 2 presents the item-level reliability and validity results for the items included in the final PEMAT.

The final PEMAT reflects the expert panel's agreement that items that achieved a kappa < 0.40 should be dropped to provide a more reliable and parsimonious instrument, with one exception Actionability item 20 was retained because there was a strong theoretical foundation for the item. For a list of the items that were dropped, see the Supplemental Material attached.

3.7. Final version of the PEMAT

The final PEMAT consists of 26 items and two scales: understandability (19 items) and actionability (7 items). There are two versions: the PEMAT-P for printable materials (understandability = 17 items and actionability = 7 items) and the PEMAT-A/V for audiovisual materials (understandability = 13 items and actionability = 4 items). The PEMAT scores materials on a scale of 0–100. The PEMAT and Users' Guide, along with a PEMAT Auto-Scoring Form that will automatically calculate PEMAT scores once you enter ratings, are available at: http://www.ahrq.gov/pemat.

4. Discussion and conclusion

4.1. Discussion

This article reports on the development and reliability and validity testing of the Patient Education Materials Assessment Tool (PEMAT), a new instrument to assess the understandability and actionability of both printable and audiovisual patient education materials on diverse topics. The development of the PEMAT was guided by earlier work defining and measuring desirable characteristics of patient education materials and a multidisciplinary expert panel. The PEMAT had strong internal consistency for both scales: understandability and actionability. The inter-rater reliability agreement (external consistency), while moderate as measured by kappa, demonstrated acceptably strong agreement when calculated by Gwet's AC1, and has comparable reliability to a similar instrument that used trained raters in testing [24].

The PEMAT signals advancement for the field of health literacy and health communications on several fronts. First, the PEMAT has undergone psychometric testing in an iterative

manner that is more extensive than the other available instruments [37], and its internal and external consistency is strong.

Second, it is the first time an instrument of its kind has been tested with consumers to establish construct validity. Most developers compare their instrument to existing instruments or to experts' judgment. In consumer testing, we observed significant differences in the materials the PEMAT indicated were actionable versus those it did not, though for understandability we only observed a difference for one of the two topics on one of the two consumer testing metrics. We found, however, strong negative correlations between the PEMAT's understandability scale and readability scores, an established though limited metric related to understandability. The qualitative consumer testing results, which found examples of the importance of several of the concepts reflected in the PEMAT (e.g., the importance of a caption with picture to ensure consumers can understand it), further bolsters the evidence on the performance of the understandability scale.

Third, the PEMAT is the first instrument that measures actionability. Actionability is an increasingly emphasized aim of patient education materials, making the significant consumer testing results particularly noteworthy. Fourth, the PEMAT can reliably assess audiovisual materials. Despite the increasing availability and use of these types of materials, most instruments were not specifically developed to assess audiovisual materials [37].

Fifth, the PEMAT allows a user to assess a material with only the material itself and no other information (e.g., how it was developed, who it was for). Although such information may be helpful in assessing materials, the reality is that this information is often not readily available. Finally, unlike most instruments (e.g., [24]), the reliability of the PEMAT was established using lay professionals who were not trained to use the instrument. While training raters is likely to enhance the uniformity of assessments, the fact that the PEMAT was tested by untrained raters makes its use by the general public more likely, and also supports its ease of use.

As with all newly-developed instruments, the PEMAT could be improved with further development and use in the field. This includes further validation with a larger sample of materials and consumers with limited health literacy, comparing the PEMAT results to similar instruments' results [17,24,25] empirically defining the threshold (i.e., PEMAT score) for a material to be considered understandable or actionable, and using the PEMAT to assess materials in languages other than English.

The limitations of our study should be considered in interpreting the results. The relatively small, convenience sample of materials for IRR and consumer testing (46 materials for IRR, 12 materials for consumer testing) may have driven results and should not be generalized to all patient education materials.

Additionally, several materials scored near the cutoff between understandable and poorly understandable and between actionable and poorly actionable. This may have affected our ability to detect an effect between understandable and poorly understandable and actionable and poorly actionable materials.

We also had a relatively small, convenience sample of consumers ($n = 47$). Despite our recruitment of individuals with a high school diploma or less, every consumer had adequate health literacy skills according to the STOFHLA. Research has shown that the STOFHLA has a ceiling effect, especially with younger respondents. Although we did not observe a correlation between age and STOFHLA scores, the limited sample and more narrow age range likely explain this.

The weak consumer testing results for understandability may be an artifact of the higher than expected health literacy of consumer testers. These consumers may have been able to retrieve information from materials rated poorly understandable that might have stymied consumers with inadequate health literacy. Furthermore, the artificial testing situation encouraged consumers to make the effort to tease out the responses to comprehension questions even if it was difficult to do so. We did not place any time constraints on consumer responses, but in real world situations consumers might not take the time required to comprehend lower quality materials.

The PEMAT does not and did not intend to assess accuracy, comprehensiveness, or cultural appropriateness, or to perform readability tests. These are important criteria to consider assessing patient education materials on, but are beyond the scope of the PEMAT.

4.2. Conclusion

The PEMAT can help lay and health professionals select patient education materials that reduce health literacy demands. It is an internally consistent and reliable instrument with evidence of construct validity that can be used to assess the understandability and actionability of printable and audiovisual materials on diverse topics. As such, the PEMAT advances the *National Action Plan* goal to “develop and disseminate health and safety information that is accurate, accessible and actionable” [10].

4.3. Practice implications

The PEMAT is a user-friendly tool that does not require training. Anyone (e.g., clinician, medical librarian, patient educator) can use it to identify understandable and actionable patient education materials. A web-based, simple, user-centered interface allows for ease of use and immediate, auto-calculated results to further assist lay professionals in their assessment of materials. We expect the PEMAT functionality to continue to evolve as it is disseminated.

Acknowledgments

We would like to acknowledge the raters from Abt Associates, AHRQ, Massachusetts General Hospital, and Northwestern University who helped to establish the reliability of the PEMAT, Allyson Ross Davies for her guidance on instrument development, and Ken Carlson and Mark Spranca from Abt Associates for their valuable engagement with the reliability and validity testing of the PEMAT.

We would like to thank the technical expert panel who helped to shape this instrument by providing guidance and feedback at critical points in the development process: Geri Lynn Baumblatt, MS; Cynthia Baur, PhD; Patricia Brennan, RN, PhD; Darren DeWalt, MD, MPH; Robert Mayes, MS, RN; Michael Paasche-Orlow, MD, MPH; Eva Powell, MSW, CPHQ; Dean Schillinger, MD; Josh Seidman, PhD, MHS; and Paul Smith, MD.

Funding

The information upon which this publication is based was performed under Contract #HHSA290200900012I, TO 4 "*Improving EHRs Patient Education Materials*" funded by the Agency for Healthcare Research and Quality (AHRQ), Department of Health and Human Services. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The author assumes full responsibility for the accuracy and completeness of the ideas presented. Financial support for this study was provided by AHRQ under contract #HHSA290200900012I, TO 4.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pec.2014.05.027.

References

1. U.S. Department of Health and Human Services. Healthy people 2010. 2nd. U.S. Government Printing Office; Washington, DC: 2000.
2. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011; 155:97–107. [PubMed: 21768583]
3. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med. 2004; 19:1228–39. [PubMed: 15610334]
4. Keller DL, Wright J, Pace HA. Impact of health literacy on health outcomes in ambulatory care patients: a systematic review. Ann Pharmacother. 2008; 42:1272–81. [PubMed: 18648014]
5. Nielsen-Bohlman, L.; Panzer, AM.; Kindig, DA., editors. Health literacy: a prescription to end confusion. Washington, DC: National Academies Press; 2004.
6. U.S. Department of Health and Human Services. Research-based web design and usability guidelines. Washington, DC: U.S. Government Printing Office; 2006. Available from http://www.usability.gov/sites/default/files/documents/guidelines_book.pdf?post=yes [accessed 23.08.13]
7. Brach, C.; Keller, D.; Hernandez, LM.; Baur, C.; Parker, R.; Dreyer, B., et al. Ten attributes of a health literate health care organization. Washington, DC: Institute of Medicine; 2012.
8. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med. 2006; 21:878–83. [PubMed: 16881951]
9. Rudd, R. Communicating health: priorities and strategies for progress. Washington, DC: U.S. Department of Health and Human Services; 2003. Objective 11-2 Improvement of health literacy.
10. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. National action plan to improve health literacy. 2010. Available from http://www.health.gov/communication/hlactionplan/pdf/Health_Literacy_Action_Plan.pdf [accessed 23.08.13]
11. Pfizer Clear Health Communication Initiative. Pfizer principles for clear health communication. 2nd2004. Available from http://www.pfizerhealthliteracy.com/asset/pdf/PfizerPrinciples.pdf [accessed 23.08.13]
12. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Quick guide to health literacy. 2010. Available from http://www.health.gov/communication/literacy/quickguide/ [accessed 23.08.13]
13. Centers for Medicare & Medicaid Services. The toolkit for making written material clear and effective. 2012. Available from http://www.cms.gov/WrittenMaterialsToolkit/ [accessed 23.08.13]
14. National Institutes of Health, National Cancer Institute. Clear & simple: developing effective print materials for low-literate readers. 2013. Available from http://www.cancer.gov/cancertopics/cancerlibrary/clear-and-simple [accessed 23.08.13]
15. National Institutes of Health, U.S. National Library of Medicine. How to write easy-to-read health materials. 2013. Available from http://www.nlm.nih.gov/medlineplus/etr.html [accessed 23.08.13]
16. Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav. 2006; 33:352–73. [PubMed: 16699125]

17. Doak, CC.; Doak, LG.; Root, JH. Teaching patients with low literacy skills. 2nd. Philadelphia: Lippincott; 1996.
18. Helitzer D, Hollis C, Cotner J, Oestreicher N. Health literacy demands of written health information materials: an assessment of cervical cancer prevention materials. Cancer Control. 2009; 16:70–8. [PubMed: 19078933]
19. Clayton L. TEMPtEd: development and psychometric properties of a tool to evaluate material used in patient education. J Adv Nurs. 2009; 65:2229–38. [PubMed: 19686403]
20. 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 Commun Health. 1999; 53:105–11.
21. Northwest Territories Literacy Council. A plain language audit toolkit. 2008. Available from http://www.nald.ca/library/learning/nwt/auditool/audit.pdf [accessed 23.08.13]
22. Smith, S. Reviewer's guide to evaluating health information materials. 2008. Available from http://www.beginningsguides.com/upload/Reviewers-Guide.pdf [accessed 23.08.13]
23. Literacy Partners of Manitoba. ClearDoc Index. 2004. Available from http://www.nald.ca/library/research/defining/33.htm [accessed 23.08.13]
24. Kaphingst KA, Kreuter MW, Casey C, Leme L, Thompson T, Cheng M-R, 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(Suppl. 3):203–21. [PubMed: 23030571]
25. Centers for Disease Control and Prevention. CDC clear communication index. 2013. Available from http://www.cdc.gov/healthcommunication/ClearCommunicationIndex/ [accessed 23.08.13]
26. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971; 76:378–82.
27. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–74. [PubMed: 843571]
29. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2010; 61:29–48.
30. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990; 43:543–9. [PubMed: 2348207]
31. Nurss, JR.; Parker, RM.; Williams, MV.; Baker, DW. Short test of functional health literacy in adults (STOFHLA). Hartford, MI: Peppercorn; 2001.
32. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med. 1995; 10:537–41. [PubMed: 8576769]
33. Stata 11.2, MP-Parallel ed [Computer software], College Station, TX: StataCorp; 2011.
34. Davis TC, Wolf MS, Bass PF, Middlebrooks M, Kennen E, Baker DW, et al. Low literacy impairs comprehension of prescription drug warning labels. J Gen Intern Med. 2006; 21:847–51. [PubMed: 16881945]
35. Stenner, AJ.; Horabin, I.; Smith, DR.; Smith, M. The Lexile Framework. Durham, NC: MetaMetrics; 1988.
36. Wolf MS, Shekelle P, Choudry NK, Agnew-Blais J, Parker RM, Shrank WH. Variability in pharmacy interpretations of physician prescriptions. Med Care. 2009; 47:370–3. [PubMed: 19194338]
37. Finnie RKC, Felder TM, Linder SK, Mullen PD. Beyond reading level: a systematic review of the suitability of cancer education print and web-based materials. J Cancer Educ. 2010; 25:497–505. [PubMed: 20237884]

Table 1

Summary of raters, material types, and PEMAT version for each round of reliability testing.

Round	Raters (#)	Material characteristics			PEMAT characteristics				
		Materials[a] (#)	Topics (#)	Language(s)	Total items	Understandability items	Actionability items	Response scale	Item instruction and examples
1	8	16 (*P* = 11, A/V = 5)	3	English and Spanish[b]	32	26	6	4 point	None
2	12	12 (*P* = 8, A/V = 4)	2	English	32	24	8	2 point	Few items had examples
3	2	46 (*P* = 23, A/V = 23)	33	English	32	25	7	2 point	Full user's guide
4	2	46 (*P* = 23, A/V = 23)	33	English	32	25	7	2 point	Full user's guide

[a] *P* = print materials, A/V = audiovisual materials; in rounds 1 and 2, Print refers to both websites and pdfs.

[b] Spanish materials were not tested after round 1 because the reliability results were poor and we needed to limit the points of variation in our testing approach.

Patient Educ Couns. Author manuscript; available in PMC 2016 October 28.

Table 2

Item-level reliability and validity results for items in the final PEMAT.

		Content validity[a]	External consistency (inter-rater reliability)			Internal consistency		Construct validity
Item #	**Item (relevant for printable [*P*] and/or audiovisual [A/V] materials)**	**Average (1–3)**	**% Agree**	**Kappa**	**Gwet's AC1**	**Cronbach's α**	**Item-total correlation**	**Qualitative Findings[b]**
	Understandability							
1	The material makes its purpose completely evident (*P* and A/V)	1.00	88	0.71	0.80	0.68	0.62	(+)
2	The material does not include information or content that distracts from its purpose (*P*)	1.25	81	0.48	0.70	0.80 (*P*)[c]	0.62 (*P*)	(+)
3	The material uses common, everyday language (*P* and A/V)	1.00	81	0.57	0.66	0.66	0.70	(+)
4	Medical terms are used only to familiarize audience with the terms. When used, medical terms are defined (*P* and A/V)	1.25	81	0.54	0.68	0.66	0.73	(+)
5	The material uses the active voice (*P* and A/V)	1.00	88	0.40	0.85	0.75	0.31	
6	Numbers appearing in the material are clear and easy to understand (*P*)	1.00	81	0.55	0.76	0.79 (*P*)	0.79 (*P*)	
7	The material does not expect the user to perform calculations (*P*)	1.00	86	0.69	0.74	0.82 (*P*)	0.32 (*P*)	(+)
8	The material breaks or "chunks" information into short sections (*P* and A/V)	1.00	79	0.56	0.72	0.70	0.55	(+)
9	The material's sections have informative headers (*P* and A/V)	1.00	83	0.70	0.77	0.69	0.60	(+)
10	The material presents information in a logical sequence (*P* and A/V)	1.25	86	0.42	0.81	0.67	0.68	
11	The material provides a summary (*P* and A/V)	1.75	69	0.46	0.58	0.73	0.41	(+)
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)	1.00	74	0.51	0.65	0.71	0.52	(+)
13	Text on the screen is easy to read (A/V)	*e*	71	0.40	0.63	0.77 (A/V)	0.50 (A/V)	(+)
14	The material allows the user to hear the words clearly (e.g., not too fast, not garbled) (A/V)	1.00	86	0.66	0.82	0.76 (A/V)	0.53 (A/V)	(+)
15	The material uses visual aids whenever they could make content more easily understood (e.g., illustration of healthy portion size) (*P*)	1.00	76	0.48	0.56	0.80 (*P*)	0.59 (*P*)	(+)
16	The material's visual aids reinforce rather than distract from the content (*P*)	1.00	86	0.70	0.81	0.81 (*P*)	0.56 (*P*)	(+)

Item #	Item (relevant for printable [*P*] and/or audiovisual [A/V] materials)	Content validity[a]	External consistency (inter-rater reliability)			Internal consistency		Construct validity
		Average (1–3)	% Agree	Kappa	Gwet's AC1	Cronbach's α	Item-total correlation	Qualitative Findings[b]
17	The material's visual aids have clear titles or captions (*P*)	1.25	90	0.84	0.86	0.81 (*P*)	0.51 (*P*)	(+)
18	The material uses illustrations and photographs that are clear and uncluttered (*P* and A/V)	1.25	88	0.67	0.85	0.74	0.34	
19	The material uses simple tables with short and clear row and column headings (*P* and A/V)	1.00	86	0.49	0.80	*d*	*d*	
	Actionability							
20	The material clearly identifies at least one action the user can take (*P* and A/V)	1.00	74	0.35	0.67	0.71	0.88	(+)
21	The material addresses the user directly when describing actions (*P* and A/V)	1.75	88	0.60	0.86	0.85	0.81	
22	The material breaks down any action into manageable, explicit steps (*P* and A/V)	1.25	76	0.52	0.68	0.69	0.89	(+)
23	The material provides a tangible tool (e.g., menu planners, checklists) whenever it could help the user take action (*P*)	1.50	86	0.71	0.72	0.83 (*P*)	0.84 (*P*)	(+)
24	The material provides simple instructions or examples of how to perform calculations (*P*)	1.50	86	0.76	0.80	*d*	*d*	(+)
25	The material explains how to use the charts, graphs, tables, or diagrams to take actions (*P* and A/V)	1.25	81	0.47	0.77	*d*	*d*	
26	The material uses visual aids whenever they could make it easier to act on the instructions (*P*)	1.25	81	0.64	0.74	0.85 (*P*)	0.77 (*P*)	(+)

[a] Average from 1 to 3; Yes = 1, Maybe = 2, and No = 3.

[b] (+) indicates qualitative findings from consumer testing that supported the item/criterion.

[c] All internal consistency results presented are for both material types unless specified: print (*P*) or audiovisual (A/V).

[d] Internal consistency could not be calculated because too few materials received ratings on the item.

[e] Item was added after content validation was completed.

Patient Educ Couns. Author manuscript; available in PMC 2016 October 28.

Table 3

Round 4 external consistency (inter-rater reliability) and internal consistency results for the final PEMAT.

Items/Scales	External consistency (inter-rater reliability)									Internal consistency					
	Average percentage agreement			Average Cohen's Kappa			Average Gwet's AC1			Average Cronbach's α			Average Item-Total Correlation		
Material type	All (%)	A/V (%)	Print (%)	All	A/V	Print	All	A/V	Print	All	A/V	Print	All	A/V	Print
Understandability items/scale	82	79	84	0.57	0.49	0.58	0.74	0.71	0.76	0.70	0.77	0.81	0.55	0.56	0.54
Actionability items/scale	82	83	80	0.58	0.53	0.54	0.75	0.72	0.72	0.75	0.74	0.84	0.86	0.85	0.81
All items/scales	82	80	83	0.57	0.50	0.57	0.74	0.71	0.75	0.71	0.76	0.82	0.62	0.62	0.60

Patient Educ Couns. Author manuscript; available in PMC 2016 October 28.

Table 4

Demographic characteristics of consumers.

Consumers (*n* = 47)	**Number (%)**
Gender	
Male	22 (46.8%)
Female	25 (53.2%)
Age mean (±SD)	36.4 (±13.2)
Age range	
19–24	10 (21.3%)
25–39	17 (36.2%)
40–49	11 (23.4%)
50–64	8 (17.0%)
65+	1 (2.1%)
Hispanic/Latino	
Yes	19 (40.4%)
No	28 (59.6%)
Race[a]	
White/Caucasian	18 (38.3%)
Black/African American	15 (31.9%)
Asian	1 (2.1%)
American Indian or Alaska Native	2 (4.3%)
Other	17 (36.2%)
Highest grade	
High school or GED	33 (70.2%)
Some high school	14 (29.8%)
Health literacy (STOFHLA)	
Adequate (range 23–36)	47 (100.0%)

[a]Consumers could indicate more than one race, so this number does not add to 47 or 100%; “Other” race indicated was most commonly a Latino ethnicity (e.g., Puerto Rican).

Table 5

Consumer Testing results for understandability.

Understandability						
	Average comprehension score (%)			**Average numeric score[a] (#)**		
	Understandable materials[b] (%)	**Poorly understandable materials[c] (%)**	***p***	**Understandable materials[b]**	**Poorly understandable materials[c]**	***p***
All materials	76	75	0.37	8.3	8.4	0.13
Material topic						
Colonoscopy	75	79	0.37	8.3	9.1	0.12
Inhaler/asthma	76	72	0.53	8.6	6.9	0.01**
Material type						
Audiovisual	79	77	0.72	9.0	8.7	0.71
Printable	73	73	0.98	8.1	7.6	0.41
Actionability						
	Average comprehension score (%)			Average numeric score[a] (#)		
	Actionable materials[b] (%)	Poorly actionable materials[c] (%)	*p*	Actionable materials[b]	Poorly actionable materials[c]	*p*
All materials	76	63	0.03**	8.9	7.7	0.03**
Material topic						
Colonoscopy	76	62	0.07*	8.8	8.3	0.37
Inhaler/asthma	75	63	0.16	9.1	7.1	0.05*
Material type						
Audiovisual	78	63	0.05*	9.3	7.0	0.02**
Printable	74	63	0.22	8.6	8.5	0.92

*$p < 0.1$

**$p < 0.05$ *t*-test, no directional assumption.

[a]Rating scale: 1 = very difficult to understand/act on, 10 = very easy to understand/act on.

[b]Understandable = PEMAT score ≥ 70% and Actionable = PEMAT score ≥ 70%.

[c]Poorly understandable = PEMAT score < 70% and poorly actionable = PEMAT score < 70%.

Table 6

Correlation between PEMAT scale results and consumer testing results.

Pearson's product-moment correlation (*r*)			
Material type	**All**	**Audiovisual**	**Print**
Understandability			
PEMAT and comprehension scores	0.12	0.12	0.39
PEMAT and rating scores	0.21	−0.04	0.72
Actionability			
PEMAT and comprehension scores	0.87**	0.88	0.96**
PEMAT and rating scores	0.61	0.91*	−0.10

* $p < 0.1$.

** $p < 0.05$.

Table 7

Comparison of consumer testing results and average grade level.

Material #, type, and topic	Consumer testing results		Readability
Measure	Average comprehension score (%)	Average numeric score [a]	Average grade level
Material #1: print-colonoscopy	79	9	10.9
Material #2: print-colonoscopy	72	7	11.5
Material #3: print-colonoscopy	77	9.3	8.4
Material #4: A/V-colonoscopy	74	8.5	11.3
Material #5: A/V-colonoscopy	81	9.9	8.1
Material #6: A/V-colonoscopy	76	8.5	11.6
Material #7: print-inhaler/asthma	67	6.2	13.5
Material #8: print-inhaler/asthma	64	7.9	10.7
Material #9: print-inhaler/asthma	80	8.4	7.7
Material #10: A/V-inhaler/asthma	51	7.6	10.3
Material #11: A/V-inhaler/asthma	96	N/A	11.9
Material #12: A/V-inhaler/asthma	84	9.4	9.1

[a]Numeric score: 1 = hardest to understand/least actionable, 10 = easiest to understand/most actionable.