



# HHS Public Access

Author manuscript

*Patient Educ Couns.* Author manuscript; available in PMC 2019 March 01.

Published in final edited form as:

*Patient Educ Couns.* 2018 March ; 101(3): 490–496. doi:10.1016/j.pec.2017.09.003.

## Interrater Reliability of the Patient Education Materials Assessment Tool (PEMAT)

Julia Vishnevetsky<sup>a</sup>, Chasity Burrows Walters<sup>a</sup>, and Kay See Tan<sup>b</sup>

<sup>a</sup>Patient and Caregiver Engagement, Memorial Sloan Kettering Cancer Center, New York USA

<sup>b</sup>Epidemiology-Biostatistics, Memorial Sloan Kettering Cancer Center, New York USA

### Abstract

**Objective**—To assess the interrater reliability (IRR) and usability of the Patient Education Materials Assessment Tool (PEMAT) and the relationship between PEMAT scores and readability levels.

**Methods**—One hundred ten materials (80 print, 30 audiovisual) were evaluated, each by two raters, using the PEMAT. IRR was calculated using Gwet's AC1 and summarized across items in each PEMAT domain (understandability and actionability) and by material type. A survey was conducted to solicit raters' experience using the PEMAT. Readability of each material was assessed using the SMOG Index.

**Results**—The median IRR was 0.92 for understandability and 0.93 for actionability across all relevant items, indicating good IRR. Eight PEMAT items had Gwet's AC1 values less than 0.81. PEMAT and SMOG Index scores were inversely correlated, with a Spearman's rho of  $-0.20$  ( $p = 0.081$ ) for understandability and  $-0.15$  ( $p = 0.194$ ) for actionability. While 92% of raters agreed the PEMAT was easy to use, survey results suggested specific items for clarification.

**Conclusion**—While the PEMAT demonstrates moderate to excellent IRR overall, amendments to items with lower IRR may increase the usefulness of the tool.

**Practice Implications**—The PEMAT is a useful supplement to reading level alone in the assessment of educational materials.

### Keywords

PEMAT; patient education material assessment tool; evaluation; patient education; readability; understandability; actionability

---

Corresponding author at 405 Lexington Ave, New York, NY 10174, USA, 646-888-5672, vishnevj@mskcc.org.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 1. Introduction

Recognized by Healthy People 2020 as a national priority in the United States, the development of health information that is accurate, accessible, and actionable is essential in the delivery of quality, safe health care [1]. Despite the growing attention of policymakers and healthcare providers, however, it is evident that health education materials remain too complex for many to comprehend [2]. The assessment of patients' comprehension of educational materials in the healthcare setting is a challenge, and is amplified when considering the 80% of Americans who search for health information on the internet independently [3]. The inability of those developing health information to determine the health literacy level of their consumers has prompted the Agency for Healthcare Research and Quality to recommend the adoption of a "universal precautions" approach [4].

A central construct in the application of universal precautions is readability, which refers to the ease by which the reader is able to read and understand text [5]. The concept of readability has long been integral to the process of development and evaluation of educational materials. The extent to which text is considered readable is measured mathematically, commonly resulting in a score representative of a grade level in the US school system. While readability formulas are critical and widely used to evaluate the reading difficulty of educational material, scores vary widely [6, 7] and overlook other factors known to impact one's ability to comprehend the information provided [7]. Thus, those developing educational materials must go beyond readability levels alone [4]. Indeed multiple resources exist to guide those developing patient education materials, and several tests are available to determine the appropriateness of patient education materials for diverse audiences [8, 9].

Developed to address the shortcomings of readability formulas alone, the Patient Education Material Assessment Tool (PEMAT) assesses the domains of understandability (when consumers of diverse backgrounds and varying levels of health literacy can process and explain key messages) and actionability (when consumers of diverse backgrounds and varying levels of health literacy can identify what they can do based on the information presented) [10]. In addition, the PEMAT is the only tool that includes a measure of objective assessment of audiovisual (A/V) materials.

To date there are few publications reporting the interrater reliability (IRR) of the PEMAT. Several studies have reported the IRR of one or both domains [11–14], and one study reported the IRR of the items grouped by topic [15]. Only one publication was found reporting the IRR at the item level, however this was in an evaluation of clinical summaries therefore it included only those items in the print version of the PEMAT [16]. The objective of this study was to evaluate the IRR of the individual PEMAT items on both print and A/V patient education materials.

## 2. Methods

### 2.1 Evaluation of patient education materials

**2.1.1 Evaluation using the PEMAT**—We used the PEMAT to evaluate the actionability and understandability of 110 patient education materials disseminated through the Patient & Caregiver Education Department of a National Cancer Institute (NCI)-Designated Comprehensive Cancer Center. All of the materials were oncology related, created by Health Education Specialists, and included both print and A/V materials (videos).

Drawing from a sample of 17 raters, two raters assessed each educational material using the PEMAT. As the PEMAT is intended for use by both professionals and the lay public [2], the raters included both cancer center staff and former patients. The raters were instructed to read the PEMAT user manual, read or watch the patient education materials, and complete a PEMAT using the accompanying Excel spreadsheet for the score calculations.

The PEMAT is comprised of two domains, understandability and actionability. Each domain consists of criteria statements, organized by topic, to which the rater will either agree or disagree, scoring them as 1 or 0, respectively. In addition, some items include a not applicable (NA) option. These items, along with the PEMAT domain and topic that they appear in, are shown in Table 1. Scores are added to create a percentage understandability and actionability score, ranging from 0 to 100%.

**2.1.2 User experience using the PEMAT**—We used an electronic survey to measure raters' experience using the PEMAT after all evaluations were completed. The survey included an overall question "The PEMAT was easy to use" with Likert-type answers (strongly agree, agree, disagree, or strongly disagree). In addition, each of the 26 PEMAT items were presented with answer choices about how clear they were (very clear, somewhat clear, or not clear). We also solicited comments for each item and for the PEMAT as whole using a free text box.

**2.1.3 Evaluation using readability formula**—We used the SMOG Index [17], which has been shown to have consistent results in healthcare [6], to measure readability for each print material. The SMOG Index, often referred to as the Simple Measure of Gobbledygook, provides a numerical value of a grade level, ranging from fourth grade to a graduate education (grades 4–18) [17], with higher numbers indicating materials that are more difficult to read.

### 2.2 Statistical analysis

**2.2.1 Summary scores**—We present the median and interquartile range (IQR) of the PEMAT scores for the understandability and actionability domains for print and A/V materials. We also calculated the readability scores for each print material.

**2.2.2 Interrater reliability of the PEMAT**—We used three measures to assess the IRR of the PEMAT, including percent raw agreement (proportion of exact matches in responses among all pairs of ratings), and two chance-corrected agreement measures: Fleiss Kappa and Gwet's AC1.

Percent raw agreement was used to calculate what percentage of the time the raters agreed on their scores on a PEMAT item. When both raters provided the same answer (either 1, 0, or NA) on the same item for the same patient education material, it was considered a match. Any other combination of answers was considered a mismatch. NA was considered an actual response, meaning if rater A answers 1 while rater B answers NA, it would be considered a mismatch.

Fleiss' Kappa is an extension of the more commonly reported Cohen's kappa. While Cohen's kappa requires the same two raters to evaluate every material, Fleiss' Kappa allows for the design of two raters selected from a pool of potential raters [18]. We also present the IRR measured by Gwet's AC1, which is more appropriate than Fleiss' Kappa in this study. Gwet's AC1 was used as it takes into account prevalence-bias, which occurs when one response category is extremely common [19]. For each of the three approaches, IRR is estimated per-item across all applicable material and then summarized as median (IQR) across all items-level IRRs.

We summarized all agreement statistics across items by material type (print items only, A/V items only, and all items) and by domain (individual understandability and actionability items and the summary scales for both). We also calculated IRR for the summary scales of understandability and actionability.

Both Fleiss' Kappa and Gwet's AC1 range from -1 to 1. Following the guidelines from Landis, et al, a kappa statistic of 0.00–0.20 is considered slight agreement, 0.21–0.40 has fair agreement, 0.41–0.60 has moderate agreement, 0.61–0.80 has substantial agreement, and values of 0.81–1.00 have almost perfect agreement [20]. Agreements less than 0.00 are considered poor. This study sought to identify items demonstrating IRR less than 0.81 using Gwet's AC1. All statistics were calculated using R 3.1.1., using the *IRR* package, and Gwet's user-written R functions [21].

**2.2.3 Correlation between readability and PEMAT measures**—We used Spearman's correlation to assess the relationship between the PEMAT measures of understandability and actionability with the SMOG Index for the print materials.

### 3. Results

A total of 110 materials were evaluated using the PEMAT, each by two of 17 reviewers. Of those 110, 80 were print materials and 30 were A/V.

#### 3.1 Summary scores

The median and IQR of PEMAT scores, separated by domain and material type, are presented in Table 2.

The average readability score on the SMOG Index was 7.4 (range 4.1–10.2) for the print materials, indicating that these materials were understandable for people who had a 7<sup>th</sup> grade education. This is in accordance with the National Library of Medicine, which indicates that patient education materials should be written at a 7<sup>th</sup> or 8<sup>th</sup> grade reading level [22].

**3.1.1 Correlation between PEMAT and SMOG index**—Spearman’s rho was  $-0.20$  ( $p = 0.081$ ) for understandability and  $-0.15$  ( $p = 0.194$ ) for actionability, indicating that readability and the PEMAT measures are inversely correlated among the print materials in this sample. However, these observations were not statistically significant.

### 3.2 Interrater reliability of the PEMAT

The degree to which two reviewers agreed in their assessment of the resources are presented as percent raw agreement, Fleiss Kappa and Gwet’s AC1. Table 3 shows the median IRR for the domain measures, stratified by material type. The IRRs calculated at the item-level are summarized as median (IQR) across items by domains.

Tables 4–6 show the IRR measures for the individual items of the PEMAT, separated by resource type (print, A/V and both combined).

Eight items with Gwet’s AC1 scores less than 0.81 were identified, including print items 6 and 11 (see Table 4) and A/V items 8, 9, 11, 12, 13, 18, and 25 (see Table 5). Of those items present in both the print and A/V versions, items 11, 18, and 25 were below the threshold of 0.81 when those data were combined (see Table 6).

### 3.3 User experience with the PEMAT

Fourteen out of 17 raters (82%) responded to the survey assessing their experience using the PEMAT. While 13 (93%) agreed or strongly agreed that the PEMAT was easy to use, eight PEMAT items (3, 4, 5, 8, 14, 15, 23, and 26) were perceived as unclear to at least one rater.

Additionally, raters provided qualitative feedback regarding PEMAT items. For example, addressing item 4 (use of medical terms), one rater stated:

“Most education documents need to use medical terms not just to familiarize the audience but to actually educate them and ensure they can utilize the terminology when speaking with their HCPs. The use of medical terms is unavoidable and important and should not be penalized, provided that all terms are defined.”

Additionally, several raters commented on item 15 (contribution of visual aids to understandability), as follows:

“I think this item is unclear only because it can be interpreted differently by different reviewers, especially the sentence ‘If you can think of a meaningful visual aid that could have been added to clarify to meaning of text, you should disagree with this item.’”

“The option of N/A should be available for this question.”

“the ‘whenever they could make content more easily understood’ caveat should be more emphasized.”

“Uncertain how to score this item if visual aids are not included because they would not be helpful.”

Regarding item 26 (contribution of visual aids to actionability), one rater noted:

“How would we answer if we don’t identify any opportunities for visual aids, but feel that a visual aid already included doesn’t make it easier to act on the instructions?”

## 4. Discussion and conclusion

### 4.1 Discussion

**4.1.1 Summary of findings**—While the PEMAT is emerging as a tool in the literature evaluating educational materials, this is the first report to examine the IRR of individual items in the PEMAT on both print and A/V patient education materials. Our findings indicate that while IRR is high across all items, IRR between PEMAT items varies considerably.

The use of the PEMAT is described in the extant literature as a means to measure the quality of a variety of patient education materials. It has been used to assess a wide range of resources including materials on vocal cord paralysis [11], end-of-life decisions [23], chronic kidney disease [15], discharge instructions and clinical summaries [16, 24], and Zenker’s diverticulum [12]. While the presentation of overall scores in these reports provide meaningful information on the quality of those materials, their intent was not to evaluate the tool itself.

Other work has examined the IRR of the two PEMAT domains of understandability and actionability. These studies found that there was no significant difference between the PEMAT scores of two raters on webpages on tonsillectomy [25] or maritime health related information [26] at the  $p < 0.05$  level. In addition, evaluations of patient education materials on c. Diff and surgical site infections had kappa values of 0.80 when comparing understandability and actionability scores among 3 raters [13, 14].

Looking at the sections that make up the overarching domains, Morony et al., (2017) used percent agreement to assess the IRR of the PEMAT topics [15]. The IRR ranged from 72% agreement (organization) to 90% agreement (use of numbers). Comparatively, when using percent raw agreement for print material topics in the current study, IRR ranged from 85% (use of numbers) to 99% (content).

In the only other published study addressing IRR at the item level, Sarzynski et al., (2017) examined IRR in their evaluation of 100 clinical summaries (50 from each of two vendors) [16]. After reviewing the PEMAT instructional guide, two non-clinician reviewers evaluated clinical summaries extracted from patients’ charts using the print version of the PEMAT. Items 6, 7, 8, 12, and 24 consistently had kappa values of less than 0.81, while items 1, 2, 9, 10, 11, 19, and 26 had kappa values of 1.00. The IRR on the two sets of clinical summaries was 0.55 and 0.72 for understandability and 0.56 and 0.76 for actionability.

**4.1.2 Recommendations for item clarification**—Our study found variation in IRR across PEMAT items. Furthermore, inconsistencies between IRR and raters’ perceptions of the clarity of items suggest a need for clarification. These issues are discussed in detail

below and recommendations for amendments and additions to the PEMAT tool are presented in Table 7.

The item with the lowest IRR in this study was item 6, which states that numbers appearing in the material are clear and easy to understand. The PEMAT mentions that times and dates should not be considered numbers for this item, leaving raters to question whether other numbers, such as addresses and telephone numbers, should be considered. This lack of clarity led some raters to score this item numerically while others used NA. Notably, despite their lack of agreement in scores, all raters described this item as being very clear or somewhat clear.

Item 4 discusses the use of medical terms, recommending that medical terms should only be used to familiarize the audience. The explanation in the PEMAT user's manual further states that "Even when there are not obvious plain language substitutes for a medical term, a material that uses medical terms will not be easily understood. You should disagree with this item if the material uses medical terms other than to introduce them." However, health education materials are often required to include medical terms so that patients can communicate effectively about their conditions, such using as the word "radiation" for a person undergoing cancer treatment. As a result, if certain terms are central to the understanding for one's health, they should be allowed to be used, as long as they are defined.

Item 11, which asks whether the material has a summary, was among the items demonstrating low IRR in both the print and A/V versions. The discrepancies among the raters prompted an independent review by this study team that found summaries in none of the materials evaluated. The failure to include a summary was in line the findings of other studies [11, 27], with Balakrishnan et al. (2016) indicating they were uncertain regarding the value of a summary [27]. More research is needed to determine the usefulness for a summary, particularly in materials that have information on multiple topics or explain the steps of a procedure.

Some of the items that assess A/V materials could benefit from further clarification. Items 8 and 9, which explain chunking of content, do not apply to A/V items that are shorter than one minute. However, videos between one and two minutes may be too short to include meaningful content chunks. More research is needed for recommendations for timing of subsections of videos. In addition, item 13 could benefit from additional instructions as to what constitutes text on screen and what can be omitted from the assessment. Finally, item 18 discusses assessing the clarity and focus of illustrations and photographs, however no instructions specific to videos are given. Further elaboration on how to assess the understandability of the video on screen would be helpful, such as the ability to visualize the actions being demonstrated.

The phrase "whenever it could," such as in items 15 and 23 which ask if certain characteristics are in place whenever they could make the material more understandable or actionable, emerged as a source of confusion for raters. The raters identified this wording as

being too subjective. Furthermore, there was uncertainty as to how to rate the item if the material didn't have visual aids or tools but the raters felt that they were not needed for it.

**4.1.3 Limitations**—This study has several limitations. Scores on the educational materials assessed in this project were generally higher than those in the published literature. This may be a result of the materials being developed by trained Health Education Specialists with the use of a style guide specifically addressing many of the items in the tool. Next, while the PEMAT was designed to be used by lay people and health professionals alike, the majority of raters in this study were health professionals. As a result, they may come into the rating with their own knowledge, which may make it difficult for them to assess the documents as a lay person would. Finally, because 30 A/V materials in contrast to 80 print materials were evaluated in this study, the A/V IRR results should be interpreted with more caution.

## 4.2 Conclusion

This study is the first comprehensive study of IRR of the PEMAT since the publication of Shoemaker et al.'s original article [2]. While further studies are needed to assess a wider range of materials as well as use raters from different backgrounds, these findings suggest the PEMAT makes a valuable contribution to the assessment of patient education materials.

## 4.3 Practice implications

It is essential that healthcare professionals provide patient education materials that are not only readable but also actionable and understandable. The use of the PEMAT to evaluate educational materials may not only help those developing them, but also healthcare providers looking for available materials to meet the learning needs of their patients or the public.

It is vital that raters thoroughly read the PEMAT user's manual before beginning an assessment and refer to it as needed going forward. If more than one rater are reviewing materials, our findings suggest it may be beneficial for raters to discuss their scores after doing their first few analyses to ensure that everyone interpreted the PEMAT the same way.

## Acknowledgments

The authors would like to thank Christopher Brooks, Kristen Carotenuto, Deirdre Casey, Meagan Harrington, Marisol Hernandez, Jean Kotkiewicz, Jacqueline LaGrassa, Laura Paloubis, Brieyona Reaves, Allison Reichel, Anna Skripnik, and Inderani Walia for evaluating the patient education materials using the PEMAT.

The authors would also like to thank Jennifer Wang for evaluating the patient education materials and collecting survey data.

### Funding

Funding was provided by the P30 Cancer Center Support Grant (CCSG) (P30 CA008748).

## References

1. U.S. Department of Health and Human Services. National Action Plan to Improve Health Literacy. 2010 Washington, DC.



2. 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; 96:395–403. [PubMed: 24973195]
3. Fox, S. *Online Health Search 2006*. Pew Research Center; 2006.
4. Brega AG, Freedman MA, LeBlanc WG, Barnard J, Mabachi NM, Cifuentes M, Albright K, Weiss BD, Brach C, West DR. Using the Health Literacy Universal Precautions Toolkit to Improve the Quality of Patient Materials. *J Health Commun*. 2015; 20(Suppl 2):69–76.
5. DuBay WH. *The Principles of Readability*, Impact Information Costa Mesa, CA. 2004
6. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm*. 2013; 9:503–16. [PubMed: 22835706]
7. Centers for Medicare & Medicaid Services. Toolkit for Making Written Material Clear and Effective. 2012. <<https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/index.html?redirect=/WrittenMaterialsToolkit>>accessed December 27 2016
8. Doak, CC., Doak, LG., Root, JH. *Teaching patients with low literacy skills*. 2nd. J.B. Lippincott; Philadelphia: 1996.
9. Kaphingst KA, Kreuter MW, Casey C, Leme L, Thompson T, Cheng MR, Jacobsen H, Sterling R, Oguntimein J, Filler C, Culbert A, Rooney M, Lapka C. 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]
10. Shoemaker, SJ., Wolf, MS., Brach, C. *The Patient Education Materials Assessment Tool (PEMAT) and User's Guide*. 2013. <[https://www.ahrq.gov/sites/default/files/publications/files/pemat\\_guide.pdf](https://www.ahrq.gov/sites/default/files/publications/files/pemat_guide.pdf)>
11. Balakrishnan V, Chandy Z, Hseih A, Bui TL, Verma SP. Readability and Understandability of Online Vocal Cord Paralysis Materials. *Otolaryngology-Head and Neck Surgery*. 2016; 154:460–4. [PubMed: 26861227]
12. Balakrishnan V, Chandy Z, Verma SP. Are Online Zenker's Diverticulum Materials Readable and Understandable? *Otolaryngol Head Neck Surg*. 2016; 155:758–63. [PubMed: 27329416]
13. Zellmer C, Zimdars P, Parker S, Safdar N. How well do patient education materials for *Clostridium difficile* infection score? A systematic evaluation. *International Journal of Infection Control*. 2015; 11
14. Zellmer C, Zimdars P, Parker S, Safdar N. Evaluating the usefulness of patient education materials on surgical site infection: a systematic assessment. *Am J Infect Control*. 2015; 43:167–8. [PubMed: 25541334]
15. Morony SS. 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? *Journal of health communication*. 2017; 22:163–70. [PubMed: 28121226]
16. Sarzynski E, Hashmi H, Subramanian J, Fitzpatrick L, Polverento M, Simmons M, Brooks K, Given C. Opportunities to improve clinical summaries for patients at hospital discharge. *Brit Med J Qual Saf*. 2017; 26:372–80.
17. Mc Laughlin GH. SMOG Grading-a New Readability Formula. *Journal of Reading*. 1969; 12:639–46.
18. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971; 76:378–82.
19. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008; 61:29–48. [PubMed: 18482474]
20. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
21. Gwet, KL. R functions for calculating agreement coefficients. 2010. <[http://www.agreestat.com/r\\_functions.html](http://www.agreestat.com/r_functions.html)>accessed January 31.2017
22. MedlinePlus. *How to Write Easy-to-Read Health Materials*. 2016. <<https://medlineplus.gov/etr.html>>accessed May 16, 2017.2017

23. White B, Willmott L, Tilse C, Wilson J, Lawson D, Pearce A, Dunn J, Aitken JF, Feeney R, Jowett S. Community knowledge of law at the end of life: availability and accessibility of web-based resources. *Aust Health Rev.* 2017
24. Unaka NI, Statile A, Haney J, Beck AF, Brady PW, Jerardi KE. Assessment of readability, understandability, and completeness of pediatric hospital medicine discharge instructions. *J Hosp Med.* 2017; 12:98–101. [PubMed: 28182805]
25. Arsenault M, Blouin MJ, Guitton MJ. Information quality and dynamics of patients' interactions on tonsillectomy web resources. *Internet Interventions.* 2016; 4(Part 2):99–104.
26. Guitton MJ. Online maritime health information: an overview of the situation. *Int Marit Health.* 2015; 66:139–44. [PubMed: 26394313]
27. Zellmer C, Zimdars P, Safdar N. Usefulness of patient education materials for central line associated blood stream infection prevention. *International Journal of Infection Control.* 2016; 12

### Highlights

- The PEMAT adds to the objective evaluation of patient education materials
- The PEMAT demonstrates good IRR overall
- Clarification of certain items may increase IRR
- Raters found the PEMAT easy to use

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

The items on the PEMAT

<b>Domain: Understandability</b>
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: 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)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Median and IQR of PEMAT scores across all educational materials

	<b>Understandability</b>	<b>Actionability</b>
Median (IQR) of Print Materials	92.3% (88.7, 93.8)	100.0% (100, 100)
Median (IQR) of A/V Materials	83.1% (78.6, 90.8)	100.0% (100, 100)
Median (IQR) of all Materials	92.0% (84.6, 93.4)	100.0% (100, 100)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Median (IQR) IRR of PEMAT scores on patient education materials

	<b>Percent Raw Agreement (IQR)</b>	<b>Fleiss Kappa (IQR)</b>	<b>Gwet's AC1 (IQR)</b>
<b>Print materials</b>			
Understandability	0.95 (0.7, 1)	0.72 (-0.01, 1)	0.94 (0.45, 1)
Actionability	0.94 (0.85, 1)	0.61 (-0.01, 1)	0.93 (0.81, 1)
<b>A/V materials</b>			
Understandability	0.87 (0.6, 1)	0.41 (-0.12, 1)	0.84 (0.38, 1)
Actionability	0.97 (0.73, 1)	0.02 (-0.02, 1)	0.97 (0.69, 1)
<b>All Materials</b>			
Understandability	0.93 (0.67, 1)	0.62 (-0.01, 1)	0.92 (0.45, 1)
Actionability	0.94 (0.82, 1)	0.5 (-0.01, 1)	0.93 (0.78, 1)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4**

Agreement on print materials on the PEMAT (n = 80 materials)

PEMAT Item number	Percent Raw Agreement	Fleiss Kappa (95% CI)	Gwet's AC1 (95% CI)
1	0.99	-0.01 (-0.02, 0.01)	0.99 (0.96, 1)
2	0.99	-0.01 (-0.02, 0.01)	0.99 (0.96, 1)
3	1	1 (1, 1)	1 (1, 1)
4	0.95	0.57 (0.19, 0.96)	0.94 (0.89, 1)
5	1	1 (1, 1)	1 (1, 1)
6	0.70	0.34 (0.12, 0.56)	0.45 (0.24, 0.66)
7	1	1 (1, 1)	1 (1, 1)
8	1	1 (1, 1)	1 (1, 1)
9	0.99	0.85 (0.56, 1)	0.99 (0.96, 1)
10	0.98	-0.01 (-0.03, 0.01)	0.97 (0.94, 1)
11	0.75	0.2 (-0.05, 0.46)	0.7 (0.58, 0.83)
12	0.94	0.51 (0.13, 0.89)	0.93 (0.86, 0.99)
15	0.94	0.51 (0.13, 0.89)	0.93 (0.86, 0.99)
16	0.91	0.82 (0.69, 0.95)	0.83 (0.7, 0.95)
17	0.86	0.74 (0.59, 0.88)	0.81 (0.71, 0.92)
18	0.92	0.85 (0.73, 0.97)	0.90 (0.82, 0.98)
19	0.90	0.72 (0.55, 0.9)	0.88 (0.79, 0.96)
20	0.99	-0.01 (-0.02, 0.01)	0.99 (0.96, 1)
21	1	1 (1, 1)	1 (1, 1)
22	0.99	0.79 (0.39, 1)	0.99 (0.96, 1)
23	0.88	0.22 (-0.12, 0.55)	0.85 (0.75, 0.95)
24	0.94	0.61 (0.32, 0.91)	0.93 (0.87, 0.99)
25	0.85	0.61 (0.41, 0.82)	0.81 (0.71, 0.92)
26	0.89	0.25 (-0.11, 0.6)	0.87 (0.78, 0.96)

**Table 5**

Agreement on A/V materials the PEMAT (n = 30 materials)

PEMAT Item number	Percent Raw Agreement	Fleiss Kappa (95% CI)	Gwet's AC1 (95% CI)
1	1	1 (1, 1)	1 (1, 1)
3	0.87	0.26 (-0.29, 0.81)	0.84 (0.66, 1)
4	0.87	-0.07 (-0.15, 0)	0.85 (0.68, 1)
5	1	1 (1, 1)	1 (1, 1)
8	0.77	0.41 (0.02, 0.80)	0.71 (0.50, 0.92)
9	0.70	0.44 (0.11, 0.77)	0.59 (0.35, 0.83)
10	1	1 (1, 1)	1 (1, 1)
11	0.73	0.08 (-0.33, 0.49)	0.69 (0.47, 0.91)
12	0.60	-0.12 (-0.45, 0.21)	0.38 (0, 0.76)
13	0.67	0.24 (-0.13, 0.61)	0.57 (0.33, 0.82)
14	1	1 (1, 1)	1 (1, 1)
18	0.63	0.13 (-0.23, 0.5)	0.54 (0.28, 0.79)
19	0.90	0.52 (0.03, 1)	0.89 (0.76, 1)
20	0.97	-0.02 (-0.05, 0.02)	0.97 (0.89, 1)
21	1	1 (1, 1)	1 (1, 1)
22	0.97	-0.02 (-0.05, 0.02)	0.97 (0.89, 1)
25	0.73	0.06 (-0.33, 0.45)	0.69 (0.47, 0.91)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 6**

Agreement on both print and A/V PEMAT materials (N = 110 materials)

PEMAT Item number	Percent Raw Agreement	Fleiss Kappa (95% CI)	Gwet's AC1 (95% CI)
1	0.99	0 (-0.01, 0)	0.99 (0.97, 1)
2 <sup>a</sup>	0.99	-0.01 (-0.02, 0.01)	0.99 (0.96, 1)
3	0.96	0.78 (0.57, 0.99)	0.96 (0.91, 1)
4	0.93	0.39 (0.05, 0.73)	0.92 (0.86, 0.98)
	1	1 (1, 1)	1 (1, 1)
6 <sup>a</sup>	0.70	0.34 (0.12, 0.56)	0.45 (0.24, 0.66)
7 <sup>a</sup>	1	1 (1, 1)	1 (1, 1)
8	0.94	0.61 (0.33, 0.88)	0.93 (0.88, 0.98)
9	0.91	0.66 (0.46, 0.86)	0.89 (0.83, 0.96)
10	0.98	0.49 (-0.13, 1)	0.98 (0.95, 1)
11	0.75	0.17 (-0.04, 0.39)	0.70 (0.59, 0.81)
12	0.85	0.62 (0.45, 0.78)	0.81 (0.72, 0.90)
13 <sup>b</sup>	0.67	0.24 (-0.13, 0.61)	0.57 (0.33, 0.82)
14 <sup>b</sup>	1	1 (1, 1)	1 (1, 1)
15 <sup>a</sup>	0.94	0.51 (0.13, 0.89)	0.93 (0.86, 0.99)
16 <sup>a</sup>	0.91	0.82 (0.69, 0.95)	0.83 (0.70, 0.95)
17 <sup>a</sup>	0.86	0.74 (0.59, 0.88)	0.81 (0.71, 0.92)
18	0.85	0.70 (0.57, 0.83)	0.79 (0.70, 0.89)
19	0.90	0.69 (0.53, 0.86)	0.88 (0.81, 0.95)
20	0.98	-0.01 (-0.02, 0)	0.98 (0.96, 1)
21	1	1 (1, 1)	1 (1, 1)
22	0.98	0.66 (0.21, 1)	0.98 (0.95, 1)
23 <sup>a</sup>	0.88	0.22 (-0.12, 0.55)	0.85 (0.75, 0.95)
24 <sup>a</sup>	0.94	0.61 (0.32, 0.91)	0.93 (0.87, 0.99)
25	0.82	0.50 (0.30, 0.69)	0.78 (0.68, 0.87)
26 <sup>a</sup>	0.89	0.25 (-0.11, 0.60)	0.87 (0.78, 0.96)

<sup>a</sup>Items that are only for print materials (n = 80 materials)<sup>b</sup>Items that are only for A/V materials (n = 30 materials)

**Table 7**

## Recommendations for item clarification

<b>PEMAT Item</b>	<b>Recommendation</b>
Item 4: Medical terms are used only to familiarize audience with the terms. When used, medical terms are defined (P and A/V)	Expand explanation to include “Medical terms that are necessary to understand the condition or treatment being discussed in the material are defined at first use.”
Item 5: The material uses the active voice (P and A/V)	Include a reminder that one should agree with the item if this characteristic happens at least 80% of the time.
Item 6: Numbers appearing in the material are clear and easy to understand	Rephrase to read “Choose ‘N/A’ if the material has no numbers or if the only numbers in the material are used to identify things such as times, dates, telephone numbers, and addresses.”
Item 13: Text on the screen is easy to read (A/V)	Clarify which text the item is referring to and which can be omitted for the assessment.
Item 15: The material uses visual aids whenever they could make content more easily understood (e.g., illustration of healthy portion size) (P)	Clarify the instructions to read “Choose ‘agree’ if the item doesn’t have characteristic but doesn’t need it.” Or, add a N/A response.
Item 18: The material uses illustrations and photographs that are clear and uncluttered (P and A/V)	Change the item to say “visual aids” instead of “illustrations and photographs” Expand explanation to include understandability of A/V materials, such as video quality and focus.
Item 23: The material provides a tangible tool (e.g., menu planners, checklists) whenever it could help the user take action (P)	Clarify the instructions to read “Choose ‘agree’ if the item doesn’t have characteristic but doesn’t need it.” Or, add a N/A response
Item 26: The material uses visual aids whenever they could make it easier to act on the instructions (P)	Clarify the instructions to read “Choose ‘agree’ if the item doesn’t have characteristic but doesn’t need it.” Or, add an N/A response