Practical Advice Regarding the Reliability of the Patient Educational Materials Assessment Tool for Health Educators

Joseph G. L. Lee, PhD, MPH[1]
Mahdi Sesay[1]
Paula A. Acevedo[1]
Zachary A. Chichester[1]
Beth H. Chaney, PhD, MCHES[1]

The quality of patient education materials is an important issue for health educators, clinicians, and community health workers. We describe a challenge achieving reliable scores between coders when using the Patient Educational Materials Assessment Tool (PEMAT) to evaluate farmworker health materials in spring 2020. Four coders were unable to achieve reliability after three attempts at coding calibration. Further investigation identified improvements to the PEMAT codebook and evidence of the difficulty of achieving traditional interrater reliability in the form of Krippendorff's alpha. Our solution was to use multiple raters and average ratings to achieve an acceptable score with an intraclass correlation coefficient. Practitioners using the PEMAT to evaluate materials should consider averaging the scores of multiple raters as PEMAT results otherwise may be highly sensitive to who is doing the rating. Not doing so may inadvertently result in the use of suboptimal patient education materials.

Keywords: *reproducibility of results; patient education as topic; pamphlets*

BACKGROUND

There are profound health inequities for migrant and seasonal farmworkers ("farmworkers"). Across the country, health educators for migrant health centers and community-based organizations conduct outreach to farmworkers and provide health education. Yet there is no single location where materials for farmworker health outreach are housed. As part of a larger project funded by the National Library of Medicine, we systematically identified patient education materials for use in farmworker outreach. By May 2020, we had identified over 600 materials (Lee, 2020).

There has been increasing recognition that patient education materials must be designed to make their findings understandable and actionable (U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion, 2010). A number of checklists, tools, and ratings schemes have been developed to help public health practitioners develop and assess patient education materials, such as the Centers for Disease Control and Prevention Clear Communication Index (Baur & Prue, 2014), the Suitability Assessment of Materials (Doak et al., 1996), and the Patient Educational Materials Assessment Tool (PEMAT; Shoemaker et al., 2014). These tools are designed to leverage best practices in health communication for patient education materials.

[1]*East Carolina University, Greenville, NC, USA*

Authors' Note: *The authors thank Israel M. Mendez for his help with coding. Research reported in this article was supported by the National Library of Medicine of the National Institutes of Health under Award No. G08LM013198. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Address correspondence to Joseph G. L. Lee, East Carolina University, 1000 E 5th Street, Mail Stop 529, Greenville, NC 27858, USA; e-mail: leejose14@ecu.edu.*

Health Promotion Practice
January 2022 Vol. 23, No. (1) 17–19
DOI: 10.1177/1524839920984790
Article reuse guidelines: sagepub.com/journals-permissions

The PEMAT is recognized for overcoming limitations of previous tools (Beaunoyer et al., 2017). The PEMAT was developed for the Agency for Healthcare Research and Quality to ensure that health education materials are understandable and actionable following recommendations from the National Action Plan to Improve Health Literacy (Shoemaker et al., 2014; U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion, 2010). Furthermore, developers of the PEMAT note it was created specifically to address limitations with other patient education material assessments (Shoemaker et al., 2014). Limitations of previous assessments include being tested with a specified topic in mind, not achieving interrater reliability, use of readability formulas, and being evaluated using only raters trained for using the tool. Importantly, the PEMAT was developed to be used by untrained practitioners for assessing patient education materials (Shoemaker et al., 2014).

Four authors and a graduate student thus used the PEMAT on the educational materials for farmworkers. The coders were an associate professor of health education experienced in quantitative content analysis and coding reliability, three undergraduate students, and a physician assistant graduate student. For the purposes of assessing coding described here, we used only English-language versions of materials.

First, three coders (P.A.A., J.G.L.L., M.S.) reviewed the codebook and independently coded five materials. To calculate reliability, we used Krippendorff's alpha (Hayes & Krippendorff, 2007), which we have successfully used in other studies. We achieved $\alpha = .42$ for the PEMAT's understandability score and $\alpha = .19$ for its actionability score, which indicate unacceptably low reliability.

Then, the four coders (P.A.A., Z.A.C., J.G.L.L., M.S.) reviewed divergent coding, discussed it, calibrated on materials together, and then independently coded 15 systematically selected materials. We achieved $\alpha = .34$ for understandability and $\alpha = .07$ for actionability. This process was repeated again with 10 systematically selected materials. We achieved $\alpha = .32$ for understandability and $\alpha = .18$ for actionability. In each wave, we failed to achieve acceptable reliability (defined in our case as $>.65$) on the overall understandability and actionability scores, as well as on almost all individual items.

Use of the PEMAT by practitioners without attention to these issues of reliability may lead to use of less effective health education materials. The aim of this article is to describe how we overcame these problems with reliability of the PEMAT to help practitioners avoid inadvertently selecting poor quality materials.

PROBLEM

We sought to further investigate the reliability of the PEMAT. We found that the original publications about the PEMAT never calculated the reliability of its overall score, which is what is used by practitioners. Instead, the reliability of each item was calculated (Shoemaker et al., 2014); by some conventional standards, the reliability for these items was not ideal (mean Cohen's kappa of 0.57 and range between 0.35 and 0.84). Given the limited reliability of many items, the overall reliability is likely unacceptable. We also identified an article on PEMAT reliability that found similar results and recommended changes to the codebook, some of which we independently arrived at as described below (Vishnevetsky et al., 2018). Finally, we found examples in the literature where researchers come to consensus, ignore reliability altogether, or find challenges in achieving reliability like we did (e.g., Lipari et al., 2019; Salama et al., 2020). For example, in one recent study, 45% to 67% of PEMAT ratings were due to the variability introduced by the raters (Salama et al., 2020). This is a problem for practitioners using the PEMAT. Scoring by any one person may give a substantially different result than when scored by another person.

In order to increase the reliability of scoring using the PEMAT tool, we annotated the PEMAT codebook to minimize ambiguous terminology. For example, Item 6 asks about numbers but is not clear if telephone numbers are included. Item 12 asks if "[t]he material uses visual cues (e.g., arrows, boxes, bullets, bold, larger font, highlighting) to draw attention to key points." This item did not clarify how fotonovela-style materials should be scored; therefore, the group edited the guidelines for the item to instruct scorers to assign a score of 1 to all fotonovela-style material. However, this did not solve our problems.

PATHWAY FORWARD

We went back into the literature on reliability, and we identified a solution for our project. Our initial approach was to ensure that coders would code the same document similarly. Then, one coder would code each piece of health education material. However, another approach is to calculate the reliability of a group of raters' *combined* ratings. This is more similar to scoring in some sports where a panel of judges each provides a rating, and the average is taken. This approach of using the average of scores between multiple coders instead of the similarity of scores between raters is an option in intraclass correlation measures of reliability (Koo & Li, 2016), which can easily be implemented in SPSS software. Using this averaging approach, we achieved acceptable reliability for the total score, with intraclass

correlations, respectively, of .76 and .73 for understandability and actionability. Specific combinations of raters had even higher reliability scores.

It is certainly possible that a different team of coders could have achieved acceptable reliability with extensive training and calibration. However, we think it concerning that our team, with past experience and interest in the topic, was unable to do so. Our inability to do so, even if others could do better, raises an important issue for practice.

IMPLICATIONS FOR PRACTICE

Some practitioners may be using the PEMAT to assess potential materials, as it was created for novice users and in response to limitations noted with other assessment tools (Shoemaker et al., 2014). Our experience suggests that PEMAT assessment should be conducted by two or more raters and the results should be averaged. Any one rater may introduce a substantial degree of variability from others rendering results unreliable. Practitioners should also consult with prior work suggesting clarifications to the PEMAT's codebook (Vishnevetsky et al., 2018). This is not to say the PEMAT and similar tools like the Centers for Disease Control and Prevention Clear Communication Index are not valuable. Rather, our experience suggests that use of the PEMAT and other ratings tools by practitioners without attention to these issues of reliability could produce results that may lead to use of less effective health education materials.

ORCID iD

Joseph G. L. Lee https://orcid.org/0000-0001-9698-649X

REFERENCES

Baur, C., & Prue, C. (2014). The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information. *Health Promotion Practice*, *15*(5), 629–637. https://doi.org/10.1177/1524839914538969

Beaunoyer, E., Arsenault, M., Lomanowska, A. M., & Guitton, M. J. (2017). Understanding online health information: Evaluation, tools, and strategies. *Patient Education and Counseling*, *100*(2), 183–189. https://doi.org/10.1016/j.pec.2016.08.028

Doak, C., Doak, L., & Root, J. (1996). *Teaching participants with low-literacy skills*. Lippincott Williams & Wilkins.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lee, J. (2020, May 26). *Education materials for farmworker health: A resource list*. https://doi.org/10.15139/S3/F1M9KC

Lipari, M., Berlie, H., Saleh, Y., Hang, P., & Moser, L. (2019). Understandability, actionability, and readability of online patient education materials about diabetes mellitus. *American Journal of Health-System Pharmacy*, *76*(3), 182–186. https://doi.org/10.1093/ajhp/zxy021

Salama, A., Panoch, J., Bandali, E., Carroll, A., Wiehe, S., Downs, S., Cain, M. P., Frankel, R., & Chan, K. H. (2020). Consulting "Dr. YouTube": An objective evaluation of hypospadias videos on a popular video-sharing website. *Journal of Pediatric Urology*, *16*(1), 70.e71–70.e79. https://doi.org/10.1016/j.jpurol.2019.11.011

Shoemaker, S. J., Wolf, M. S., & Brach, C. (2014). Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Education and Counseling*, *96*(3), 395–403. https://doi.org/10.1016/j.pec.2014.05.027

U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2010). *The national action plan to improve health literacy*. https://health.gov/sites/default/files/2019-09/Health_Literacy_Action_Plan.pdf

Vishnevetsky, J., Walters, C. B., & Tan, K. S. (2018). Interrater reliability of the Patient Education Materials Assessment Tool (PEMAT). *Patient Education and Counseling*, *101*(3), 490–496. https://doi.org/10.1016/j.pec.2017.09.003