

ORIGINAL ARTICLE

The patient engagement evaluation tool was valid for clinical practice guideline development

Ainsley Moore^{a,1}, Yin Wu^{b,c}, Linda Kwakkenbos^d, Kyle Silveira^e, Sharon Straus^{e,*},
Melissa Brouwers^f, Roland Grad^g, Brett D. Thombs^{b,c,h,i,j,k,l}

^aDepartment of Family Medicine, McMaster University, McMaster, Ontario, Canada

^bLady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

^cDepartment of Psychiatry, McGill University, Montreal, Quebec, Canada

^dDepartment of Clinical Psychology, Behavioral Science Institute, Radboud University, Nijmegen, The Netherlands

^eKnowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario

^fSchool of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

^gDepartment of Family Medicine, McGill University, Montreal, Quebec, Canada

^hDepartment of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

ⁱDepartment of Medicine, McGill University, Montreal, Quebec, Canada

^jDepartment of Psychology, McGill University, Montreal, Quebec, Canada

^kDepartment of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

^lBiomedical Ethics Unit, McGill University, Montreal, Quebec, Canada

Accepted 24 November 2021; Available online 28 November 2021

Abstract

Objective: To evaluate reliability and validity of the six and 12 item Patient Engagement Evaluation Tool (PEET) to inform guideline developers about the quality of patient and public involvement activities.

Study Design and Setting: PEET-12 and three embedded validation questions were completed by patients and members of the public who participated in developing 10 guidelines between 2018 and 2020. Confirmatory factor analysis (CFA) was used to assess the validity of a single-dimension factor structure. Cronbach's alpha and Pearson correlations were calculated for internal consistency reliability. Concurrent validation was used to test the construct validity.

Results: A total of 290 participants completed the PEET-12. To improve tool efficiency, based on results indicating redundancy from initial item analysis and experts' review, six of 12 items were included in the final tool (PEET-6). For the PEET-6, CFA supported the single-factor structure ($\chi^2(15) = 5173.4$, $P < 0.001$, Tucker-Lewis Index = 1.00, Comparative Fit Index = 0.99, Root Mean Square Error of Approximation = 0.08). Correlation between the total score for the 3 validation questions and the PEET-6 total score was 0.71, 95% CI [0.65, 0.77], supporting construct validity.

Conclusion: PEET-6 and 12 are valid tools to measure patient and public involvement within settings of clinical practice guideline development. © 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords: Clinical practice; engagement; guideline development; patient involvement; public involvement; engagement assessment

Conflict of interest: Dr. Ainsley Moore is Vice-Chair of the Canadian Task Force on Preventive Health Care. Dr. Kyle Silveira is a research coordinator with the Knowledge Translation Program at St. Michael's Hospital, Toronto, Ontario, and conducts patient engagement and other knowledge translation activities on behalf of the Canadian Task Force on Preventive Health Care. Dr. Sharon Straus leads the Knowledge Translation Program at St. Michael's Hospital, Toronto, Ontario and leads patient engagement activities and other knowledge translation activities on behalf of the Canadian Task Force on Preventive Health Care. Dr. Melissa Brouwers is director of the School of Epidemiology and Public Health at the University of Ottawa and leads evidence review and knowledge synthesis work on behalf of the Canadian Task Force on Preventive Health. Dr. Roland Grad is a member of the Canadian Task Force on Preventive Health Care. Dr. Brett Thombs is Chair of the Canadian Task Force on Preventive Health Care.

Funding: The Public Health Agency of Canada supported this work through an agreement with the Sir Mortimer B. Davis Jewish General Hospital. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor or subcontractor. Dr. Wu was supported by Fonds de recherche du Québec – Santé (FRQ-S) Postdoctoral Training Fellowship. Drs. Straus and Thombs were supported by Tier 1 Canada Research Chairs.

¹ Dr. Moore died in 2021 after this manuscript was submitted for review.

* Correspondence author. Tel.: 416 864 3068, Fax: 416 864 6035.

E-mail address: Sharon.straus@utoronto.ca (S. Straus).

What is new?

Key findings

- The PEET-6 and the PEET-12 were validated to assess patient and public engagement in clinical practice guideline development.
- The PEET was developed as a theory-informed measure of the extent to which criteria for successful engagement are met across domains including trust, respect, fairness, competency, legitimacy and accountability from a participant's perspective.
- To minimise response burden, guideline developers may prefer PEET-6.

1. Introduction

Meaningful patient and public involvement (PPI) in guideline development is an ethical imperative for developing trustworthy guidance. It is stipulated by the Guidelines International Network [1] and the Institute of Medicine-US (Now the National Academy of Medicine) [2] and emphasized in guideline quality appraisal standards (e.g., The Appraisal of Guidelines for Research & Evaluation Instrument) [3]. Guidelines developed with patient involvement are more likely to address patient preferences, provide recommendations that are better tailored to individual needs, and better support clinical decision making, particularly when practitioners perceive incongruency between patient preference and the guideline recommendations [4,5].

Guideline developers worldwide, including the Canadian Task Force on Preventive Health Care (CTFPHC) [6], United States Preventive Services Task Force [7], Scottish Intercollegiate Guidelines Network [8], and National Institute for Health and Care Excellence [9], undertake strategies to involve patients and the public in guideline development. Some criticized such strategies as tokenistic in some cases and potentially contributing to inequity in guideline recommendations [10,11], emphasizing the need for guideline developers to evaluate the quality of their engagement activities [12].

The Patient Engagement Evaluation Tool (PEET) was developed as a theory-informed measure of the extent to which criteria for successful engagement are met across domains (trust, respect, fairness, competency, legitimacy, accountability) from a participant's perspective [13]. PEET was applied to evaluate knowledge user engagement during the development of a systematic review of geriatrician-led models of care [12] and during guideline development by the CTFPHC [7], which produces clinical practice guidelines on primary preventive health care. The objectives of this project were to evaluate the reliability and validity of the PEET and to determine if it could be shortened without substantively changing measurement characteristics.

2. Methods

This cross-sectional study evaluated factor structure, reliability, and validity of the 12 item PEET, the selection of items for a shortened six item version, and similar testing with the six item version. Data were collected from members of the public who provided input into the development of 10 CTFPHC guidelines and completed the 12 PEET items between 2018 and 2020.

2.1. Participants and engagement activities

Between 10 and 26 individuals were recruited per guideline with attempts to include people from each Canadian province and territory. Participants were recruited through advertisements on public websites (e.g., Kijiji, Craigslist), the CTFPHC website, the website of the Knowledge Translation Program (KTP) of St. Michael's Hospital (Toronto, Ontario, Canada), and from a KTP database of individuals who had expressed interest in providing feedback on CTFPHC guidelines and tools [7,13]. People expressing interest completed an online eligibility survey containing demographic, health, health equity, and conflict of interest questions.

Participants representing the guideline target population were engaged at two stages of guideline development. The 12-item PEET was completed after each stage. In stage 1, participants used the Grading of Recommendations Assessment, Development and Evaluation outcome rating approach [14] to rate the extent to which a series of pre-defined screening outcomes (benefits and harms) were either; *not important* (rating 1–3), *important* (rating 4–6), or *critical* (rating 7–9) for making decisions relevant to the guideline topic. For example, reduced risk of infection transmission due to screening for chlamydia and gonorrhea was an outcome rated by participants. They were also asked to list other outcomes they deemed important. This was followed by an online moderated focus group where participants discussed their outcome ratings.

For stage 2 of each guideline input process, a different group of participants were provided with the evidence summary from the systematic review to evaluate their preferences when considering undergoing a screening intervention (or not) for a specific health condition (such as colon cancer). Participants used a 9-point scale via an online survey to rate the extent to which each outcome would influence their decision to be screened for the health condition (1= *This isn't important for my decision at all* to 9= *This is very important for my decision*) [15]. Consent was obtained and participants were engaged in a 60-minute moderated, recorded focus group via teleconference, which included a CTFPHC content expert to answer any questions, to discuss the survey outcomes and general screening preferences. One week after the focus group, these participants completed an online survey to assess their engagement (PEET) and experience with this project stage.

Table 1. Demographic characteristics ($n = 290$)

Demographic Variable	PEET Group
Age in years, mean (SD) ^a	45.0 (17.9)
Woman, n (%)	210 (72.4)
Residence, n (%) ^a	15.8 (3.5)
Rural	35 (12.4)
Suburban	76 (27.0)
Urban	171 (60.6)
Highest education level, n (%) ^a	
Less than high school	2 (0.7)
high school	36 (12.8)
College diploma or bachelor's degree	164 (58.2)
Graduate or professional degree	80 (28.4)
Annual household income, n (%) ^b	
Less than \$25,000	0 (0.0)
\$25,000-\$29,999	73 (26.0)
\$30,000-\$39,999	29 (10.3)
\$40,000-\$49,999	25 (8.9)
\$50,000-\$59,999	33 (11.7)
\$60,000-\$69,999	21 (7.5)
\$70,000-\$99,999	55 (19.6)
\$100,000 or more	45 (16.0)
Race/ethnicity, n (%) ^c	
White	92 (66.7)
Asian	22 (15.9)
Indigenous Canadian	14 (10.1)
Black	4 (2.9)
Hispanic	4 (2.9)
Arabic	2 (1.4)

Due to missing data:

^a $n = 282$

^b $n = 281$

^c $n = 152$

Details about data collection are available in previous publications [6,13].

2.2. Measures

2.2.1. Demographic Characteristics

Participants provided demographic data including age, gender (woman, man, other), education level (less than high school, high school, college diploma or bachelor's degree, graduate or professional degree), race/ethnicity, place of residence (rural, urban, suburban) annual household income level (less than \$25,000, \$25,000-\$29,999, \$30,000-\$39,999, \$40,000-\$49,999, \$50,000-\$59,999, \$60,000-\$69,999, \$70,000-\$99,999, \$100,000 or more).

2.2.2. The PEET

The PEET tool was designed to quantify the level of participant engagement in clinical practice guideline devel-

opment using theory-informed meta-criteria, or domains, from a stakeholder perspective [13]. The meta-framework was based on democratic participation principles [16]. The tool gauges participants' opinions regarding the extent to which each attribute was present during their engagement activity across six domains: trust, respect, accountability, legitimacy, competency, and fairness [13].

The original 12 item PEET tool included two items for each of the six domains except for fairness, which has three items, and trust, which has one item. Items were rated on a seven point adjectival Likert scale (ranging from 1 = no extent to 7 = very large extent). Survey items (see Appendix A) were tailored to the engagement activities employed. For example, "To what extent do you believe that your ideas were heard during the engagement process?" Respondents were asked to explain their choices if they rated any item one to four (text entry). The score for the

Table 2. Score and response distribution for all PEET items (possible item scores 1–7)

PEET Questions	Score	Item Responses ^a						
	Mean (SD)	1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)	6 n (%)	7 n (%)
1. To what extent do you believe that your ideas were heard during the engagement process?	5.3 (1.1)	1 (0.3)	1 (0.3)	13 (4.5)	53 (18.3)	107 (36.9)	68 (23.4)	47 (16.2)
2. To what extent did you feel comfortable contributing your ideas to the engagement process?	5.3 (1.1)	1 (0.3)	1 (0.3)	9 (3.1)	53 (18.3)	100 (34.5)	81 (27.9)	45 (15.5)
3. To what extent do you believe organizers took your contributions to the engagement process seriously?	5.5 (1.1)	1 (0.3)	0 (0.0)	9 (3.1)	44 (15.2)	100 (34.5)	72 (24.8)	64 (22.1)
4. To what extent do you believe that your input will influence final decisions that underlie the engagement process?	4.4 (1.2)	1 (0.3)	15 (5.2)	40 (13.8)	103 (35.5)	83 (28.6)	33 (11.4)	15 (5.2)
5. To what extent were you able to clearly express your viewpoints?	5.1 (1.1)	3 (1)	10 (3.4)	36 (12.4)	103 (35.5)	91 (31.4)	37 (12.8)	10 (3.4)
6. To what extent were organizers neutral in their opinions (regarding topics) during the engagement process?	5.5 (1.2)	2 (0.7)	1 (0.3)	12 (4.1)	42 (14.5)	92 (31.7)	61 (21)	80 (27.6)
7. To what extent did all participants have equal opportunity to participate in discussions?	5.4 (1.2)	1 (0.3)	3 (1)	12 (4.1)	41 (14.1)	100 (34.5)	59 (20.3)	74 (25.5)
8. To what extent did you clearly understand your role in the process?	5.4 (1.1)	0 (0.0)	0 (0.0)	16 (5.5)	39 (13.4)	106 (36.6)	77 (26.6)	52 (17.9)
9. To what extent was information made available to you either prior or during the engagement process so as to participate knowledgeably in the process?	5.5 (1.1)	0 (0.0)	1 (0.3)	6 (2.1)	38 (13.1)	112 (38.6)	67 (23.1)	66 (22.8)
10. To what extent were the ideas contained in the information material easy to understand?	5.3 (1.2)	0 (0.0)	2 (0.7)	19 (6.6)	57 (19.7)	87 (30)	77 (26.6)	48 (16.6)
11. To what extent did you clearly understand what was expected of you during the engagement process?	5.4 (1.1)	0 (0.0)	1 (0.3)	15 (5.2)	38 (13.1)	99 (34.1)	81 (27.9)	56 (19.3)
12. To what extent did you clearly understand what the goals of the engagement process were?	5.2 (1.2)	0 (0.0)	1 (0.3)	19 (6.6)	54 (18.6)	91 (31.4)	79 (27.2)	45 (15.5)
Validation Questions								
1. To what extent do you believe that your values and preferences will be included in the final health advice from this process?	4.5 (1.1)	0 (0.0)	3 (1)	15 (5.2)	74 (25.5)	90 (31)	74 (25.5)	34 (11.7)

(continued on next page)

Table 2 (continued)

PEET Questions	Score Mean (SD)	Item Responses ^a						
		1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)	6 n (%)	7 n (%)
2. To what extent would you follow health advice from the Canadian Task Force on Preventive Health Care (if it related to your health condition)?	5.0 (1.2)	2 (0.7)	4 (1.4)	15 (5.2)	76 (26.2)	88 (30.3)	63 (21.7)	42 (14.5)
3. To what extent would you advise others to follow health advice from the Canadian Task Force on Preventive Health Care (if it related to their health condition)?	5.0 (1.4)	9 (3.1)	5 (1.7)	16 (5.5)	71 (24.5)	89 (30.7)	56 (19.3)	44 (15.2)

^a On a 7-point scale, 1 = not at all to 7 = very large extent.

scale was the total of all items, with higher scores reflecting greater engagement.

2.2.3. Validation Questions

Three validation items embedded in the survey (see Appendix A) evaluated concurrent validation [17] by assessing convergence between the overall construct, degree of engagement, and the extent to which participants believed that: (1) Their values were reflected in the final conclusions of the patient and public engagement activity; Their degree of 'buy in' with the engagement process as measured by their intent to (2) Follow the health recommendations for which they participated in developing, and (3) Advise others to follow those health recommendations. For consistency, validation items were also rated on a seven point adjectival scale (ranging from 1 = no extent to 7 = very large extent). We hypothesized that high levels of overall meaningful engagement (total PEET scores) would be associated with high scores for these validation items.

2.3. Measure Shortening

Two investigators with experience in guideline development and patient engagement (AM, RG) initially selected one item of each of the six PEET domains from the 12 item version (to retain one from item each domain) for inclusion in the shortened six item version. Items selected were deemed to have better face validity and were discussed and agreed upon via a consensus process with other research team members. We created and tested a shortened version in response to patient suggestions to consider response burden.

2.4. Statistical Analysis

All analyses (descriptive statistics, reliability assessment, factor analysis, and assessment of concurrent valid-

ity), were carried out for both the 12 item and shortened six item PEET after item reduction.

Means and standard deviations (SDs) summarized continuous demographic variables, and percentages were used for categorical variables. For each PEET item, means, standard deviations, frequency of endorsement of each response option, and corrected item-total correlations were calculated. Means and SDs were also calculated for total scale scores. Floor and ceiling effects were examined, defined as $\geq 15\%$ of the participants having the lowest or highest possible score, respectively [18,19].

Inter-item correlations were calculated, and Cronbach's alpha was used to assess the internal consistency of the PEET. We planned a priori to consider item reduction to improve tool efficiency and decrease participant burden [18,19] if the internal consistency of the 12-item version was greater than 0.95, signaling item redundancy [17].

Construct validity was assessed using Confirmatory Factor Analysis (CFA) and concurrent validation methods. CFA was selected to confirm the validity of a unidimensional structure of item responses, as identified a priori by the developers. Unidimensionality was proposed because PEET domains and items were closely related, and they all measured an overall engagement construct. CFA used the weighted least squares estimator with a diagonal weight matrix, robust standard errors, and a mean-and-variance-adjusted chi-square statistic with delta parameterization in MPlus seven [20]. Model adequacy was assessed using a chi-square goodness-of-fit test and three fit indices, including the Tucker-Lewis Index (TLI) [21], the Comparative Fit Index (CFI) [22], and the Root Mean Square Error of Approximation (RMSEA) [23]. Since the chi-square test is susceptible to sample size and can lead to the rejection of well-fitting models, practical fit indices (TLI, CFI, RMSEA) were emphasized [24]. Models with a TLI and CFI close to 0.95 or higher and RMSEA close to 0.06 or lower are representative of good fitting models [25]. Since RMSEA is calculated partially based on chi-square,

Table 3. Characteristics of the patient engagement evaluation tool

Question ^a	Corrected Item-total Correlation: 12 item	Corrected Item-total Correlation: 6 item	CFA ^b Factor Loading:12 item model	CFA ^b Factor Loading:6 item model
Questions Included in 6 question PEET				
1. To what extent do you believe that your ideas were heard during the engagement process?	0.82	0.83	0.89	0.92
3. To what extent do you believe organizers took your contributions to the engagement process seriously?	0.83	0.82	0.90	0.91
4. To what extent do you believe that your input will influence final decisions that underlie the engagement process?	0.59	0.59	0.73	0.75
5. To what extent were you able to clearly express your viewpoints?	0.79	0.78	0.86	0.86
7. To what extent did all participants have equal opportunity to participate in discussions?	0.78	0.76	0.84	0.85
9. To what extent was information made available to you either prior or during the engagement process so as to participate knowledgeably in the process?	0.77	0.69	0.84	0.78
Questions in the 12 question PEET removed after first CFA				
2. To what extent did you feel comfortable contributing your ideas to the engagement process?	0.80	————	0.87	————
6. To what extent were organizers neutral in their opinions (regarding topics) during the engagement process?	0.75	————	0.82	————
8. To what extent did you clearly understand your role in the process?	0.76	————	0.84	————
10. To what extent were the ideas contained in the information material easy to understand?	0.76	————	0.86	————
11. To what extent did you clearly understand what was expected of you during the engagement process?	0.82	————	0.91	————
12. To what extent did you clearly understand what the goals of the engagement process were?	0.79	————	0.90	————

^a Question numbers were consistent with that in Table 2.^b On a 4-point scale, 1, 2, 3, and 4 were combined into one category.

a RMSEA of 0.08 or more [26] may also be considered to represent a reasonably acceptable model fit. Item response categories were combined in cases where the spread of the distribution of responses was too sparse, including having no responses, across one or more categories for CFA modelling [20,27]. Previous studies have found that collapsing categories with few responses in CFA leads to scales with roughly equivalent psychometric properties, including factor structure [27].

Pearson's correlations (r) with 95% confidence intervals were used to assess the strength of the relationships between participant ratings of embedded validation questions and total PEET scores [17]. Generally, correlations greater than 0.40 suggest construct validation of the instrument, in this case reflecting participant buy-in and potential uptake of the recommendation [17].

The 95% CIs for the difference between correlations (Δr) with each of the three validation items were calculated [28] to compare differences in r between the 12 item versus the shortened 6-item tool.

3. Results

3.1. Participant characteristics

A total of 304 members of the public provided input on CTFPHC guideline development during the 2-year study period, of whom 299 (98%) completed the PEET. Of these, nine participants who submitted responses for the PEET did not answer all items. Therefore, 290 participants (95%) with complete PEET responses were included. Most of the respondents were women (72%), attained college diplomas or bachelor's degrees (58%), were living in urban areas (61%) and self-identified as white (67%). The mean age was 45 (SD = 18) years (Table 1).

3.2. Evaluation of the 12 item version of the PEET

3.2.1. Item Statistics and Reliability

For the 12 item PEET, mean (SD) total score was 63.3 (11.1) (median = 63.0, range 29.0 to 84.0, skewness = -0.08, kurtosis = -0.53). Mean item scores ranged from 4.4 for Item 4 to 5.5 for Items 3, 7 and 10 (Table 2). Cronbach's alpha of the 12 item PEET was 0.95, indicating redundancy among items. Responses for validation items are also shown in Table 2.

Correlations between item scores ranged from $r = 0.40$ ($P < 0.01$, Items 4 and 8) to $r = 0.80$ ($P < 0.01$, Items 11 and 12) (Appendix B1). In addition, the correlations between Items 1 and 2 ($r = 0.78$), 1 and 3 ($r = 0.78$), 2 and 3 ($r = 0.76$), 1 and 5 ($r = 0.75$), 2 and 5 ($r = 0.75$), 3 and 7 ($r = 0.72$), 8 and 11 ($r = 0.73$), and 10 and 11 ($r = 0.72$), were all > 0.70 (all $P < 0.01$) (Appendix B1). Corrected item-total correlations ranged from $r = 0.59$ (Item 4) to $r = 0.83$ (Item 3) (Table 3). No participants had the lowest possible total score (12.0) on the scale, and seven (2.4%)

had the highest possible score (84.0), suggesting that there were no floor or ceiling effects.

3.2.2. Confirmatory Factor Analysis

Given the sparse responses in lower response categories, we collapsed categories (Table 2) [26] and modelled with responses 1-4 in a single category, leaving four response categories (1-4, 5, 6, 7). Model fit for a one-factor solution was good based on the CFI and TLI, although sub-optimal based on the RMSEA ($\chi^2(66) = 13360.7$, $P < 0.001$; CFI = 0.98; TLI = 0.98; RMSEA = 0.13). Factor loadings range from 0.73 (Item 4) to 0.91 (Item 11), see Table 3.

3.3. Item reduction and 6 item Patient Engagement Evaluation Tool

Considering the high degree of inter-item correlations and internal consistency ($\alpha = 0.95$) found for the 12 item tool, items 2, 6, 8, 10, 11 and 12 were removed, and a six item version (one item for each domain) of the tool (Fig. 1) was selected for testing [29].

3.4. Evaluation of the 6 item version of the PEET

3.4.1. Item Statistics and Reliability

For the PEET-6, mean (SD) total score was 31.2 (5.7) (median = 30.0, range 14.0 to 42.0, skewness = -0.02, kurtosis = -0.53). Mean item scores ranged from 4.4 for Item four to 5.5 for Item nine (Table 2). Corrected item-total correlations ranged from $r = 0.59$ (Item 4) to $r = 0.83$ (Item 1) (Table 3). Correlations between items ranged from $r = 0.40$ ($P < 0.01$, Items 4 and 8) to $r = 0.76$ ($P < 0.01$, Items 1 and 3) (Appendix B2).

Cronbach's alpha for six item PEET was 0.92, reflecting good internal consistency across scores. No participants had the lowest possible total score (6.0) on the scale, and 9 (3.1%) had the highest possible score (42.0), suggesting that there were no floor or ceiling effects.

3.4.2. Confirmatory Factor Analysis

Confirmatory factor analysis was also performed on the 6 items to confirm the unidimensional construct of the instrument (Table 3). Inspection of the indices indicated good model fit based on the CFI and TLI, and acceptable based on RMSEA ($\chi^2(15) = 5173.4$, $p < 0.001$, TLI = 1.00, CFI = 0.99, RMSEA = 0.08). All factor loadings were adequate, with factor loadings ranging from 0.75 (Item 4) to 0.92 (Item 1), see Table 3.

3.5. Concurrent Construct Validity

The correlation between total score for the three validation questions and the 12 item version total score (r) was 0.70 (95% CI 0.63, 0.75) vs. 0.71 (95% CI 0.65, 0.77) for the six item version. Both were greater than 0.40, which

Please respond to each of the following statements using the scales provided.

If you select 1–4 for any question, please explain your rating in the space below the question.

1. To what extent do you believe that your ideas were heard during the engagement process?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

2. To what extent do you believe organizers took your contributions to the engagement process seriously?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

3. To what extent do you believe that your input will influence final decisions that underlie the engagement process?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

Fig. 1. Final 6 item Patient Engagement Evaluation Tool.

supported the construct validation of the instruments [17]. The correlation between the validation question scores and the 12 question total score was slightly lower than that for the six question version, but the difference was not statistically significant ($\Delta r = -0.018$, 95% CI -0.076 to 0.004).

4. Discussion

Patients and members of the public who provided input into CTFPHC guideline development reported high levels of engagement. Clustering of responses was noted at the upper end of the scale, but neither a ceiling nor floor effect for total scores was found. High levels of inter-item correlations and internal consistency for the 12 item PEET suggested item redundancy, specifically potential conceptual overlap between questions. Consequently, a shorter 6-item version of the tool was developed, with similarly good reliability. CFA found a good model fit for both versions

of the tool and identified a single dimension to the data. Good measures of concurrent validation were found for both tools with no difference between versions. Considering decreased respondent burden and similar reliability and validity measures, the more economical six item tool is preferred.

Limitations have been identified with the growing number of tools available to evaluate patient and public engagement in health care policy and research development. These include the lack of validation and evaluation of measurement properties (92% of tools did not report reliability measures), lack of a theory-based framework [30], and lack of specification of purpose and context of the engagement activity for which the tool is intended. [31] The PEET-6 is an efficient tool that addresses these gaps, has good measurement properties, and is theoretically informed and specifically intended to support patient and public involvement in the context of guideline development activities.

4. To what extent were you able to clearly express your viewpoints?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

5. To what extent did all participants have equal opportunity to participate in discussions?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

6. To what extent was information made available to you either prior or during the engagement process so as to participate knowledgeably in the process?

No Extent	Very Small Extent	Small Extent	Fair Extent	Moderate Extent	Large Extent	Very Large Extent
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Explanation:

 Canadian Task Force on Preventive Health Care.

Fig. 1. Continued

Guideline developers face challenges in stakeholder engagement throughout guideline development. They adopted various approaches to incorporate the perspectives of patients and the public. Some have criticized such efforts as tokenistic, identifying lack of participant remuneration, failure to prepare participants adequately (e.g., materials, knowledge), and other barriers to meaningful engagement [11]. Such limitations have been identified as ultimately “fueling inequity in guidelines” [11]. To our knowledge, the PEET is the first tool designed and evaluated for reliability and validity in the context of guideline development. Similar tools in other contexts include a generic 21 item instrument developed by Abelson et al., which is much longer than the PEET-6 [31]. It is intended for broad application in healthcare organizations, provides a qualitative assessment of the engagement process, and is supported by face, content [2], and usability testing [30]. Item generation was based on literature review and consensus among engagement experts. Stocks et al. [32] also developed a generic tool, in this case, to support healthcare researchers

by providing quantitative outcomes of the quality of the engagement process. This theory-informed, 24-item tool is supported by acceptable to good internal consistency (Cronbach’s α 0.74–0.81), with discriminatory ability to measure decreasing scores in engagement quality over time (within-subject test, re-test). Still, it is limited by a ceiling effect to measure improved engagement experience over time. It is also much longer than the PEET-6.

There are limitations to consider in interpreting our results. Future analyses by other guideline developers should consider a test, re-test reliability (post-engagement) analysis to explore stability of responses within participants. Inter-rater reliability could be done to understand the tool’s capacity to discriminate between types of engagement activities (e.g., focus groups, interviews, surveys) and stages of engagement (e.g., outcome prioritization, recommendation formulation, dissemination activities). Such findings may identify optimal strategies for engagement during guideline development. Reliability and validity testing in other guideline development groups is encouraged to con-

firm confidence with the unidimensionality of the construct and internal consistency of the items.

5. Conclusion

Both PEET-12 and PEET-6 provide guideline developers with a measure of the overall quality of their patient and public engagement activities, ultimately supporting the development of implementable, meaningful and equitable clinical practice guideline recommendations. To minimize response burden, guideline developers may prefer PEET-6.

Author Contributions

Ainsley Moore: Conceptualization, Study design, Data collection and cleaning, Interpretation of the analysis, Writing – original draft, review & editing. Yin Wu: Conceptualization, Study design, Data analysis, Interpretation of the output, Writing – original draft, review & editing. Linda Kwakkenbos: Data analysis, Writing – review & editing. Kyle Silveira: Data collection and cleaning, Writing – review & editing. Sharon Straus: Conceptualization, Study design, Data collection and cleaning, Writing – review & editing. Melissa Brouwers: Conceptualization, Study design, Data collection and cleaning, Writing – review & editing. Roland Grad: Interpretation of the analysis,

Writing – review & editing. Brett D. Thombs: Conceptualization, Study design, Data analysis, Interpretation of the output, Writing – review & editing.

Acknowledgments

We would like to thank Ms. Danica Buckland for contributing to the data collection and refinement of the 12-item PEET.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Appendix A: Patient Engagement Evaluation Tool (12-item with 3 validation items)

Please respond to each of the following statements using the scales provided.

Please respond to each question using the following ratings: 1: Not at all (no extent) 2: Very small extent 3: Small extent 4: Fair extent 5: Moderate extent 6: Large extent 7: Very large extent.

If you select 1-4 for any question, please explain your rating in the space below the question.

Appendix B1. Inter-item Correlation Matrix for the 12-item PEET.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12
Item 1	1.00	0.75**	0.75**	0.58**	0.72**	0.65**	0.70**	0.59**	0.60**	0.60**	0.61**	0.59**
Item 2		1.00	0.70**	0.48**	0.75**	0.64**	0.62**	0.60**	0.60**	0.65**	0.63**	0.60**
Item 3			1.00	0.53**	0.67**	0.70**	0.72**	0.63**	0.66**	0.63**	0.68**	0.66**
Item 4				1.00	0.55**	0.42**	0.43**	0.38**	0.40**	0.42**	0.46**	0.49**
Item 5					1.00	0.57**	0.64**	0.58**	0.57**	0.60**	0.64**	0.66**
Item 6						1.00	0.59**	0.62**	0.66**	0.64**	0.61**	0.59**
Item 7							1.00	0.63**	0.64**	0.60**	0.67**	0.64**
Item 8								1.00	0.67**	0.63**	0.73**	0.68**
Item 9									1.00	0.69**	0.70**	0.63**
Item 10										1.00	0.72**	0.62**
Item 11											1.00	0.80**
Item 12												1.00

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix B2. Inter-item Correlation Matrix for the 6 item PEET.

	Item 1	Item 3	Item 4	Item 5	Item 7	Item 9
Item 1	1.00	0.76**	0.57**	0.75**	0.70**	0.61**
Item 3		1.00	0.54**	0.68**	0.72**	0.67**
Item 4			1.00	0.59**	0.44**	0.42**
Item 5				1.00	0.65**	0.57**
Item 7					1.00	0.64**
Item 9						1.00

**. Correlation is significant at the 0.01 level (2-tailed).

- 1) To what extent do you believe that your ideas were heard during the engagement process?
- 2) To what extent did you feel comfortable contributing your ideas to the engagement process?
- 3) Did organizers take your contributions to the engagement process seriously?
- 4) To what extent do you believe that your input will influence final decisions that underlie the engagement process?
- 5) To what extent do you believe that your values and preferences will be included in the final health advice from this process?
- 6) To what extent were you able to clearly express your viewpoints?
- 7) How neutral in their opinions (regarding topics) were organizers during the engagement process?
- 8) Did all participants have equal opportunity to participate in discussions?
- 9) How clearly did you understand your role in the process?
- 10) To what extent was information made available to you either prior or during the engagement process so as to participate knowledgeably in the process?
- 11) To what extent were the ideas contained in the information material easy to understand?
- 12) How clearly did you understand what was expected of you during the engagement process?
- 13) How clearly did you understand what the goals of the engagement process were?
- 14) To what extent would you follow health advice from the Canadian Task Force on Preventive Health Care (if it related to your health condition)?
- 15) To what extent would you advise others to follow health advice from the Canadian Task Force on Preventive Health Care (if it related to their health condition)?

References

- [1] Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: Toward International Standards for Clinical Practice Guidelines. *Annals Int Med* 2012;157(7):525–31. doi:10.7326/0003-4819-156-7-201204030-00009.
- [2] Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust*. Graham R, Mancher M, Wolman Miller, et al., editors. Washington: National Academies Press (US); 2011.
- [3] Brouwers M, Kho ME, Browman GP, Cluzeau F, Feder G, Fervers B, et al. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Can Med Assoc J* 2010;182(18):E839–42. doi:10.1503/cmaj.090449.
- [4] Murad MH, Montori VM, Guyatt GH. Incorporating patient preferences in evidence-based medicine. *JAMA* 2008;300(21):2483–4. doi:10.1001/jama.2008.730.
- [5] Barratt A. Evidence based medicine and shared decision making: The challenge of getting both evidence and preferences into health care. *Patient Educ Couns* 2008;73(3):407–12. doi:10.1016/j.pec.2008.07.054.
- [6] Buckland D, Bashir N, Moore JE, Straus S, et al. CTFPHC patient engagement protocol. Toronto: Li Ka Shing Institute, St. Michael's Hospital. Available at: <http://canadiantaskforce.ca/methods/patient-engagement-protocol>. Accessed 22 April 2021.
- [7] Procedure Manual Section 9. Engagement With the Public, Stakeholders, and Partners. Available at: <https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-section-9-engagement-public-stakeholders-and-partners>. Accessed 11 Nov 2021.
- [8] SIGN Patient and public involvement. Available at: <https://www.sign.ac.uk/patient-and-public-involvement>. Accessed 22 Apr 2021.
- [9] NICE Patient and public involvement policy. Available at: <https://www.nice.org.uk/about/nice-communities/nice-and-the-public/public-involvement/public-involvement-programme/patient-public-involvement-policy>. Accessed 22 Apr 2021.
- [10] Légaré F, Boivin A, van der Weijden T, Pakenham C, Burgers J, Légaré J, et al. Patient and public involvement in clinical practice guidelines: a knowledge synthesis of existing programs. *Med Decis Making* 2013;31(6):E45–72. doi:10.1177/0272989X11424401.
- [11] Lang E, da Silva SA, Persaud N. Are guidelines fueling inequity? A call to action for guideline developers and their panelists. *Chest* 2021;159(2):465–6. doi:10.1016/j.chest.2020.10.036.
- [12] Soobiah C, Straus SE, Manley G, Marr S, Jenssen EP, Teare S, et al. Engaging knowledge users in a systematic review on the comparative effectiveness of geriatrician-led models of care are possible: A cross-sectional survey using the Patient Engagement Evaluation Tool. *J Clin Epidemiol* 2019;113:58–63. doi:10.1016/j.jclinepi.2019.05.015.
- [13] Moore A, Brouwers M, Straus SE, Tonelli M. Advancing patient and public involvement in guideline development. Ottawa: ON: Canadian Taskforce on Preventative Health Care; 2015.
- [14] Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. editors. The GRADE Working Group; 2013. Available at guidelinedevelopment.org/handbook Accessed 24 Apr 2021.
- [15] Fitch K, Bernstein S, Aguilar M, Burnand B, LaCalle J, Lazaro P, et al. The RAND/UCLA Appropriateness Method User's Manual, Santa Monica, CA: RAND Corporation; 2001. Available at: https://www.rand.org/pubs/monograph_reports/MR1269.html Also available in print form.
- [16] Deverka PA, Lavalley DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *J Comp Eff Res* 2012;1(2):181–94. doi:10.2217/ce.12.7.
- [17] Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press; 2015.
- [18] Terwee CB, Bot SDM, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012.
- [19] Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN standards and criteria for evaluating the content validity of health-related Patient-Reported Outcome Measures: a Delphi study. *Qual Life Res* 2018;27(5):1159–70. doi:10.1007/s11336-018-1829-0.
- [20] Kwakkenbos L, Jewett LR, Baron M, Bartlett SJ, Furst D, Gottesman K, et al. The Scleroderma Patient-centered Intervention Network (SPIN) Cohort: Protocol for a cohort multiple randomised controlled trial (cmRCT) design to support trials of psychosocial and rehabilitation interventions in a rare disease context. *BMJ Open* 2013;3(8):e003563. doi:10.1136/bmjopen-2013-003563.
- [21] Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 1973;38(1):1–10. doi:10.1007/BF02291170.

- [22] Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107(2):238. doi:[10.1037/0033-2909.107.2.238](https://doi.org/10.1037/0033-2909.107.2.238).
- [23] Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav Res* 1990;25(2):173–80. doi:[10.1207/s15327906mbr2502_4](https://doi.org/10.1207/s15327906mbr2502_4).
- [24] Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychol Bull* 1993;114:552. doi:[10.1037/0033-2909.114.3.552](https://doi.org/10.1037/0033-2909.114.3.552).
- [25] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6(3):1–55. doi:[10.1037/0033-2909.114.3.552](https://doi.org/10.1037/0033-2909.114.3.552).
- [26] Browne MW, Cudeck R, Bollen KA, Long JS. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. *Testing structural equation models*. Beverly Hills (CA): Sage; 1993. p. 136–62.
- [27] Colvin KF, Gorgun G. Collapsing scale categories: Comparing the psychometric properties of resulting scales. *Pract Assess Res* 2020;25(1):6.
- [28] El-Baalbaki G, Lober J, Hudson M, Baron M, Thombs BD. Measuring pain in systemic sclerosis: Comparison of the short-form McGill Pain Questionnaire versus a single-item measure of pain. *Rheumatology* 2011;38(12):2581–7. doi:[10.3899/jrheum.110592](https://doi.org/10.3899/jrheum.110592).
- [29] Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill; 1978.
- [30] Boivin A, L'Espérance A, Gauvin FP, Dumez V, Macaulay AC, Lehoux P, et al. Patient and public engagement in research and health system decision making: A systematic review of evaluation tools. *Health Expect* 2018;21(6):1075–84. doi:[10.1111/hex.12804](https://doi.org/10.1111/hex.12804).
- [31] Abelson J, Humphrey A, Syrowatka A, Bidonde J, Judd M. Evaluating patient, family and public engagement in health services improvement and system redesign. *Health Q* 2018;21(SP):61–7. doi:[10.12927/hcq.2018.25636](https://doi.org/10.12927/hcq.2018.25636).
- [32] Stocks S, Giles S, Cheraghi-Sohi S, Campbell S. Application of a tool for the evaluation of public and patient involvement in research. *BMJ Open* 2015;5:e006390. doi:[10.1136/bmjopen-2014-006390](https://doi.org/10.1136/bmjopen-2014-006390).