Supplementary File 1:

1a. Description of the Patient and Public Involvement and Engagement (PPIE) process

1b. Description of the cognitive interview process

1a. Description of the Patient and Public Involvement and Engagement (PPIE) process

Purpose of the Focus Group

A PPIE focus group was conducted with the RUG in September 2018 to seek feedback on the first draft of the ATEMPT. The aim of this meeting was to identify any elements of the questionnaire that patients may find unclear, difficult to understand, or ambiguous, before it was distributed to a large sample of physiotherapy patients to gather data for further testing. Feedback from the RUG was sought as the ATEMPT was designed to be used as a self-completion questionnaire. Consequently, patients had assumed an equal role alongside researchers and physiotherapists in the creation of statements that were developed into items for the measure. While this process has been used in order to improve the validity of the final tool and its relevance to the patients who would eventually use it, inevitably, the final items were written by the research team. The RUG was consulted to ensure that the comprehensibility, relevance, acceptability, and ease of completion of the items, had not been compromised during the development stage and they were still appropriate to be used with lay individuals such as patients. The RUG’s feedback was also sought on elements of the questionnaire’s structure where guidance in the literature had been equivocal.

Content of the Focus Group

The PPIE focus group was attended by six RUG members, all of whom had experience of living with an MSK pain condition. DB led and facilitated the meeting, while AB took notes. A presentation was given to the group, outlining the project so far and the aims of the focus group. The group were then invited to complete the draft ATEMPT and take part in a series of activities designed to encourage feedback from the members on the following aspects of the draft measure:

- Whether items would allow patients to respond negatively as well as positively
• Whether the meaning of each item was clear
• Whether all the items could be answered satisfactorily, i.e., were the response options adequate and appropriate?
• Whether any words used were unclear or could have multiple meanings
• Which introductory question from a selection provided was clearest
• Which five or seven option Likert scale from a selection provided was most appropriate for the items in the measure
• Whether ‘Don’t know’, ‘Not applicable’ or ‘Not sure’ response options were needed

Feedback from the Focus Group

Item Feedback
RUG members were encouraged to verbalise any comments they had about the draft measure and DB facilitated discussion ensuring all members were given the opportunity to share their ideas. The RUG raised concerns in the following categories relating to the items in ATEMPT:

• Words that may have different meanings to different people.
  o For example, the term ‘exercise specialist’ was felt to be confusing and associated more with exercise instructors in a gym setting than with HCPs. This term was initially adopted so that the measure could be used with patients who had been advised exercises from any kind of HCP, not just physiotherapists.

• Words that were not familiar or appropriate.
  o For example, the term ‘prescription’ was disliked and had been used multiple times in the context of ‘exercise prescription’. Members felt that the term was associated with, and specific to, medication prescription and should therefore be replaced when used in the context of exercise prescription.

• Terms that were not appropriate for a lay reader.
  o For example, in the item ‘I can assess whether I am doing my exercises correctly’, it was noted that a term more suitable for lay patients than ‘assess’ should be used.
• Items that seemed to be asking the same thing as other items.
  o For example, the item ‘I am motivated to do my exercises’ was deemed similar to the item ‘I have the motivation needed to undertake my exercises’.

• Phrasing which was ambiguous
  o For example, in the item ‘I complete my exercises on a regular basis’, the term ‘regular’ was deemed to be too vague to determine an appropriate response to the item.

Following a review of the feedback obtained from the members of the RUG during the PPIE focus group, eight out of the 56 items presented to the RUG were removed for one or more of the following reasons: the item contained words that were interpreted inconsistently; the item contained words that had unclear meanings; the item referred to a specific situation that would not be relevant to all patients; the item lacked clarity regarding the timeframe it referred to; the meaning of the item was generally not understood; the item asked the same thing as another item; and the item contained medical terminology for which a consensus on an acceptable lay term could not be reached.

Consequently, the 56-item version of ATEMPT was reduced to a 48-item version.

Response Option Feedback
The RUG members were presented with five different Likert response options. These ranged from five to seven response option variants, some of which included ‘not applicable’, ‘don’t know’ or ‘not sure’ response options. Members reported that the additional response options available in the seven-option variant were unnecessary as it was not possible to differentiate between the addition levels of agreement available. It was also decided by the members that a ‘not applicable’, ‘don’t know’ or ‘not sure’ response option was not required as for all but one or two of the items in ATEMPT, the members were able to exercise a level of agreement or disagreement. Where some members felt that items may not be relevant to them, they instinctively adopted the ‘neither agree nor disagree’ response as a surrogate for ‘not applicable’. None of the members reported any frustration at the lack of a ‘not applicable’ or similar response option in the draft version of the ATEMPT. All felt that the
The five-option Likert scale proposed in the draft version was more suitable than any of the presented alternatives.

It was suggested that some respondents may feel a Likert response was superfluous for some items as only a ‘Yes’ or ‘No’ response option was necessary. As previously discussed, there is a risk that the anchors of a Likert scale can generate opinions or attitudes in the respondent that were previously absent. However, reducing the response options to a dichotomous categorical nature, where scaled attitudes or opinions may exist, reduces the amount of information that could be potentially gathered and the subsequent reliability and efficiency of the measure (Streiner 2015). Less efficient measures require more respondents to demonstrate an effect (Suissa 1991) so may be less responsive, particularly when used in trials with small sample sizes. Additionally, varying the response option for each item may increase the burden for the respondent and increase the complexity of interpreting and scoring the tool.

**General Layout Feedback**

RUG members were presented with five different options for the instruction statement that would begin the ATEMPT and direct respondents as to how they should approach answering the items within the measure. Members selected one statement that they felt provided the clearest and most suitably worded guidance for completing the ATEMPT. Members preferred the instructions that did not contain formal phrases such as, ‘when thinking about’ and ‘with respect to’. They also preferred the instructions that contained the word, ‘please’.

RUG members reported that they felt including section headings in the questionnaire was helpful and increased motivation by breaking the task up into sections. Based on this feedback, the three parts of the questionnaire: 1. About your MSK pain, 2. About your exercise, and 3. About you, were clearly compartmentalised with separate headings.
1b. Description of the cognitive interview process

Introduction

Following the feedback obtained from the PPIE focus group, various changes were made to the draft version of the ATEMPT as described above. Although the language of the items in the first draft measure of the ATEMPT was grounded in the statements originally proposed by stakeholders during the generation stage of the concept mapping, the PPIE consultation had identified elements of the questionnaire that may affect respondent’s interpretation. These issues may have arisen because the majority of the stakeholders were either researchers or physiotherapists, and therefore more likely to communicate their ideas in ways that are not always clear to lay people. Furthermore, the development of the statements into items was conducted by researchers who were also clinicians, which may have obscured potentially ambiguous terminology. Two main reasons have been identified for respondent’s interpreting questionnaire items incorrectly, 1. the items themselves, and 2. the process of answering them (Streiner et al. 2015). Feedback from the PPIE group helped identify issues with the items themselves, and cognitive interviews were used to identify why respondents may have problems answering them. The different techniques used were: rephrasing, probing, double interviewing, and thinking aloud. As the items were derived from statements created by participants in the concept mapping study rather than the author or the research team, the determination of whether the item was understood by the participant was based on the consistency of interpretation between participants, rather than by comparison to the way the interviewer perceived the item should be interpreted.

The specific aims of the cognitive interviews were as follows:

1. To identify if any items should be removed due to their inconsistent interpretation between patients.

2. To identify any modifications necessary to be make the language of the items more easily understood by patients.
3. To confirm the suitability of the response options and acceptability of the time taken to complete the questionnaire.

4. To identify any items that may demonstrate ceiling effects as they are only likely to be answered on the negative or positive end of the Likert scale.

5. To identify whether any items could be removed due to duplication.

Methods

To effectively assess the comprehensibility, relevance, and completeness of a questionnaire, the intended target population should be used (de Vet et al. 2011). This required involving patients who were currently receiving exercise recommendation for an ongoing MSK pain problem, so it was decided to invite patients who had already taken part in the concept mapping study. While these patients were not naïve to the development process of the ATEMPT and therefore may have demonstrated more understanding of the items than typical MSK pain patients, their input was considered to be valuable in the further development of the ATEMPT. Although the intended participants had already been involved in the ATEMPT project previously and had given consent to be contacted about taking part in further studies, they had not consented to take part in cognitive interviews. Consequently, a substantial amendment was required to the original ATEMPT project NHS research ethics committee (REC) and Health Research Authority (HRA) approval. Modifications were made to the ATEMPT project patient information sheet and study protocol. A participant invitation letter, cognitive interview protocol, and consent form were created and sent out with the draft ATEMPT questionnaire. Once ethical approval was provided by the HRA and REC (amendment number 1-26/07/2018, 5th December 2018), all eight patients from the generation stage focus group were invited to provide feedback during a telephone call on the new draft measure of adherence.

To reduce the burden on individual participants, the draft ATEMPT was split into three sections of 16 items each, meaning a minimum of three participants were required for the cognitive interviews. The interviews were conducted via telephone to facilitate participation. Following commencement of the telephone call, the purpose of the interview and the expectations of the participant were explained.
It was confirmed that the participant had had the opportunity to read the participant information sheet and had provided a signed consent form. The participant was then asked to read the first of 16 questions that had been assigned to them. A combined approach to interviewing was used, with one of the different methods for cognitive interviews described above selected for each item (van der Veer 2013). The style of question used was based on two factors. Firstly, items involving memory are more suited to thinking aloud or double interviewing to ensure respondents are able to recall the information, whereas establishing that the participant has understood the terminology of the item is best assessed using specific probing prompts (Streiner et al. 2015). Secondly, it was observed that some participants responded better to certain types of questioning and gave more detailed answers. In these cases, the style of questioning was influenced by the previous responses of the participant.

In addition to asking the participants to rephrase the question, think aloud while responding, and using double-interview prompts, the following probes were created a priori based on the aims of the cognitive interview pilot testing process:

- Are you unsure of the meaning of any of the words or terms in the question?
- Do you feel the question has suitable and adequate response options?
- Do you feel that patients would be able to answer the question positively or negatively if required?
- Do you feel that the question is the same as any of the other questions or that it is asking about a similar thing?
- Do you feel that the meaning of the question is clear?
- Do you feel it took too long to be able to answer the question?

Once the respondents had been questioned about the items, they were asked if they had any more comments about the questionnaire, or if they felt anything was missing. Finally, they were thanked for their time and the interview was concluded.

The interviews were not audio recorded due to the need for additional equipment and project time constraints. As such, it was not possible to create transcriptions of the interviews and independently code and extract the data obtained. This may have resulted in bias when interpreting the findings.
from the interviews, as interpretation was conducted by DB who was also the interviewer. To reduce bias, field notes were taken during the interview by DB that included verbatim quotes of the participants’ views on the items and example responses to questions. The notes were then discussed and reviewed with the research team (who had not been involved in the interviews) before any conclusions were drawn.

Results

Five of the eight patients invited took part in the cognitive telephone interviews. The first three patients who took part were each questioned on 16 different items of the 48-item draft version of ATEMPT. All items that participants raised issues with were then selected to be included in the cognitive interviews with the fourth and fifth participants. In this way, potentially problematic items identified by the first round of cognitive interviews were then double assessed using at least one further participant. While not all items were assessed using two participants, a level of sampling redundancy was achieved as no new issues were raised during the second round of interviews. The cognitive interviews each took between 50 minutes and one hour to complete.

The following is a summary of the issues identified during this phase of the pilot testing and the resulting changes made to the ATEMPT questionnaire.

Examples of items not functioning effectively

Draft items presented to the participant during the cognitive interview are shown in bold, the revised version of the item following feedback from participants is shown in italics.

I can see the benefit of doing my exercises (original)

I can see or feel the benefit of doing my exercises (revised)

Participants were unsure if ‘seeing’ the benefit was the same as ‘feeling’ the benefit. For example, one participant explained that you can see if you are able to walk better or raise your arm higher when you condition improves, but you can only feel an improvement if you have pain. Another participant felt that respondents may only be able to report they are seeing the benefit of their exercises if they
were having regular follow ups with a physiotherapist who could assess them and determine this. Consequently, the term feel was added to the item.

**I am willing to prioritise my exercises (original)**

**I am willing to prioritise my exercises over other activities/responsibilities (revised)**

This item was clarified as participants reported they were unclear whether it meant prioritising one exercise over another, completing the exercises in a particular order, or prioritising the exercises over other responsibilities.

**I do more exercise now compared to before my exercises were recommended to me (original)**

**I do more exercise now compared to before the assessment when my exercises were recommended to me (revised)**

One participant reported that they had always exercised regularly, but the onset of their pain condition had made it impossible to continue doing so. Following their assessment, they had begun the exercises as recommended by the physiotherapist. They were now exercising more than they had been since the onset of their pain condition, but not to the same extent as before their pain condition started. The participant was therefore unclear as to the frame of reference for assessing their current exercise frequency. As the measure aims to assess adherence to exercise recommended for MSK pain, the point of assessment and exercise recommendation was determined as the most appropriate cut off point to determine a change in exercise frequency.

**Examples of items functioning effectively**

The following examples demonstrate how participants made appropriate context-based judgements when answering the items and describe adherent behaviours when selecting agree or strongly agree response options. Insight is also provided on how the participants approached items that may not have been applicable to them and how they anticipated non-adherent respondents may answer.

**I know when I need to rest or not do my exercises**

One participant reported that she would choose ‘strongly agree’ for this question as the physiotherapist had explained the need to rest if she was experiencing any strain when doing the
exercises. The participant also felt that because she had always exercised regularly prior to her problem, she would instinctively know when to stop her exercises or have a day off if the symptoms were bad.

**I am sticking to a routine with my exercises**

One participant reported that they would choose ‘strongly agree’ for this item as they can’t exercise after eating, when certain things are going on in the house, or when she is performing carer duties for her husband. Consequently, she had to complete her exercises at specific times to fit with her routine.

**I feel my exercises are individualised to me, not just a ‘one size fits all’ plan**

One participant described how they would choose ‘agree’ for this item, as their physiotherapist had been very knowledgeable about their condition and the relevant anatomy involved. The physiotherapist recommended exercises based on this thorough understanding and ensured the movements required would not trigger any symptoms. The exercises were demonstrated and then the participant was able to practise them under the guidance of the physiotherapist. The participant went on to say, that had a generic set of exercise simply been printed off a computer and handed to him, he would choose ‘disagree’ or ‘strongly disagree’ for this item. No participant reported any difficulty in interpreting the phrase ‘one size fits all’.

**I am getting feedback on whether I am doing my exercises correctly**

One participant reported that they would ‘neither agree nor disagree’ as to date they had not had a follow up appointment with their physiotherapist since the exercises were recommended to them. However, as she anticipated that she would have an appointment with the physiotherapist at some point in the future where she would receive feedback, she did not select ‘disagree’ or strongly disagree’.

**I can adapt the exercise targets as my life’s demands change**

One participant reported that they would ‘strongly agree’ with this statement because their exercises didn’t take long and could be done even when they were at someone else’s house. The same participant also reported that if their exercises required any kind of equipment, or took a long time to
complete, this would not provide the flexibility necessary to complete the exercises as life’s demands changed and would therefore answer ‘disagree’ or ‘strongly disagree’.

**I attend my exercise classes as recommended**

One participant reported that they don’t go to exercise classes, despite being offered to them. They chose to undertake their exercises at home and as such the question did not apply to them. The participant would therefore choose ‘neither agree nor disagree’.

**I am completing my exercises at the appropriate tempo or speed**

Participants reported that speed had not been a specific parameter of their exercise recommendations, therefore they would choose ‘neither agree nor disagree’. They instinctively chose this option and did not feel that the response options were inadequate without a ‘not applicable’ option.

The participants unanimously reported that the instructions for the questionnaire were clear and that the response options were adequate. All participant reported that the questions could be answered in under 30 seconds each, although one participant commented that due to the length of the questionnaire, some respondents may give up before completing it.

**Summary**

In summary, items for the ATEMPT were developed based on selected concept mapping statements, these were reviewed through PPIE using the RUG and via cognitive interviewing with MSK pain patients. A three-section questionnaire resulted: section 1 included questions concerning the details of the respondent’s MSK problem, and the exercises recommended for it, section 2 included the 48-item ATEMPT, and section 3 included demographic questions. The questionnaire was therefore ready for further testing according to classical test theory, which necessitated data from a large number of respondents with MSK pain being recommended exercise for their condition.