



OPEN ACCESS

A new measure of exercise adherence: the ATEMPT (Adherence To Exercise for Musculoskeletal Pain Tool)

Daniel Leslie Bailey ¹, Annette Bishop,¹ Gareth McCray,¹ Nadine E Foster,^{1,2} Melanie A Holden¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2022-106347>).

¹Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, UK
²Surgical Treatment and Rehabilitation Service (STARS) Education and Research Alliance, The University of Queensland and Metro North Health, Herston, Queensland, Australia

Correspondence to

Dr Daniel Leslie Bailey, Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, ST5 5BG, UK; d.bailey2@keele.ac.uk

Accepted 9 October 2023

ABSTRACT

Objectives This study aimed to (1) develop a new measure of adherence to exercise for musculoskeletal (MSK) pain (Adherence To Exercise for Musculoskeletal Pain Tool: ATEMPT) based on previously conceptualised domains of exercise adherence, (2) report the content and structural validity, internal consistency, test–retest reliability, and measurement error for the ATEMPT outcome measure in patients managed with exercise for MSK pain.

Methods ATEMPT was created using statements describing adherence generated by patients, physiotherapists and researchers, with content validity established. Baseline and retest questionnaires were distributed to patients recommended exercise for MSK pain in 11 National Health Service physiotherapy clinics. Items demonstrating low response variation were removed and the following measurement properties assessed: structural validity, internal consistency, test–retest reliability and measurement error.

Results Baseline and retest data were collected from 382 and 112 patients with MSK pain, respectively. Confirmatory factor analysis established that a single factor solution was the best fit according to Bayesian Information Criterion. The 6-item version of the measure (scored 6–30) demonstrated optimal internal consistency (Cronbach's Alpha 0.86, 95% CI 0.83 to 0.88) with acceptable levels of test–retest reliability (intraclass correlation coefficient 0.84, 95% CI 0.78 to 0.88) and measurement error (smallest detectable change 3.77, 95% CI 3.27 to 4.42) (SE of measurement 2.67, 95% CI 2.31 to 3.16).

Conclusion The 6-item ATEMPT was developed from the six domains of exercise adherence. It has adequate content and structural validity, internal consistency, test–retest reliability and measurement error in patients with MSK pain, but should undergo additional testing to establish the construct validity and responsiveness.

INTRODUCTION

Musculoskeletal (MSK) pain is highly prevalent and burdensome globally.^{1,2} In a systematic review of interventions for MSK pain, exercise had the strongest evidence for reducing pain and improving function³ and is consistently recommended as a core treatment in international clinical guidelines.^{4–9} However, the average clinical effect sizes of exercise are often small compared with non-exercise controls and can diminish over time.¹⁰ This may be due to suboptimal levels of exercise adherence.¹¹ Consequently, it is important to be able to measure exercise adherence in both clinical practice and research, using a measurement tool with adequate

WHAT ARE THE FINDINGS?

⇒ Adherence to exercise for musculoskeletal (MSK) pain is an important component of clinical trials that use exercise as an intervention. Valid and reliable measurement of adherence to exercise for MSK pain is currently limited by the lack of a standardised tool.

WHAT THIS STUDY ADDS?

⇒ The Adherence to Exercise for Musculoskeletal Pain Tool (ATEMPT) is the first measure of adherence to exercise for MSK pain that was conceived by relevant stakeholders with adequate content and structural validity, internal consistency, test–retest reliability and measurement error.

HOW MIGHT IT IMPACT ON CLINICAL PRACTICE IN THE FUTURE?

⇒ The ability to measure exercise adherence for MSK pain will help clinicians measure their patient's responses to exercise-based treatment and take the appropriate actions to improve outcomes. As a research tool, ATEMPT will enable the consistent evaluation of exercise interventions for MSK pain and strategies for improving exercise adherence.

psychometric properties, including validity and reliability; that is, it should measure the construct it purports to measure, and the values obtained should have an acceptable level of measurement error.¹² Five systematic reviews have established that there is currently no valid and reliable measure of exercise adherence for MSK pain.^{13–17} They highlighted current measurement inconsistency, with 234 methods, including 49 separate questionnaires identified in the literature.¹³ This may be due to the absence of an agreed definition of exercise adherence on which to develop a measure.¹⁸

Since the reviews, a self-report measure of exercise adherence has been developed, the Exercise Adherence Rating Scale (EARS), a 6-item self-report measure of adherence to exercise for any health condition, not specifically MSK pain.¹⁹ The EARS development study did not conceptualise or define exercise adherence at the outset but used focus groups with people with chronic low back pain and physiotherapists to develop the items. Eleven of the 15 items generated were found to be unsuitable for exploratory factor analysis (EFA) due to their high variance, suggesting the underlying construct of the measure was not well defined. When testing the



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Bailey DL, Bishop A, McCray G, et al. *Br J Sports Med* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjsports-2022-106347

EARS, questionnaires were issued to patients up to 6 weeks after treatment completion, increasing the likelihood of recall bias, and only 30 participants completed the retest questionnaire, fewer than recommended for assessing test–retest reliability.²⁰ The measurement error and smallest detectable change (SDC) were also not calculated. As such, there remains no robust tool to assess adherence to exercise for MSK pain.

As a first step to robustly measure exercise adherence, we recently undertook a concept mapping study (involving the synthesis of qualitative and quantitative data from patients, physiotherapists and researchers) to conceptualise exercise adherence for MSK pain. Adherence was found to consist of six domains: communication with expert; targets; how exercise is prescribed; patient knowledge and understanding; motivation and support; and psychological approach and attitude.²¹

Building on these findings, we aimed to: (1) develop a new measure of adherence to exercise for MSK pain (ATEMPT) based on previously conceptualised domains of exercise adherence, (2) report the content and structural validity, internal consistency, test–retest reliability, and measurement error for the ATEMPT outcome measure in patients managed with exercise for MSK pain.

METHODS

Equity, diversity and inclusion statement

The authors include three women, two men and consists of junior, mid-career and senior researchers from different disciplines from two countries. Our study population included different ages and genders. In discussing the generalisability of our results, we acknowledge that further research in different healthcare settings, including in other countries, and with participants exhibiting more varied ethnicity is warranted.

Development of the draft version of ATEMPT

During our previous study,^{21 22} 56 items across 6 domains of exercise adherence for MSK pain were identified as important for inclusion in a new measure. These 56 items were discussed for readability and comprehensibility within a patient and public involvement and engagement workshop, including 6 patients with MSK pain. Response options were also discussed and agreed. Following the workshop, 8 items were removed due to issues with comprehensibility and the remaining 48 items were coupled with a 5-option Likert response (1=strongly disagree to 5=strongly agree).

The 48 items were further tested with 5 patients with MSK pain during pilot interviews to establish content validity (online supplemental file 1). This led to changes to three items (online supplemental file 2) resulting in the draft ATEMPT (online supplemental file 3).

Testing the draft version of ATEMPT

The draft ATEMPT, questions on personal characteristics, MSK pain, current exercise recommendation and a consent form for further contact, were included in a cross-sectional questionnaire survey of adults (aged 18 years and over) with MSK pain who had been recommended exercise at 1 of 11 physiotherapy clinics across 4 National Health Services (NHS) trusts in the Midlands region of England. This encompassed urban, suburban and rural areas. Questionnaires were distributed to current patients who met the inclusion criteria by clinic staff, as well as being advertised and made available in waiting areas between October 2019 and March 2020. Questionnaires were returned via post directly to the study team. A retest questionnaire (draft ATEMPT only)

was posted to consenting participants immediately on receipt of the baseline questionnaire.

Sample size

Based on the intended analysis method of confirmatory factor analysis (CFA), a minimum of 200 responses were required,²³ with a further 100 responses to assess test–retest reliability via the retest questionnaire.²³

Participants

Any adult aged 18 years or over that had been recommended exercise for any existing MSK pain as part of their physiotherapy treatment. Participants were required to be able to read and write in English.

Data entry and analysis

Data from returned questionnaires was manually entered into a customised Excel spreadsheet. Random samples of 1-in-10 participants' data were checked by a second coder for accuracy. Due to the low number of missing values, an imputation method was used by calculating the mean item scores for the participant.^{24 25} Frequencies, means and standard deviation (SD) were calculated for participant demographics, exercise and MSK pain data, and for each item response. Ceiling and floor effects were considered present if more than 15% of participants achieved the lowest or highest score possible (48 or 240).²⁶ The statistical analysis and presentation are consistent with the Checklist for statistical Assessment of Medical Papers.²⁷

Item reduction

Items with fewer than six responses were removed as they demonstrated low levels of variation and would be less useful to differentiate between responders.²⁶ This decision was made pragmatically based on the results as it was not possible to model sparse data.²⁸ In determining this cut-off point, consideration was given to the impact on the total number of items remaining, the number of items per domain, the SD of the individual items removed, and the conceptual content of the item removed.

Measurement property analysis

The following measurement properties identified by the CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN)¹² were assessed: structural validity—the degree to which the scores of a measure are an adequate reflection of the dimensionality of the construct of interest; internal consistency—how well the different questions measure the same construct; test–retest reliability—the reliability of the measure repeated at different times; and measurement error—the difference between the measured value and the true value.

Structural validity

CFA was conducted on the items using an iterative process to establish the model with the best fit rather than confirming a single model's fit. Models with one to eight factors were analysed as these had previously been identified via hierarchical clustering in the concept mapping study.²¹ The factor analysis used Multidimensional Item Response Theory²⁹ via the Quasi-Monte Carlo Estimation Method³⁰ using the graded response model³¹ as the item link and was undertaken in RStudio.³² The fit of the competing models was compared using the Bayesian Information Criterion (BIC) along with interpretation of the loading values for each item. BIC is an estimate of a function of the probability

of a model being the best fit, assuming such a model is among the available candidates.³³ The lower the value of BIC, the better the data fit to the model. EFA was not used, as the factor structure was predetermined by the findings from the previous concept mapping study.²¹

Model fit statistics were calculated as recommended by COSMIN.¹² These included root mean square error of approximation (RMSEA), standardised root mean square residual (SRMSR), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI). These values indicate the goodness of fit of the model,

with acceptable values being: RMSEA<0.06, SRMSR<0.08, CFI>0.95 and TLI>0.95.³⁴

Internal consistency

Cronbach's alpha coefficient was calculated using the baseline data to determine internal consistency, with a minimum acceptable value being 0.7 and the maximum 0.9, as above this value it is perceived there is redundancy or duplication in the items.²⁶

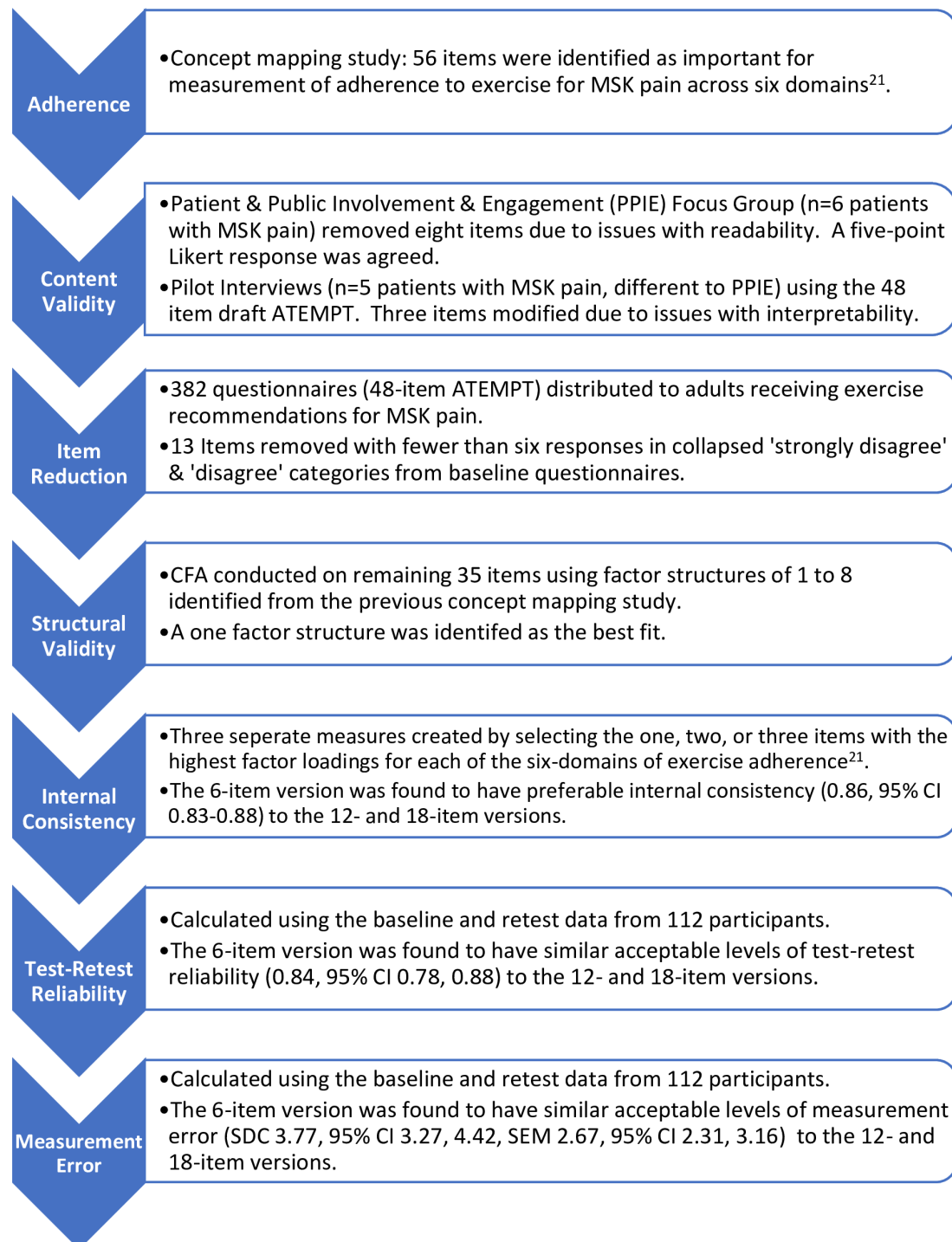


Figure 1 Flow chart of the measurement property being assessed, participants involved and the number of items in Adherence to Exercise for Musculoskeletal Pain Tool (ATEMPT).

Table 1 Participant characteristics

Variable		n=382
Sex	Female	253 (66%)
	Male	128 (34%)
Age group	18–25 years old	11 (2.9%)
	26–35 years old	14 (3.7%)
	36–45 years old	35 (9.2%)
	46–55 years old	75 (19.8%)
	56–65 years old	89 (23.5%)
	66–75 years old	120 (31.7%)
	76–85 years old	31 (8.2%)
Ethnicity	86 years or over	4 (1%)
	White	345 (90.5%)
	Asian/Asian British	19 (5%)
	Other	11 (2.9%)
	Black/African/Caribbean	3 (0.8%)
Employment	Mixed ethnicity	3 (0.8%)
	Retired	189 (50.1%)
	Employed usual job	113 (30%)
	Employed sick leave	33 (8.8%)
	Employed light duties	23 (6.1%)
Symptom duration	Unemployed	19 (5%)
	0–6 weeks	23 (6.1%)
	7–12 weeks	34 (9%)
	3–6 months	66 (17.3%)
	7–12 months	74 (19.5%)
	13 months–2 years	52 (13.7%)
MSK symptom location*	Over 2 years	131 (34.4%)
	Knee	135 (35.3%)
	Shoulder	113 (29.6%)
	Back	111 (29.1%)
	Hip	82 (21.5%)
	Neck	56 (14.7%)
	Other	45 (11.8%)
	Foot or ankle	44 (11.5%)
	Wrist or hand	37 (9.7%)
	Elbow	15 (3.9%)
Pain intensity	Specific condition†	11 (2.9%)
	Mean (SD)	5.25 (2.48)
Exercise type*	Joint specific	347 (90.8%)
	Postural	86 (22.5%)
	Exercises in water	70 (18.3%)
	General physical activity	65 (17.0%)
	Other	48 (12.6%)
	Pilates	26 (6.8%)
	Yoga	22 (5.8%)
	Time since exercises recommended	
	Less than a week	21 (5.5%)
	2–6 weeks	177 (46.6%)
	7–12 weeks	85 (22.4%)
	3–6 months	54 (14.2%)
	7 months–1 year	23 (6.1%)
	More than 1 year	20 (5.2%)

Totals may not equal 382 due to missing data.

*Participants were allowed to select multiple responses; hence, totals do not sum to 100%.

†For example, fibromyalgia.

MSK, musculoskeletal.

Test–retest reliability

The test–retest reliability of the measure was assessed using intra-class correlation coefficient (ICC) on data from participants who completed baseline and retest questionnaires. ICC was calculated by mean squares obtained through analysis of variance. A

two-way mixed effect, single measurement, absolute agreement ICC model was used. The ICC values were interpreted according to published guidance for reliability at a group level rather than an individual level as individual patients are not normally compared with each other, and the SE of measurement (SEM) is a more useful value for clinicians.³⁵ Values less than 0.5 indicate poor reliability, 0.5 to 0.75 moderate reliability, 0.75 to 0.9 good reliability and greater than 0.9 excellent reliability.³⁶ The time between completion was calculated from consent form dates.

Measurement error

Measurement error was assessed by calculating the SEM and SDC at an individual level. SEM was analysed using the following formula: $SEM = SD \times \sqrt{(1 - ICC)}$, in which SD. The SDC was analysed as follows: $SDC = 1.96 \times \sqrt{2} \times SEM$.²⁰

RESULTS

A flow chart summarising each study phase is presented in figure 1.

Survey response

Two thousand questionnaires were distributed, and 382 completed baseline questionnaires returned. A response rate could not be calculated as undistributed questionnaire packs could not be collected due to clinic closures during the COVID-19 pandemic. Retest questionnaire packs were posted to all 234 consenting participants, with 112 returned (48% response rate). The mean time between completion of the baseline and retest questionnaires was 17 days (SD=10, range 5–63).

Data integrity

The data entry error rate was $\leq 1\%$. There were 78 missing responses to items in the baseline questionnaire (0.4%) and 27 missing responses to items in the retest questionnaire (0.5%) (online supplemental file 4). Nine of the 382 respondents (2.4%) achieved the highest score possible (240), no respondents received the lowest possible score (48); consequently, no floor or ceiling effects were observed.

Participant demographics

Most participants were female (66.4%) with the most common age range 66–75 years (31.7%). The majority were white (90.6%) and half were retired (50%). The average pain intensity score was 5.25 out of 10 (SD 2.48), the knee was the most common site of pain (35.3%) and 34.3% of participants had experienced their symptoms for over 2 years. Joint-specific exercises, such as strengthening and stretching, were the most common kind of exercise recommended (90.8%) and 2–6 weeks the most common duration since the exercise recommendation (46.3%) (table 1).

Item reduction

The mean baseline score for all participants for 48 items was 197 (SD 23, range 123–240). Thirteen items with fewer than six responses in the response categories ‘strongly disagree’ or ‘disagree’ were removed. These items were also the only ones to have a SD of less than 0.7 (online supplemental file 4).

Structural validity

CFA was conducted on the remaining 35 items using factor structures of 1–8 previously identified in the concept mapping study²¹ (online supplemental file 5). The BIC value was smallest for the one factor solution (24 258.93) (online supplemental file 6).

Table 2 Cronbach's alpha values and model fit statistics for the 6-item, 12-item and 18-item versions of Adherence to Exercise for Musculoskeletal Pain Tool

	Cronbach's alpha values				Model fit statistics (acceptable values ³²)			
	Raw alpha	Standardised alpha	Lower 95% CI	Upper 95% CI	RMSEA (<0.06)	SRMR (<0.08)	TLI (>0.95)	CFI (>0.95)
6 items	0.85	0.86	0.83	0.88	0.03	0.04	1.0	1.0
12 items	0.91	0.91	0.9	0.92	0.12	0.08	0.93	0.95
18 items	0.93	0.93	0.92	0.94	0.13	0.09	0.92	0.93

CFI, Comparative Fit Index; RMSEA, root mean square error of approximation; SRMR, standardised root mean squared residual; TLI, Tucker-Lewis Index.

Internal consistency

Although a one factor structure was identified as the best fit, it was felt that to maintain content validity, the final measure should contain items representative of the six domains of exercise adherence previously identified by stakeholders.²¹ To achieve this while also assessing reliability, 6-item, 12-item and 18-item measures were created by selecting the items with the highest factor loadings (on the unidimensional model) from each of the 6 domains (online supplemental file 7). Consequently, the measures contained the 6, 12 or 18 items that best correlated (as determined by CFA) with each of the six domains of exercise adherence. This allowed for the shortest measure to be identified exhibiting satisfactory levels of reliability while retaining content from all six domains. Cronbach's alpha values and model fit statistics for the three measures are presented in table 2. Figure 2 shows a patch diagram of the 6-item tool with loadings and variance explained for each item.

Test-retest reliability

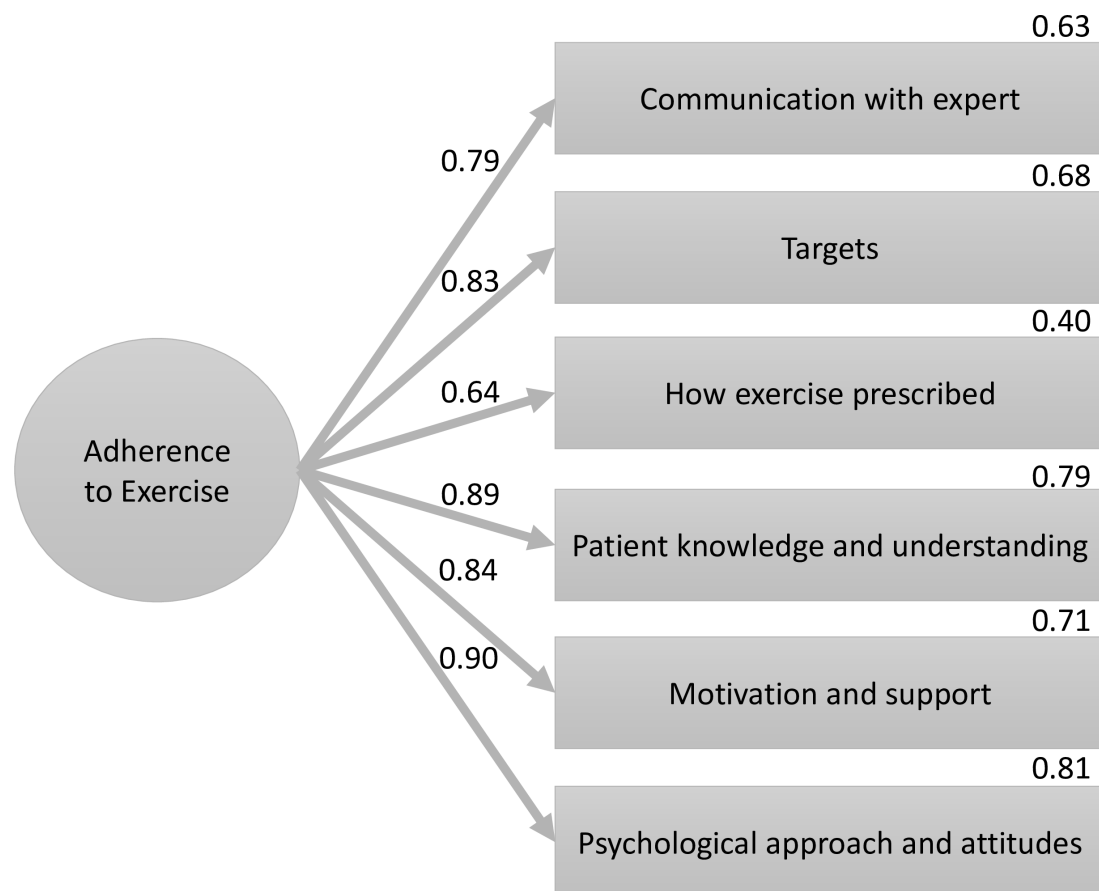
ICC values for the three measures are presented in table 3.

Measurement error

Measurement error for the three measures is presented in table 4. For the 12-item and 18-item versions, the point estimate and CIs were scaled down (divided by 2 and 3, respectively), to account for the increased maximum possible scores.

DISCUSSION

The 6-item version of the measure shown in figure 3, and online supplemental file 8, was shown to have acceptable model fit statistics and preferable internal consistency, while demonstrating similar acceptable levels of test-retest reliability and measurement error as the 12-item and 18-item versions when examined in patients recommended exercise for MSK pain. As



The numbers next to the arrows are the factor loadings and the numbers above the boxes are the variance explained, of the observed variable, by the construct.

Figure 2 Patch diagram of the 6-item Adherence to Exercise for Musculoskeletal Pain Tool (ATEMPT).

Table 3 Intraclass correlation coefficient values and means scores for the 6-item, 12-item and 18-item measures

	Point estimate	Lower 95% CI	Upper 95% CI	Mean baseline score (SD)	Mean follow-up score (SD)	Mean change score (SD)
6 items	0.84	0.78	0.88	25.41 (3.40)	25.27 (3.39)	0.14 (4.63)
12 items	0.84	0.79	0.88	49.63 (6.72)	49.08 (6.96)	0.53 (9.40)
18 items	0.84	0.79	0.88	74.04 (9.67)	73.26 (10.00)	0.77 (13.53)

a shorter measure is less burdensome for patients, the 6-item version is the preferable version of ATEMPT. ATEMPT is scored by summing all responses as follows: strongly disagree=1, disagree=2, neither agree nor disagree=3, agree=4, strongly agree=5. The total score is therefore between 6 and 30 with a change in score of 4 or more required to indicate a change beyond measurement error. ATEMPT is suitable for further psychometric testing, to include responsiveness and construct validity via hypothesis testing.

The structural validity, internal consistency and measurement error of ATEMPT were tested using data from 382 patients, 112 of these were also used to assess the test-retest reliability. The patients were all recommended exercise for MSK pain at 11 NHS physiotherapy clinics. CFA established that ATEMPT functions as a unidimensional measure of exercise adherence. However, the items retain the original six-domain conceptualisation of exercise adherence to ensure content validity.²¹ The mean time between completion of the baseline and retest questionnaires (17 days) was considered short enough to expect the participants' exercise adherence to be similar at the two time-points,³⁷ with time being considered the most appropriate indicator of adherence stability based on existing research.¹⁰

Clinical and research implications

The structural validity assessment of ATEMPT indicates that its score is an adequate reflection of the dimensionality of the construct of interest¹² (exercise adherence for MSK pain as described in our development study²¹). Furthermore, the good internal consistency (0.85) suggests that all items are measuring this construct. The SDC of 3.77 indicates that a change in ATEMPT score of 4 or more out of a possible maximum of 30, represents a change greater than measurement error. The test-retest reliability of ATEMPT is good, meaning it should measure adherence consistently. ATEMPT would therefore be suitable for monitoring a patient's adherence to their exercise recommendations during a course of treatment, whether that be in a clinical setting or as part of a research study. A reduction in ATEMPT scores may reflect reduced exercise adherence, which could indicate a patient is not adhering to their exercise programme as well as they were. This might help to identify patients who may benefit from adherence enhancing strategies, such as goal setting³⁸ or to consider alternate treatment plans. In a research setting, changes in ATEMPT scores when interpreted with additional patient reported outcome measures (PROMs) for pain and functional outcomes, may assist researchers in interpreting whether differences between patient groups are due to adherence

levels or other variables, such as treatment efficacy. This would reduce the potential for incorrectly concluding lack of effectiveness of exercise interventions that is actually due to declining adherence levels. ATEMPT would also facilitate the comparison of interventions designed to improve exercise adherence due to the current lack of consistency in adherence measurement.

Strengths and limitations

This study had a number of strengths; the study builds on the first conceptualisation of adherence specific to exercise for MSK pain.²¹ The study used data obtained from multiple physiotherapy clinics in the UK, incorporating diverse areas from urban to rural. The population included patients with a diverse range of ages, MSK pain location, duration and intensity. Participants had been recommended a variety of exercises over a range of time periods. Furthermore, combining the results of the previous concept mapping study²¹ with CFA, meant potential models were derived directly from the stakeholders' conceptualisation of adherence, thereby, arguably enhancing the validity of the models and subsequent tool. Finally, in addition to its favourable measurement properties, ATEMPT demonstrates good utility; with only six items, it can be completed quickly and easily.

However, a number of limitations need to be taken into consideration. It is possible that patients who were willing to respond to the survey were also more likely to have completed their exercises. This may have been a source of sampling bias, as patients who were less likely to have adhered to their exercises may have responded differently, meaning that some or all the 13 items removed due to low response rates would have been retained for CFA.

No additional PROMs measuring pain or disability were used in conjunction with ATEMPT. This was because the variety of MSK pain presentations included in the study would make selecting appropriate PROMs difficult and additional questionnaires may have led to reduced participation due to participant burden. Consequently, future studies should assess the responsiveness (including the minimally important difference) and construct validity of ATEMPT via hypothesis testing using expected correlates with exercise adherence in an independent sample using longitudinal research designs.

As there is no available method for ensuring adherence stability, it was not possible to guarantee the participants' stability during the period between the test and retest questionnaires; however, time has been shown to be the most appropriate indicator of adherence stability based on existing research.¹⁰ Future studies should attempt to control for clinical stability in participants

Table 4 SE of measurement (SEM) and smallest detectable change (SDC) at an individual level for the 6-item, 12-item and 18-item versions of the measure (min-max scores for each measure)

	Point estimate (scaled)		Lower 95% CI (scaled)		Upper 95% CI (scaled)	
	SDC	SEM	SDC	SEM	SDC	SEM
6 items (6–30)	3.77	2.67	3.27	2.31	4.42	3.16
12 items (12–60)	7.45 (3.73)	5.27 (2.64)	6.45 (3.23)	4.57 (2.29)	8.53 (4.27)	6.04 (3.02)
18 items (18–90)	10.72 (3.57)	7.59 (2.53)	9.29 (3.10)	6.57 (2.19)	12.28 (4.10)	8.68 (2.89)

ATEMPT

Adherence To Exercise for Musculoskeletal Pain Tool

The following statements describe aspects of the exercise programme that has been recommended to you. Please indicate the extent to which you agree with each of the statements by **crossing one of the boxes**. For example, if you agreed strongly with a statement, you would cross the box as shown below:

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

1. I feel my individual needs were understood when the exercise therapist recommended my exercises

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. I am doing enough exercise to produce a positive change in my musculoskeletal pain

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. I am doing my exercises as instructed

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. I understand how my exercises will help with my musculoskeletal pain

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. I understand the consequences of not doing my exercises

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6. I believe in the exercises that have been recommended to me

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Score each question as follows and sum out of 30: Strongly disagree = 1, Disagree = 2, Neither agree nor disagree = 3, Agree = 4, Strongly Agree = 5. A change in score of 4 or more indicates a change in exercise adherence.

Copyright © 2022 Keele University

Figure 3 The final 6-item Adherence To Exercise for Musculoskeletal Pain Tool (ATEMPT).

when assessing test–retest reliability. Future comparison of the EARS and ATEMPT measures may also be useful to explore overlap and progress testing in this area.

CONCLUSION

ATEMPT is a unidimensional measure that contains six items from the six domains of exercise adherence. ATEMPT demonstrates

adequate content and structural validity, acceptable internal consistency, good test–retest reliability, satisfactory measurement error and is the only measure of exercise adherence to be based on a conceptualisation of the construct from the perspective of relevant stakeholders. It is therefore suitable to measure patient's adherence to exercise for MSK pain but should undergo additional testing to further establish the construct validity and responsiveness.

Acknowledgements Dr Kirstie Haywood, Warwick Research in Nursing, Warwick Medical School, Warwick University, for her contribution to the methodological design of the study.

Contributors All authors contributed to the design of the study. DLB and AB conducted the focus groups. DLB analysed the data and developed the first draft of this manuscript. All authors interpreted the data, contributed to the critical revision of the manuscript, and approved the final version. DLB is acting as guarantor.

Funding DLB was supported for this research through a Keele University, Research Institute for Primary Care and Health Sciences, ACORN PhD Studentship. NEF and AB were supported through a National Institute for Health Research (NIHR) Research Professorship awarded to NEF (NIHR-RP-011-015). Professor NEF is a NIHR Senior Investigator. The study was funded through a NIHR Clinical Research Network (CRN) 'Fast Track' grant.

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by London, Surrey Research ethics committee, the Health Research Authority (HRA) and Health and Care Research in Wales (HCRW) approved the study (REC Reference: 19/LO/0903; IRAS Reference: 257591), July 2019. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Keele University is a member of the UK Reproducibility Network and committed to the principles of the UK Concordat on Open Research Data. The School of Medicine has a longstanding commitment to sharing data from our studies to improve research reproducibility and to maximise benefits for patients, the wider public, and the health and care system. Deidentified individual participant data (IPD) that underlie the results from this study are securely stored on servers approved by a government-backed cyber security scheme and made available to bona-fide researchers upon reasonable request via our controlled access procedures. Unless there are exceptional circumstances, data will be available upon publication of main study findings or within 18 months of study completion (whichever is earlier) and with no end date. Data requests and enquiries should be directed to medicine.dashsharing@keele.ac.uk. We encourage collaboration with those who collected the data, to recognise and credit their contributions. The data generated from this study will remain the responsibility of the Sponsor. Release of data will be subject to a data use agreement between the sponsor and the third party requesting the data. Deidentified IPD will be encrypted on transfer.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Daniel Leslie Bailey <http://orcid.org/0000-0001-9323-1884>

REFERENCES

- GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.
- Institute for Health Metrics and Evaluation (IHME). GBD arrow diagram. Seattle, WA IHME, University of Washington; 2020. Available: <http://vizhub.healthdata.org/irank/arrow.php> [Accessed Oct 2020].
- Babatunde OO, Jordan JL, Van der Windt DA, et al. Effective treatment options for musculoskeletal pain in primary care: a systematic overview of current evidence. *PLoS One* 2017;12:e0178621.
- National Institute for Health and Care Excellence (NICE). Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. NICE guideline [NG193]. 2021. Available: <https://www.nice.org.uk/guidance/ng193> [Accessed Oct 2021].
- National Institute for Health and Care Excellence (NICE). Rheumatoid arthritis in adults: management NICE guideline [NG100]. 2018. Available: <https://www.nice.org.uk/guidance/ng100> [Accessed Oct 2021].
- National Institute for Health and Care Excellence (NICE). Low back pain and sciatica in over 16s: assessment and management NICE guideline [NG59]. 2016. Available: <https://www.nice.org.uk/guidance/ng59> [Accessed Oct 2021].
- National Institute for Health and Care Excellence (NICE). Osteoarthritis: care and management [CG177]. 2014. Available: <https://www.nice.org.uk/guidance/cg177> [Accessed Oct 2021].
- British Pain Society (BPS). Guidelines for pain management programmes for adults. 2013. Available: https://www.britishtainsociety.org/static/uploads/resources/files/pmp2013_main_FINAL_v6.pdf [Accessed 05 Jul 2017].
- Vliet Vlieland TP, Fernandes L, Bijlsma JW, et al. Response to Dr Forestier's eLetter on EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Ann Rheum Dis* 2014;73:e7.
- Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American college of rheumatology/ arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. *Arthritis Rheumatol* 2020;72:220–33.
- McLean SM, Klaber Moffett JA, Sharp DM, et al. A randomised controlled trial comparing graded exercise treatment and usual physiotherapy for patients with non-specific neck pain (the GET UP neck pain trial). *Man Ther* 2013;18:199–205.
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- McLean S, Holden MA, Potia T, et al. Quality and acceptability of measures of exercise adherence in musculoskeletal settings: a systematic review. *Rheumatology (Oxford)* 2017;56:426–38.
- Uzawa H, Davis S. Outcome measures for adherence to home exercises among patients with chronic low back pain: a systematic review. *J Phys Ther Sci* 2018;30:649–53.
- Frost R, Levati S, McClurg D, et al. What adherence measures should be used in trials of home-based rehabilitation interventions? A systematic review of the validity, reliability, and acceptability of measures. *Arch Phys Med Rehabil* 2017;98:1241–56.
- Bollen JC, Dean SG, Siegert RJ, et al. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercises programmes, and their psychometric properties. *BMJ Open* 2014;4:e005044.
- Hall AM, Kamper SJ, Hernon M, et al. Measurement tools for adherence to non-pharmacologic self-management treatment for chronic musculoskeletal conditions: a systematic review. *Arch Phys Med Rehabil* 2015;96:552–62.
- Bailey DL, Holden MA, Foster NE, et al. Defining adherence to therapeutic exercise for musculoskeletal pain: a systematic review. *Br J Sports Med* 2020;54:326–31.
- Newman-Beinart NA, Norton S, Dowling D, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the exercise adherence rating scale (EARS). *Physiotherapy* 2017;103:180–5.
- Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- Bailey D, Bishop A, Foster NE, et al. Conceptualising adherence to exercise for musculoskeletal pain: a concept mapping study. *Physiother Res Int* 2023;28:e1989.
- Kane M, Trochim WMK. *Concept Mapping for Planning and Evaluation*. Applied Social Research Methods Series. Sage Publications London, 2007.
- Myers ND, Ahn S, Jin Y. Sample size and power estimates for a confirmatory factor analytic model in exercise and sport. *Res Q Exerc Sport* 2011;82:412–23.
- Christiansen DH, McCray G, Winding TN, et al. Measurement properties of the musculoskeletal health questionnaire (MSK-HQ): a between country comparison. *Health Qual Life Outcomes* 2020;18:200.
- Kroenke K, Spitzer RL, Williams JBW, et al. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry* 2010;32:345–59.
- de Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in Medicine*. NY: Cambridge University Press, 2011.
- Mansournia MA, Collins GS, Nielsen RO, et al. A checklist for statistical assessment of medical papers (the CHAMP statement): explanation and elaboration. *Br J Sports Med* 2021;55:1009–17.
- DiStefano C, Shi D, Morgan GB. Collapsing categories is often more advantageous than modelling sparse data: investigations in the CFA framework. *Structural Equation Modeling: A Multidisciplinary Journal* 2021;28:237–49.
- Chalmers RP. MIRT: A multidimensional item response theory package for the R environment. *J Stat Softw* 2012;48:1–29.
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 1969;34:1–97.

- 31 Cai L. High-dimensional exploratory item factor analysis by a metropolis-hastings Robbins-Monro algorithm. *Psychometrika* 2010;75:33–57.
- 32 RStudio Team. Rstudio: integrated development environment for R. Boston, MA, Available: <http://www.rstudio.com>
- 33 Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012;17:228–43.
- 34 Prinsen CAC, Vohra S, Rose MR, *et al.* "How to select measurement instruments for outcomes included in a "core outcome set" – a practical guideline. *Trials* 2016;17:449.
- 35 Geerinck A, Alekna V, Beaudart C, *et al.* Standard error of measurement and smallest detectable change of the Sarcopenia quality of life (Sarqol) questionnaire: an analysis of subjects from 9 validation studies. *PLoS ONE* 2019;14:e0216065.
- 36 Koo TK, Li MY. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
- 37 Allen MJ, Yen WM. *Introduction to measurement theory*. Long Grove, IL: Waveland Press, 2002.
- 38 Jordan JL, Holden MA, Mason EE, *et al.* Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. *Cochrane Database Syst Rev* 2010;2010:CD005956.