


RESEARCH ARTICLE

Examining the factor structure, reliability, and validity of the main outcome measure used in mentalization-based therapy skills training

Adam Polnay^{1,2,3}  | Lindsey G. McIntosh^{3,4} | Aileen Burnett⁴ |
Andrea Williams⁵ | Catherine Cahill⁶ | Peter Wilkinson⁷ | Fiona Mohammad⁸ |
Jon Patrick¹

¹Medical Division, The State Hospital, Carstairs, UK

²Psychotherapy Department, Royal Edinburgh Hospital, Edinburgh, UK

³Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK

⁴Psychological Therapy Service, The State Hospital, Carstairs, UK

⁵Personality Disorder and Homelessness Team, Glasgow City Health and Social Care Partnership, Glasgow, UK

⁶Lansdowne Psychotherapy Service, Glasgow, UK

⁷South Lanarkshire Psychotherapy Service, Carluke, UK

⁸Rowanbank Clinic, Glasgow, UK

Correspondence

Adam Polnay, Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK.

Email: adam.polnay2@nhs.scot

Abstract

Background and Aims: Multiple national guidelines stress the importance for clinicians to possess good therapeutic skills for working with patients with significant relational difficulties (who may receive a diagnosis of personality disorder). Training clinicians in mentalization-based treatment skills (MBT-S) is one approach to address this. The main outcome measure used in MBT-S studies is the *Knowledge and Application of MBT Questionnaire* (KAMQ). However, an absence of research into the properties and validity of the KAMQ has limited the methodological quality of MBT-S evaluations so far. The aim of this study was therefore to investigate the factor structure, internal consistency, reliability, and validity of the KAMQ.

Methods: Using an existing multiprofessional sample of 217 clinicians from 2014 to 2016, we undertook exploratory factor analysis to determine the factor structure and internal consistency of the KAMQ. Convergent validity of the measure with the Attitudes to Personality Disorder Questionnaire (APDQ) was assessed in a subset of this dataset where both questionnaires had been administered ($n = 92$). Additionally, by recruiting a new sample of 70 clinicians, we assessed the measure's test-retest reliability.

Results: Factor analysis found three factors underlying 17 of the 20 KAMQ items, relating to therapeutic skills in mentalizing, beliefs about applying MBT in practice, and specific MBT knowledge. The KAMQ was revised following the factor analysis to form the KAMQ-2 with 17 items. Internal consistency ($\alpha = .85$, 95% confidence interval [CI] = 0.80-0.89) and test-retest reliability (ICC = 0.84, 95% CI = 0.73-0.91) were good. In correlation analyses, the KAMQ-2 showed convergent validity with the main factor from the APDQ ($n = 48$; $r_s = 0.39$, $P < .01$).

Conclusion: The KAMQ-2 provides a short, reliable self-report instrument which probes clinicians' knowledge about mentalizing skills, and beliefs about using these.

Adam Polnay and Lindsey G. McIntosh are joint lead authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Health Science Reports* published by Wiley Periodicals LLC.

There was preliminary evidence for validity. The properties of the KAMQ-2 mean that more robust evaluation and development of MBT-S is now possible.

KEYWORDS

factor analysis, MBT, mentalizing skills, outcome measure, reliability

1 | INTRODUCTION

It is well recognized that working with patients with significant relational difficulties can feel troubling and stressful for their treating clinicians, with clinicians potentially acting in unhelpful ways toward patients.¹⁻³ In complicated and emotionally demanding clinical encounters, such as working with a patient who is expressing suicidal thoughts yet refusing care, a clinician's own ability to remain reflective is challenged. For example, feelings of confusion or frustration experienced by a clinician in such a situation may lead to the clinician assuming a nonreflective stance such as a judgmental attitude toward the patient that brings about an escalation of the situation.^{4,5} A cycle of rejection can emerge between patients and staff⁶ that inadvertently repeats patients' expectations and experiences in relationships.⁷ This process can be fundamental to the maintenance of patients' difficulties. Accordingly, multiple national guidelines have stressed the importance for multidisciplinary clinicians to have appropriate training to ensure they possess good therapeutic skills when working with patients with significant relational difficulties (who may receive a diagnosis of personality disorder).^{2,8-10}

A widely used approach to such training is to take the core principles of established models of therapy and apply and adapt these so as to be useful for general clinicians in their everyday interactions with patients with relational difficulties, that is, to train general clinicians in applied therapeutic skills as opposed to "pure" therapy. While there is clearly a role for "formal" therapy delivered by specialists for patients, many if not most patient encounters and interactions occur not with a specialist psychological therapist but with a range of clinical staff in the multidisciplinary team.¹¹ Interventions under investigation include cognitive behavioral therapy skills training,¹² dialectical behavioral Therapy skills training,¹³ using cognitive analytic therapy principles to support team working,¹⁴ and mentalization-based treatment skills training (MBT-S).^{11,15}

Alongside MBT constituting a specific therapy delivered by specialist therapists for people with significant interpersonal difficulties,¹⁶⁻¹⁸ there is growing interest in teaching the core skills of MBT to general clinicians from the multidisciplinary team. A 2-day skills course, MBT-S, has been developed for this purpose.¹⁹ Investigating a measurement tool to evaluate MBT-S is the focus of the current article.

A consistent finding in reviews of therapy skills training studies is that short courses or workshops such as MBT-S are likely to bring about significant improvements in participants' knowledge about the new mode of working, with possible short-term benefits in observable therapeutic skills. However, short courses alone are very unlikely to

bring about observable improvements in clinicians' therapeutic skills in the long term—this latter aim requires ongoing, regular, supervision, or consultation in order to build on knowledge and skills taught in an initial short course.²⁰⁻²² Hence, the 2-day MBT-S course is best understood as the first part of a longer training program. In our everyday practice, we stress the importance for services of ongoing supervision for staff as a follow-up to the MBT-S course, as opposed to engaging in a "drive-by" training where effects are likely to be short-lived.

The capacity to "mentalize" is defined as the ability to understand mental states in ourselves and others and to consider how actions relate to these mental states.^{17,18} According to MBT theory, the ability to mentalize develops in the context of secure-enough relationships to our attachment figures (ie, parents or carers). In people with significant relational difficulties, the development of mentalizing is disrupted, related to a number of possible factors: insecurity in early attachment relationships; constitutional factors; and childhood psychological trauma.²³

MBT-S provides a framework to help clinicians understand their attitudes, feelings, and intentions that emerge in clinical interactions. Using a mixture of didactic teaching, video clips, and role play, the MBT-S approach trains clinicians in a helpful stance to take in clinical encounters: namely an interested, nonjudgmental, open-minded, and transparent approach.¹⁵ Furthermore, MBT-S teaches clinicians ways to restore mentalizing in both the clinician and the patient during tricky interactions when intense feelings can make it hard to reflect. These techniques include paying attention to the clinicians' own facial expression and the need for facial communications to be active and attuned the patient's mental state (but not overwhelmed by it) and not "blank." MBT-S also teaches clinicians to be aware of when the clinician themselves is struggling to reflect and "mentalize," and be prepared to slow down ("stop and stand") the interaction to prevent escalating misunderstandings.^{19,23,24} MBT-S is considered highly acceptable to the professionals undertaking the training.²⁵

The first quantitative study to evaluate the effect of a mentalizing skills intervention was an evaluation of a brief teaching program in mentalizing skills for psychiatry trainees.²⁴ Trainees' knowledge of MBT theory and practice was assessed using the *Knowledge and Application of MBT Questionnaire* (KAMQ).²⁶ The study found clinically relevant improvements in participants' KAMQ scores as well as improvements in attitudes. Subsequent research in a multisite study of MBT-S for mental health professionals found a large increase in clinicians' KAMQ scores from baseline to end-of-program as measured by KAMQ scores and a small improvement in participants' attitudes toward people with relational difficulties.¹⁵ A third study evaluating MBT-S with subsequent MBT supervision

for staff and longer term follow-up is in progress.²⁷ A qualitative evaluation of MBT-S for psychiatric nurses found that nurses experienced MBT-S to be a straightforward, empowering skill set which contributed to attitudinal change toward patients with relational difficulties.¹¹

In parallel with this research, MBT-S is being used in practice. Since 2014, MBT-S has been delivered in at least 44 training courses to over 600 clinicians within Scotland, from a wide range of professional groups, including psychiatric nurses, psychiatrists, psychologists, psychotherapists, allied health professionals (AHPs), prison officers, and criminal justice social workers (personal communication, MBT Scotland).

The main outcome measure used to evaluate MBT-S is the KAMQ. This is a 20-item self-report questionnaire, with items related to key aspects of MBT theory and practice. Thus far, there has been no research into the properties and validity of the KAMQ tool. While the above study findings are encouraging as to the effect of MBT-S, without any data on the properties of the KAMQ, there remains uncertainty as to the reliability and validity of the results. Developing the KAMQ as an outcome measure would allow better interpretation of existing findings. Furthermore, should the properties of the tool prove favorable, this would allow for more robust evaluations of MBT-S in the future. To our knowledge, there exists no other quantitative self-report tool that measures general clinicians' knowledge about core MBT-S theory and practice.

Given the pressing clinical issue and the promising MBT-S approach, investigating the properties and validity of the KAMQ is important for both research and applied work. The aim of this study was therefore to investigate the factor structure, internal consistency, reliability, and validity of the KAMQ. As this is a first investigation of the measure, we also aimed to refine the tool and identify where subsequent research on the KAMQ is needed.

2 | METHODS

2.1 | Procedure

Internal consistency and test structure were explored in Sample 1 and informed revision of the KAMQ to a final version, henceforth referred to as the KAMQ-2. Convergent validity of the KAMQ-2 with the Attitudes to Personality Disorder Questionnaire (APDQ)²⁸ was assessed in a subset of Sample 1 where both questionnaires had been administered. Test-retest reliability of the KAMQ-2 was established by recruiting a convenience sample of general and forensic mental health practitioners (Sample 2).

2.2 | Materials

KAMQ²⁶—the KAMQ is a 20-item questionnaire that measures knowledge of MBT-S theory and practice. Respondents rate their agreement with a series of statements (5-point Likert scaling; strongly

disagree = 1; strongly agree = 5). Items are scored so that higher scores indicate better knowledge of mentalizing concepts and MBT techniques, with some items having the scale reversed to allow this.

The KAMQ was developed by an expert panel of senior MBT Practitioners (study authors Williams, Cahill, and Patrick). To develop items, the panel drew on the following sources:

- Statements about MBT¹⁶⁻¹⁸ contained in Anna Freud Centre training course materials.
- Core theory and practice elements of the MBT-S¹⁹ course manual itself.
- Personal clinical experience of using mentalizing skills in everyday psychiatric work.
- Personal experience of recurring key themes and teaching points that emerge in delivering MBT-S to clinicians.

The subject of interest (knowledge about core MBT-S theory and practice) was a clearly defined and articulated area, with helpful existing sources to draw from to guide generation of items. From the above, AW and CC drafted initial items to pertain to key MBT-S theory and practice:

1. Knowledge about the clinical stance (eg, “The professionals might use a ‘stop and stand’ technique if they get muddled in mentalizing”).
2. Underlying theory (eg, “I think childhood experiences can have a profound impact on adult relationships”).
3. Awareness of key helpful and unhelpful staff actions (eg, “A therapist using mentalizing skills will be expressionless” [item scoring reversed]).
4. Staff beliefs and confidence about being able to apply mentalizing skills in practice (eg, “I think providing mentalizing skills requires a specialist psychotherapist” [item scoring reversed]).

AW and JP refined initial items to improve clarity and readability and reduce duplication. See Table 1 for full item list. The KAMQ has previously been used to measure the effects of MBT-S courses.^{15,24,27}

APDQ—the APDQ is a widely used questionnaire with good internal consistency (Cronbach's alpha = .94) and test-retest reliability ($r = 0.71$).²⁸ It is comprised of 37 Likert-scaled items involving attitudes toward people who have a diagnosis of personality disorder. APDQ scores can be represented using the five subscales (Enjoyment, Security, Acceptance, Purpose, Enthusiasm) or by the total score (range 37-222). The APDQ authors interpret higher scores as indicating more favorable attitudes. The rationale for using the APDQ to assess the concept validity of the KAMQ is the overlap between the core MBT-S stance and the core content of the APDQ. The Enjoyment subscale of the APDQ is the largest subscale and accounts for the most individual variance in APDQ total scores. The Enjoyment subscale reflects clinicians' “warmth and liking for, and interest in contact” with patients with personality disorder.²⁸ This concept overlaps with the core “mentalizing” stance taught by MBT-S, that is, a stance of support, empathy, and interest toward patients.

2.3 | Sample size determination

Sample size calculations completed using G*Power²⁹ were used to determine the necessary sample size for the planned analyses. Exploratory factor analysis (EFA) of the 20-item KAMQ required a minimum of 100 participants, using a commonly used minimum 5:1 subject-to-item ratio.^{30,31} Correlational analyses between the KAMQ-2 and the APDQ required a sample of 67 to detect a moderate, positive correlation ($r > 0.30$) between the measures ($\beta = .80$; α [one-tailed] = .05). Establishing test-retest reliability required a sample of 46 (ICC > 0.90; $\beta = 80\%$; $\alpha = .05$).

2.4 | Participants

Sample 1. The first sample ($n = 217$) is a multiprofessional group of staff working in general or forensic mental health settings, collated

from existing anonymized datasets. The sample comprises “pre-” training data from 2014 to 2016 from: a previous research study on staff training ($n = 42$) (“Tricky Interactions”²⁷) and ($n = 175$) from MBT Skills courses delivered in NHS Lothian,¹⁵ NHS Greater Glasgow & Clyde, and NHS State Hospitals Board for Scotland.³² Some information on professional affiliation was available: $n = 105$ nursing, $n = 12$ psychologists, $n = 10$ psychiatrists, $n = 7$ occupational therapists, $n = 8$ other (largely AHPs), $n = 75$ no information recorded.

Sample 2. A second sample ($n = 70$) comprised of mental health professionals of multiple disciplines working in either general or forensic mental health settings who were prospectively recruited by to establish test-retest reliability. Participants gave informed consent to complete the KAMQ-2 twice over a 2- to 4-week period. Stratified sampling was used to achieve approximately equal representation across four professional groups: nursing, psychiatry, psychology, and other clinical staff. To ensure

TABLE 1 Summary of exploratory factor analysis results for the Knowledge and Application of MBT Questionnaire using principal axis factoring and direct oblimin rotation ($n = 217$)

Item	Factor loadings		
	1	2	3
In mentalizing, professionals do not consider their own feelings	.68		
A therapist using mentalizing skills will be expressionless	.64		
Mentalizing avoids exploring the current therapeutic relationship with the person you are working with	.60		
I think providing mentalizing skills requires a specialist psychotherapist	.43		
I think mentalizing skills have a theoretical basis		.61	
I think mentalizing skills help promote therapeutic communication with people who have problems like borderline and antisocial personality disorder		.56	
I believe treating people using psychological techniques is a poor use of resources		.54	
Mentalization therapies do not allow patients to take medication	.38	.40	-.25
As mentalizing is straightforward, there is no need for supervision	.28	.40	
I think childhood experiences can have a profound impact on adult relationships		.39	
Strong disagreements between professionals about a person's management may be an indication of nonmentalizing	-.22	.37	.26
A partner being uncertain about what people are thinking, without checking, is an example of psychic equivalence			.67
The professionals might use a “stop and stand” technique if they get muddled in mentalizing			.57
A key component of mentalizing is thinking about people's attachment relationships			.51
Using mentalizing, you can ask questions to promote exploration		.28	.38
A person saying that a visit in the middle of the night from her partner was the only way she was reassured he loved her is an example of a teleological stance			.38
Breaks in therapy (or from relationships with people) are not considered important by MBT ^a	.31		.31
I understand when to apply/use MBT skills ^b		-.25	.26
% variance accounted for	12	11	11
Cumulative % variance	12	24	35
Correlations			
Factor 2		.55	
Factor 3		.48	.31

Note: Primary loadings are presented in boldface. Factor loadings < .30 are suppressed.

^aItem wording revised for KAMQ-2. Now reads “Breaks in contact with staff or key figures (eg, if a key-worker or partner goes away on holiday) are not considered important by MBT.”

^bItem was excluded from revised KAMQ-2 as it failed to load on any factor above the specified .30 threshold.

independent completion of the KAMQ, the first completed measure was submitted immediately so that it was not in the possession of the participant during their second completion of the measure. A total of 48 participants completed both sets of questionnaires and were included in the test-retest analysis.

2.5 | Research approvals

The study was reviewed by the State Hospital Research Committee and received managerial approval from the NHS State Hospitals Board for Scotland, NHS Lothian, and NHS Greater Glasgow and Clyde prior to commencing. The study was exempt from NHS ethical review as it did not involve patient data.

2.6 | Statistical analysis

Data were entered into an electronic spreadsheet and analyzed using R version 3.0.2,³³ with the *psych* package.³⁴ Missing items were imputed and pairwise complete cases were used. Shapiro-Wilk's test indicated most KAMQ and APDQ items were not normally distributed, and therefore nonparametric statistical tests were used. Internal consistency was assessed using Cronbach's alpha coefficient.³⁵ The KAMQ test structure was examined using EFA with oblique rotation as it was conceivable any underlying factors may be conceptually related to one another. Test-retest reliability of the KAMQ-2 total score over the two time-points was analyzed using the ICC calculated in a two-way random effects model.³⁶ Convergent validity between the KAMQ-2 and the APDQ was assessed using Spearman's rho (r_s) correlation. All statistical tests were two-tailed with $\alpha = .05$.

3 | RESULTS

3.1 | Measure completion

There were little missing data across the samples: 2.05% of KAMQ and 1.20% of APDQ observations in Sample 1 were missing. Forty-eight of the 70 participants (68.6%) recruited in Sample 2 completed the KAMQ-2 at both time-points. In the final test-retest sample ($n = 48$), there were no missing data at time 1, and 0.2% of observations were missing at time 2.

3.2 | Test structure and scale revision

The factorability of the 20 KAMQ items was examined using several indicators. Examination of the correlation matrix showed that 18 of the 20 KAMQ items correlated with another item at least $>.30$, suggesting reasonable factorability. All off-diagonal coefficients were below 0.70, and the determinant was 0.0003,

indicating an absence of multicollinearity between the variables. Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA³⁷) across all KAMQ variables was .81. MSAs of each variable were also considered. Two items, "I believe mentalizing is simple to put into practice" and "I feel confident about applying MBT skills into working practice" had unacceptable MSA values ($<.50$) indicating limited variance in respondents' scores and in association with the other items. These items were discarded from further analysis. Consequently, the revised MSA value for the remaining 18 KAMQ items increased to .85, indicating meritorious sampling adequacy.³⁸ Bartlett's Test of Sphericity on the 18-item KAMQ dataset was significant, $\chi^2(153) = 1023.91, P < .001$. These tests indicated the data were suitable for EFA.

A range of tests, the Scree test,³⁹ Velicer's Minimum Average Partial (MAP) test,⁴⁰ and parallel analysis⁴¹ were used to determine the number of factors to retain during EFA. The "elbow" in scree plot was observed at four factors (Figure 1), MAP test (completed with oblimin rotation and principal axis factoring [PAF] method) achieved a minimum value of 0.015 with one factor, and parallel analysis (using PAF) found three factors to be significant. Together these tests suggested one to four factors may be suitably retained in these data. Four separate models were then specified using PAF and oblimin rotation. Model selection was carried out on the basis of comparing absolute fit with the standardized root mean square residual index (SRMR) and root mean squared error of approximation (RMSEA⁴²), incremental fit using the Tucker-Lewis Index (TLI⁴³), and comparative fit using the Bayesian Information Criterion (BIC⁴⁴). The four-factor model reached the best fit across multiple indicators but included an invalid factor which contained a single item only, and therefore the four-factor model was disregarded. Of the remaining models, the three-factor model demonstrated the best fit, reaching acceptable or

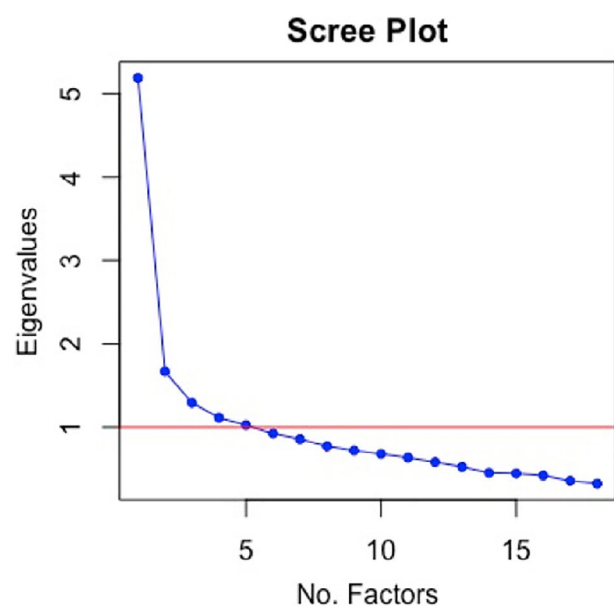


FIGURE 1 Scree plot showing Eigen decomposition for KAMQ items ($n = 217$)

KAMQ-2	
Factors	Example Items
1 Therapeutic relationship & techniques	"In mentalizing, professionals do not consider their own feelings" *
2 Beliefs about applying MBT in practice	"I believe treating people using psychological techniques is poor use of resources" *
3 Specific MBT knowledge	"A key component of mentalizing is thinking about people's attachment relationships" *

* Scoring reversed for these items

FIGURE 2 Three-factor structure of the KAMQ-2

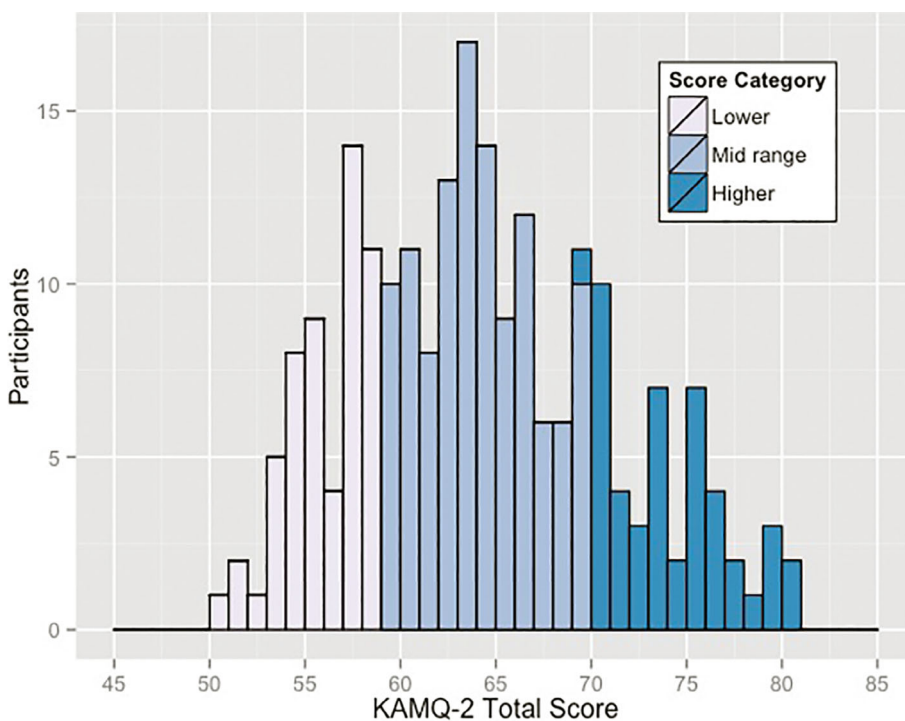


FIGURE 3 Frequency distribution of KAMQ-2 total scores in sample 1 ($n = 217$), by score category

good fit on all the fit indices used (RMSEA 0.052; SRMR 0.046; TLI 0.908; BIC -394 [smallest of the models tested]). One further item ("I understand when to apply/use MBT skills") failed to load onto any of the three factors above the prespecified 0.30 threshold and was therefore removed. Moderate-to-large positive correlations were observed between factors.

The three-factor KAMQ structure, item loadings, and inter-factor correlations are presented in Table 1 and depicted in Figure 2. The conceptual meaning of the three factors was considered. The highest loading items on Factor 1 relate to skills and techniques possessed by and characteristics of a practitioner engaged in mentalizing. The highest loading items on Factor 2 relate to beliefs and attitudes relevant for the use of mentalizing skills and MBT in clinical practice. Finally, several Factor 3 items including the highest loading items relate to knowledge of specific MBT theory and terminology. It is noted that the lowest loading item on

Factor 3, relating to breaks in therapy, has a nearly equal cross-loading with Factor 1. There several salient cross-loadings, particularly on Factors 1 and 2, consistent with the moderate correlation between these factors ($r = 0.55$), and this indicates that these two factors in particular share a degree of conceptual similarity. In addition to the removal of three items as described above, one item was reworded for clarity. The form was otherwise unchanged, and the resultant 17-item revised version is referred to as the KAMQ-2 (see Supplementary File).

3.3 | Reliability

Both internal consistency ($n = 217$, $\alpha = .85$, 95% confidence interval [CI] = 0.80-0.89) and test-retest reliability ($n = 48$, ICC = 0.84, 95% CI = 0.73-0.91) of the KAMQ-2 total score were "good."^{35,36}

TABLE 2 Correlation coefficients (Spearman's rho [r_s]) for associations between Attitudes to Personality Disorder (APDQ) and Knowledge and Application of MBT Questionnaire, revised (KAMQ-2) subscale and total scores ($n = 92$)

KAMQ		APDQ					Total
		Enjoyment	Security	Acceptance	Purpose	Enthusiasm	
Factor 1	Therapeutic relationships and techniques	.30**	-.20	-.10	.02	-.23*	.04
Factor 2	Specialized knowledge	.22*	-.18	-.13	.05	-.20	-.004
Factor 3	Beliefs about application	.41**	.01	.06	.19	-.03	.27**
	Total	.39**	-.16	-.06	.10	-.18	.13

* $P < .05$; ** $P < .01$.

3.4 | Score interpretation

The distribution of KAMQ-2 total scores in sample 1 ($n = 217$) approximated a normal distribution (skewness = 0.32, SE = 0.17; kurtosis = -0.53, SE = 0.33). The total KAMQ-2 score sample mean, with possible range of 17 to 85, was 63.79 (SD = 6.73). The scores corresponding to the 25th and 75th percentile were 58.8 and 69.0, respectively. These scores may be considered as potential cut-offs for “lower” and “higher” KAMQ-2 scores; however, as described further in the discussion, it would be useful to test the validity of these cut-offs by examination of KAMQ-2 scores in a sample of individuals with established skill levels for MBT. The frequency distribution of the KAMQ-2 scores with shading representing resultant score category following application of suggested cut-offs is provided in Figure 3.

3.5 | Validity

Spearman's rho coefficient was used to examine the association between the KAMQ-2 and APDQ total and subscales. Six of the 24 correlations (25%) reached statistical significance. Correlation coefficients are presented in Table 2. The analysis found statistically significant positive correlations between all KAMQ-2 factors and the APDQ “Enjoyment” subscale. The positive correlation between KAMQ-2 scores and the APDQ “Enjoyment” provides very provisional evidence for the validity of the KAMQ-2 measure. All significant associations were positive except for a small negative correlation between KAMQ-2 Factor 1 and APDQ “Enthusiasm” scores.

4 | DISCUSSION

Interventions to develop clinicians' skills for working with patients with relational difficulties are a clinical priority. Rigorous evaluation and development of such interventions depend on having high-quality measurement tools. MBT-S is one approach to support generalist clinicians' capacity to act reflectively and therapeutically in their everyday interactions with patients. The present study was carried out to provide researchers and clinicians using MBT-S with important information about the main outcome measure used—the KAMQ.

4.1 | Strengths and limitations

There are a number of strengths of the study. The revised KAMQ-2 was derived using a considered and empirically driven approach. Sample size requirements, determined using power analysis and best practice guidance, were met for each analysis, conferring confidence in the results. Study participants represented a broad range of staff across general psychiatry and forensic settings, rendering the findings relevant to a variety of mental health professionals. To our knowledge, this is the first study to investigate the properties and validity of a measure that assesses clinicians' knowledge of core MBT-S theory and practice, appropriate for evaluating the outcome of MBT-S courses.

In terms of limitations of this study, there remains a degree of uncertainty as to what score represents a “good” level of MBT-S knowledge. The KAMQ-2 scores corresponding to the 25th and 75th percentile in the present study's large sample of clinicians may serve as general guidance for interpreting KAMQ-2 scores; however, examining KAMQ-2 scores in samples of clinicians with and without established skill levels for MBT would add confidence in the interpretation of respondents' scores. Planning of future research employing such a “known groups” design to further assess the meaning of KAMQ-2 scores is underway; this planned work would also build on this article in further investigating the KAMQ-2's validity. Secondly, as mentioned in the introduction, we acknowledge the tenuous correlation between a participant's score on a self-report questionnaire following a skills workshop and actual observable clinical skills.²⁰ Accordingly, if a service is considering introducing mentalizing skills as a general approach for clinical staff to employ with patients, we suggest the KAMQ-2 be used in conjunction with other sources of information, in particular direct observation of clinicians' behavior in clinical practice. Other sources of information include observation of clinicians in role-plays, information from ongoing clinical supervision of MBT skills, patient measures such as the Service Attachment Questionnaire,⁴⁵ and routinely acquired indicators of safety (eg, frequency of patients' self-harm or violence as reported by Datix systems). Lastly, we realize that there would be benefit from further data to confirm the factor structure of the questionnaire—we plan to repeat the factor analysis with new data as this is collected in future MBT-S courses.

4.2 | Possible ceiling effects

We note that the mean score from a range of 17 to 85 is fairly high (63.79). Given this is a pretraining sample, this raises the possibility that ceiling effects could be observed. Indeed, in a multisite study of MBT-S for mental health professionals by Welstead et al, doctors and psychologists' KAMQ scores were higher at baseline compared with nurses, but their KAMQ scores increased to a lesser degree post MBT-S course as compared with nurses¹⁵—these results are consistent with ceiling effects. Possible ceiling effects mean it is not clear if the questionnaire will be able to discriminate between the most able participants, and this limitation is important for note for clinicians considering using the KAMQ-2 in practice. The possible ceiling effects may be connected to the fact that KAMQ-2 items have been designed to probe MBT-S content which is about *core* skills in everyday clinical practice (not “advanced” formal therapy skills). Notwithstanding ceiling effects for the most able participants, in the study by Welstead et al, there were large changes (effect size 1.2) in KAMQ scores from baseline to post MBT-S course, suggesting the measure is sensitive to change.

4.3 | Convergent validity and a new research avenue

Alongside the positive correlations between all KAMQ-2 factors and the APDQ “Enjoyment” subscale, we note the statistically significant negative correlation between the KAMQ-2 Factor 1 and the two-item APDQ “Enthusiasm” subscale. APDQ “Enthusiasm” items relate to feeling “frustrated” or “drained” by patients. Participants who endorse these two statements more strongly are scored lower (ie, worse) on this APDQ subscale. However, clinicians with better MBT skills and knowledge may be more likely to endorse these two statements in the APDQ (leading to the negative correlation between the two factors) due to higher awareness that frustration and “feeling drained” are normal and common experiences for clinicians. Furthermore, clinicians with better MBT knowledge would be expected to believe in the importance of acknowledging such responses so they can process these (or “mentalize” about them) for their own well-being and for the benefit of their patients. To address this assertion that being aware of a range of feelings is both helpful and normal for clinicians when working in disturbing clinical situations, we are currently developing a related self-report tool that assesses staff awareness of their emotional responses (both “positive” and “negative” feelings), with items scored higher (ie, better) when clinicians are more aware of such responses.⁴⁶

Further evidence to support the validity of the KAMQ is available from a recent study that used the KAMQ as an outcome measure of MBT-S training.¹⁵ Welstead et al observed that the KAMQ total score increased (effect size Cohen's $d = 1.2$) after a 2-day MBT-S training in a multiprofessional clinical staff group. There would be benefit from future assessment of convergent validity against a range of related measures, such as the Barrett-Lennard Relationship Inventory scale⁴⁷ and the Toronto Empathy Questionnaire.⁴⁸

4.4 | Wider context and implications of the study

The potential role of the KAMQ-2 following a skills course needs to be set in the context of the wider literature on therapeutic skills acquisition. As mentioned in the introduction, systematic reviews of therapist training studies have concluded that knowledge about a new therapeutic approach is likely to improve immediately following a skills workshop like MBT-S, but observable improvement in clinical skills is unlikely from a workshop alone.²⁰⁻²² Furthermore, even when skills do initially increase after a workshop, without additional regular supervision or consultation, skills tend to decrease over time and become comparable to those in untrained groups after several months.²⁰ Therefore, while the KAMQ-2 (which primarily assesses knowledge) is appropriate for what could be reasonably expected as an immediate outcome of the 2-day MBT-S course, it would be a misunderstanding to conflate a participant's good score on the KAMQ-2 following an MBT-S course as indicating that good mentalizing skills have been learned and applied in clinical practice. Rather, the MBT-S course and desired gain in knowledge as assessed by the KAMQ-2 should be seen as a first step only for participants, albeit an important one on the way to acquiring new skills. In order for clinicians to translate potential knowledge gain from a 2-day MBT-S course into new psychotherapeutic skills that are sustained over time, the existing literature on skills training would strongly suggest that regular supervision would be required over an extended follow-up phase.²⁰⁻²² In this follow-up phase of training, additional outcome measures would be needed to directly assess and observe clinicians' therapeutic skills and the impact on patients (see limitations section above for suggestions of outcome measures). The KAMQ-2 might potentially have a role in this follow-up phase in assessing participants' retention or otherwise of core MBT-S knowledge.

Within the identified limitations, our results suggest that the KAMQ-2 provides a reliable self-report tool, with very provisional evidence to support its validity. The internal factor structure makes clinical sense, with items clustering into three groups with the highest loading items relating to therapeutic relationship and techniques (Factor 1), beliefs about applying MBT in practice (Factor 2), and specific MBT knowledge (Factor 3). The final questionnaire is concise, making it practical for assessing the MBT-S course. It has face validity, given items were derived directly from the core content of MBT-S.

Establishment of the properties of the KAMQ-2 now allows for more robust evaluation of brief training courses such as MBT-S which are designed as the first step in training general staff in MBT skills for use in everyday clinical practice. The present study also increases confidence in the reliability and validity of existing studies of MBT-S that used the first version of the measure.^{15,24} Furthermore, we believe the KAMQ-2 may be useful in cross-sectional studies exploring staff factors which may affect staff burnout⁴⁹ and absenteeism⁵⁰ as well as patient outcomes including inpatient aggression⁵¹—that is, exploring whether clinicians' scores on the KAMQ-2, as a measure of core MBT-S knowledge, correlate with these important outcomes. Of course, development of this measurement tool provides but one piece of the puzzle—researchers and clinicians now need to work out how best to deliver MBT-S to teams and services, embed initial training into everyday practice, and sustain the impact through ongoing

supervision. Researchers now have an adequate tool to help with this task and assess the effect of the MBT-S course.

ACKNOWLEDGMENTS

We thank the anonymous reviewers, whose close reading and suggestions have significantly enhanced this article. We thank Hannah Welstead, Gary Cooney, Catherine Mulvenna, Sarah Beedie, and Chris Gray for their assistance gathering and entering data in NHS Lothian and NHS Greater Glasgow & Clyde; Claire Maclean for her role in organizing and delivering MBT-S training in NHS Lothian; Jamie Pitcairn and The State Hospital Research Committee for their support and comments on the original proposal; Tom Russ for discussion about ceiling effects which informed the revision. We also extend thanks to those individuals who participated in the study.

CONFLICT OF INTEREST

The Research Committee at The State Hospital provided input to study design. Dr Jon Patrick has received payment from the University of the West of Scotland (UWS) and NHS Education for Scotland (NES) for teaching MBT skills. Neither UWS or NES had any involvement in this article. Dr Jon Patrick receives personal fees for delivering MBT supervision. Dr Andrea Williams and Dr Jon Patrick are Anna Freud Centre accredited MBT supervisors.

AUTHOR CONTRIBUTIONS

Conceptualization: Adam Polnay, Lindsey G. McIntosh, Jon Patrick

Data curation: Lindsey G. McIntosh

Formal analysis: Lindsey G. McIntosh

Investigation: Adam Polnay, Lindsey G. McIntosh, Aileen Burnett, Andrea Williams, Catherine Cahill, Peter Wilkinson, Fiona Mohammad, Jon Patrick

Methodology: Adam Polnay, Lindsey G. McIntosh

Project administration: Adam Polnay, Lindsey G. McIntosh

Resources: Adam Polnay, Lindsey G. McIntosh, Aileen Burnett, Andrea Williams, Catherine Cahill, Peter Wilkinson, Fiona Mohammad, Jon Patrick

Supervision: Adam Polnay, Jon Patrick

Visualization: Lindsey G. McIntosh

Writing - original draft preparation: Adam Polnay, Lindsey G. McIntosh, Jon Patrick

Writing - review and editing: Adam Polnay, Lindsey G. McIntosh, Aileen Burnett, Andrea Williams, Catherine Cahill, Peter Wilkinson, Fiona Mohammad, Jon Patrick

All authors have read and approved the final version of the manuscript.

Lindsey G. McIntosh, manuscript guarantor, had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

TRANSPARENCY STATEMENT

The lead authors confirm that manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies

from the study as planned (and, if relevant, registered) have been explained.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Lindsey G. McIntosh (joint lead author) via the corresponding author upon reasonable request.

ORCID

Adam Polnay  <https://orcid.org/0000-0001-6309-0814>

REFERENCES

1. Craissati J, Minoudis P, Shaw J, et al. *Working with Offenders with Personality Disorder: A Practitioners Guide*. London: Ministry of Justice, National Offender Management Service; 2015.
2. NICE. *Borderline Personality Disorder: Treatment and Management. NICE Guideline CG78*. Leicester, England: The British Psychological Society and the Royal College of Psychiatrists [Full guideline]; 2009.
3. Hughes P, Kerr I. Transference and countertransference in communication between doctor and patient. *Adv Psychiatr Treat*. 2000;6:57-64.
4. Bateman A, Bolton R, Fonagy P. Antisocial personality disorder: a mentalizing framework. *FOCUS J Lifelong Learn Psychiatry*. 2013;11:178-186.
5. Watts D, Morgan G. Malignant alienation: dangers for patients who are hard to like. *Br J Psychiatry*. 1994;164:11-15.
6. National Institute for Mental Health in England. *Personality Disorder: No Longer a Diagnosis of Exclusion*. London: NIMHE; 2003.
7. Hinshelwood RD. Abusive help - helping abuse: the psychodynamic impact of severe personality disorder on caring institutions. *Crim Behav Ment Health*. 2002;12:S20-S30.
8. Simonsen S, Bateman A, Bohus M, et al. European guidelines for personality disorders: past, present and future. *Borderline Personal Disord Emot Dysregul*. 2019;6:9.
9. Oldham JM, Gabbard GO, Goin MK, et al. *Practice Guideline for the Treatment of Patients with Borderline Personality Disorder*. Washington, DC: American Psychiatric Association; 2001.
10. National Health and Medical Research Council. *Clinical Practice Guideline - Borderline Personality Disorder*. Melbourne, Australia: NHMRC; 2013.
11. Warrender D. Staff nurse perceptions of the impact of mentalization-based therapy skills training when working with borderline personality disorder in acute mental health: a qualitative study. *J Psychiatr Ment Health Nurs*. 2015;22:623-633.
12. Shanks C, Pfohl B, Blum N, Black DW. Can negative attitudes toward patients with borderline personality disorder be changed? The effect of attending a STEPPS workshop. *J Pers Disord*. 2011;25:806-812.
13. Haynos AF, Fruzzetti AE, Anderson C, Briggs D, Walenta J. Effects of dialectical behavior therapy skills training on outcomes for mental health staff in a child and adolescent residential setting. *J Hosp Adm*. 2016;5:55-61.
14. Kirkland J, Baron E. Using a cognitive analytic approach to formulate a complex sexual and violent offender to inform multi-agency working: developing a shared understanding. *J Sex Aggress*. 2015;21:394-405.
15. Welstead H, Patrick J, Russ T, et al. Mentalizing skills in generic mental healthcare settings: can we make our day-to-day interactions more therapeutic? *BJPsych Bull*. 2018;42:102-108.
16. Bateman A, Fonagy P. Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *Am J Psychiatry*. 2009;166:1355-1364.
17. Bateman A, Fonagy P. Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry*. 1999;156:1563-1569.

18. Bateman A, Fonagy P. Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *Am J Psychiatry*. 2001;158:36-42.
19. Bateman A, Treliving L. *Mentalizing for Mental Health Professionals*. London: Anna Freud Centre; 2012.
20. Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: a review and critique with recommendations. *Clin Psychol Rev*. 2010;30:448-466.
21. Rakovshik SG, McManus F. Establishing evidence-based training in cognitive behavioral therapy: a review of current empirical findings and theoretical guidance. *Clin Psychol Rev*. 2010;30:496-516.
22. Frank HE, Becker-Haimes EM, Kendall PC. Therapist training in evidence-based interventions for mental health: a systematic review of training approaches and outcomes. *Clin Psychol Sci Pract*. 2020;27:e12330.
23. Bateman A, Fonagy P. Mentalization based treatment for borderline personality disorder. *World Psychiatry*. 2010;9:11-15.
24. Polnay A, MacLean C, Lewington E, Patrick J. A pilot before-and-after study of a brief teaching programme for psychiatry trainees in mentalizing skills. *Scott Med J*. 2015;60:185-191.
25. MacLean C. *Report for Forensic Psychological Interventions Team*. Edinburgh: NHS Education Scotland; 2014.
26. Williams A, Cahill C, Patrick J. The knowledge and application of MBT questionnaire. Personal communication. Reproduced in Polnay A, MacLean C, Lewington E, Patrick J. A pilot before-and-after study of a brief teaching programme for psychiatry trainees in mentalizing skills. *Scott Med J*. 2015;60:185-191.
27. Williams A, Beedie S, Gumley A. Tricky interactions: exploring mental health staff responses to stressful interpersonal sequences following mentalizing skills training. *British and Irish Group for the Study of Personality Disorders 18th Annual Conference, Inverness, Scotland*; 2017.
28. Bowers L, Allan T. The attitude to personality disorder questionnaire: psychometric properties and results. *J Pers Disord*. 2006;20:281-293.
29. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175-191.
30. Gorusch R. *Factor Analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1983.
31. Hatcher L. *A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute; 1994.
32. McIntosh L, Judge J, Robertson L, et al. Evaluation of brief teaching in mentalizing skills to the multi professional team in a high secure setting. *The State Hospital Research and Clinical Effectiveness Conference, Carstairs*; 2016.
33. R Core Team. *R Version 3.0.2: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (2013).
34. Revelle W. *Psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, IL, <http://CRAN.R-project.org/package=psych> (2014).
35. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297-334.
36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
37. Cerny BA, Kaiser HF. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivar Behav Res*. 1977;12:43-47.
38. Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974;39:31-36.
39. Cattell RB. The scree test for the number of factors. *Multivar Behav Res*. 1966;1:245-276.
40. Velicer WF. Determining the number of components from the matrix of partial correlations. *Psychometrika*. 1976;41:321-327.
41. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behav Res Methods Instrum Comput*. 2000;32:396-402.
42. Steiger J, Lind J. Statistically-based tests for the number of common factors. *Paper Presented at the Annual Spring Meeting of the Psychometric Society, Iowa City, IA*; 1980.
43. Tucker LR, Lewis CA. Reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973;8:1-10.
44. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464.
45. Goodwin I, Holmes G, Cochrane R, Mason O. The ability of adult mental health services to meet clients' attachment needs: the development and implementation of the service attachment questionnaire. *Psychol Psychother Theory Res Pract*. 2003;76:145-161.
46. Polnay A, Walker H, Gallacher C. *The Relational Aspects of CarE Scale (TRACE); Final Report to The State Hospital Research Committee*. Carstairs: The State Hospital; 2019.
47. Barrett-Lennard G. *The Relationship Inventory: A Complete Resource and Guide*. West Sussex: Wiley-Blackwell; 2015.
48. Spreng RN, McKinnon MC, Mar RA, et al. The Toronto empathy questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J Pers Assess*. 2009;91:62-71.
49. Bowers L, Nijman H, Simpson A, Jones J. The relationship between leadership, teamworking, structure, burnout and attitude to patients on acute psychiatric wards. *Soc Psychiatry Psychiatr Epidemiol*. 2011;46:143-148.
50. Zogmaister C, Roccato M, Borra L. Health practitioners' implicit and explicit attitudes toward drug addicts predict absenteeism and extra work. *Basic Appl Soc Psychol*. 2013;35:191-199.
51. Bowers L, Allan T, Simpson A, Jones J, Van der Merwe M, Jeffery D. Identifying key factors associated with aggression on acute inpatient psychiatric wards. *Issues Ment Health Nurs*. 2009;30:260-271.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Polnay A, McIntosh LG, Burnett A, et al. Examining the factor structure, reliability, and validity of the main outcome measure used in mentalization-based therapy skills training. *Health Sci Rep*. 2021;4:e245. <https://doi.org/10.1002/hsr2.245>