A Validation Study of the Mindfulness-Based Interventions Teaching Assessment Criteria for Assessing Mindfulness-Based Intervention Teacher Skill: Inter-Rater Reliability and Predictive Validity

Frederick M Hecht, MD[1] ⓘ**, Rebecca S Crane, PhD[2]** ⓘ**, Patricia Moran, PhD[1],
Willem Kuyken, PhD[3]** ⓘ**, Wendy Hartogensis, PhD[1], and Judson Brewer, MD, PhD[4]**

Abstract

Background: Prior data suggests the Mindfulness-Based Interventions: (MBI) Teaching Assessment Criteria (MBI:TAC) has good inter-rater reliability, but many raters knew teacher experience level.

Objective: We sought to further evaluate the MBI-TAC's inter-rater reliability and obtain preliminary data on predictive validity.

Methods: We videorecorded 21 MBSR teachers from academic and community settings. We trained 19 experienced MBI teachers in using the MBI:TAC. MBSR teachers were rated by three assessors; teachers and their assessors did not know one another. To assess predictive validity, MBSR students in courses taught by 18 of the MBSR teachers were invited to complete PROMIS-29 measures before the MBSR course, at the end of the course (month 2), and month 4.

Results: Intraclass correlation coefficients (ICCs) representing a single rater ranged from 0.33 to 0.56 on the 6 MBI:TAC domains. Using an average of two raters, ICC estimates ranged from 0.48 to 0.71 and ICCs generalizing to an average of three raters ranged from 0.6 to 0.8. Among n = 152 participating MBSR students, we found improvements from baseline to 2 months and 4 months in PROMIS measures of Anxiety, Depression, Fatigue, Sleep, and Social Role function (range in improvement 2.3 to 6.3, $P < 0.0001$ for all comparisons except Social Role at 2 months, $P = 0.007$). Higher MBI:TAC ratings were associated with greater improvements in anxiety among MBSR students from baseline to 2 months, with a -0.31 lower participant anxiety score per 1 unit increase in MBI:TAC composite teaching rating (95% CI -0.58, -0.05, $P = 0.019$), but we did not find statistically significant relationships with improvements in other PROMIS-29 domains.

Conclusions: ICCs indicated good reliability using an average of three ratings, but inter-rater reliability was only fair using a single rater. We found initial validation that higher MBI:TAC ratings predicted greater improvements in anxiety symptoms in MBSR participants.

[1]Osher Center for Integrative Medicine, University of California, San Francisco, CA, USA
[2]Centre for Mindfulness Research and Practice, Bangor University, Bangor, UK
[3]Oxford Mindfulness Centre, Oxford University, Oxford, UK
[4]Mindfulness Center at Brown, Brown University, Providence, RI, USA

Corresponding Author:
Frederick M Hecht, MD, Osher Center for Integrative Health, UCSF, UCSF Box 1726, San Francisco, CA 94143-1726, USA.
Email: rick.hecht@ucsf.edu

Mindfulness-based approaches such as Mindfulness Based Stress Reduction (MBSR) have gained significant empirical support, showing benefits for the treatment of chronic pain,[1] substance use disorders,[2,3] anxiety disorders,[4] and depression.[5] Teacher skill is likely critical to the quality of mindfulness-based interventions (MBIs). In psychotherapy, there are clear indications that therapist skill influences outcomes.[6] In the case of MBIs, teaching competency may be even more important in influencing outcomes. MBI delivery has emphasized the centrality of the teacher's capacity to embody mindfulness through their way of being in the teaching space, rather than conveying concepts cognitively.[7]

Defining which teacher-related factors can both be feasibly and reliably assessed and shown to predict participant outcomes is potentially important in selection of teachers for MBI research studies and monitoring of intervention delivery quality.[8] Identifying teacher factors that influence outcomes may also be important for supporting the integrity of implementation of evidence-based MBIs,[9] and for strengthening teacher training for research and clinical programs.[10] The Mindfulness-Based Intervention Teacher Assessment Criteria (MBI:TAC) instrument may be a useful tool in assessing teaching competence in MBIs. We use "competence" in a specific way in this context: the knowledge, skills and attitudes relevant to leading MBIs.[11] Assessing the competence of MBI teachers is challenging because it is multidimensional and the key aspects need to be defined and validated. The MBI:TAC development involved a close analysis of the MBI teaching process by a group of teacher trainers from three university training centers.[12] These teacher trainers conducted a series of developmental stages in which the face and content validity of the tool were tested by practical application in training and research contexts.[12]

In initial assessment of the reliability and validity of the MBI:TAC, the internal consistency was high ($\alpha = .94$), as was the inter-rater reliability, with an overall intraclass correlation coefficient = .81 (range = .60-.81).[12] (There are several limitations to this prior work, however. First, inter-rater reliability was tested using 16 assessors from three centers, some of whom developed the instrument. For broader dissemination, it important to know whether training of a broader, more diverse group of assessors is feasible and still results in good inter-rater reliability. Second, the validity of the MBI:TAC in distinguishing experienced from novice teachers was tested in situations in which the assessor was usually aware of the experience of the teacher. Blinding assessors to the experience and background of teachers being

rated provides a more rigorous assessment of whether more experienced teachers are rated as having greater skill. Third, there has been limited evaluation of whether teacher skill as measured by the MBI:TAC is related to participant benefit from mindfulness-based programs (predictive validity). We carried out the PrOMPT (Predictors of Outcomes in MBSR Participants from Teacher Factors) study, which we report on here, to address these three issues. The study aimed to assess the inter-rater reliability of the MBI:TAC when evaluating teachers who were not known to the assessor, using a panel of recently trained assessors from a variety of centers.[13] We also asked MBSR students in courses taught by the teachers being evaluated in the PrOMPT study to participate in surveys before and after the course so that we could obtain preliminary predictive validity data on whether teacher skill, as measured by the MBI:TAC, was associated with the amount of change in validated measures of outcomes such as depression, anxiety, and stress.

Methods

Participants and Study Procedures

This study was reviewed and approved by the Institutional Review Board of University of California, San Francisco, and all participants provided written, informed consent. We recruited MBSR teachers from several sites that agreed to provide information about the study to MBSR teachers and students. These sites consisted of MBSR programs at academic medical centers (University of California, San Francisco and University of Massachusetts Medical School); and community programs in Florida, New York, North Carolina, Florida, Texas, and in Canada. The sites were selected to include both academic medical centers and community-based programs. MBSR teachers were eligible to be included in the study if they agreed to participate and were not participating in a training on the MBI:TAC instrument as assessors for this study. Participating teachers agreed to video record their 8-week MBSR courses. Recordings were made with the video camera pointing toward the teacher (and away from students). Students in the MBSR course were informed about the recording process and the procedures in place to protect students and teachers; these included use of a secure, password-protected server for digital storage, and the fact that recordings would only be used for research and carefully selected training purposes. MBSR students were also told that they could sit in locations that were not adjacent to the teacher

to avoid having their faces included in the recording. Three MBSR teachers we recruited had video recordings of courses they previously taught that we used for assessment of inter-rater reliability in this study. As they were not teaching courses in which we could recruit participants at the time we were enrolling MBSR participants these three teachers were not included in the predictive validity component of the current study.

Once an MBSR teacher agreed to study participation and provided written informed consent, students who registered for their MBSR courses were provided with study information including links to a study website with further information about the study. Inclusion criteria for MBSR students were: (1) enrollment in an MBSR classes taught by a participating MBSR teacher during the study period, (2) providing consent to participate in the study after receiving detailed information about what participation involved, and (3) age 18 years or older. Those who were interested in enrolling were asked to view a video describing the study and the importance of follow-up if they enrolled. After viewing the video, students who were interested in participating signed an online consent form and complete enrollment online. If preferred, potential research participants could also request to be contacted by phone, or could call a study coordinator directly to learn more about the study, and could complete study enrollment and assessments steps in person at University of California San Francisco or the Center for Mindfulness at University of Massachusetts rather than online.

Measures

We used a panel of assessors who were trained in using the MBI:TAC instrument to rate teaching skill in six domains. The training process has been described in detail previously.[13] In brief we conducted a 7-session course to train experienced MBI teachers in using the MBI:TAC;18 assessors provided ratings in the current study. Three assessors, none of whom knew the teacher, then rated each MBSR teacher. We randomly selected two recorded sessions from each teacher for rating, with one selected from the first four classes and the second recording from the last four classes of the teacher's 8-week MBSR course. Assessors assigned an initial rating after viewing the first session, then made a final rating of the teacher after watching the second session. The MBI:TAC instrument has six domains: (1) coverage, pacing, organization; (2) relational skills; (3) embodying mindfulness; (4) guiding mindfulness practices; (5) conveying course themes through interactive inquiry and didactic teaching; and (6) holding the group learning environment.[12] Each of these domains is rated on a six-point scale, ranging from "1: incompetent" to "6: advanced." There is also a summary score across all six domains.

MBSR students who agreed to be part of the study were asked to fill out an online survey three times: (1) prior to starting the MBSR course, (2) 2 months later (immediate post-MBSR), and (3) 4 months later (post-MBSR follow-up). The survey included the PROMIS-29 profile v1.0,[14] which has sub-scales for measuring fatigue, depression, anxiety, sleep disturbance, physical function, social role function, pain interference, and pain intensity. We also included the Perceived Stress Scale 4-itemshort form[15] to measure perceived stress.

Analysis

We used Stata (version 16) for statistical analyses, We evaluated inter-observer variability using the intraclass-correlation coefficient (ICC).[2] ICC provides a measure of how strongly ratings of the same item (e.g., a domain of the MBI:TAC) by different raters resemble each other, with values ranging from zero0 (only random agreement) to 1 (perfect agreement). We calculated the ICC for each of the six domains in MBI:TAC, using the ratings from the different assessors. We also calculated a composite score on the MBI:TAC by the summing the six MBI:TAC domains, and calculated the ICC for the composite score. We made use of the pool of three different assessors' rating of each teacher in several different ways. We calculated the ICC using each assessor separately to estimate the ICC if a single assessor is used to rate a teacher. As averaging the ratings from three different assessors should provide a higher ICC that reduces variability due to individual assessor diffferences, we also used the ratings obtained from three different reviewers' rating of each component of the MBI:TAC to model the ICC if ratings from three different assessors were obtained and averaged. As each teacher was rated by three different assessors, but the combination of assessors differed for each teacher, this is meant to provide an estimate of the ICC if a panel of three assessors is used to evaluate a teacher and the rating of each MBI:TAC domain is averaged. Third, to provide a robust estimate of the ICC values if we obtained ratings from two assessors, we made use of ratings from all three assessors and implemented a resampling procedure with 10,000 reps in which we randomly selected 2 of 3 ratings for each teacher to obtain estimates of ICCs if the average of 2 assessors was used for ratings. The recommended practice for MBI:TAC ratings has been to assess two classes before making a rating. To assess whether reviewing two MBSR classes rather than a single class improved the inter-rater reliability, we asked assessors to provide an initial rating after reviewing a single video recording of an MBSR course and calculated ICC values. We then compared the ICC after viewing a single class to the ICC values after viewing two classes. Based on prior examples, we pre-defined ICC values of at least 0.6 to be good agreement, with 0.75 or greater considered excellent agreement.[16] To assess whether MBI:TAC teacher ratings predicted change in MBSR student outcome measures, we used linear mixed effects models with participant PROMIS-29 measures as outcomes (one model

Table 1. Baseline Characteristics of MBSR Teachers in the PrOMPT-F Study.

Teacher Characteristic	% (n/N) or mean (SD) (n = 18 in outcomes study)	% (n/N) or mean (SD) (n = 21 in rater reliability study)
Age, years (mean, SD)	59.1 (10.5)	58.7 (10.2)
Gender		
Female	88.9% (16/18)	81.0% (17/21)
Male	11.1% (2/18)	19.0% (4/21)
Race/ethnicity		
Non-Hispanic White	88.9% (16/18)	90.5% (19/21)
Latino/Latina/Latinx	5.6% (1/18)	4.8% (1/21)
Other[a]	5.6% (1/18)	4.8% (1/21)
Education		
College Graduate	22.2% (4/18)	19.0% (4/21)
Master's level Degree	55.6% (10/18)	61.9% (13/21)
Doctoral level Degree	22.2% (4/18)	19.0% (4/21)
Years Teaching MBSR Courses (mean, SD)	5.3 (4.9)	5.1 (4.6)
Number of MBSR Courses Taught ever (mean, SD)	14.9 (15.3)	10.5 (11.4)
Enrolled Participants per teacher (mean, SD)	8.7 (4.7)	--

[a]n = 1 teacher checked only "other" for race/ethnicity, and in the follow up text box listed "Chinese and White European".

per outcome), and MBI:TAC rating, time point, and their interaction, as predictors, and with random effects for students nested within teachers. Models were used to estimate marginal slopes of teacher ratings on PROMIS-29 outcome measures at 2 and 4 months. Pearson correlation coefficients were calculated for the correlation between participant outcome measures and teacher MBI:TAC ratings at follow up time points.

Results

We enrolled 21 MBSR teachers in the study. The average age of the teachers was 59 years old, and they were predominately female (Table 1). The length of experience teaching MBSR courses ranged from 1 to 20 years, with an average of 5 years. We enrolled 152 students in the recorded courses from 18 of the 21 teachers (average number of participants per teacher = 8.7, range 2 to 22). Participant data was not available for the remaining three teachers because they provided video recordings from courses taught prior to the start of participant enrollment, and thus only contributed to the inter-rater reliability assessment. The average age of students was 49 years old, and 78% were female (Table 2). From a list of main reasons for taking the course the four most common reasons reported were to "become more mindful," "reduce anxiety," "improve quality of life," and "stress reduction" (each 17% to 19%). Less frequently endorsed reasons included physical health problems (4%), professional interest (8%), and coping with pain (3%). Students were also asked about "other reasons" that they chose to enroll in the course, and 28.3% endorsed depression, and 52.6% endorsed reducing anxiety as a secondary reason for enrollment. PROMIS 29 T-scores are calibrated so that a score of 50 corresponds to a United States general population. A score of 55 - 60 on a PROMIS 29 sub-scale

represents mild symptoms (except for the physical function sub-scale, in which a lower score of 40 - 45 represents mild symptoms). The PROMIS-29 baseline T-scores for participants were all close to the normal range, except for the Anxiety sub-scale, with a score of 58.6 (Table 2).

We evaluated inter-rater reliability using ICCs for a single assessor after viewing two MBSR classes (Table 3). The ICC ranged from 0.33 for domain 2 (relational skills) to 0.56 for domain 3 (embodiment of mindfulness). When we calculated ICCs generalizing to the use of an average MBI:TAC score across ratings by three different assessors, ICCs were higher, and ranged from .60 to .80 on the six MBI:TAC domains after reviewing two sessions. Similar to the ICCs for a single assessor, the ICC was best for domain 3 (embodiment of mindfulness, .80) and lowest for domain 2 (relational skills, .60). The ICC estimates based on resampling for the use of two assessors (generalizing to using the average across two assessors) were between the ICC's based on using a single assessor and the ICC's using the average of three reviewers, and ranged from 0.48 to 0.71.

To assess how much reviewing two MBSR classes rather than a single class improved the inter-rater reliability, we asked assessors to provide an initial rating after reviewing a single video recording of an MBSR course (Table 3). The ICCs of ratings done after reviewing only one session were substantially lower when using a single assessor, with the highest ICC being 0.37. When we used the average of three assessors viewing a single class session, ICCs were lower than after viewing two sessions, but were still above 0.5 for all domains except domain 2 (relational skills, ICC = 0.32).

Next we assessed whether MBI:TAC ratings of teaching skill were related to how much students in the MBSR course improved in different outcomes assessed by the PROMIS-29 measure (Table 4). For the PROMIS anxiety scale, the

Table 2. Baseline Characteristics of PrOMPT-F MBSR students in Courses Taught by Teachers who Were Rated using the MBI:TAC Instrument.

Student Characteristic (n = 152)	% (n/N) or mean (SD)
Age, years (mean)	49.0 (14.1)
Gender	
Female	78.3% (119/152)
Male	21.1% (32/152)
Transgender	0.7% (1/152)
Race/Ethnicity	
White	84.2% (128/152)
Asian	3.3% (5/152)
Hispanic/Latino/a	5.9% (9/152)
Black	3.3% (5/152)
Other[a]	3.3% (5/152)
Education	
High School graduate or GED	11.9% (18/152)
College graduate	33.6% (51/152)
Master's degree	41.4% (63/152)
Doctoral Degree (e.g. PhD or equivalent)	13.2% (20/152)
Employment Status	
Full time, 35+hrs	55.3% (84/152)
Part time, <=34 hrs	20.4% (31/152)
Unemployed, <1mo	5.2% (8/152)
Not working (e.g., student, homemaker)	19.1% (29/152)
Income, annual	
<$25k	7.2% (11/152)
$25k-$45k	9.9% (15/152)
$45k-$70k	16.4% (25/152)
$70k-$125k	23.0% (35/152)
>$125k	27.0% (41/152)
Do not know/decline to answer	16.5% (25/152)
Main reason for enrolling in MBSR class	
Become more mindful	19.08% (29/152)
Reduce anxiety	19.08% (29/152)
Improve Quality of Life	19.08% (29/152)
Stress Reduction	17.11% (26/152)
Professional Interest	7.24% (11/152)
Physical Health Problem	3.95% (6/152)
Reduce Depression	1.32% (2/152)
Improve concentration/focus	1.97% (3/152)
Improve sleep	0.66% (1/152)
Adjust to life changes	6.58% (10/152)
Cope with pain	2.63% (4/152)
Other	1.32% (2/152)
PROMIS 29 Measures T scores	
Physical function	53.2 (6.4)
Anxiety	58.6 (7.7)
Depression	52.6 (8.4)
Fatigue	52.4 (9.2)
Pain Interference	49.1 (8.5)
Sleep Disturbance	51.9 (6.9)
Social Role Function	48.8 (8.3)

[a]Among 5 participants listed as race/ethnicity = "other," n = 2 identified as both White and as having Hispanic/Latino ethnicity (for their self-reported "other" race, n = 1 self-reported "Nicaraguan"; and n = 1 self-reported "Hispanic" in a text follow up field). One of the 5 who selected "other" identified as White but did not identify as having Hispanic/Latino ethnicity, and this participant self-reported "Cape Verdean" as their other race/ethnicity. The remaining 2 participants who listed race/ethnicity as "other" did not select any other race category, and both identified as having Hispanic/Latino ethnicity; both also self-reported "Hispanic" in the "other" race text follow up field.

composite score on the MBI:TAC (the sum of all six MBI:TAC domains) was inversely associated with anxiety at 2 months (end of MBSR course), with a −0.31 lower participant anxiety score per 1 unit increase in MBI:TAC composite teaching rating (95% CI −0.58, −0.05, $P = 0.019$; Figure 1). By 4 months this association of MBI:TAC scores with anxiety was no longer significant (−0.01, 95% CI −0.29, 0.27, $P = 0.96$). On individual MBI:TAC domains, all the domains had statistically significant associations with anxiety scores at 2 months except for Domain 6, but no

Table 3. Intraclass Correlation Coefficients (ICCs) for MBI:TAC Domains When Rating Mindfulness-Based Stress Reduction Teachers.

MBI:TAC Domain	Measurement Type	ICC of final rating	ICC after 1 Video
1 (Coverage, organization)	Individual	0.45	0.35
	Average of 2 assessors	0.60	
	Average of 3 assessors	0.71	0.62
2 (Relational skills)	Individual	0.33	0.14
	Average of 2 assessors	0.48	
	Average of 3 assessors	0.60	0.32
3 (Embodiment of mindfulness)	Individual	0.56	0.30
	Average of 2 assessors	0.71	
	Average of 3 assessors	0.80	0.57
4 (Guiding practices)	Individual	0.43	0.37
	Average of 2 assessors	0.59	
	Average of 3 assessors	0.70	0.64
5 (Interactive inquiry and didactics)	Individual	0.51	0.31
	Average of 2 assessors	0.67	
	Average of 3 assessors	0.76	0.57
6 (Group learning environment)	Individual	0.44	0.27
	Average of 2 assessors	0.59	
	Average of 3 assessors	0.70	0.53

Notes: ICC = intraclass correlation coefficient. The final rating represents the rating after the standard process of three assessors viewing two MBSR classes. We also asked assessors to make a rating after viewing the first class (1 video rating). The individual ICC represents the ICC if using a rating from a single assessor. The Average represents the ICC if using an average of 3 assessors. The estimates based on 2 assessors were generated by a random resampling procedure with 10,000 reps.

Table 4. Association of Composite MBI:TAC Teacher Rating With Change in MBSR Participant PROMIS Measures.

PROMIS Measure	Month	Correlation coefficient	95% CI for correlation coefficient (lower, upper)	Slope	Slope 95% Confidence Intervals (lower, upper)	P-value
Anxiety	2	−0.21	−0.38,-0.03	−0.31	−0.58, −0.05	0.019
	4	0.03	−0.16,0.21	−0.01	−0.29, 0.27	0.96
Depression	2	−0.03	−0.21,0.16	0.01	−0.29, 0.32	0.92
	4	0.01	−0.18,0.19	0.03	−0.29, 0.35	0.84
Fatigue	2	−0.13	−0.30,0.05	−0.17	−0.46, 0.13	0.27
	4	−0.04	−0.23,0.15	0.00	−0.31, 0.32	0.99
Pain interference	2	0.00	−0.18,0.18	−0.02	−0.30, 0.26	0.90
	4	0.03	−0.16,0.22	0.04	−0.26, 0.33	0.80
Physical function	2	0.05	−0.14,0.23	0.08	−0.13, 0.29	0.45
	4	0.05	−0.14,0.24	0.10	−0.12, 0.32	0.36
Sleep disturbance	2	0.01	−0.17,0.19	0.03	−0.22, 0.27	0.83
	4	0.14	−0.05,0.32	0.19	−0.07, 0.45	0.15
Social role function	2	0.20	0.01,0.36	0.25	−0.03, 0.52	0.079
	4	0.07	−0.13,0.25	0.01	−0.29, 0.30	0.97

Notes: We report Pearson correlation coefficients for the association of participant outcomes (PROMIS-29 measures) at 2 and 4 months with teacher's mean MBI: TAC composite score (the composite score was defined as the sum of scores across 6 domains; these scores were averaged across three assessors). The 95% confidence intervals for the correlation coefficient were based on Fisher's transformation. The slopes of MBT:TAC teacher ratings on outcome measures at 2 and 4 months (with 95% confidence intervals and associated P-values) were derived from linear mixed effects models, and represent the change in participant outcomes with each one unit increase in the teacher's mean MBI:TAC composite score.
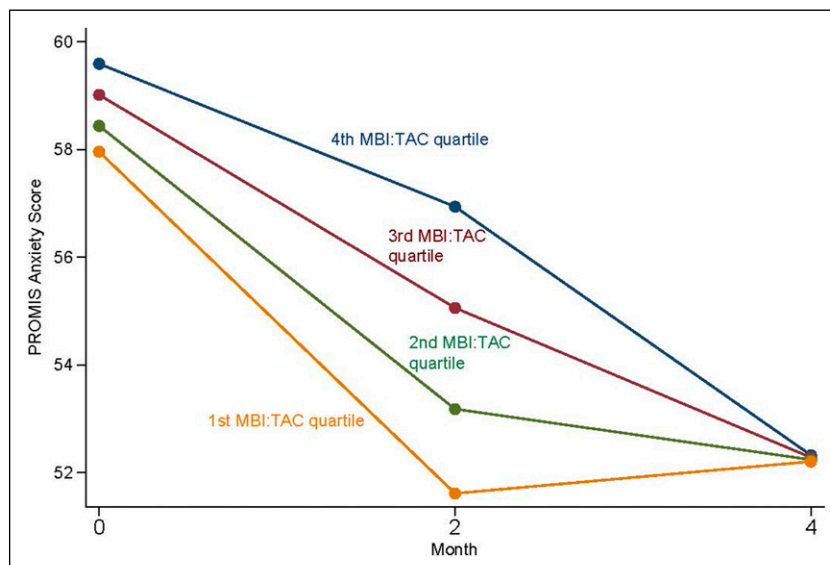
Figure 1. Title: MBSR participant Mean PROMIS Anxiety Score Over Time by Teacher MBI:TAC Rating. Legend: The y-axis shows the mean MBSR participant's score on the PROMIS Anxiety measure at baseline (0 months), 2 months (end of MBSR course), and 4 months (two months after the end of course). Participants are divided into four quartiles based on the average composite score of their MBSR teacher across all six MBI:TAC domains. Participants with teachers in the 1st quartile of MBI:TAC ratings (highest rating) had the greatest decrease in PROMIS Anxiety scores, followed in order by each of the remaining quartiles (P = 0.019, linear mixed model of MBI:TAC score predicting change in PROMIS Anxiety measure). The change in PROMIS Anxiety score by teacher MBI:TAC rating was no longer statistically significant at 4 months (P = 0.96).

domains were significantly associated with anxiety scores at 4 months. For the remaining PROMIS measures, there were no statistically significant associations between the composite MBI:TAC measure with depression, fatigue, pain interference, physical function, sleep disturbance, or social role function student outcomes, nor with any of the individual MBI:TAC domains at 2 months (Table 4).

Discussion

We had several important findings in this study of the MBI:TAC instrument that may be particularly relevant for its use in research studies, but also have implications for its use in other settings in which it is used to evaluate teaching competence. Overall, we found good inter-rater reliability for the average of ratings from three assessors, with ICC's ranging from 0.60 to 0.80 on different domains after viewing two MBSR sessions. This helps to further validate the instrument. However, ICC's were lower when using only one assessor, with ICC's above .50 for only three of the six domains, indicating limits on inter-rater reliability for several of the MBI:TAC domains when using a single assessor. This suggests that for purposes where a high degree of inter-rater reliability is needed with the MBI:TAC, averaging across several assessors is optimal. ICCs for an average of two assessors were lower than for an average of three assessors, but found good agreement ($\geq$0.6) for three domains (domains 1, 3 and 5), with ICCs just below our threshold for two other

domains (ICCs for domains 4 and 6 = 0.59). This suggests that two assessors may provide adequate inter-rater reliability for many purposes.

The ICCs in this study are lower than in a prior report, where the ICC using a single assessor ranged between 0.60 and 0.81.[12] Several differences between the methods used in the current study and those used in earlier studies may account, at least in part, for the lower ICCs we observed. First, in the prior study, the teachers being rated were typically known to the assessor, including their level of experience in teaching. Knowledge of the teacher's background may have provided additional information that assessors used when rating the teacher, resulting in more consistent or potentially biased ratings. In contrast, in the present study we selected assessors who did not know or recognize the teacher being assessed. Second, we used assessors who had gone through a standardized training in use of the MBI:TAC and were experienced teachers themselves, but in general assessors in our study had less experience using the MBI:TAC than in prior studies, and may have had less opportunity to develop a scoring approach that was closely calibrated to other assessors in the study.[13] This was planned intentionally to represent inter-rater reliability that might be obtained after training new assessors. Third, our assessors came from multiple countries and may have had greater diversity in their training experiences and approach to MBSR teaching as well as diversity in cultural and language backgrounds than assessors in earlier studies, most of whom were trained in the UK. Of note, however, assessors used in the current study had

substantially higher ICCs when rating selected test videos at the end of their training, when ICC's ranged between 0.67 and 1.0. The lower ICC in the current study using many of the same assessors might be due, in part, to greater challenges in rating the videos used in this study. The earlier test videos were shorter, focused on specific sections of a class, and selected to check calibration of ratings after training. It is possible that the longer and more "real life" MBSR class sessions, with greater diversity of class activities being evaluated in this study were more challenging to evaluate consistently. The ratings for this study were also done at least 6 months after the training was completed, and it is possible there was some loss of shared calibration of ratings over this time.

We also evaluated whether assessing a single MBSR session rather than viewing two MBSR sessions, as has been standard practice, yielded ICC's that were fairly similar to those from watching two sessions. If reliable ratings could be obtained after viewing a single session, this could reduce the time needed to obtain an MBI:TAC rating nearly in half. Unfortunately, we found that ICC's were substantially lower after viewing a single session, suggesting that viewing two sessions is more optimal for inter-rater reliability.

There are several implications of this study for the use of the MBI:TAC in research studies. We believe our results suggest that two or more assessors should usually be used to get good inter-rater reliability for fidelity assessments in research studies. For good inter-rater reliability, two class sessions need to be viewed rather than one. This is labor intensive and thus resource intensive (e.g., about $250 or more per teacher evaluated by one assessor) and relies on the availability of trained assessors. This project has expanded the pool of trained assessors, and developed materials that can be used for future training of assessors, making it more feasible to have trained assessors available. The cost of obtaining ratings from skilled assessors may still make it challenging to use the MBI:TAC in studies with limited resources, however, such as in pilot studies. An important future direction may be the development and validation of instruments for teacher rating by participants, which could offer less expensive, if potentially less accurate, measures of teacher skill. Checklists of elements of teaching that can be evaluated by study staff may also provide an important measure of fidelity with lower cost.

This study was also intended to assess feasibility of a study design to assess whether MBI:TAC ratings predict participant benefit on selected outcomes, and to obtain preliminary data on associations between MBI:TAC teaching ratings and outcomes in students taking MBI courses in typical academic and community-based centers. While there was some loss of follow-up in our design—which we believe might be improved in future research—overall retention was adequate. We found that higher MBI:TAC teacher ratings predicted greater improvements in anxiety at 2 months, the end of the MBSR course, suggesting that teaching skill as rated by the MBI:TAC was important for reducing anxiety in course participants (Figure 1). By 2 months, average PROMIS Anxiety scores were no longer in the elevated range (< 55) in participants in MBSR courses taught be teachers in the upper half of MBI:TAC ratings (1st and 2nd quartiles). However, by 4 months these differences based on the MBI:TAC were no longer statistically significant. This occurred despite slight overall improvements in average PROMIS anxiety scores between 2 and 4 months. By month 4, average PROMIS Anxiety scores were no longer elevated in participants across all four MBI:TAC quartiles of MBSR teachers, however. This could be consistent with more gradual improvements in anxiety after the MBSR course that were less dependent on teacher skill.

We did not find clear evidence of a relationship between participant improvements on other PROMIS-29 domains and MBI:TAC ratings. One factor may be that participants in this study appeared to have greater elevation of anxiety than other PROMIS 29 scales, which was the only scale elevated above the normal range, while other scales were in a generally normal range (between 45 and 55), leaving less room for improvement (floor effects). The importance of anxiety to the population we studied was also reflected in the answers to why participants had chosen to take the MBSR course: "reduce anxiety" and "stress reduction" were two of the four most common reasons for taking the course. In contrast, only 3% of participants noted pain as a reason for taking the course, and the PROMIS pain interference score (49) was slightly below average for a US population. The modest room for improvement in pain interference limited the utility of this measure in assessing predictive validity of the MBI:TAC for this outcome in the population we studied.

An important limitation of the current study in assessing whether MBI:TAC teacher ratings predict participant outcomes is that it was designed to collect preliminary data for designing future studies of the relationship between teaching skill and outcomes in MBIs, and was not designed to collect definitive data. For future studies of the relationship between teaching skill and MBI outcomes, our study suggests it may make sense to restrict the outcomes studied to those that are central for the population studied (e.g., assess pain in a population seeking the program for a pain issue). Alternatively, if a general population taking MBSR is studied, this may require a large population in which sub-sets of individuals seeking the program for specific reasons, such as chronic pain, can be used in testing the relationship between teacher skill and pain outcomes. Another potential limitation is that we evaluated seven different outcome measures from the PROMIS-29 and did not adjust for multiple comparisons. We did not plan adjustment for multiple comparisons for several reasons, including that we were collecting preliminary data, some of the outcome measures (e.g., anxiety and depression) are correlated which is not optimal for the assumptions of most multiple comparison adjustments, and the need for multiple comparison adjustments in this type of

study is controversial.[17] The consistency of finding that five out of the six domains on the MBI:TAC had statistically significant associations with anxiety at 2 months provides additional reassurance in these associations. Nevertheless, further validation of these findings is needed for greater confidence in their implications. Another limitation of our study is that MBSR teachers and participants were mostly college graduates or had more advanced degrees, and were mostly white. This limited diversity reflects, in part, the demographics of current MBSR teachers and participants, but is an important limitation of the current study that the authors hope can be better addressed in future studies.

The experience of training MBI:TAC assessors in this study led to significant learning and subsequent adjustments to future training implementation practice.[13] This included recognizing the diversity of motivations for engaging in training to use the MBI:TAC and creating tailored trainings for these various aims.[18] Trainee motivations include building skills in training MBI teachers,[19] supervising MBI teachers,[20] conducting assessments of MBI teachers,[21] and as an informal tool to enable personal reflection on MBI teaching skills.[22]

In summary, this study provides further data on the inter-rater reliability of the MBI:TAC instrument. Our data suggest that if higher ICCs are important, averaging ratings from more than one assessor is optimal, and that reviewing two MBSR course sessions rather than one provides a higher ICC. We found preliminary data that greater teaching skill, as measured by the MBI:TAC, predicts greater improvements in student anxiety at the end of the MBSR course, providing initial predictive validity of the instrument. Further research is needed to assess the relationship of MBI:TAC ratings to other student outcomes. Future evaluations of the relationship between teaching skill and participant outcome need to select research measures that are sensitive to the particular issues that are meaningful and relevant to the population in question.

The MBI field holds significant promise for addressing wellbeing in individuals and groups. Realizing this potential requires that the 'thorny issue of clinician training'[23] and subsequent teaching skill is engaged with and is folded into the research journey going forward.

Acknowledgments

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Frederick M Hecht https://orcid.org/0000-0002-5782-1171
Rebecca S Crane https://orcid.org/0000-0003-3605-0256
Willem Kuyken https://orcid.org/0000-0002-8596-5252

References

1. Kabat-Zinn J, Lipworth L, Burney R. The clinical use of mindfulness meditation for the self-regulation of chronic pain. *J Behav Med*. 1985;8:163-190.
2. Bowen S, Witkiewitz K, Clifasefi SL, et al. Relative efficacy of mindfulness-based relapse prevention, standard relapse prevention, and treatment as usual for substance use disorders: a randomized clinical trial. *JAMA Psychiatry* 2014;71:547-556.
3. Brewer JA, Mallik S, Babuscio TA, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. *Drug Alcohol Depend* 2011;119:72-80.
4. Hoge EA, Bui E, Mete M, et al. Mindfulness-based stress reduction vs escitalopram for the treatment of adults with anxiety disorders: a randomized clinical trial. *JAMA Psychiatry* 2023;80:13-21.
5. Kuyken W, Warren FC, Taylor RS, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. *JAMA Psychiatry* 2016;73:565-574.
6. Trepka C, Rees A, Shapiro DA, et al. Therapist competence and outcome of cognitive therapy for depression. *Cogn Ther Res* 2004;28:143-157.
7. Crane RS, Kuyken W, Hastings RP, Rothwell N, Williams JMG. Training teachers to deliver mindfulness-based interventions: learning from the UK experience. *Mindfulness N.* 2010;1:74-86.
8. Crane RS, Hecht FM. Intervention integrity in mindfulness-based research. *Mindfulness*. 2018;9:1370-1380.
9. Rycroft-Malone J, Gradinger F, Owen Griffiths H, et al. Mind the gaps': the accessibility and implementation of an effective depression relapse prevention programme in UK NHS services: learning from mindfulness-based cognitive therapy through a mixed-methods study. *BMJ Open*. 2019;9: e026244.
10. Crane RS, Kuyken W. The mindfulness-based interventions: teaching assessment criteria (MBI:TAC): reflections on implementation and development. *Curr Opin Psychol*. 2019;28: 6-10.
11. Crane RS, Kuyken W, Williams JM, et al. Competence in teaching mindfulness-based courses: concepts, development and assessment. *Mindfulness N* 2012;3:76-84.
12. Crane RS, Eames C, Kuyken W, et al. Development and validation of the mindfulness-based interventions – teaching assessment criteria (MBI:TAC). *Assessment*. 2013;20:681-688.

13. Crane RS, Hecht FM, Brewer J, et al. Can we agree what skilled mindfulness-based teaching looks like? Lessons from studying the MBI:TAC. *Glob Adv Health Med*;9:2164956120964733.
14. Health Measures. *PROMIS Adult Profile Instruments*; 2021. https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Adult_Profile_Scoring_Manual.pdf. Accessed July 11, 2024.
15. Cohen S. Perceived stress in a probability sample of the United States. In: S Spacapan, eds. *The Social Psychology of Health*. Newbury Park, CA: Sage Publications; 1988:31-67.
16. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol* 2012;8:23-34.
17. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1:43-46.
18. *Training to Use the MBI:TAC | Mindfulness Teaching Skills*. Bangor: Bangor University. https://mbitac.bangor.ac.uk/training.php.en. Accessed April 13, 2023).
19. Griffith G, Crane R, Baer R, et al. Implementing the mindfulness-based interventions; teaching assessment criteria (MBI:TAC) in mindfulness-based teacher training. *Glob Adv Health Med*. 2021;10:2164956121998340.
20. Evans A, Griffith GM, Crane RS, Sansom SA. Using the mindfulness-based interventions: teaching assessment criteria (MBI:TAC) in supervision. *Glob Adv Health Med*. 2021;10:2164956121989949.
21. Crane R, Koerbel L, Sansom S, Yiangou A. Assessing mindfulness-based teaching competence: good practice guidance. *Glob Adv Health Med*. 2020;9:2164956120973627.
22. Griffith G, Crane R. Introducing the mindfulness-based interventions: teaching and learning companion (The TLC). *Glob Adv Health Med*. 2021;10:21649561211056883.
23. Dimidjian S, Segal ZV. Prospects for a clinical science of mindfulness-based intervention. *Am Psychol*. 2015;70:593-620.