

# Reliability, Validity, and Factor Structure of the Current Assessment Practice Evaluation-Revised (CAPER) in a National Sample

**Aaron R. Lyon, PhD**  
**Michael D. Pullmann, PhD**  
**Shannon Dorsey, PhD**  
**Prerna Martin, MPH**  
**Alexandra A. Grigore, BA**  
**Emily M. Becker, PhD**  
**Amanda Jensen-Doss, PhD**

## Abstract

*Measurement-based care (MBC) is an increasingly popular, evidence-based practice, but there are no tools with established psychometrics to evaluate clinician use of MBC practices in mental health service delivery. The current study evaluated the reliability, validity, and factor structure of scores generated from a brief, standardized tool to measure MBC practices, the Current Assessment Practice Evaluation-Revised (CAPER). Survey data from a national sample of 479 mental health clinicians were used to conduct exploratory and confirmatory factor analyses, as well as reliability and validity analyses (e.g., relationships between CAPER subscales and clinician MBC attitudes). Analyses revealed competing two- and three-factor models. Regardless of the model used, scores from CAPER subscales demonstrated good reliability and convergent and*

---

Address correspondence to Aaron R. Lyon, PhD, Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, 6200 NE 74th Street, Suite 100, Seattle, WA 98115, USA Phone: (206)221-8604 E-mail: lyona@uw.edu.

Michael D. Pullmann, PhD, Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, Seattle, WA, USA.

Alexandra A. Grigore, BA, Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, Seattle, WA, USA.

Shannon Dorsey, PhD, Department of Psychology, University of Washington, Seattle, WA, USA.

Prerna Martin, MPH, Department of Psychology, University of Washington, Seattle, WA, USA.

Emily M. Becker, PhD, Department of Psychology, University of Miami, Coral Gables, FL, USA.

Amanda Jensen-Doss, PhD, Department of Psychology, University of Miami, Coral Gables, FL, USA.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11414-018-9621-z>) contains supplementary material, which is available to authorized users.

*Journal of Behavioral Health Services & Research*, 2018. 1–21. © 2018 National Council for Behavioral Health. DOI 10.1007/s11414-018-9621-z

*divergent validity with MBC attitudes in the expected directions. The CAPER appears to be a psychometrically sound tool for assessing clinician MBC practices. Future directions for development and application of the tool are discussed.*

## Introduction

### Measurement-based care

Measurement-based care (MBC) is increasingly recognized as an evidence-based practice for mental and behavioral health [1] and as one that may qualify as a low-burden, scalable *minimal intervention needed for change*. [2,3] MBC is defined as the use of data collected throughout treatment to drive clinical decisions<sup>3</sup> and is largely synonymous with terms such as “routine outcome monitoring.” [4] Core MBC practices include the following: (a) collection of assessment data, (b) review of the data, and (c) feedback/discussion of the data with clients. [5] Evidence supports the positive impact of MBC practices on adult and youth mental health treatment outcomes. [6–8] MBC is also perceived by clients to be valuable, [9] may enhance communication between therapists and clients, [7] and guides therapy to more rapidly focus on critical issues. [10]

### Evaluating MBC practices

Despite growing attention to MBC in the mental health services literature, studies consistently indicate that MBC practices are used infrequently or suboptimally by community-based providers. [9,11–13] While some providers do not routinely conduct assessments at all, others may collect assessment data, but fail to make use of those data. In one study, even when scored assessments were provided directly to clinicians, results were rarely used for the purposes of monitoring client progress or ongoing treatment planning. [14] Clearly, there remains much to learn about how MBC practices occur in community settings. However, most studies treat assessment practices as a relatively unitary construct, without sufficient attention to the *individual practices* that comprise MBC (e.g., assessment administration, review, client feedback), [5] and without explicit consideration of both standardized and individualized types of assessment (both described in detail below).

To better characterize and understand how clinicians use MBC, valid measures that assess a wider range of MBC practices are needed. Although sophisticated tools have recently emerged to assess clinician attitudes toward different elements of MBC (e.g., use of different assessment types, client feedback processes etc.), [15–17] there are no tools with established psychometrics to evaluate clinician use of MBC practices. Even as increasing numbers of initiatives are being designed to increase use of MBC, [5,18,19] the lack of established tools to measure MBC practices limits MBC research. Additionally, because of the lack of tools, no benchmarks exist for MBC practices in usual care. Such benchmarks would provide a useful point of comparison to identify the need for quality improvement initiatives or to evaluate the impact of MBC implementation efforts. Given literature indicating that attitudes toward MBC practices are important, but frequently insufficient, to predict use of MBC practices, studies are needed that identify predictors of clinician behavior and the relationships between MBC practices and attitudes. [17,20] This paper reports on an effort to advance measurement of MBC practices by developing a standardized and normed tool that can be used to evaluate the different types of assessment (including both standardized and individualized approaches) relevant to providing clinical care. Measure development has been directly informed by contemporary emphases on pragmatic measurement in light of growing interest in pragmatic studies of implementation. [21,22] The development of rigorous, locally relevant, and low-burden (i.e., brief) instruments [23] to evaluate clinician practice also addresses the challenge that many implementation assessment instruments are rarely feasible to use in real-world settings and often have low relevance to stakeholders. [24]

## **Standardized (nomothetic) and individualized (idiographic) assessment**

Assessments conducted in the context of MBC may be either standardized (i.e., nomothetic or inter-individual) or individualized (i.e., idiographic or intra-individual). [25–28] There are compelling reasons to incorporate both into MBC practices. Standardized assessment focuses on information that allows a specific individual's results to be compared to aggregate data across multiple individuals on a construct of interest (e.g., depression symptoms, work functioning). The Patient Health Questionnaire (PHQ-9 [29]) is a prominent example of a standardized instrument. In contrast, individualized assessment focuses on intra-individual variability and is primarily concerned with tracking changes in a specific behavior (e.g., getting into fights) or problem (e.g., not getting along with a spouse) for a specific individual across contexts or time.

The focus of MBC in mental health services has predominantly been on standardized assessment; however, there is growing interest in the application of individualized approaches. [25,27,30] Individualized assessments tend to yield data that are more specific than their standardized counterparts and may be more likely to focus on issues that align with client priorities, enhance client engagement, and fit into everyday practice. [31] A series of recent studies have documented that both clinicians and clients may value individualized assessment as much, or more than, standardized assessment. [32–34] In addition, recent meta-analysis results indicate that individualized assessment targets may demonstrate larger effect sizes than standardized assessments in response to treatment. [35] Given the benefits of individualized and standardized approaches for informing clinical practice and measuring client progress, use of both approaches may provide the best compromise of rigor and relevance. [26,27]

Aside from clinician attitudes, there is little known about the variables that are differentially associated with use of each type of assessment. Some prior work has evaluated general practitioner characteristics (e.g., gender, age, experience) and practice characteristics (e.g., caseload demographics, practitioner role in conducting intakes, etc.) as they relate to the implementation of evidence-based practices more generally, [36,37] and assessment practices in particular. [11,38] Virtually no work, however, has evaluated how these characteristics—or others of potential importance—relate to standardized and individualized approaches.

## **Current Assessment Practice Evaluation-Revised (CAPER)**

Development of psychometrically valid measures has been noted as a critical issue in implementation science. [24] The Current Assessment Practice Evaluation (CAPE [19]) was developed to characterize clinicians' use of MBC practices. The CAPE and the revised version evaluated in this study (CAPER) appear to be the only tools designed specifically for this goal. Prior to revision, the CAPE assessed MBC practices across different phases of intervention including at intake, during treatment, and at treatment termination, but was focused exclusively on standardized assessments. The CAPE captures use of standardized assessment tools, providing assessment-based feedback to clients (e.g., adults, children, or families), incorporating assessment results into treatment planning, and using ongoing assessment to inform changes to treatment plans. In one study, the CAPE was used to evaluate clinicians' use of standardized assessment tools following participation in a statewide child and adolescent-focused cognitive-behavioral therapy (CBT) training and consultation program that emphasized routine assessment. [19] Clinician attitudes toward assessment were associated with CAPE scores, and CAPE scores significantly improved from pre-training to post-consultation. Another study used the CAPE to examine MBC practices among school-based mental health clinicians [34] and found that, on average, clinicians engaged in MBC practices with less than half of their caseload, were more likely to administer assessments at intake, and least likely to adjust treatment plans based on assessment results. Although prior research using the CAPE has documented acceptable inter-item reliability, [19] no large-scale evaluations of the instrument have been conducted.

The goal of revising the original CAPE measure was to include individualized assessment approaches while maintaining the brief self-report format. The purpose of the current study was to evaluate the internal reliability, convergent and discriminant validity, and factor structure of therapist scores on the CAPER tool. The CAPER and associated measures were administered to a large, national sample of clinicians in the USA.

## Method

### Participants

The participants for this study were mental health professionals recruited through mailing lists from three national professional organizations (American Mental Health Counselors Association, American Association for Marriage and Family Therapy, National Association of Social Workers). These organizations were prioritized based on evidence indicating that masters-level clinicians tend to provide the bulk of mental health services. [39] Contact information for a random, nationally representative sample of members who work as mental health clinicians was provided by each organization. The survey was mailed to 1200 providers (400 from each professional organization) with a pre-addressed, stamped envelope. From these 1200 individuals, 15 could not be reached due to undeliverable addresses, 104 declined participation, and 461 did not respond. Of the 620 responders (52.3%—consistent with prior applications of this approach), [40] 115 were excluded because they did not actively conduct or supervise therapy. One respondent only held a bachelor's degree and was also excluded, yielding a sample of 504. An additional 25 participants were excluded from the current analyses due to missing data on the entirety of the CAPER, yielding a final sample of 479 clinicians. Representativeness of the sample was supported by geographic and demographic information. Geographically, surveys were returned from 46 out of 50 states in the USA, with the four most populous states (California, Texas, Florida, New York) among the five states with the most survey respondents (Florida, California, New York, Washington, Texas). Consistent with other national samples, [41] this sample was primarily female (73.1%) and Caucasian (82.7%), with a mean age of 56 and age range from 28 to 82 years. Participants were generally masters-level clinicians (84.8%). Table 1 details the demographic, professional, and practice characteristics of participants.

### Procedures

The CAPER and other measures used to establish its validity (see below) were included as part of a larger survey developed to assess different facets of clinician perspectives on the use of assessment in practice. [15] Using the Tailored Design Method, [42] the survey was developed, piloted, and iteratively revised with a sample of six mental health providers to determine readability and comprehension. These participants provided feedback and ideas for improving the survey during a semi-structured interview. The survey was revised in an iterative fashion throughout these interviews before it was distributed broadly.

All study procedures were approved by the Institutional Review Board at the coordinating site. Participants received up to four separate mailings. The first, sent to all 1200 providers, consisted of a personally addressed, hand-signed, pre-notice letter informing them of the upcoming survey. The second, also sent to all participants, included a personalized cover letter, non-contingent \$2 bill reinforcement, [40] survey, and pre-addressed, stamped return envelope. The third was a signed postcard thanked respondents and reminded non-respondents to return the survey. The fourth was sent to non-respondents only and included a personalized cover letter, a copy of the survey, and a stamped return envelope.

**Table 1**Sample descriptive characteristics for study participants ( $n = 479$ )

Variable	<i>N</i>	%
Sex		
Male	126	26.3
Female	350	73.1
Missing	3	0.6
Race		
Caucasian	396	82.7
Hispanic/Latino	15	3.1
Black/African American	16	3.3
Asian/Pacific Islander	6	1.3
Mixed/other	9	1.9
Missing	37	7.7
Work setting <sup>1</sup>		
Private practice	295	61.6
Community agency or other outpatient clinic	82	17.1
School—elementary, middle or high	19	4.0
Higher education setting	9	1.9
Hospital or medical center or other inpatient location	26	5.4
Day treatment facility	3	0.6
Residential facility or group home	3	0.6
Other work environment	19	4.0
Missing	44	9.1
Highest degree obtained		
Doctoral level	70	14.6
Masters level	406	84.8
Missing	3	0.6
	<i>M</i>	<i>SD</i>
Age (range 28–82)	56	11.7

<sup>1</sup>Adds to more than 100% because participants could choose multiple settings

## Measures

*Demographic, professional, and practice characteristics* Participants completed open-ended items describing their age, gender, ethnicity, work setting, and theoretical orientation. For analysis purposes, work setting was coded as Private Practice = 1, Other = 0, and theoretical orientation was coded into Cognitive-Behavioral = 1, Other = 0. Participants were also provided a range of degree options and asked to check all that apply; highest degrees were grouped into Master's (0) and Doctoral (1) for analysis. Clinicians also indicated whether seeing children, adults, ethnic minority clients, and low-income clients was a major part of their work, a minor part, or not at all a part of their work. These variables were coded as Major Part = 1, Minor or Not at all = 0 for analysis. Amount of work with adults was excluded from analysis due to low variability (92% reported that it was a major aspect of their work). Participants also indicated their roles (i.e., whether they conducted intakes, supervised intakes, supervised therapy), and whether they routinely collected child, adult, and family assessments.

*Current Assessment Practice Evaluation-Revised (CAPER)* The CAPER includes the original four-item CAPE measure [19] plus three new items to assess use of individualized assessment. The CAPER assesses clinician self-reported use of the full range of MBC practices including administration of individualized/standardized assessments at different points during treatment (e.g., "...with what percentage of your total caseload did you administer a standardized assessment measure?"), client feedback on assessment results (e.g., "...with what percentage of your clients did you give feedback about/discuss a systematically-tracked individualized outcome variable with your client or your client's family/caregiver?"), and using assessment results to inform treatment (e.g., "with what percentage of your clients did you alter or change your specific plan or activities for a single session based on standardized or individualized assessment data?"). Importantly, to improve the validity of reporting, the CAPER focuses on well-operationalized MBC practices and assesses the extent to which clinicians have engaged in them in a recent, specified period of time (i.e., the last month [for items focused on assessments at intake] or the last week [for all other MBC practices]). All items are scored on a 1–4 scale, with anchors signifying percent of clients they used the practice with: "1 = None (0%), 2 = Some (1–39%), 3 = Half (40–60%), 4 = Most (61–100%)". The original CAPE has previously been applied in samples of community- and school-based clinicians [19,34] and scores demonstrated acceptable inter-item reliability ( $\alpha = 0.72$ ) and sensitivity to change. The CAPER is included in Appendix 1.

*Attitudes toward Standardized Assessment Scale-Monitoring and Feedback (ASA)* The ASA used in the current project is a modification of the original Attitudes toward Standardized Assessment Scales. [17] The revised 18-item instrument focuses explicitly on the use of standardized progress measures in treatment. Standardized measures are defined as "client self-report measures with standard items and scoring procedures." Items are rated on a scale from 1 (strongly disagree) to 5 (strongly agree) and load on three factors: ASA Clinical Utility (8 items), ASA Treatment Planning (5 items), and ASA Practicality (5 items). Prior research using the current sample has supported the factor structure of the measure and indicated acceptable reliability for all subscale scores ( $\alpha = .81-.85$ ). [16]

*Attitudes toward Individualized Assessment Scale-Monitoring and Feedback (AIA)* The AIA includes items from the ASA, modified for the larger project that contained the current study to reference "individualized assessment" instead of "standardized assessment." Individualized assessment is defined as "client-specific measures of treatment progress, such as having clients rate the severity of self-identified problems on a scale from 1-10, tracking number of self-injury incidents in the past week, or tracking frequency of office discipline referrals for a child client." The AIA factor structure mirrors that of the ASA (above) and includes three subscales: AIA Clinical Utility, AIA Treatment Planning, and AIA Practicality. Alphas for each subscale range from  $\alpha = .84$  to  $\alpha = .89$ . [15]

*Monitoring and Feedback Attitudes Scale (MFA)* The MFA [15] assesses clinician attitudes toward key MBC practices, including routine progress monitoring—defined as "administering measures to your therapy clients every 1-2 sessions in order to monitor their progress in treatment"—and providing feedback to clients about treatment progress—defined as "discussing data that have been gathered through routine progress monitoring with clients." The MFA was developed and evaluated as a component of the larger research project mentioned above. Fourteen items are rated on a scale from 1 (strongly disagree) to 5 (strongly agree) and load on two subscales: MFA Benefit (10 items) and MFA Harm (4 items). Internal consistencies for both subscales are acceptable (MFA Benefit  $\alpha = .87$ , MFA Harm  $\alpha = .87$ ). [15]

## Analyses

Inter-item correlations were run among all items. Split-half randomization was used to divide the sample prior to conducting principal components analysis (PCA) and exploratory factor analysis (EFA) with one half and confirmatory factor analysis (CFA) with the second half. PCA was chosen because it maximizes item variance in describing the internal structure of the data. To ensure randomization adequacy, the split-half samples were compared using *t* tests of CAPER item scores. Number of components for the EFA was chosen through running multiple models and using Kaiser's stopping rule, an examination of the scree chart/% of cumulative variance for a natural break, and examining rotated loading matrices for number of non-trivial factors. [43] Varimax rotation with Kaiser normalization was used to interpret the PCA. In order to confirm PCA findings, three additional EFAs were computed using maximum likelihood, generalized least squares, and principal axis factoring extraction methods. The CFA was computed using MPlus, and fit was evaluated using the model comparative fit index (CFI; values greater than .95 are acceptable), Tucker-Lewis Index (TLI; values greater than .95 are acceptable), [44,45] and the root mean square error of approximation (RMSEA values less than .07 are acceptable). [46] The fit of a model specifying two factors was compared with a model specifying three factors using a likelihood ratio deviance test.

For the EFA and CFA, cases with missing or not applicable values were retained in analyses and only excluded during pairwise computations. This approach was taken instead of an alternative, such as listwise elimination or imputation, for two reasons. First, six of the seven items on the CAPER had very limited missing data (<1.3%), so imputation would have negligible impact. Second, the item with a large amount of missing data (item 1, missing  $n = 43$ , 8.9%) was not missing at random, because respondents who reported "not applicable" had not had any new cases in the prior month. Therefore, imputation was not appropriate, and listwise elimination would have removed important information from the analyses. Pairwise elimination permitted the inclusion of these cases while avoiding artificially inflated estimates.

Correlations among indicated factors were then computed. Factor internal consistency was computed using Cronbach's  $\alpha$ . To test the validity of the measure scales, analyses using *t* tests and Pearson correlations were conducted. It was anticipated that there would be no relationship between the CAPER and therapist gender or age, but that participants who are more involved in intakes and those who regularly collect assessments would have higher CAPER scale scores. It was anticipated that all CAPER scales would be positively correlated with all ASA and AIA scales and the MFA benefit scale, and negatively correlated with the MFA harm scale. No predictions were made about possible relationships between CAPER scale scores and years of experience or caseload characteristics (portion of work with children, ethnic minorities, and low-income clients), but these relationships were examined in exploratory analyses.

## Results

Item means and standard deviations for the CAPER are displayed in Table 2. Average scores for all items ranged from 1.5 to 2.2, indicating that clinicians generally applied MBC practices with well under half of their caseloads. The highest rated item was "Percent of clients given feedback about systematically tracked individualized outcome variable in last week" and the lowest item being "Percent of clients whose overall treatment plan was altered based on either assessment data in last week." Standard deviations ranged from 0.7 to 1.3, with the item with the most variability being "Percent of intakes administered standardized assessments during first 1-2 meetings in last month" and the item with the least variability being "Percent of clients whose overall treatment plan was altered based on either assessment data in last week." A crosstab comparison of CAPER Item 2 (use of standardized assessment) and Item 4 (use of individualized assessment) revealed that

**Table 2**

Current assessment practice evaluation-revised subscales, and individual items

Item	Mean <sup>1</sup>	SD	$\alpha$	Principal components analysis		Confirmatory factor analysis
				PCA component 1: Individualized Assessment and Treatment Modification <i>M</i> = 1.74 <i>SD</i> = 0.87 $\alpha$ = .82	PCA component 2: Standardized Assessment <i>M</i> = 1.84 <i>SD</i> = 0.73 $\alpha$ = .80	
<b>CFA Factor 1: Standardized Assessment</b>						
Item 1: % intakes administered standardized assessments during first 1–2 meetings in last month	1.74	0.87	.82			
Item 2: % caseload administered standardized assessment measure in last week	2.12	1.26		.031	.813	.764
Item 3: % clients given feedback about standardized measure from client or family in last week	1.56	0.85		.214	.884	.810
<b>CFA Factor 2: Individualized Assessment Subscale</b>						
Item 4: % caseload systematically tracked individualized outcome variable in last week	1.60	0.89		.228	.828	.838
Item 5: % clients given feedback about systematically tracked individualized outcome variable in last week	2.12	1.03	.82		.052	.791
Item 6: % clients whose overall treatment plan altered based on either assessment data in last week	2.08	1.15		.808	.016	.897
<b>CFA Factor 3: Treatment Modification Subscale</b>						
Item 7: % specific plan for single session altered due to either assessment data in last week	2.15	1.09		.891		
Item 8: % clients whose overall treatment plan altered based on either assessment data in last week	1.57	0.61	.72	.643	.354	.789
Item 9: % specific plan for single session altered due to either assessment data in last week	1.51	0.66		.692	.285	.749

Italicized values indicate major loadings

<sup>1</sup>CAPER item values 1 = None (0%), 2 = Some (1–39%), 3 = Half (40–60%), 4 = Most (61–100%)



70.7% of those who used a standardized measure in the last week also used an individualized measure, while 46.5% of those who used an individualized measure also used a standardized measure.

Initial analyses indicated that the data were suitable for factor analysis. Items had sufficiently normal distributions based on skewness and kurtosis statistics. Assessment of inter-item correlations revealed a high presence of coefficients .3 and above. The Kaiser-Meyer-Olkin value was .72, exceeding the suggested value of .6, [<sup>47,48</sup>] and significance was reached in Bartlett's test of sphericity ( $\chi^2_{(21)} = 613.22, p < .001$ ). Items were multivariate normal and scatterplots indicated that relationships among variables were linear. Inter-item correlations were then computed, but are described later to facilitate interpretation of the subscales extracted from the EFA and CFA analyses. No statistically significant differences between the two split-half randomized groups on CAPER scores were found.

### Exploratory factor analysis

A PCA was computed on half of the sample ( $n = 241$ , 4 items had no missing cases, 2 items had 1 missing case, 1 item had 3 missing cases, and 1 item had 25 "not applicable" cases). The PCA uncovered two components, explaining 46.7 and 21.5% of the variance after extraction, respectively, accounting for a total of 68.2% of the variance after extraction. Two components were retained for further investigation due to three reasons: (1) assessment of the scree plot revealed a point of inflection at the third factor, (2) inspection of the rotated factor loadings matrix showed strong loadings on 2 components, and (3) adding a third component would only have added 10.7% additional variance explained, which would have decreased the efficiency of this seven-item measure. Table 2 presents the component loading matrix. Component 1 consisted of four variables pertaining to tracking individualized outcomes and altering treatment based on assessment data. Component 2 consisted of three variables related to the administration of standardized assessment measures. Results from additional EFAs using maximum likelihood, generalized least squares, and principal axis factoring extraction methods were highly consistent with the PCA and are therefore not described here.

Therefore, results of the PCA and EFAs supported the formation of two subscales among the 7 items of the CAPER "Standardized Assessment" ( $M = 1.74, SD = .87$ , skewness = 1.01, kurtosis =  $-.01$ ), and "Individualized Assessment and Treatment Modification" ( $M = 1.84, SD = .73$ , skewness = .63, kurtosis =  $-.45$ ). A paired samples  $t$  test revealed that ratings of usage of Individualized Assessment and Treatment Modification was significantly greater than Standardized Assessment ( $t_{478} = -2.36, p = .019$ ).

### Confirmatory factor analysis

A two-factor CFA using maximum likelihood extraction was computed on half of the sample ( $n = 241$ , 3 items had no missing cases, 2 items had 1 missing case, 2 items had 3 missing cases, and 1 item had 18 "not applicable" cases), specifying items into the two factors uncovered in the EFA. Model fit was considered acceptable from some measures and borderline or poor from others (CFI = .93; TLI = .89; RMSEA = .12). Because fit was not strong across all fit statistics, the theoretical rationale supported a three-factor model, and inter-item correlation tables were consistent with this theoretical rationale, a three-factor CFA was computed that separated the second factor into two factors; items 1, 2, and 3 were specified as one factor, 4 and 5 as a second factor, and 6 and 7 as a third factor. Model fit was acceptable for all indices (CFI = .98; TLI = .97; RMSEA = .068). Likelihood ratio deviance tests revealed that the three-factor model was a significantly better fit than the two-factor model (difference tests:  $\chi^2[2] = 35, p < .001$ ; Akaike

Information Criterion  $AIC_2 = 31.02$ ,  $p < .001$ ; Bayesian information criterion  $BIC_2 = 24.05$ ,  $p < .001$ ).

Therefore, results of the CFA appear to be an elaboration of the EFA results, and supported the formation of three subscales, “Standardized Assessment” (the same factor described in the EFA,  $M = 1.74$ ,  $SD = .87$ , skewness = 1.01, kurtosis =  $-.01$ ), and two 2-item sub-factors of the EFA’s “Individualized Assessment and Treatment Modification”, split into “Individualized Assessment” ( $M = 2.12$ ,  $SD = 1.03$ , skewness = .60, kurtosis =  $-.86$ ) and “Treatment Modification” ( $M = 1.57$ ,  $SD = 0.61$ , skewness = 1.05, kurtosis = 1.17). A Repeated Measures ANOVA indicated that the mean scores on these three factors significantly differed ( $F_{2,956} = 82.61$ ,  $p < .001$ ). Therefore, clinicians report most often tracking and providing feedback about individualized assessments ( $M = 2.12$ ), followed by standardized assessments ( $M = 1.74$ ) and modifying treatment based on assessment data ( $M = 1.57$ ).

Based on the results of the EFA and CFA, the data were analyzed (below) examining four scores: two primary scales (Standardized Assessment scale and Individualized Assessment and Treatment Modification scale) and two subscales from the second scale (Individualized Assessment subscale and Treatment Modification subscale).

### Internal consistencies and correlations

Inter-item correlations for the entire sample are presented in Table 3. All items were significantly correlated, and items within the two primary scales were correlated ( $r > .40$ ). There was a weak positive correlation between the two primary scale total scores ( $r = .35$ ,  $p < .01$ ), indicating that only 12% of the variance in one scale was explained by the other. This suggests a lack of unidimensionality in the overall items and contraindicates the use of a “total score” for the measure. The two subscales were correlated at  $r = .55$  ( $p < .001$ ), providing sufficient overlap to suggest that they could be combined (consistent with the EFA results), but sufficient difference to suggest they could be separated (consistent with the CFA results).

Cronbach’s alpha obtained from the item scores for the Standardized Assessment scale was .82, and the Individualized Assessment and Treatment Modification scale was .80. The alpha did not improve with the removal of any item from either CAPER scale. The lowest and highest item-to-total correlations for the CAPER Individualized Assessment and Treatment Modification scale items were .57 and .74, respectively. The lowest and highest item-to-total correlations for the CAPER Standardized Assessment scale items were .65 and .74, respectively. The subscales also had acceptable alphas (Individualized Assessment  $\alpha = .82$ , Treatment Modification  $\alpha = .72$ ). Because these subscales only had two items each, alpha if item removed could not be calculated and item-total correlations were approximately the same as item-item correlations shown in Table 3.

### Convergent and divergent validity

Tables 4 and 5 depict the results from a series of  $t$  tests analyzing the mean CAPER scale and subscale scores associated with a variety of categorical variables that describe the therapists’ involvement in intakes and collection of assessments for children, adults, and families at intake, during treatment, and at termination. Nearly, all tests were statistically significant in the anticipated direction for the CAPER primary scale and minor subscale scores: in general, participants had higher scores if they conducted intakes, supervised intakes, and reported collecting assessments at intake, during treatment, and at termination. On the Standardized Assessment scale, participants had higher scores if their workplace dictated use of assessments and lower scores if they worked in private practice. On the other hand, participants had higher scores on the Individualized Assessment and Treatment modification scale and the Individualized Assessment minor subscale if their primary theoretical orientation was CBT.

**Table 3**  
CAPER Inter-item Pearson correlations

Item	Standardized Assessment Scale			Individualized Assessment and Treatment Modification Scale			
	1	2	3	4	5	6	7
2	.62	1					
3	.59	.71	1				
4	.22	.28	.26	1			
5	.22	.22	.26	.71	1		
6	.30	.34	.35	.40	.53	1	
7	.28	.27	.31	.44	.49	.55	1

All correlations were significant at the 0.01 level (two-tailed). To ease interpretation, all values > .4 are italicized

Table 6 depicts correlations between CAPER scales and ASA scales, AIA scales, and MFA scales. All CAPER scales were significantly correlated with all scales on the ASA, AIA, and MFA in anticipated directions, with negative relationships for MFA harm scale and positive relationships with all other scales. Consistent with convergent validity, most correlations were significantly higher between Standardized Assessment and ASA subscales than AIA subscales. Results of Williams' *t* tests for comparing the difference in magnitude between two non-independent correlations are depicted in the table as matching superscript letters when significant. Results indicated that the CAPER Standardized Assessment scales had significantly stronger correlations with all of the ASA scales when compared to their parallel AIA scales (Clinical Utility  $t = 2.2, p = .024$ ; Treatment Planning  $t = 2.6, p = .010$ ; Practicality  $t = 3.6, p < .001$ ). The CAPER Individualized Assessment and Treatment Modification scale had a significantly higher correlation with the AIA vs. ASA Clinical Utility scale ( $t = -3.1, p = .001$ ) a statistically borderline higher correlation with the AIA vs. ASA Treatment Planning scale ( $t = -1.9, p = .052$ ), but no significant difference between the correlations with AIA vs. ASA Practicality ( $t = -1.6, p = .111$ ). The Individualized Assessment minor subscale had significantly higher correlation with the AIA vs. ASA Clinical Utility scales ( $t = -3.3, p < .001$ ) and Practicality scales ( $t = -2.4, p = .016$ ), and a statistically borderline higher correlation with the AIA vs. ASA Treatment Planning scale ( $t = -2.0, p = .051$ ). Though the Treatment Modification subscale is not specific to individualized assessments or standardized assessments, there was a significantly higher correlation between Treatment Modification and AIA vs. ASA Clinical Utility scale ( $t = -2.1, p = .034$ ); however, there were no significant differences between Treatment Modification and AIA vs. ASA Treatment Planning ( $t = -1.2, p = .216$ ) or Practicality ( $t = .4, p = .704$ ). Supportive of both convergent and divergent validity, all CAPER scales were significantly positively related to the MFA benefit scale and negatively correlated with the MFA Harm scale.

### Exploratory analyses

There were no significant associations between the CAPER scales and gender, age, years of experience, or whether they provided supervision. Table 6 indicates that portion of work with

**Table 4**  
*t* tests for convergent validity for primary CAPER scales

Variable	<i>n</i>	CAPER Scale			Standardized Assessment			Individualized Assessment and Treatment Modification		
		<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	
Gender										
	Female	1.70	.84	-1.86 (477)	.06	1.81	.73	-1.71 (477)	.09	
	Male	1.86	.91			1.94	.73			
Conducts intakes	Yes	1.84	.88	-11.44 (255)	<.001	1.91	.73	-5.09 (97)	<.001	
	No	1.16	.33			1.47	.61			
Supervises intakes	Yes	2.14	.96	-5.22(160)	<.001	2.06	.80	-3.59 (164)	<.001	
	No	1.61	.79			1.76	.68			
Supervises therapy	Yes	1.83	.93	-1.59 (441)	.11	1.88	.76	-1.01 (441)	.31	
	No	1.69	.83			1.80	.69			
Never collects child assessment	Yes	1.58	.79	3.51 (458)	<.001	1.67	.69	4.54 (467)	<.001	
	No	1.85	.89			1.97	.72			
Child assessment at intake	Yes	1.93	.92	-4.28 (414)	<.001	1.99	.72	-4.32 (467)	<.001	
	No	1.59	.79			1.71	.70			
Ongoing child assessment	Yes	1.85	.89	-2.36 (467)	.02	2.00	.74	-4.20 (467)	<.001	
	No	1.66	.84			1.72	.68			
Child assessment at termination	Yes	2.07	1.02	-3.80 (135)	<.001	2.24	.73	-6.47 (467)	<.001	
	No	1.65	.80			1.73	.68			
Never collects adult assessment	Yes	1.27	.64	4.46 (47)	<.001	1.35	.63	4.37 (446)	<.001	
	No	1.78	.87			1.89	.72			
Adult assessment at intake	Yes	1.85	.89	-4.89 (271)	<.001	1.90	.72	-2.82 (446)	.01	
	No	1.45	.72			1.69	.73			
Ongoing adult assessment	Yes	1.79	.87	-1.72 (445)	.09	1.93	.72	-3.87 (445)	<.001	
	No	1.63	.86			1.64	.69			
Adult assessment at termination	Yes	1.94	.95	-3.92 (337)	<.001	2.05	.74	-5.19 (446)	<.001	

**Table 4**  
(continued)

Variable	n	CAPER Scale		Standardized Assessment			Individualized Assessment and Treatment Modification		
		M	SD	t(df)	p	M	SD	t(df)	p
Never collects family assessment	No	1.60	.77			1.70	.68		
	Yes	1.48	.71	4.32 (256)	<.001	1.56	.63	5.27 (420)	<.001
Family assessment at intake	No	1.85	.91			1.96	.72		
	Yes	1.89	.91	-3.65 (420)	<.001	1.99	.73	-4.22 (420)	<.001
Ongoing family assessment	No	1.59	.80			1.70	.68		
	Yes	1.81	.93	-1.50 (420)	.14	2.02	.74	-4.86 (420)	<.001
Family assessment at termination	No	1.69	.82			1.69	.66		
	Yes	2.04	.98	-3.32 (125)	<.001	2.11	.77	-3.87 (420)	<.001
Workplace dictates use of assessment	No	1.67	.82			1.78	.69		
	Yes	1.95	.91	-4.72 (467)	<.001	1.90	.72	-1.69 (467)	.091
CBT theoretical orientation	No	1.58	.79			1.79	.71		
	Yes	1.83	.88	-1.85 (448)	.052	1.95	.73	-2.71 (448)	.007
Works in private practice	No	1.67	.85			1.77	.71		
	Yes	1.67	.84	3.31 (434)	<.001	1.78	.73	1.92 (434)	.055
	No	1.96	.91			1.93	.68		

Italicized *p* values indicate significant findings at  $p < .05$

**Table 5**  
*t* tests for convergent validity for CAPER minor subscales

Variable	<i>n</i>	CAPER subscale				Individualized Assessment				Treatment Modification			
		CAPER subscale		Individualized Assessment		Individualized Assessment		Treatment Modification		Individualized Assessment		Treatment Modification	
		<i>M</i>	<i>SD</i>	<i>t</i> ( <i>df</i> )	<i>p</i>	<i>M</i>	<i>SD</i>	<i>t</i> ( <i>df</i> )	<i>p</i>	<i>M</i>	<i>SD</i>	<i>t</i> ( <i>df</i> )	<i>p</i>
Gender	352	2.08	1.03	-1.30 (477)	.19	1.54	.61	-1.84 (477)	.06				
	Female												
	Male	127	2.23	1.03		1.65	.62						
Conducts intakes	405	2.20	1.03	-3.79 (469)	<.001	1.62	.63	-5.70 (118)	<.001				
	Yes												
	No	66	1.69	.94		1.27	.42						
Supervises intakes	110	2.37	1.05	-3.08 (439)	<.001	1.75	.71	-3.45 (154)	<.001				
	Yes												
	No	331	2.02	1.00		1.49	.54						
Supervises therapy	166	2.14	1.03	-.59 (441)	.55	1.60	.65	-1.36 (308)	.17				
	Yes												
	No	277	2.08	1.01		1.52	.56						
Never collects child assessment	205	1.87	.98	4.68 (467)	<.001	1.47	.58	2.95 (467)	<.001				
	Yes												
	No	264	2.31	1.02		1.63	.61						
Child assessment at intake	210	2.32	1.01	-3.91 (467)	<.001	1.67	.62	-3.65 (467)	<.001				
	Yes												
	No	259	1.96	1.00		1.47	.58						
Ongoing child assessment	200	2.35	1.03	-4.34 (467)	<.001	1.65	.63	-2.74 (467)	<.001				
	Yes												
	No	269	1.95	.99		1.49	.58						
Child assessment at termination	101	2.64	1.03	-5.79 (153)	<.001	1.83	.66	-5.14 (467)	<.001				
	Yes												
	No	368	1.98	.97		1.49	.57						
Never collects adult assessment	36	1.49	.94	3.99 (446)	<.001	1.21	.40	5.13 (51)	<.001				
	Yes												
	No	412	2.19	1.02		1.56	.61						
Adult assessment at intake	325	2.19	1.00	-1.89 (446)	.05	1.62	.61	-3.51 (446)	<.001				
	Yes												
	No	123	1.98	1.09		1.39	.55						
Ongoing adult assessment	314	2.27	1.03	-4.25 (445)	<.001	1.59	.60	-2.07 (445)	.03				
	Yes												
	No	133	1.82	.97		1.47	.61						
Adult assessment at termination	183	2.39	1.06	-4.57 (367)	<.001	1.70	.62	-4.42 (446)	<.001				
	Yes												
	No	265	1.95	.96		1.45	.57						

**Table 5**  
(continued)

Variable	<i>n</i>	CAPER subscale		Individualized Assessment			Treatment Modification		
		<i>M</i>	<i>SD</i>	<i>t</i> ( <i>df</i> )	<i>p</i>	<i>M</i>	<i>SD</i>	<i>t</i> ( <i>df</i> )	<i>p</i>
Never collects family assessment	Yes	1.74	.90	5.33 (226)	<.001	1.38	.53	4.29 (237)	<.001
	No	2.28	1.02			1.64	.62		
Family assessment at intake	Yes	2.29	1.02	-3.54 (420)	<.001	1.68	.64	-4.11 (420)	<.001
	No	1.95	.99			1.44	.54		
Ongoing family assessment	Yes	2.38	1.02	-5.02 (414)	<.001	1.66	.65	-3.09 (400)	<.001
	No	1.90	.95			1.48	.55		
Family assessment at termination	Yes	2.48	1.03	-3.64 (420)	<.001	1.73	.71	-2.58 (121)	.01
	No	2.04	.99			1.52	.57		
Workplace dictates use of assessments	Yes	2.19	1.03	-1.54 (467)	.123	1.60	.62	-1.25 (467)	.211
	No	2.04	1.01			1.53	.59		
CBT theoretical orientation	Yes	2.29	1.07	-2.80 (402)	.005	1.62	.60	-1.61 (448)	.107
	No	2.01	.99			1.52	.60		
Works in private practice	Yes	2.04	1.03	1.86 (434)	.063	1.53	.59	1.33 (434)	.280
	No	2.23	.98			1.61	.61		

Bolded *p* values indicate significant findings at  $p < .05$

**Table 6**

Pearson correlations for CAPER scales and subscales and participant work experience

	<i>n</i>	Standardized Assessment	Individualized Assessment and Treatment Modification	Individualized Assessment	Treatment Modification
		<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
ASA utility	454	.37** <sup>a</sup>	.234** <sup>d</sup>	.20** <sup>c</sup>	.226** <sup>g</sup>
ASA treatment planning	459	.38** <sup>b</sup>	.206**	.18**	.194**
ASA practicality	461	.38** <sup>c</sup>	.307**	.28** <sup>f</sup>	.267**
AIA utility	455	.29** <sup>a</sup>	.344** <sup>d</sup>	.31** <sup>c</sup>	.298** <sup>g</sup>
AIA treatment planning	466	.29** <sup>b</sup>	.266**	.24**	.235**
AIA practicality	462	.26** <sup>c</sup>	.360**	.36** <sup>f</sup>	.261**
MFA benefit	454	.17**	.314**	.32**	.211**
MFA harm	465	-.17**	-.209**	-.23**	-.111**
Participant age	469	-.09	-.01	.00	-.02
Years of full-time experience	467	-.09	.06	.07	.01
What portion of therapists' work is with...					
Children	466	.06	.11**	.11*	.08
Ethnic minorities	469	.24**	.18**	.17**	.13*
Low-income clients	475	.19**	.15**	.13*	.13*

\* $p < .01$ ; \*\* $p < .001$ a, b, c, d, e, f, g Williams *t* tests indicate significant differences between the magnitude of correlations with matching superscript letters at  $p < .05$  (only the ASA, AIA, and CAPER scale correlations were tested)

children was positively associated with the Individualized Assessment and Treatment Modification scale and the Individualized Assessment minor subscale. Work with ethnic minorities and with low-income clients were positively associated with all scales and subscales.

## Discussion

The current study assessed the psychometrics of the CAPER instrument, the first standardized measure of mental health clinician MBC practices. Analyses revealed a lack of unidimensionality for the tool, suggesting that the use of a total score is likely contraindicated. Exploratory and confirmatory factor analyses provided support for two distinct, but related, factor structures; a two-factor solution and a three-factor solution in which one of the two factors (Individualized Assessment and Treatment Modification) was split into two sub-factors. Regardless of the solution, scores on all CAPER factors were found to have adequate to good inter-item reliabilities. Further analyses of both sets of factors yielded evidence for convergent and divergent validity. As predicted, measures of attitudes toward standardized assessment were most strongly related to the CAPER Standardized Assessment scale. Measures of individualized assessment attitudes were



most strongly related to the CAPER Individualized Assessment and Treatment Modification scales. Additionally, all CAPER scales were positively related to measures of clinician attitudes indicating MBC practices were beneficial, and negatively related to those indicating that MBC practices were harmful. Furthermore, overall use of both kinds of assessment was low, with average ratings indicating that respondents reported using standardized and individualized with well under 40% of their caseloads. The implications of the identified two- and three-factor solutions, the relevance of the findings to advancing measurement of both standardized and individualized assessment approaches, and the importance of the identified correlates of MBC practices are described below.

### **Two- and three-factor solutions**

Although the two-factor solution was supported by the EFA, the CFAs revealed somewhat better model fit statistics for the three-factor solution. The three-factor solution is also more consistent with the theoretical underpinnings of the project, which suggest that individualized and standardized assessments reflect distinct practices [15,27] and that there is a meaningful difference between conducting initial assessments and incorporating them into intervention planning. [9,14,19] Nevertheless, the two-factor solution is more consistent with contemporary measurement conventions, [43] which recommend at least three items per subscale.

Based on these results, both the two- and three-factor CAPER factor structures have merit, and either one may be used with an understanding of their specific benefits and limitations. For example, the three-factor solution, due to its greater level of specificity, may ultimately yield more actionable information when evaluating clinicians' MBC practices as part of an effort to improve service quality. Given the persistent gap in community practice between assessment administration and incorporation of the results into services, separate evaluation of those practices is likely to be most informative when attempting to identify specific areas of MBC practice that need more support. Future revisions of the CAPER tool may also consider incorporating additional items related to individualized assessment (e.g., explicit items about clinicians' review of the data in supervision or prior to delivering feedback to clients) and treatment modification (e.g., items specific to altering treatment duration based on assessment results) to increase confidence in the stability of the factor solution and potentially enhance subscale variability to improve its sensitivity to change.

### **Standardized and individualized assessment**

The results uphold important distinctions between standardized and individualized assessment practices, with the use of one type of assessment often occurring somewhat independently of the other. Although a majority (70.7%) of those using standardized tools also reported using individualized ones, the reverse was not true (i.e., 46.5% of those who used an individualized measure in the past week also used a standardized measure). Emerging evidence suggests that they may differ in their intuitive appeal to clinicians and service recipients alike. [33,34] It is therefore unsurprising that clinicians across sectors report using individualized methods more frequently than standardized ones. [15,32] Understanding and independently measuring clinician use of both types of assessment practices—and how clinicians use each assessment type to inform treatment planning and adaption—may be important for specific studies. With its two- and three-factor solutions, the CAPER provides some flexibility depending on the research question.

### **Characteristics associated with use of MBC practices**

To evaluate correlates of use and explore the functioning of the measure, the current study also identified multiple clinician and practice characteristics that were associated with the use of MBC practices (as measured using either the two- or three-factor solutions). Specifically, conducting or

supervising intakes, working with children, and working with ethnic/racial minority or low-income clients were all associated with one or more CAPER subscales. Among these, participation in frequent intakes may have the most straightforward connection to MBC practices, given the amount of assessment that occurs in those initial meetings. Although intakes do not require systematic assessment or identification of standardized or individualized treatment targets, the initial information gathering process in an intake lends itself to the incorporation of these assessment methods. As noted previously, research has suggested that the use of standardized assessment of client problems at intake may be more malleable than routine monitoring to inform the course of treatment. [19] In the current study, however, all CAPER subscales (including Treatment Modification) were associated with conducting intakes. This is encouraging, as it suggests that MBC practices may not be limited to intake sessions in the current sample.

Also notable is the fact that clinicians who indicated working with children and adolescents were more likely to report using individualized assessments on the CAPER. There are a number of reasons why this might be the case. First, youth mental health services are more likely to be initiated when a child or adolescent's problems are interfering with their successful functioning in a key development context, such as school. [49] These types of functional issues may be less likely to map onto the traditional symptom structures that characterize the majority of standardized tools used in mental health treatment. Second, the very nature of youth service provision may lend itself most readily to individualized approaches. Weisz and colleagues [31] have suggested that individualized assessment might be particularly applicable to youth mental health, where services often involve both youth and their caregivers. Engagement of both groups may be enhanced by methods that identify and incorporate the most important presenting issues into the services provided. [31] Third, providers primarily working with children in the current study were slightly less likely to be in private practice than providers primarily working with adults (61 vs. 70%), and private practice was associated with lower scores on assessment use.

Finally, the relationship between service provision to ethnic and economic minority youth and all CAPER subscales is potentially encouraging. While extensive research has underscored that minority populations are less likely to receive mental health services, [50] very little is known about potential disparities in the quality of the services received by individuals once they engage in treatment. Although concerns have been raised that such disparities could arise due to inequities in evidence-based practice implementation, [51] the findings from the current study suggest that, at least for MBC, this may not be true. Nevertheless, considering that MBC use was also related to other workplace characteristics—such as a lower likelihood of being in private practice—it may be that this finding is as much a function of the service setting as it is of the population served.

## Limitations

Despite important implications for the measurement of MBC practices in usual care settings, the current study also has several limitations. First, the CAPER is a self-report measure and literature exists to suggest that clinicians may over-report use of evidence-based practices. [52,53] The CAPER was designed to address some of the known limitations of clinician self-report by (1) gathering information only about a recent, circumscribed period of time and (2) providing specific behavioral anchors that relate to the percentage of the clinician's caseload. In addition, research suggests that clinicians can be reliable reporters of their own clinical behaviors in situations where practices are relatively simple and clearly articulated, [54,55] so responses on the CAPER may be less vulnerable to these biases. Additional research is currently underway to compare CAPER reports to other sources of information, such as clinical case note content.

Second, because the CAPER was conceptualized as a pragmatic and low-burden measurement tool to enhance its utility in evaluating clinician MBC practices in community contexts, [23] a parsimonious set of items was critical. However, the emphasis on maintaining a small number of

items simultaneously limited the feasibility of using factor analysis for item or scale reduction. Nevertheless, factor analysis serves additional important purposes that were leveraged in the current study, such as identifying unobserved constructs. Measure parsimony also may have contributed to ambiguity about the relative merits of the two- and three-factor solutions. As the CAPER is applied in additional studies, it may be important to continue to evaluate the potential value of additional items, especially those that would be expected to load on the Individualized Assessment and the Treatment Modification subscales.

Finally, although the CAPER evaluates adherence to core aspects of MBC, it is not a measure of MBC competence or quality. Clinicians who consistently collect data, provide feedback, and make treatment adjustments may still engage in such practices in suboptimal ways, and the CAPER was not designed to index this information (e.g., whether the instruments they use have strong psychometrics, whether treatment adjustments made in response to assessment data are clinically appropriate). Although they are likely to be considerably longer than the CAPER, opportunities to develop MBC competence measures remain.

### **Implications for Behavioral Health**

Use of MBC in routine clinical practice is notoriously limited. The findings of this study support the structural, convergent, and divergent validity of the CAPER for assessing MBC practices. Given the relatively nascent literature on ways to evaluate and support MBC in mental health service delivery, the CAPER reflects a useful advancement, especially with its explicit attention to both standardized and individualized assessment practices. In practice, the CAPER might be used by supervisors, trainers, and organizations to understand individual clinicians' MBC practices and identify areas where individualized supports might be provided. For example, given its pragmatic nature (i.e., brief, can be quickly administered), a supervisor may use the CAPER with all supervisees and learn that standardized assessments are used routinely at the beginning of treatment, but that feedback is infrequently provided. The supervisor might then build recommendations for measure interpretation, feedback to clients, and collaborative decision-making into supervision. At an aggregate level, clinics or service systems might collect CAPER data from all clinicians to plan larger professional development initiatives and/or monitor the success of ongoing efforts to increase the use of MBC in their setting.

Next steps for research on the CAPER include linking the tool to more objective indicators of MBC practices, such as direct observations or, as mentioned above, review of clinical records. This research should also be conducted across multiple settings, clinicians with varying professional backgrounds, and diverse client populations to evaluate potential variability and establish more specific benchmarks. There is considerable potential for the CAPER to be applied in other contexts where MBC is becoming increasingly popular, such as integrated mental health services in primary care settings or the mental health services delivered by the Department of Veterans Affairs. Additional applications of the tool may be most appropriate in the context of MBC implementation initiatives, where the CAPER could be used to evaluate baseline MBC practices and track changes over time. Based on prior research, [19] it might be anticipated that the majority of participants in an MBC implementation project will need the greatest amount of support surrounding treatment modification, highlighting the value of the CAPER's individual subscale and the three-factor model.

Individual assessment is also highly consistent with the current trajectory in contemporary healthcare toward personalized and precision medicine. Evaluating client-reported symptoms and functioning as well as contextual data are cornerstones of a recently articulated framework for precision mental health. [4] Prior research also suggests that individualizing intervention targets and ways to monitor those targets is likely to improve the data's utility for treatment planning. [31] As such, using a tool like the CAPER to monitor clinicians' evaluation of both standardized *and* individualized targets has great potential to improve client-centered precision mental health.

# Compliance with Ethical Standards

*Conflict of Interest* The authors have no conflicts of interest to report.

## References

1. Substance Abuse and Mental Health Services Administration. Partners for Change Outcome Management System (PCOMS): International Center for Clinical Excellence.
2. Glasgow RE, Kessler RS, Ory MG, et al.. Conducting rapid, relevant research: lessons learned from the My Own Health Report project. *American Journal of Preventive Medicine* 2014;47(2):212–219.
3. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. *Cognitive Behavioral Practice*. 2015;22(1):49–59. <https://doi.org/10.1016/j.cbpra.2014.01.010>
4. Bickman L, Lyon AR, Wolpert M. Achieving precision mental health through effective assessment, monitoring, and feedback processes. *Administration & Policy in Mental Health & Mental Health Services Research*. 2016;43:271–276. <https://doi.org/10.1007/s10488-016-0718-5>
5. Lewis CC, Scott K, Marti CN, et al. Implementing measurement-based care (iMBC) for depression in community mental health: a dynamic cluster randomized trial study protocol. *Implementation Science*. 2015;10(1):127.
6. Bickman L, Kelley SD, Breda C, et al.. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatric Services*. 2011;62(12):1423–1429. <https://doi.org/10.1176/appi.ps.002052011>
7. Carlier IVE, Meuldijk D, Van Vliet IM, et al.. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of Evaluation in Clinical Practice*. 2012;18(1):104–110. <https://doi.org/10.1111/j.1365-2753.2010.01543.x>
8. Lambert MJ, Whipple JL, Hawkins EJ, et al.. Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science & Practice*. 2003;10(3):288–301. <https://doi.org/10.1093/clipsy.bpg025>
9. Guthrie D, McIntosh M, Callaly T, et al.. Consumer attitudes towards the use of routine outcome measures in a public mental health service: A consumer-driven study. *International Journal of Mental Health Nursing*. 2008;17(2):92–97.
10. Douglas SR, Jonghyuk B, Andrade ARV de, et al.. Feedback mechanisms of change: How problem alerts reported by youth clients and their caregivers impact clinician-reported session content. *Psychotherapy Research*. 2015;25(6):678–693. <https://doi.org/10.1080/10503307.2015.1059966>
11. Hatfield DR, Ogles BM. The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research & Practice*. 2004;35(5):485–491.
12. Ionita G, Fitzpatrick M. Bringing science to clinical practice: A Canadian survey of psychological practice and usage of progress monitoring measures. *Canadian Psychology*. 2014;55(3):187.
13. Palmiter DJ. A survey of the assessment practices of child and adolescent clinicians. *American Journal of Orthopsychiatry*. 2004;74(2):122.
14. Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: What's the use? *Journal of Behavioral Health Services & Research*. 2003;30(4):393–405.
15. Jensen-Doss A, Smith AM, Becker-Haimes EM, et al. Individualized progress measures are more acceptable to clinicians than standardized measures: Results of a national survey. under review.
16. Jensen-Doss A, Haimes EMB, Smith AM, et al. Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. *Administration & Policy in Mental Health & Mental Health Services Research*. 2016:1–14.
17. Jensen-Doss A, Hawley KM. Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology*. 2010;39(6):885–896. <https://doi.org/10.1080/15374416.2010.517169>
18. Lyon AR, Pullmann MD, Whitaker K, et al. A digital feedback system to support implementation of measurement-based care by school-based mental health clinicians. *Journal of Clinical Child & Adolescent Psychology*. 2017:1–12.
19. Lyon AR, Dorsey S, Pullmann M, et al. Clinician use of standardized assessments following a common elements psychotherapy training and consultation program. *Administration & Policy in Mental Health & Mental Health Services Research*. 2015;42(1):47–60.
20. Close-Goedjen JL, Saunders SM. The effect of technical support on clinician attitudes toward an outcome assessment instrument. *Journal of Behavioral Health Services & Research*. 2002;29(1):99–108.
21. Berwick DM. Broadening the view of evidence-based medicine. *BMJ Quality & Safety*. 2005;14(5):315–316. <https://doi.org/10.1136/qshc.2005.015669>
22. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Education & Behavior*. 2013;40(3):257–265. <https://doi.org/10.1177/1090198113486805>
23. Glasgow RE, Riley WT. Pragmatic measures: What they are and why we need them. *American Journal of Preventive Medicine*. 2013;45(2):237–243. <https://doi.org/10.1016/j.amepre.2013.03.010>
24. Martinez RG, Lewis CC, Weiner BJ. Instrumentation issues in implementation science. *Implementation Science*. 2014;9(1):118.
25. Borntrager C, Lyon AR. Client progress monitoring and feedback in school-based mental health. *Cognitive & Behavioral Practice*. 2015;22(1):74–86. <https://doi.org/10.1016/j.cbpra.2014.03.007>
26. Cone JD. Idiographic, nomothetic, and related perspectives in behavioral assessment. *Conceptual Foundations of Behavioral Assessment*. 1986:111–128.

27. Lyon AR, Connors E, Jensen-Doss A, et al. Intentional research design in implementation science: implications for the use of nomothetic and idiographic assessment. *Translational Behavioral Medicine*. 2017;1–14.
28. McLeod BD, Jensen-Doss A, Ollendick TH. *Diagnostic and Behavioral Assessment in Children and Adolescents: A Clinical Guide*. Guilford Press; 2013. [https://books.google.com/books?hl=en&lr=&id=Ja8YAAAAQBAJ&oi=fnd&pg=PP2&dq=Diagnostic+and+behavioral+assessment+in+children+and+adolescents:+A+clinical+guide&ots=X\\_Qv66A4ly&sig=mLeKpn1V-zNc-Ga-PYN\\_ubghbtII](https://books.google.com/books?hl=en&lr=&id=Ja8YAAAAQBAJ&oi=fnd&pg=PP2&dq=Diagnostic+and+behavioral+assessment+in+children+and+adolescents:+A+clinical+guide&ots=X_Qv66A4ly&sig=mLeKpn1V-zNc-Ga-PYN_ubghbtII). Accessed May 10, 2017.
29. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*. 2002;32(9):509–515. <https://doi.org/10.3928/0048-5713-20020901-06>
30. Haynes SN, Mumma GH, Pinson C. Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review*. 2009;29(2):179–191.
31. Weisz JR, Chorpita BF, Frye A, et al. Youth Top Problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting & Clinical Psychology*. 2011;79(3):369.
32. Connors EH, Arora P, Curtis L, et al. Evidence-based assessment in school mental health. *Cognitive Behavioral Practice*. 2015;22(1):60–73. <https://doi.org/10.1016/j.cbpra.2014.03.008>
33. Duong MT, Lyon AR, Ludwig K, et al. Student perceptions of the acceptability and utility of standardized and idiographic assessment in school mental health. *International Journal of Mental Health Promotion*. 2016;18(1):49–63. <https://doi.org/10.1080/14623730.2015.1079429>
34. Lyon AR, Ludwig K, Wasse JK, et al. Determinants and functions of standardized assessment use among school mental health clinicians: A mixed methods evaluation. *Administration & Policy in Mental Health & Mental Health Services Research*. 2015;43(1):122–134. <https://doi.org/10.1007/s10488-015-0626-0>
35. Lindhiem O, Bennett CB, Orimoto TE, et al. A meta-analysis of personalized treatment goals in psychotherapy: A preliminary report and call for more studies. *Clinical Psychology: Science & Practice*. 2016;23(2):165–176.
36. Becker EM, Smith AM, Jensen-Doss A. Who's using treatment manuals? A national survey of practicing therapists. *Behaviour Research & Therapy*. 2013;51(10):706–710.
37. Brookman-Frazee L, Haine RA, Baker-Ericzen M, et al. Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration & Policy in Mental Health & Mental Health Services Research*. 2010;37(3):254–269.
38. Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. *Administration & Policy in Mental Health & Mental Health Services Research*. 2007;34(3):283–291.
39. Hyde PS. Report to Congress on the Nation's substance abuse and mental health workforce issues. *US Department of Health & Human Services Substance Abuse & Mental Health Services*. Jan 2013. 2013;10. [http://www.cimh.org/sites/main/files/file-attachments/samhsa\\_bhwork\\_0.pdf](http://www.cimh.org/sites/main/files/file-attachments/samhsa_bhwork_0.pdf). Accessed May 10, 2017.
40. Hawley KM, Cook JR, Jensen-Doss A. Do noncontingent incentives increase survey response rates among mental health providers? A randomized trial comparison. *Administration & Policy in Mental Health & Mental Health Services Research*. 2009;36(5):343–348.
41. Glisson C, Landsverk J, Schoenwald S, et al. Assessing the Organizational Social Context (OSC) of mental health services: Implications for research and practice. *Administration & Policy in Mental Health & Mental Health Services Research*. 2008;35(1–2):98. <https://doi.org/10.1007/s10488-007-0148-5>
42. Dillman DA, Smyth JD, Christian LM. *Internet, Mail, and Mixed Mode Survey: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons; 2009.
43. Tabachnick BG, Fidell LS. *Using Multivariate Statistics* (5th Edition). Needham Heights: Allyn & Bacon, Inc.; 2007.
44. Hooper D, Coughlan J, Mullen M. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*. 2008;6(1):53–60.
45. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling Multidisciplinary Journal*. 1999;6(1):1–55.
46. Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. *Personality & Individual Differences*. 2007;42(5):893–898. <https://doi.org/10.1016/j.paid.2006.09.017>
47. Kaiser HF. A second generation little jiffy. *Psychometrika*. 1970;35(4):401–415. <https://doi.org/10.1007/BF02291817>
48. Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974;39(1):31–36. <https://doi.org/10.1007/BF02291575>
49. Bradshaw CP, Buckley JA, Ialongo NS. School-based service utilization among urban children with early onset educational and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly*. 2008;23(2):169–186. <https://doi.org/10.1037/1045-3830.23.2.169>
50. Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*. 2002;159(9):1548–1555.
51. Cabassa LJ, Baumann AA. A two-way street: Bridging implementation science and cultural adaptations of mental health treatments. *Implementation Science*. 2013;8(1):90.
52. Hurlburt MS, Garland AF, Nguyen K, et al. Child and Family Therapy Process: Concordance of Therapist and Observational Perspectives. *Administration & Policy in Mental Health & Mental Health Services Research*. 2009;37(3):230–244. <https://doi.org/10.1007/s10488-009-0251-x>
53. Nakamura BJ, Selbo-Bruns A, Okamura K, et al. Developing a systematic evaluation approach for training programs within a train-the-trainer model for youth cognitive behavior therapy. *Behavioral Research & Therapy*. 2014;53:10–19. <https://doi.org/10.1016/j.brat.2013.12.001>
54. Chapman JE, McCart MR, Letourneau EJ, et al. Comparison of youth, caregiver, therapist, trained, and treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting & Clinical Psychology*. 2013;81(4):674.
55. Hogue A, Dauber S, Henderson CE, et al. Reliability of therapist self-report on treatment targets and focus in family-based intervention. *Administration & Policy in Mental Health & Mental Health Services Research*. 2014;41(5):697–705.