Improving Measurement Feedback Systems for Measurement-Based Care

Alexandra Paige Peterson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Corey Fagan, Chair

Ronald Smith

Aaron Lyon

Program Authorized to Offer Degree:

Department of Psychology

University of Washington

Abstract

Improving Measurement Feedback Systems for Measurement-Based Care

Alexandra Paige Peterson

Chair of the Supervisory Committee:

Corey Fagan

Department of Psychology

Measurement-based care (MBC) is increasingly recognized as a beneficial addition to behavioral healthcare but is nevertheless underutilized by clinicians. Measurement feedback systems (MFS) are a class of health information technologies developed to increase adoption and facilitate MBC. These systems often include a diverse array of features, though there is little knowledge about the influence they might have on MBC. To that end, this vignette-based study tested the impact of four MFS features (progress graph, expected change trajectory, alert, clinical decision support) on key, clinician-driven MBC processes (progress assessment accuracy and treatment adjustments) in the context of different clinical scenarios (patient deterioration, no progress, and remission). Analyses revealed MFS features differentially impacted clinicians' progress assessment accuracy, how likely they were to make a treatment change, and their specific treatment choices. However, which feature was most impactful varied depending on the clinical

scenario. When asked to reflect on their answers, clinicians reported the graphs influenced their progress assessments and treatment choices significantly more than the other three features. However, when asked which MFS feature(s) they would prefer to use in their own clinical work, the majority stated they would like to use all of them.

Improving Measurement Feedback Systems for Measurement-Based Care

Measurement-based care (MBC), or using patient-reported outcome measures (PROMs) to track patient progress (e.g., symptoms) or process (e.g., therapeutic alliance) and inform treatment decisions, is increasingly recognized as a beneficial addition to behavioral healthcare. Shifts in policy, concerns about quality assurance, and increased value placed on patients' own views have accompanied MBC's rise in general and empirical discourse (Carman et al., 2013; Gaebel et al., 2015; Porter, Larsson, & Lee, 2016). Numerous reviews show MBC improves outcomes beyond usual care, particularly for patients who worsen at some point in treatment (Gondek, Edbrooke-Childs, Fink, Deighton & Wolpert, 2016; Lambert, Whipple, & Kleinstäuber, 2018; Lewis et al., 2018; Tam & Ronan, 2017), and the practice's trans-diagnostic and trans-theoretical nature make it broadly applicable (Scott & Lewis, 2015). Nevertheless, MBC is underutilized by clinicians worldwide (Ionita & Fitzpatrick, 2014; Macdonald & Fugard, 2015; Patterson, Matthey, & Baker, 2006; Zimmerman & McGlinchey, 2008). The most recent survey of psychologists and masters level providers in the U.S. indicated only 5.2% engage in MBC regularly (Jensen-Doss et al., 2016).

Low MBC usage can be explained by the myriad of ways implementation can go awry. Barriers to use are well documented and multilevel, ranging across the patient (e.g., not enough time to complete measures), practitioner (e.g., concern about how the data will be used), organization (e.g., limited resources for training), and system (e.g., lack of incentives from third party payers) (Boswell, Kraus, Miller, & Lambert, 2015; Lewis et al., 2018). No 'silver bullet' can guarantee successful MBC implementation: instead, "multifaceted or blended strategies tailored to target local barriers" are most likely to result in success (Lewis et al., 2018). At the same time, the discrete strategies that comprise a comprehensive implementation plan must be closely scrutinized for effectiveness to ensure they are not the ‘thorn in the side’ of what would otherwise be a successful MBC implementation effort.

One such strategy that needs further investigation and refinement encompasses a class of health information technologies called measurement feedback systems (MFS; Lewis et al., 2018; Lyon et al., 2015; Lyon, Lewis, Boyd, Hendrix & Liu, 2016). MFS support MBC implementation by facilitating the delivery and automatic scoring of PROMs, as well as generating immediate feedback, often in the form of graphs (Bickman, 2008; Bickman, Kelley, & Athay, 2012; Lyon, Lewis, Boyd, Hendrix & Liu, 2016). Reference to these systems has increased dramatically in academic literature because MFS can address important MBC barriers (e.g., automatically scoring PROMS instead of requiring clinicians to hand score them) (Boswell, Kraus, Miller, & Lambert, 2015; Lyon et al., 2016). At the same time, evidence supporting their effectiveness at facilitating MBC is mixed. Access to a MFS increased clinicians’ use of MBC in a randomized controlled trial (RCT) (Lyon et al., 2017). Conversely, a MFS was the most commonly reported barrier to a different MBC implementation effort (Gleacher et al., 2016) and a MFS appeared unnecessary for successful MBC implementation in another (Persons, Koerner, Eidelman, Thomas, &Liu, 2016).

One way to understand these discrepant results is the variability in current MFS. Forty-nine different MFS for behavioral healthcare with over 50 different capabilities and characteristics were identified at the end of 2014 (Lyon, Lewis, Boyd, Hendrix & Liu, 2016). This diversity is unsurprising, given existing systems have emerged from a variety of settings and were designed to meet assorted commercial, academic, and service interests (Lyon et al., 2016). Given high variation, the research suggesting inconsistent effectiveness, and the growing focus on implementing MBC via MFS, the field could benefit from studying how to design MFS to best facilitate 'high fidelity' MBC.

High fidelity MBC includes "(1) a routinely administered symptom, outcome, or process measure…, ideally before each clinical encounter; (2) practitioner review of data; (3) patient review of data; and (4) collaborative reevaluation of the treatment plan informed by data" (Lewis et al., 2018). This definition highlights two important factors when considering MFS design: 1) the importance of understanding progress and 2) the use of progress data to inform treatment.

First, helping clinicians accurately understand patient progress over time in an objective manner is a key function of MBC, particularly because clinicians have a positive bias when conceptualizing how their patients are doing in treatment. They tend to overestimate how many of their patients improve and struggle to detect even significant deterioration (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010; Walfish, McAlister, O'Donnell, & Lambert, 2012). There is also a shortage of behavioral health providers in the U.S. (Thomas, Ellis, Konrad, Holzer, & Morrissey, 2009; US Department of Health and Human Services, 2016) and waitlists are long. Clinicians who fail to identify and respond appropriately to patients who are not improving or continue to treat patients even after they no longer have clinically significant problems could lengthen treatment and impact efficiency, which in turn limits access to treatment for those in need.

Second, this definition requires that clinicians adjust treatment plans based on what's observed in PROM data, a direct nod to a potential mechanism of action underlying MBC. A comprehensive understanding of why MBC improves outcomes remains elusive. However, a few studies suggest clinician behavior changes (either through discussion or adjusting treatment) are relevant. One RCT with youth demonstrated increased and faster discussion of clinically relevant topics when MBC was incorporated (Douglas et al., 2015). A second RCT with adults showed psychiatrists made twice as many medication adjustments with MBC compared to usual care, which resulted in greater and faster symptom reduction (Guo et al., 2015). A vignette-based study indicated psychologists were significantly more likely to say they would change treatment when a PROM indicated a patient was worsening compared to when a patient verbally reported it (Hatfield & Ogles, 2006). Finally, group leaders in a qualitative study reported PROMS made them more attentive to certain processes and individuals in the group, and led to actions they took in the following session (Whitcomb, Woodland, & Burlingame, 2018). All four of these studies suggest clinicians change their behavior in response to PROM data in an attempt to improve care.

If a clinician accurately understanding progress and adjusting treatment accordingly is considered part of ‘high fidelity’ MBC *and* a likely MBC mechanism, the impact MFS have on these processes needs to be explored. However, the high variability in current MFS also needs to be considered. MFS have largely been studied in a ‘package’ format (Bickman et al., 2016; Gleacher et al., 2016; Persons, Koerner, Eidelman, Thomas, & Liu, 2016) with minimal attention paid to the individual features (e.g., alerts to the clinician communicating critical information, corrective feedback aimed at changing a clinician’s approach) that comprise a system. This package approach makes it impossible to identify which components of the package contribute to desired outcomes. Instead, research should address which of the many features included in MFS on the market today influence key MBC processes, including accurate progress assessments and PROM-driven treatment adjustments, as this would enable developers to create simple, parsimonious systems that are maximally effective. Furthermore, the impact of specific MFS features should be explored across different clinical situations, given features might vary in usefulness depending on what is happening with the patient (e.g., deteriorating compared to making excellent progress). An extensive review of the literature revealed no research comparing the impact specific MFS features have on progress assessment or treatment adjustments in various clinical situations. The present study aims to address this gap in knowledge.

Purpose of Research

Vignette-based methodologies have been effectively used to examine clinical assessments and decisions made by behavioral health providers, and results can be highly generalizable to 'real-life' behavior (Carpenter et al., 2016; Evans et al., 2015; Hatfield & Ogles, 2006). To that end, this vignette-based study was designed to answer the following questions: (a) Does the presence of a certain MFS feature improve clinicians' accuracy when making progress assessments? (b) Does a particular feature influence the likelihood clinicians will change treatment when a change is objectively called for (e.g., when months have passed and the patient is not improving)? (c) Do the specific changes clinicians make vary depending on which MFS feature they see? (d) Are clinicians more confident in their progress assessments and treatment decisions when they see a particular feature? (e) Do any of the above vary depending on whether the patient is deteriorating, not progressing, or approaching remission, based on PROM scores over time? Finally, (f) which MFS feature(s) do clinicians consider most influential and which would they prefer to use in their own work?

Method

Participants

Participants were 299 clinicians (M_{age} = 37.3 years, SD = 10.3 years; 76.6% female; 82.3% White; 71.5% psychologists; 71.6% practicing from a cognitive behavioral/behavioral framework), recruited through clinical listserves maintained by national (e.g., Association for

Behavioral and Cognitive Therapies, Society for a Science of Clinical Psychology) and local (e.g., Seattle Psychology) groups/organizations. All participants were encouraged to share study information with other clinicians who might be interested in participating. Table 1 provides detailed participant demographic information. All participants were required to have obtained their professional degree, actively carry adults on their caseload, and have treated depression. An additional 21 participants completed the study but were excluded from analyses either for incorrectly answering questions intended to check for inattentive/inauthentic responding (e.g., a participant checked 'psychiatrist' on one item and reported highest degree was a 'B.A./B.S.' on another; *n* = 17) or for completing the survey in less than five minutes (*n* = 4; typically, participation took 10–30 min). All participants who completed the study were compensated with a $10.00 Amazon gift card.

Procedure and Materials

The University of Washington's Institutional Review Board granted approval for the study. Recruitment emails included a brief description of the study, eligibility requirements, contact information for the principal investigator, and a link to access the study. Upon clicking the link, participants were screened for eligibility, received informed consent, and completed the study procedures online via Qualtrics Survey Software, a secure program for data collection and management.

The study materials included (a) clinical vignettes, (b) MFS feature images with (c) mock clinical data, (d) progress assessment, clinical decision-making, and confidence questions, (e) process questions, and (f) demographic information questions.

Clinical vignettes. The three study vignettes used in this study were designed according to recommendations made by Evans et al. (2015) and provided a brief scenario of a continuing therapy case that featured an adult being treated for depression. In order to cleanly test the impact of the MFS features, the vignettes were designed to provide participants with a description of the patient's depression symptoms but did not provide any information on the patient's progress in therapy. Before they were presented with the vignettes and accompanying materials, clinicians were instructed to respond as themselves, describing what they *would* do in the scenario (as opposed to what they think they *should* do) and acknowledging real-life limitations (e.g., availability of specialists for referrals) (Evans et al., 2015).

MFS features. The four MFS features tested commonly occur in current systems and are frequently advocated in MBC literature. They included: 1) a graph of patient symptom change over time, 2) an expected change trajectory line indicating the rate of change a given patient's progress is projected to follow, 3) a written alert indicating how the patient is progressing, and 4) written clinical decision support that suggests next steps the clinician might consider (see Figure 1; Chorpita, Bernstein, & Daleiden, 2008; De Jong, 2016; Fortney et al., 2018; Hooke, Sng, Cunningham, & Page, 2018; Lyon, Lewis, Boyd, Hendrix, & Liu, 2016). Furthermore, all four of these features have preliminary evidence suggesting they broadly increase MBC effectiveness (Lambert et al. 2003; Lambert, Shimokawa, & Hilsenroth, 2011; Newnham & Page, 2010; Probst et al., 2013) and/or influence desired healthcare provider behavior (Douglas et al., 2015; Gerber et al., 2013; Kelley, de Andrade, Bickman, & Robin, 2012; Tasma et al., 2018). The four features were combined to create six different possible MFS feature combinations (see Table 2), which were reviewed by three experts in MBC/MFS for clarity and relevance (Evans et al., 2015; Veloski, Tai, Evans, & Nash, 2005).

Graph (G). The mock clinical data presented by the temporal line graph (G) spanned twelve weeks and were created from the Patient Health Questionnaire-9 (PHQ-9), a nine-item measure of depression. The PHQ-9 is one of the best-validated measures of depression and was chosen for this study because of its sensitivity to change and usefulness as a repeated measure of patient progress, as well as its widespread use across clinical settings (Kroenke & Spitzer, 2002; Kroenke, Spitzer, Williams, & Löwe, 2010; Löwe, Kroenke, Herzog, & Gräfe, 2004). Items directly address the frequency of symptoms of a major depressive episode (e.g., hopelessness, change in appetite) and the measure provides severity cut-off scores (mild, moderate, moderately severe, and severe depression). Labels and color coding demarcating these four cut-offs were included on the graph. All participants were provided with this information about the PHQ-9 before beginning the vignette-based portion of the study. Representing change over time is a minimal requirement for MFS (Lewis et al., 2015), therefore, the graph was considered a necessary feature and included in all conditions.

Expected change trajectory (GT). The expected change trajectory line overlaid the patient's mock clinical data on the graph (GT) and was designed to reflect a depression change trajectory reported in the research literature (Vittengl, Clark, Thase, & Jarrett, 2013). Participants whose condition included this feature received a brief description of the trajectory line explaining it was an average created from a sample of patients with similar demographics and symptom severity to their own patient at intake, an empirically supported method for creating expected change trajectories called the "nearest neighbor approach" (Lutz et al., 2005; Lutz et al., 2006).

Alert (GA). The alert was located under the graph (GA) and indicated whether the patient had reliably improved (i.e., 5 point decrease on PHQ-9), reliably deteriorated (i.e., 5 point increase on PHQ-9), or experienced no reliable change since intake according to the Reliable Change Index (Jacobson & Truax, 1991; Mcmillan, Gilbody, & Richards, 2010). The alert also indicated whether the patient's rate of change was sufficient, which is determined by the patient's degree of deviation from the expected change trajectory in the direction of higher symptom severity (Lambert et al., 2002; Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010). Participants whose condition included this feature were oriented to the Reliable Change Index and its value for the PHQ-9.

Clinical decision support (GS). The clinical decision support was located under the graph (GS) and provided a variety of clinical 'next-step' suggestions based on the patient's progress and supported by the empirical literature (Gelenberg et al., 2010; Lambert, 2011; McKay, Abramowitz, & Taylor, 2010). Participants who saw this feature were oriented to the fact that suggestions were evidence-based and relevant to the hypothetical patient's current treatment status.

Clinical data. Each participant saw three versions of graphed PHQ-9 data (one with each vignette) that indicated the following: (1) the patient is deteriorating (D), represented by an increase of six points on the PHQ-9 since beginning treatment twelve weeks ago, falling above the 'moderately severe depression' cutoff; (2) the patient is not progressing (NP), represented by a decrease of two points since beginning treatment, falling just under the 'moderate depression' cutoff; and (3) the patient is approaching remission (R), represented by a decrease of six points since beginning treatment, falling on the 'mild depression' cutoff. The order of presentation was counterbalanced across participants.

Clinical questions. These questions were based on items in previous literature examining progress assessment, clinical decision-making and clinical confidence with vignettes (Hatfield & Ogles, 2006; Waltman, Williams, & Christiansen, 2013). After each vignette, participants were asked to rate the hypothetical patient's progress (7-point scale; 1 represents 'much worse' and 7 represents ‘much improved’) and their confidence in that progress rating (percentage scale from 0 to 100, 100 represents ‘completely confident’). Participants were then asked how likely they would be to make a treatment change in the next two sessions (percentage scale from 0 to 100, 100 represents ‘I would certainly make changes’). They then chose to ‘continue with treatment as currently provided,’ ‘shift treatment as currently provided,’ or ‘move to terminate treatment’. Several more specific treatment choices were provided under each of these three options (e.g., ‘begin relapse prevention,’ ‘seek consultation regarding accuracy of diagnosis/case conceptualization and/or therapist fidelity to the treatment approach’); participants were asked to rank their top two. Participants then rated their confidence in their clinical decision (percentage scale from 0 to 100).

Process questions. After concluding the vignette portion of the study, participants rated how much each individual MFS feature included in their combination influenced their assessments of patients’ progress and treatment decisions (scale from 1 to 5, 5 representing ‘very’). Participants then selected which of the six MFS feature combinations they would prefer to use in their own clinical practice.

Demographic questions. Participants completed items describing their age, gender, race/ethnicity, highest degree obtained, years of clinical experience, hours per week conducting psychotherapy, work environment, theoretical orientation, and licensure. Participants were also asked about their attitudes towards routine and standardized assessment (Attitudes Towards Standardized Assessment Scales– Monitoring and Feedback; ASA-MF; Jensen-Doss et al., 2016), current MBC practices (Current Assessment Practice Evaluation – Revised; CAPER; Lyon et al., 2019), exposure to MFS, graph literacy (Galesic & Garcia-Retamero, 2011), and self-perceived computer skill and comfort (Computer Fluency Scale; CFS; Becker, 2012).

Design. The study was a 6 (MFS Feature(s): (1) G; (2) GT; (3) GA; (4) GS; (5) GTA; (6) GTAS) X 3 (Clinical Scenario: (1) D; (2) NP; (3) R) mixed factorial with MFS Feature as a between-subjects factor and Clinical Scenario as a within-subjects factor (see Table 2). Therefore, each participant was randomly assigned to see one MFS feature combination (e.g., GTA) across all three clinical scenarios (i.e., D, NP, and R).

Results

Demographic Items

One-way ANOVAs indicated no significant differences between conditions in clinician attitudes towards routine and standardized assessment (ASA-Total), MBC practices (CAPER-Total), self-perceived computer skill and comfort (CFS), age, years of experience, number of therapy hours per week, and graph literacy, all *ps* > .29, nor were differences found in clinician exposure to MFS , $\chi^2(5, N = 298) = 5.598, p = .252$, or title, $\chi^2(10, N = 299) = 9.095, p = .523$. Low cell counts precluded analysis of differences between conditions in clinician gender, race/ethnicity, theoretical orientation, licensure, and work environment (see Table 3 for a detailed summary of these variables by condition).

Progress Assessment Accuracy

Eight experts in MBC and MFS read the study vignettes with the graph, expected change trajectory, alert, and clinical support (GTAS; the feature combination with the most clinical information) and provided a patient progress rating for the three clinical scenarios (7-point scale; 1 represents ‘much worse’ and 7 represents ‘much improved’). This resulted in a mean expert rating of 2.13 for Deterioration, 4.19 for No Progress, and 6.13 for Remission. The accuracy of clinician ratings of patient progress was determined by examining whether the expert rating fell within the 95% confidence interval of the progress mean for a given MFS feature and clinical scenario pairing (see Table 4). When the patient was deteriorating, clinicians provided accurate assessments of patient progress when they saw the GA, GTA, and GTAS. When the patient was not making progress, clinicians were accurate when they saw the GA and GTAS. Finally, when the patient was approaching remission, clinicians were accurate regardless of which MFS feature combination they saw. Overall, the GA was the most parsimonious feature combination that was associated with accuracy across all three scenarios.

Treatment Changes

Likelihood of change. Analyses revealed a significant main effect of MFS Feature on clinicians' likelihood of making a treatment change, $F(5,288) = 4.781$, $p < .001$, $\eta_p^2 = .077$. Clinicians were more likely to say they would adjust treatment when presented with any of the following additional features - the GA, GTA[1], and GTAS ($M_{\mathrm{GA}} = 72.88$, $SD = 17.80$; $M_{\mathrm{GTA}} = 69.42$, $SD = 13.07$; $M_{\mathrm{GTAS}} = 71.53$, $SD = 17.09$), as compared to the graph only ($M_G = 60.90$, $SD = 14.46$), $p < .05$. Analyses also revealed a significant main effect of Clinical Scenario on likelihood of making a treatment change, $F(1.557,448.306) = 387.045$, $p < .001$, $\eta_p^2 = .573$. The likelihood of a clinician endorsing a treatment adjustment was significantly different between each of the three patient groups, with the deteriorating patient resulting in the highest likelihood of changing treatment strategy ($M_{\mathrm{D}} = 87.63$, $SD = 13.33$; $M_{\mathrm{NP}} = 71.78$, $SD = 21.99$; $M_{\mathrm{R}} = 40.86$, $SD = 30.64$), $p < .001$. These main effects were qualified by a significant interaction between MFS Feature and Clinical Scenario, $F(7.783, 448.31) = 5.44$, $p < .001$, $\eta_p^2 = .086$.

Simple main effects indicated MFS Feature influenced clinicians' likelihood of making a treatment change when the patient was deteriorating ($F(5, 289) = 4.107$, $p = .001$, $\eta_p^2 = .066$) and in remission ($F(5, 293) = 7.204$, $p < .001$, $\eta_p^2 = .109$), but not when the patient was staying

[1] w/outliers $p = .061$

stagnant, ($F(5, 292) = 1.603$, $p = .159$, $\eta_p^2 = .027$). When the patient was deteriorating, post-hoc Games-Howell tests revealed clinicians were more likely to say they would change treatment when presented with the GT, GA[2], and GTA[3] ($M_{GT} = 91.25$, $SD = 10.27$; $M_{GA} = 91.30$, $SD = 12.53$; $M_{GTA} = 90.39$, $SD = 10.52$) compared to when they were presented with the GS ($M_{GS} = 82.68$, $SD = 15.21$), $p < .05$. In comparison, when the patient was in remission, post-hoc Games-Howell tests indicated clinicians were more likely to endorse a treatment change when they saw the GA, GS, GTA, and GTAS ($M_{GA} = 49.93$, $SD = 33.96$; $M_{GS} = 43.85$, $SD = 30.16$; $M_{GTA} = 41.53$, $SD = 29.79$; $M_{GTAS} = 55.63$, $SD = 33.01$) than when they saw the GT ($M_{GT} = 26.02$, $SD = 20.15$), $p < .05$. Furthermore, endorsement of a treatment change was also more likely when participants were presented with the GA ($M_{GA} = 49.93$, $SD = 33.96$) and GTAS ($M_{GTAS} = 55.63$, SD = 33.01) compared to graph alone ($M_G = 31.59$, $SD = 27.12$), $p < .05$.

Exploratory analyses. Selected additional post-hoc analyses were informed by trends observed in the results reported above, all of which were significant by a fully post hoc Scheffe critical value, $p < .05$. Overall, clinicians' likelihood of making a treatment change, regardless of the clinical scenario, was higher when the MFS included an alert (GA, GTA, GTAS) compared to when the alert was absent (G, GT, GS). When considering the deterioration scenario alone, clinicians were more likely to endorse a treatment change when the MFS included a trajectory and/or an alert (GT, GA, GTA, GTAS) compared to when those features were absent (G, GS). Finally, when considering the remission scenario, clinicians' likelihood of making a treatment change was higher when the MFS included an alert and/or clinical decision support (GA, GS, GTA, GTAS) compared to when they were absent (G, GT).

---

[2] w/outliers, $p = .539$
[3] w/outliers $p = .235$

Decision to change treatment. The treatment choices clinicians made varied across MFS features when the patient was in remission, $\chi^2(5, N = 299) = 20.626$, $p = .001$. The number of clinicians who elected to change treatment (i.e., chose to "shift treatment as currently provided" or "move to terminate treatment") ranged from a low of 46% when shown the graph only to a high of 87% when they were shown all four features (see Table 5). Differences were not found when the patient's progress was stagnant, $\chi^2(5, N = 299) = 4.379$, $p = .496$. Low cell counts precluded analysis of differences in treatment choices between MFS features when the patient was deteriorating, however, it should be noted that a high degree of consistency in choice was observed across the different features in this clinical scenario. The number of clinicians who elected to change treatment ranged from 93% when shown the graph to 98% when shown the GT or GTAS, a difference smaller than that observed when the patient's progress was stagnant (77%-89% choosing a treatment change (see Table 5)).

Specific treatment choices. The specific treatment choices clinicians made varied across the different MFS feature combinations (see Table 6). When the patient was deteriorating, the most popular choice across conditions was to "discuss some/all of the following with Patient: understanding of lack of progress, motivation for change, accuracy of diagnosis/co-morbid diagnoses, therapeutic alliance, factors in Patient's environment that may be impeding progress, adherence to treatment/barriers to adherence, contingencies for continuing treatment," ranging from 52.3% of clinicians making this choice when viewing the GTA to 73.5% when viewing the GS. This was also the most popular choice when the patient was not making progress, ranging from 65.4% (GS) to 70.5% (GTAS). When the patient was in remission, clinicians most often chose to "cautiously continue with treatment plan for now, reevaluate in the future," ranging from 25.6% (GA) to 30.6% (GTA) in four of the six conditions. Comparatively, when viewing the GS, “cautiously continue…”, “make no changes in the treatment plan”, and “begin relapse prevention” tied for first choice (19.6%). When viewing the GTAS, 40.5% of clinicians chose to “begin relapse prevention.”

Given the relatively lower consensus across clinicians when the patient was in remission compared to the Deterioration and No Progress clinical scenarios, the decision was made to investigate clinicians’ second treatment choices. “Cautiously continue with treatment plan…” was the most popular second choice in the Remission scenario when clinicians viewed the G, GT, GA, and GTA, ranging from 25% (GA) to 34% (GT). “Begin tapering sessions” was the most popular second choice when clinicians viewed the GS (22%) and GTAS (24.4%).

Clinician Confidence

Clinicians were quite confident on average. When asked to reflect on their progress asssessments, they were 76% confident in their ratings when the patient was not making progress, 78% confident when the patient was deteriorating, and 82% confident when the patient was approaching remission. When reflecting on their treatment decisions, they ranged from 76% confident in their choices when the patient was not making progress to 80% confident when the patient was deteriorating or approaching remission.

Progress assessment. Analyses revealed a significant main effect of MFS Feature on clinicians’ confidence in their ratings of patient progress, $F(5,280) = 2.413$, $p < .05$, $\eta_p^2 = .041$.[4] , though post hoc Games-Howell tests failed to reveal significant pairwise comparisons. Analyses also indicated a significant main effect of Clinical Scenario, $F(2, 560) = 20.81$, $p < .001$, $\eta_p^2 = .069$. Clinicians were significantly more confident in their progress assessments when the patient was approaching remission ($M_R = 81.98$, $SD = 11.40$) compared to when the patient was

[4] w/outliers $p = .121$

deteriorating (M_{D} = 78.34, *SD* = 15.02) or not making progress (M_{NP} = 76.21, *SD* = 14.32), $p < .01$. These main effects were qualified by a significant interaction between MFS Feature and Clinical Scenario, $F(10, 560) = 2.019$, $p < .05$, $\eta_p^2 = .035$.[5]

Simple main effects indicated MFS Feature only influenced clinicians' confidence in their progress assessments when the patient was deteriorating ($F(5, 288) = 2.86$, $p < .05$, $\eta_p^2 = .047$; Remission: $F(5, 289) = 1.53$, $p = .18$, $\eta_p^2 = .026$; No Progress: $F(5,286) = 2.07$, $p = .069$, $\eta_p^2 = .035$)). Post-hoc Games-Howell tests revealed clinicians were more confident in their progress ratings when presented with the GTA (M_{GTA} = 82.94, *SD* = 13.02) compared to the graph alone (M_{G} = 72.87, *SD* = 17.05), $p = .012$.

Treatment choice. Analyses revealed a significant main effect of Clinical Scenario on clinician confidence in their chosen treatment choice, $F(2,572) = 12.289$, $p < .001$, $\eta_p^2 = .041$. Clinicians were significantly less confident in their choice when the patient was not making progress compared to when the patient was deteriorating or approaching remission (M_{NP} = 76.34, *SD* = 14.01; M_{D} = 79.82, *SD* = 14.39; M_{R} = 79.97, *SD* = 12.90), $p < .001$. Analyses did not indicate a significant main effect of MFS Feature on clinician confidence in treatment choice, $F(5,286) = 1.641$, $p = .149$, $\eta_p^2 = .028$, nor was the Clinical Scenario main effect qualified by a significant interaction between MFS Feature and Clinical Scenario, $F(10, 572) = .819$, $p = .611$, $\eta_p^2 = .014$.

Clinician Perceived Influence of and Preference for MFS Features

When the clinicians who had seen all four MFS features were asked to reflect on their answers, they reported the graphs influenced their progress assessments (*M* = 4.11, *SD* = .82; F(2.636, 118.636) = 29.441, $p < .001$, partial $\eta^2 = .395$) and treatment choices (*M* = 4.13, *SD* =

---

[5] w/outliers $p = .186$

.65; $F(2.338, 105.217) = 23.276$, $p < .001$, partial $\eta^2 = .341$) significantly more than the other three features (*Progress* - Trajectory: $M = 3.09$, $SD = 1.23$; Alert: $M = 2.78$, $SD = .1.01$; Support: $M = 2.87$, $SD = 1.05$; *Treatment Choice* - Trajectory: $M = 3.20$, $SD = 1.24$; Alert: $M = 2.83$, $SD = 1.02$; Support: $M = 3.00$, $SD = .99$). Further analyses confirmed this trend held no matter which features clinicians saw (e.g., GA, GT, etc.) – they always rated the graph as the most influential feature, $p < .001$ (see Table 7). However, when all participating clinicians were asked which MFS feature/feature combination they would prefer to use in their own clinical work, the majority (60.1%) stated they would like to have all four features (GTAS), followed by a preference for the GTA (15.1%), GS (11.7%), GT (6.4%), G (4%) and GA (2.3%) feature combinations.

Discussion

This study suggests that MFS features impact clinicians' accuracy when assessing progress and impact whether they will implement needed treatment changes, both key elements of high-fidelity MBC. Specifically, including alerts, expected change trajectories and clinical decision support with patients' graphed PROM data improved clinician accuracy and prompted treatment adjustments. However, this varied with clinical context. Which feature (or features) best influenced accuracy and adjustments depended on how the patient was progressing.

Progress Assessment

Results indicate that viewing a patient's PROM data graphed over time does not guarantee a clinician will conceptualize progress accurately. Clinicians' progress ratings tended to be overly positive when the patient was deteriorating and not progressing and were only accurate when the graph was combined with additional features; GA, GTA, and GTAS for deterioration and GA and GTAS for no change in either direction. It may be that clinician positive bias about patient outcome (Hannan et al., 2005; Walfish, McAlister, O'Donnell, & Lambert, 2012) also colors graph interpretation. While it is understood that presenting PROM data in a graph format is a necessary MFS feature (Lewis et al., 2015), it may not be sufficient to achieve an expert level understanding of progress in 'non-ideal' clinical scenarios. Comparatively, clinicians were consistently accurate (and significantly more confident in their assessments) in this study when the patient was approaching remission, even when the graph was the only source of information, suggesting graphs might be all that is needed for accuracy when progress is aligned with clinicians' positive bias. When the clinical context is less ideal, additional information like a written alert in a MFS indicating how the patient is progressing (the feature most associated with accuracy in this study) seems to be needed for accurate assessments.

Additionally, it is notable that clinicians were the least accurate and the least confident in their progress assessments when the patient was not changing. This suggests that stagnancy might be particularly hard for clinicians to conceptualize and there is a desire to see progress where there is none (clinician inaccuracy was always in the direction of thinking the patient had progressed more than the experts thought they had). Designing MFS in such a way that they facilitate high-fidelity MBC seems to be particularly important in these cases, as they seem to be the most ambiguous.

Treatment Change

Whether specific MFS features could prompt clinicians to make needed treatment adjustments was investigated through (1) a rating of how likely the clinician would be to implement a treatment change within the next two sessions and (2) a categorical choice to either change treatment or continue with care as currently provided, followed by a selection of more specific treatment decisions. Results differed across the two methods.

Likelihood of change. The expected change trajectory, alert, and clinical decision support all significantly increased clinicians' likelihood of treatment change, though this varied depending on the clinical scenario. When the patient was deteriorating, the presence of the alert and/or expected change trajectory increased the likelihood of a treatment adjustment. When the patient was approaching remission, it was the presence of the alert and/or clinical decision support that best prompted change. When the patient was not progressing, there was no observed difference across the features; they all prompted adjustment to the same degree. When the different patterns of patient progress were ignored, it was the presence of the alert that resulted in the highest change likelihood. Though the alert was most consistently effective, these results suggest there is not a 'one size fits all' feature when it comes to prompting treatment change; instead, what is most effective depends on what is happening clinically.

It is unfortunate that none of the features significantly increased change likelihood when the patient was not progressing. Clinicians were significantly less likely to make a treatment adjustment in the context of no progress compared to when the patient was deteriorating, a concerning finding given adjusting treatment is called for in both situations. Furthermore, clinicians were significantly less confident in their treatment choice in this clinical context, suggesting once again that they find no meaningful change in either direction particularly ambiguous.

Specific treatment choices.

Deterioration and No Progress. In the context of patient deterioration and no progress, clinicians were consistent in their choices to either change treatment or continue with care as currently provided; the vast majority chose to change treatment, an ideal result given these are the clinical situations when it is arguably most important for clinicians to adjust care.

Furthermore, the most popular shift clinicians wanted to make first was to discuss the clinical situation with the patient, which aligns with the research demonstrating clinicians talk about clinically relevant topics more often and more quickly when using MBC (Douglas et al., 2016). However, the categorical data also suggests that the expected change trajectory, alert, and clinical decision support did not push a clinician towards changing treatment any more than the graph did. While this aligns with the 'No Progress' likelihood data reported above, it does not align with the 'Deterioration' likelihood data indicating change is more likely with the expected change trajectory and/or alert. It is possible item structure is driving the difference in results. A clinician's rating of how likely she would be to implement a treatment change on a 0-100 scale might be more nuanced and capture subtler variability than her forced choice between two treatment options.

Remission. When the patient was approaching remission, MFS feature made a significant difference in treatment choice. Clinicians chose to change treatment most frequently in the presence of the clinical decision support (GS and GTAS). Furthermore, the specific changes clinicians made were more oriented towards termination (e.g., 'begin relapse prevention' and 'begin tapering session') when the clinical decision support was included. These results are particularly interesting because they suggest that MFS features could influence termination behaviors and could be effectively used to prompt clinicians to graduate patients, a needed outcome given the shortage of behavioral health providers in the U.S. They also demonstrate that an accurate understanding of positive progress does not guarantee clinicians will move towards graduating patients. Clinicians' progress assessments were consistently accurate when the patient was approaching remission, regardless of which MFS features they saw. However, only 46% of the clinicians said they would adjust treatment when they saw the graph alone, compared to 87%

who said they would when they saw the GTAS, suggesting that additional features might be needed to best stimulate clinicians to consider termination.

Confidence

MFS features had minimal impact on clinicians' confidence in their progress assessments and no impact on confidence in their treatment decisions. This is somewhat surprising, as it was hypothesized that explicit statements about patients' progress or recommendations to proceed with treatment in a particular way might help clinicians feel sure of their approaches. It may be these results reflect a variation of the positive bias clinicians tend to have. Research has shown that clinicians with varying credentials think highly of their clinical abilities; when providers were asked in a survey study to rate their clinical skills relative to other clinicians, not a single individual rated themselves below the 50^{th} percentile and 25% rated themselves in the 90^{th} percentile or above, numbers that defy statistical possibility (Walfish, McAlister, O'Donnell, & Lambert, 2012). Self-assessment bias might translate to high confidence once a progress assessment or treatment choice has been made.

Clinician Perceived Influence of and Preference for MFS Features

Graphs of patient's PROM data are considered a key MFS feature (Lewis et al., 2015). However, this is the first study to show how influential clinicians find them to be in the context of important MBC processes. No matter what combination of features clinicians were presented with, they consistently reported that the graph was most influential when it came to making progress assessments and treatment decisions. Interestingly, the MFS features clinicians preferred for their own clinical work suggest that they view the graphs as particularly useful, but not sufficient. The majority of the clinicians (60.1%) reported they would prefer all four features

(GTAS); only 4% stated they would only want the graph. The graph and alert was the least desired combination (2.3%).

Limitations

Replication should be pursued in a clinical context – it is possible results might differ when clinicians are interacting with their own patients or interacting with a MFS in vivo. The study sample was also heavily weighted towards psychologists practicing from a cognitive-behavioral or behavioral framework, a limitation given master's level clinicians are often the most prevalent providers of behavioral health services in the U.S. and idenfity with a variety of theoretical orientations (Garland et al., 2010). Furthermore, the patient represented in the vignette was an adult with only one clinical target – Major Depression. It would be beneficial to investigate the same research questions with youth as well as with multi-diagnositc patients.

The vignette methodology used also created some inherent limitations. First, it was not possible to include a 'no MBC' condition; providing PROM data was necessary for clinicians to complete the study procedures. If the study were to be replicated in vivo, it would be illuminating to see the progress assessments and treatment decisions clinicians make without MBC. Second, even though there were statistically significant differences in the hypothetical progress assessments and treatment decisions clinicians made, it is unclear whether those differences would meaningfully impact clinical care.

Implications & Future Directions

Despite the study's limitations, the results provide significant heuristic value, as they highlight the need for further empirical study of the specific components of MFS that might best facilitate high fidelity MBC. The four specific features tested here differentially impacted the MBC variables of interest. Continuing to only test MFS in a 'package' format could result in the maintenance of ineffective features or the omission of potentially effective ones. Both of these outcomes would slow the development of parsimonious, maximally effective MFS. Fortunately, there are methodologies already in the published literature that can go above and beyond vignettes to further elucidate which MFS features are most beneficial (Lyon et al., 2015).

Secondly, MFS may need to include information beyond patients' graphed PROM data to best facilitate high-fidelity MBC. The participating clinicians were often inaccurate when assessing progress and less likely to make needed treatment adjustments when they only saw a graph of their patient's PROM data. In some cases, additional MFS features appeared to correct for their inaccuracies and prompt treatment change, a result that should be replicated with clinicians in their clinical practices.

Thirdly, no one MFS feature 'conquered all' when it came to influencing the high-fidelity MBC processes investigated in this study. Instead, results suggest the effectiveness of a feature can be influenced by the clinical state of the patient, a finding which raises a variety of questions when it comes to MFS design. The first is whether incorporating 'smart' elements into MFS would enhance their effectiveness. Developers could create nuanced algorithms that choose which features to present to clinicians based on patients' PROM data, ensuring that features only appear when they will be most impactful (e.g., an expected change trajectory when the patient is deteriorating, decision support when a patient is approaching remission…). One risk is that such a strategy could over-complicate MFS, an important concern given complexity is theorized to negatively impact adoption (Rogers, 2010). If that were the case, the question becomes whether a 'static' MFS should include features that are most likely to be useful across clinical scenarios (e.g., alerts) or whether it should be designed to be maximally effective in a specific clinical situation (e.g., patient deterioration) because of an empirical rationale (e.g., deteriorating patients are most likely to drop out of treatment). Overall, if replication in a clinical context confirms that MFS features vary in effectiveness depending on patient progress, 'smart' MFS may be an untapped resource for facilitating high-fidelity MBC in a variety of clinical scenarios.

Next, the results also raise important questions about the balance between what clinicians say they want to use in their clinical work and what seems to best prompt desired clinical outcomes. The graph and alert was the most parsimonious feataure combination that consistently resulted in progress assessment accuracy and often lead to needed treatment changes. However, it was the least desired by clinicians for their own work. This tension must be attended to, given unappealing MFS can be barriers to the implementation of MBC (Gleacher et al., 2016). Incorporating strategies from user-centered design (UCD) is one way to address that tension. UCD is "an approach to product development that grounds the process in information collected about the individuals and settings where products will ultimately be used" (Lyon & Koerner, 2016). Knowledge gained from UCD techniques can be combined with theory-driven or mechanistic knowledge to create features that are both appealing and effective. For example, rapid prototyping (Wilson & Rosenberg, 1988) is just one of many UCD technique likely to be highly useful for MFS feature design; to use the approach, MFS developers would quickly test many versions of a single feature (e.g., alerts) and make improvements based on iterative feedback from clinicians, resulting in an end product tailored for the clinicians using it.

Finally, the results suggest the field should take a closer look at the impact of MFS when patients are ready to begin transitioning out of treatment and when patients are not progressing as expected. There is a dearth of extant studies demonstrating that MFS features can influence treatment termination behaviors. If these findings are replicated in a clinical context, that could have important implications for the long waitlists many patients face when they attempt to access care. The results also suggest that lack of progress is more ambiguous than patient deterioration; participating clinicians were less accurate in their progress assessments and less likely to adjust treatment when the patient was not changing. Future studies would do well to investigate what MFS features can best prompt treatment changes in the context of no progress, as the features investigated in this study did not have a significant impact. Furthermore, when doing so they should take into account that not all features are created equal. It may be that the same features designed from a UCD approach might be more effective than those used in this particular study.

Conclusion

Measurement feedback systems have marked potential for facilitating high-fidelity measurement-based care, which could positively impact the lives of countless patients struggling with behavioral health disorders. However, at present they remain inpreceise, 'package' systems that may or may not increase MBC adoption. By directing attention to specific features and adjusting how MFS are investigated, the field can instead create increasingly refined, sophisticated instruments. The promise of improved patient care through a health technology designed to truly guide treatment will hopefully inspire future research on which MFS features, seen under what conditions, will ultimately provide the greatest benefit to those in need.

References

Becker, E. M. (2012). *Computer fluency, access to technology, and attitudes towards technologically-based therapeutic tools among practicing clinicians* (Unpublished doctoral dissertation). University of Miami, Miami, Florida.

Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(10), 1114.

Bickman, L., Douglas, S. R., De Andrade, A. R. V., Tomlinson, M., Gleacher, A., Olin, S., & Hoagwood, K. (2016). Implementing a measurement feedback system: A tale of two sites. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 410-425.

Bickman, L., Kelley, S. D., & Athay, M. (2012). The technology of measurement feedback systems. *Couple and Family Psychology: Research and Practice*, *1*(4), 274-284.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, *25*(1), 6-19.

Carman, K. L., Dardess, P., Maurer, M., Sofaer, S., Adams, K., Bechtel, C., & Sweeney, J. (2013). Patient and family engagement: A framework for understanding the elements and developing interventions and policies. *Health Affairs*, *32*(2), 223-231.

Carpenter, A. L., Pincus, D. B., Conklin, P. H., Wyszynski, C. M., Chu, B. C., & Comer, J. S. (2016). Assessing cognitive-behavioral clinical decision-making among trainees in the treatment of childhood anxiety. *Training and Education in Professional Psychology*, *10*(2), 109.

Chorpita, B., Bernstein, F., & Daleiden, A. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health and Mental Health Services Research, 35*(1), 114-123.

De Jong, Kim. (2016). Deriving implementation strategies for outcome monitoring feedback from theory, research and practice. *Administration and Policy in Mental Health, 43*(3), 292.

Douglas, S. R., Jonghyuk, B., de Andrade, A. R. V., Tomlinson, M. M., Hargraves, R. P., & Bickman, L. (2015). Feedback mechanisms of change: How problem alerts reported by youth patients and their caregivers impact clinician-reported session content. *Psychotherapy Research*, *25*(6), 678-693.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., ... & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology*, *15*(2), 160-170.

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2018). A tipping point for measurement-based care. *Focus*, *16*(3), 341-350.

Gaebel, W., Grossimlinghaus, I., Heun, R., Janssen, B., Johnson, B., Kurimay, T., ... & Ruggeri, M. (2015). European Psychiatric Association (EPA) guidance on quality assurance in mental healthcare. *European Psychiatry*, *30*(3), 360-387.

Galesic, M. & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making, 31*(3), 444-457.

Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R. J., Haine-Schlagel, R., & Ganger, W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services, 61*(8), 788–795.

Gelenberg, A. J., Freeman, M. P., Markowitz, J. C., Rosenbaum, J. F., Thase, M. E., Trivedi, M. H., ... & Schneck, C. D. (2010). Practice guideline for the treatment of patients with major depressive disorder third edition. *The American Journal of Psychiatry*, *167*(10), 1-3, 9-11, 13-118.

Gerber, J. S., Prasad, P. A., Fiks, A. G., Localio, A. R., Grundmeier, R. W., Bell, L. M., ... & Zaoutis, T. E. (2013). Effect of an outpatient antimicrobial stewardship intervention on broad-spectrum antibiotic prescribing by primary care pediatricians: A randomized trial. *JAMA*, *309*(22), 2345-2352.

Gleacher, A., Olin, A., Nadeem, S., Pollock, S., Ringle, E., Bickman, M., . . . Hoagwood, L. (2016). Implementing a measurement feedback system in community mental health clinics: A case study of multilevel barriers and facilitators. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 426-440.

Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 325-343.

Guo, T., Xiang, Y. T., Xiao, L., Hu, C. Q., Chiu, H. F., Ungvari, G. S., ... & Feng, Y. (2015). Measurement-based care versus standard care for major depression: A randomized controlled trial with blind raters. *American Journal of Psychiatry*, *172*(10), 1004-1013.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying patients at risk for treatment failure. *Journal of Clinical Psychology: In Session*, *61*, 155–163.

Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our patients get worse? An investigation of therapists' ability to detect negative patient change. *Clinical Psychology & Psychotherapy*, *17*(1), 25–32.

Hatfield, D. R., & Ogles, B. M. (2006). The influence of outcome measures in assessing patient change and treatment decisions. *Journal of Clinical Psychology, 62*(3), 325-337.

Hooke, G. R., Sng, A. A., Cunningham, N. K., & Page, A. C. (2018). Methods of delivering progress feedback to optimise patient outcomes: The value of expected treatment trajectories. *Cognitive Therapy and Research*, *42*(2), 204-211.

Ionita, G., & Fitzpatrick, M. (2014). Bringing science to clinical practice: A Canadian survey of psychological practice and usage of progress monitoring measures. *Canadian Psychology/Psychologie canadienne, 55*(3), 187-196.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R., Lewis, C. C., Stanick, C. F., & Hawley, K. M. (2016). Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*(1), 48-61.

Kelley, S. D., de Andrade, A. R. V., Bickman, L., & Robin, A. V. (2012). The Session Report Form (SRF): Are clinicians addressing concerns reported by youth and caregivers?

Administration and Policy in Mental Health and Mental Health Services Research, *39*(1), 133-145.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509-515.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*(4), 345-359.

Lambert, M.J., Shimokawa, K., & Hilsenroth, M.J. (2011). Collecting patient feedback. *Psychotherapy, 48*(1), 72-79.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology and Psychotherapy, 9*, 149–164.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288-301.

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, *55*(4), 520.

Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., ... & Simon, G. (2018). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry, 76*(3), 324-335.

Lewis, C., Scott, K., Marti, C., Marriott, B., Kroenke, K., Putz, J., . . . Rutkowski, D. (2015). Implementing measurement-based care (iMBC) for depression in community mental health: A dynamic cluster randomized trial study protocol. *Implementation Science, 10*(1), 1-14.

Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: Sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of Affective Disorders*, *81*(1), 61-66.

Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schu¨rch, E., & Stulz, N. (2006). The probability of treatment success, failure and duration—What can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology and Psychotherapy, 13*, 223–232.

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... & Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, *73*(5), 904-913.

Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, *23*(2), 180-200.

Lyon, A. R., Lewis, C. C., Boyd, M. R., Hendrix, E., & Liu, F. (2016). Capabilities and characteristics of digital measurement feedback systems: Results from a comprehensive review. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 441-466.

Lyon, A. R., Lewis, C. C., Melvin, A., Boyd, M., Nicodimos, S., Liu, F. F., & Jungbluth, N. (2015). Health information technologies—academic and commercial evaluation (HIT-ACE) methodology: Description and application to clinical feedback systems. *Implementation Science*, *11*(1), 128.

Lyon, A. R., Pullmann, M. D., Dorsey, S., Martin, P., Grigore, A. A., Becker, E. M., & Jensen-Doss, A. (2019). Reliability, validity, and factor structure of the Current Assessment Practice Evaluation-Revised (CAPER) in a national sample. *The Journal of Behavioral Health Services & Research*, *46*(1), 43-63.

Lyon, A. R., Pullmann, M. D., Whitaker, K., Ludwig, K., Wasse, J. K., & McCauley, E. (2017). A digital feedback system to support implementation of measurement-based care by school-based mental health clinicians. *Journal of Clinical Child and Adolescent Psychology*, 1-12.

Macdonald, A. J., & Fugard, A. J. (2015). Routine mental health outcome measurement in the UK. *International Review of Psychiatry*, *27*(4), 306-319.

McKay, D. E., Abramowitz, J. S., & Taylor, S. E. (2010). *Cognitive-behavioral therapy for refractory cases: Turning failure into success*. American Psychological Association.

Mcmillan, Gilbody, & Richards (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders, 127*(1), 122-129.

Newnham, E. A., & Page, A. C. (2010). Bridging the gap between best evidence and best practice in mental health. *Clinical Psychology Review*, *30*(1), 127-142.

Patterson, P., Matthey, S., & Baker, M. (2006). Using mental health outcome measures in everyday clinical practice. *Australasian Psychiatry*, *14*(2), 133-136.

Persons, J. B., Koerner, K., Eidelman, P., Thomas, C., & Liu, H. (2016). Increasing psychotherapists' adoption and implementation of the evidence-based practice of progress monitoring. *Behaviour Research and Therapy*, *76*, 24-31.

Porter, M. E., Larsson, S., & Lee, T. H. (2016). Standardizing patient outcomes measurement. *New England Journal of Medicine*, *374*(6), 504-506.

Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic inpatient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75*(3), 255–261.

Rogers, E. M. (2010). *Diffusions of innovations* (6th ed.). New York: Free Press.

Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice, 22*(1), 49-59.

Tam, H. E., & Ronan, K. (2017). The application of a feedback-informed approach in psychological service with youth: Systematic review and meta-analysis. *Clinical Psychology Review*, *55*, 41-55.

Tasma, M., Roebroek, L. O., Liemburg, E. J., Knegtering, H., Delespaul, P. A., Boonstra, A., ... & Castelein, S. (2018). The development and evaluation of a computerized decision aid for the treatment of psychotic disorders. *BMC Psychiatry*, *18*(1), 163.

Thomas, K. C., Ellis, A. R., Konrad, T. R., Holzer, C. E., & Morrissey, J. P. (2009). County-level estimates of mental health professional shortage in the United States. *Psychiatric Services*, *60*(10), 1323-1328.

US Department of Health and Human Services (2016). National projections of supply and demand for selected behavioral health practitioners: 2013-2025.

Veloski, J., Tai, S., Evans, A. S., & Nash, D. B. (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. *American Journal of Medical Quality*, *20*, 151-157.

Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2013). Nomothetic and idiographic symptom change trajectories in acute-phase cognitive therapy for recurrent depression. *Journal of Consulting and Clinical Psychology*, *81*(4), 615.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, *110*(2), 639–644.

Waltman, S. H., Williams, A., & Christiansen, L. R. (2013). Comparing student clinician and licensed psychologist clinical judgment. *Training and Education in Professional Psychology*, *7*(1), 33.

Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*, *78*(2), 144.

Whitcomb, K. E., Woodland, S. C., & Burlingame, G. M. (2018). Do clinicians really use feedback-monitoring systems? A qualitative analysis of 16 group leaders. *Psychotherapy*, *55*(2), 132-143.

Wilson, J. & Rosenberg, D. (1988). Rapid prototyping for user interface design. In: Helander M, editor. Handbook of human-comp. interaction.Amsterdam: North-Holland; pp. 859–875.

Zimmerman, M., & McGlinchey, J. B. (2008). Why don't psychiatrists use scales to measure outcome when treating depressed patients? *The Journal of Clinical Psychiatry, 69*(12), 1916-1919.

Table 1.
Sample Demographics

Variable	% (Count) *N*=299
Gender	
Male	22.4% (67)
Female	76.6% (229)
Other	0.1% (3)
Race/Ethnicity	
American Indian/Alaska Native	0.3% (1)
Asian	8.4% (25)
Black/African American	3% (9)
Hispanic/Latino	2.3% (7)
White	82.3% (246)
More Than One Race	3.7% (11)
Title	
Masters Level Clinician	22.7% (68)
Post-doctoral Fellow	22.7% (68)
Psychologist	48.8% (146)
Theoretical Orientation	
CBT/Behavioral	71% (214)
Psychodynamic/Psychoanalytic	3% (9)
Family Systems	1.3% (4)
Humanistic/Client Centered	2% (6)
Eclectic	11% (33)
Other	11% (33)
Licensure	
Licensed & Board Certified	9.7% (29)
Licensed	63.9% (191)
Certified	1.3% (4)
Registered	3.7% (11)
Other[a]	21.1% (63)
Work Environment	
Private Practice	35.5% (106)
Community Mental Health	9.7% (29)
Higher Education Setting	6.7% (20)
Hosptial/Medical Center	35.8% (107)
Day Tx/IOP Facility	3% (9)
Residential Facility/Group Home	1.3% (4)
Other	8% (24)
Exposure to MFS	
Yes	28.2% (84)
No	71.8% (214)
	M (SD)
Age	37.3 (10.3)
Years of Experience	10.8 (8.5)
# of Therapy Hours/Week	17.5 (10.1)
ASA-MF	70.2 (9.3)
CAPER	16.7 (5.2)
CFS	30.7 (3.5)
Line Graph Fluency	
Interpretation	26.6 (114.5)
Prediction	24.5 (1.9)

[a]largely composed of "not yet licensed" post-doctoral fellows

Table 2.
Study Conditions

MFS Feature Combinations[a]	
G	Graph
GT	Graph + Expected Change Trajectory Line
GA	Graph + Alert
GS	Graph + Clinical Decision-Making Support
GTA	Graph + Expected Change Trajectory Line + Alert
GTAS	Graph + Expected Change Trajectory Line + Alert + Clinical Decision-Making Support
Clinical Scenarios[b]	
D	Deterioration
NP	No Progress
R	Remission

[a]Between subjects. [b]Within subjects.

Table 3.
Select Demographics by MFS Feature(s)

	% (Count)					
Variable	G (*N*=54)	GT (*N*=53)	GA (*N*=46)	GS (*N*=53)	GTA (*N*=47)	GTAS (*N*=46)
Gender						
Male	18.5% (10)	24.5% (13)	23.9% (11)	17% (9)	29.8% (14)	21.7% (10)
Female	79.6% (43)	75.5% (40)	71.7% (33)	83% (44)	70.2% (33)	78.3% (36)
Other	1.9% (1)	-	4.4% (2)	-	-	-
Race/Ethnicity						
American Indian/Alaska Native	-	-	-	-	2.1% (1)	-
Asian	5.6% (3)	15.1% (8)	6.5% (3)	5.7% (3)	6.4% (3)	10.9% (5)
Black/African American	3.7 % (2)	-	4.3% (2)	9.4% (5)	-	-
Hispanic/Latino	1.9% (1)	1.9% (1)	4.3% (2)	-	2.1% (1)	4.3% (2)
White	85.2% (46)	79.2% (42)	82.6% (38)	81.1% (43)	87.2% (41)	78.3% (36)
More Than One Race	3.7% (2)	3.8% (2)	2.2% (1)	3.8% (2)	2.1% (1)	6.5% (3)
Theoretical Orientation						
CBT/Behavioral	74.1% (40)	75.5% (40)	67.4% (31)	77.4% (41)	59.6% (28)	73.9% (34)
Psychodynamic/Psychoanalytic	3.7% (2)	-	4.3% (2)	3.8% (2)	2.1% (1)	4.3% (2)
Family Systems	1.9% (1)	-	2.2% (1)	-	2.1% (1)	2.2% (1)
Humanistic/Client Centered	1.9% (1)	-	6.5% (3)	1.9% (1)	2.1% (1)	-
Eclectic	9.3% (5)	5.7% (3)	15.2% (7)	9.4% (5)	17% (8)	10.9% (5)
Other	9.3% (5)	18.9% (10)	4.3% (2)	7.5% (4)	17% (8)	8.7% (4)
Licensure						
Licensed & Board Certified	14.8% (8)	13.2% (7)	10.9% (5)	7.5% (4)	-	10.9% (5)
Licensed	61.1% (33)	67.9% (36)	60.9% (28)	62.3% (33)	74.5% (35)	56.5% (26)
Certified	-	-	-	-	4.3% (2)	4.3% (2)
Registered	3.7% (2)	1.9% (1)	6.5% (3)	3.8% (2)	4.3% (2)	2.2% (1)
Other[a]	20.4% (11)	17% (9)	21.7% (10)	26.4% (14)	14.9% (7)	26.1% (12)
Work Environment						
Private Practice	27.8% (15)	30.2% (16)	41.3% (19)	35.8% (19)	42.6% (20)	37% (17)
Community Mental Health	11.1% (6)	11.3% (6)	8.7% (4)	9.4% (5)	10.6% (5)	6.5% (3)
Higher Education Setting	11.1% (6)	7.5% (4)	6.5% (3)	3.8% (2)	6.4% (3)	4.3% (2)
Hosptial/Medical Center	37% (20)	34% (18)	30.4% (14)	35.8% (19)	34% (16)	43.5% (20)
Day Tx/IOP Facility	3.7% (2)	3.8% (2)	4.3% (2)	3.8% (2)	2.1% (1)	-
Residential Facility/Group Home	-	1.9% (1)	2.2% (1)	3.8% (2)	-	-
Other	9.3% (5)	11.3% (6)	6.5% (3)	7.5% (4)	4.3% (2)	8.7% (4)

[a]largely composed of "not yet licensed" post-doctoral fellows

Table 4.

Progress Rating Confidence Intervals by MFS Feature(s) and Clinical Scenario

	G	GT	GA	GS	GTA	GTAS
Deterioration	[2.15, 2.49]	[1.78, 2.10]	[1.81, 2.23][a]	[2.21, 2.55]	[1.84, 2.20][a]	[1.86-2.23][a]
No Progress	[4.35, 4.67]	[4.32, 4.68]	[4.18, 4.53][b]	[4.31, 4.65]	[4.25, 4.57]	[4.14, 4.48][b]
Remission	[5.90, 6.18][c]	[5.90, 6.22][c]	[5.85, 6.24][c]	[5.90, 6.22][c]	[6.12, 6.44][c]	[5.99, 6.32][c]

[a]Includes expert progress rating 2.13. [b]Includes expert progress rating 4.19. [c]Includes expert progress rating 6.13.

Table 5.
Clinicians Who Changed Treatment by MFS Feature(s) and Clinical Scenario

	Clinical Scenario		
	Deterioration	*No Progress*	*Remission*
MFS Feature	# Changed Tx / Total (Percentage)		
G	50/54 (93%)	43/54 (80%)	25/54 (46%)
GT	52/53 (98%)	47/53 (89%)	29/53 (55%)
GA	43/46 (94%)	40/46 (87%)	24/46 (52%)
GS	51/53 (96%)	41/53 (77%)	30/53 (57%)
GTA	45/47 (96%)	41/47 (87%)	24/47 (51%)
GTAS	45/46 (98%)	40/46 (87%)	40/46 (87%)

Note. Changed Tx refers to “shift treatment as currently provided” or “move to terminate treatment”

Table 6.

Clinicians' First Choice Treatment Decisions by MFS Feature(s) and Clinical Scenario

		G - Frequencies			GT - Frequencies			GA - Frequencies		
		D	*NP*	*R*	*D*	*NP*	*R*	*D*	*NP*	*R*
No Change	No Change	-	-	10 (21%)	-	-	5 (10.4%)	-	2 (4.5%)	7 (18%)
	Cautiously Continue	3 (5.6%)	10 (20%)	13 (27%)	1 (2%)	5 (10%)	16 (33.3%)	2 (4.4%)	3 (7%)	10 (25.6%)
	Other	1 (2%)	1 (2%)	4 (8.3%)	-	-	-	2 (4.4%)	-	2 (5%)
Change Tx	Discuss	36 (66.7%)	34 (68%)	2 (4%)	31 (57.4%)	33 (64.7%)	1 (2%)	28 (62.2%)	27 (61.4%)	2 (5%)
	Consult	5 (9.3%)	1 (2%)	-	11 (20.4%)	6 (11.8%)	1 (2%)	7 (15.6%)	5 (11.4%)	1 (2.6%)
	Change	1 (2%)	1 (2%)	2 (4%)	3 (5.6%)	3 (6%)	6 (12.5%)	3 (6.7%)	5 (11.4%)	3 (7.7%)
	Refer - Meds/Testing	8 (15%)	2 (4%)	-	7 (13%)	3 (6%)	1 (2%)	1 (2.2%)	2 (4.5%)	-
	Conditional Terminate	-	-	-	-	-	-	2 (4.4%)	-	-
	Other	-	-	1 (2%)	-	-	-	-	-	1 (2.6%)
Terminate	Relapse Prevention	-	1 (2%)	6 (12.5%)	-	-	9 (18.8%)	-	-	6 (15.4%)
	Session Tapering	-	-	7 (14.6%)	-	-	4 (8.3%)	-	-	5 (13%)
	Graduate	-	-	3 (6.3%)	-	-	5 (10.4%)	-	-	2 (5%)
	Refer - Different Tx	-	-	-	1 (2%)	1 (2%)	-	-	-	-
	Other	-	-	-	-	-	-	-	-	-
Total		*54 (100%)*	*50 (100%)*	*48 (100%)*	*54 (100%)*	*51 (100%)*	*48 (100%)*	*45 (100%)*	*44 (100%)*	*39 (100%)*

		GS - Frequencies			GTA - Frequencies			GTAS - Frequencies		
		D	*NP*	*R*	*D*	*NP*	*R*	*D*	*NP*	*R*
No Change	No Change	-	2 (4%)	10 (19.6%)	-	-	5 (14%)	-	-	3 (7%)
	Cautiously Continue	2 (4%)	9 (17.3%)	10 (19.6%)	2 (4.5%)	4 (10%)	11 (30.6%)	1 (2.2%)	7 (16%)	4 (9.5%)
	Other	1 (2%)	1 (2%)	3 (6%)	1 (2.3%)	-	2 (5.6%)	1 (2.2%)	-	-
Change Tx	Discuss	36 (73.5%)	34 (65.4%)	5 (10%)	23 (52.3%)	28 (70%)	1 (2.8%)	31 (69%)	31 (70.5%)	1 (2.4%)
	Consult	5 (10.2%)	3 (5.8%)	-	7 (16%)	4 (10%)	-	6 (13.3%)	3 (7%)	1 (2.4%)
	Change	1 (2%)	2 (4%)	2 (4%)	4 (9%)	2 (5%)	2 (5.6%)	1 (2.2%)	2 (4.5%)	7 (16.7%)
	Refer - Meds/Testing	4 (8%)	-	-	6 (13.6%)	2 (5%)	-	5 (11%)	1 (2.3%)	-
	Conditional Terminate	-	-	-	-	-	-	-	-	-
	Other	-	1 (2%)	3 (6%)	1 (2.3%)	-	-	-	-	-
Terminate	Relapse Prevention	-	-	10 (19.6%)	-	-	9 (25%)	-	-	17 (40.5%)
	Session Tapering	-	-	2 (4%)	-	-	3 (8.3%)	-	-	7 (16.7%)
	Graduate	-	-	6 (11.8%)	-	-	3 (8.3%)	-	-	2 (4.8%)
	Refer - Different Tx	-	-	-	-	-	-	-	-	-
	Other	-	-	-	-	-	-	-	-	-
Total		*49 (100%)*	*52 (100%)*	*51 (100%)*	*44 (100%)*	*40 (100%)*	*36 (100%)*	*45 (100%)*	*44 (100%)*	*42 (100%)*

Note. Grey highlighted data = most frequently chosen choice.

Table 7.

Perceived Influence of MFS Feature(s) on Progress Assessments and Treatment Decisions

		Progress Assessment		Tx Decision	
	Feature	**Mean[a]**	***p* value**	**Mean[a]**	***p* value**
G	*Graph*	4.26	-	3.89	-
G+T	*Graph*	4.28	0.001	4.11	<.001
	Trajectory	3.68		3.42	
G+A	*Graph*	4.07	<.001	3.74	<.001
	Alert	2.93		2.87	
G+S	*Graph*	4.26	<.001	3.98	<.001
	Support	2.92		3.26	
G+T+A	*Graph*	4.15	<.001	4.09	<.001
	Trajectory	3.51		3.28	
	Alert	3.04		2.94	
G+T+A+S	*Graph*	4.11	<.001	4.13	<.001
	Trajectory	3.09		3.20	
	Alert	2.78		2.83	
	Support	2.87		3.00	

[a]Rated 0-5, 0 = not at all [influential]; 5 = very [influential].