Title
Gauging Progress: Clinical Dashboard Use by Community-Based Clinicians

Author
Brown, Todd

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Gauging Progress:

Clinical Dashboard Use by Community-Based Clinicians

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Todd Eric Brown

2019

ABSTRACT OF THE DISSERTATION

Gauging Progress:

Clinical Dashboard Use by Community-Based Clinicians

by

Todd Eric Brown

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2019

Professor Bruce Frederick Chorpita, Chair

The use of measurement feedback systems, such as clinical dashboards, has been found to improve clinical judgments and client outcomes. However, despite the evidence demonstrating the benefits of tracking and using measurements to enhance treatment, the practice remains relatively rare amongst clinicians, even after they have received training on dashboard use. These challenges highlight the need to investigate the research-practice gap around dashboard utility in order to identify both areas for improvement as well as strengths to harness further. This dissertation sought to explore clinicians' experiences with clinical dashboards through two studies. The first study examined how various dashboard components affect clinicians' attitudes towards dashboard use and their abilities to interpret dashboard data effectively. Results found that the presence of advanced dashboard components did not subjectively improve clinicians' experiences with dashboards. However, expert users were more likely to report that data-rich

dashboards were better suited for making clinical decisions. This finding points to potential benefits in varying dashboard complexity around user levels of expertise. The second study used a mixed methods approach to explore barriers and benefits to dashboard use. Qualitative data gathered from supervisor interviews was compared with quantitative data collected from a clinician survey to examine agreements and differences related to continued dashboard use. Feedback indicated broad agreement around certain challenges, such as time constraints and lack of agency support, along with discrepancies around others, with supervisors underestimating the impact on clinicians of low agency-level prioritization. Taken together, the studies comprising this dissertation suggest that dashboard implementation efforts may be improved by designing dashboards flexibly to include content that clinicians find most useful and by targeting agency-level barriers that impede ongoing use in practice.

The dissertation of Todd Eric Brown is approved.

Denise A. Chavira

Brad J. Nakamura

Tara S. Peris

Bruce Frederick Chorpita, Committee Chair

University of California, Los Angeles

2019

Dedication

To Mom

You are like a mother to me.

Table of Contents

List of Tables

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 2

List of Appendices

Chapter 1

Chapter 2

Acknowledgments

I am lucky to have had a great many people support my journey towards the completion of this work and am unlucky to have limited space to name them all. To my advisor, Dr. Bruce Chorpita, thank you for the knowledge and insights that have shaped both this work and my psychological career. To my labmate Leslie Rith-Najarian, thank you for being by my side since the beginning. To my labmate Alayna Park, thank you for being by my side before the beginning began. To my officemate Joanna Kim, I could not imagine sharing a workspace with anyone else. To my first advisor, Dr. Judy Ho, thank you for starting me down this path and becoming a dear friend in the process. And to my family, thank you for everything.

Vita

2018-2019	Child Clinical Psychology Intern NYU Langone Child Study Center/Bellevue Hospital Center New York, New York
2014-2018	Teaching Fellow Department of Psychology University of California, Los Angeles
2013-2019	Graduate Student, Clinical Psychology Ph.D. Program Department of Psychology University of California, Los Angeles
2012-2013	Research Assistant CHILD First Lab University of California, Los Angeles
2012	Master of Arts in Psychology Pepperdine University Los Angeles, California
2011-2012	Lab Manager CREATE Lab Pepperdine University
2011-2012	Graduate Teaching Assistant Graduate School of Education and Psychology Pepperdine University
2001	Bachelor of Science in Symbolic Systems, Conferred with Distinction Stanford University Stanford, California

SELECTED PUBLICATIONS

Guan, K., Kim, R. E., Rodas, N. V., Brown, T. E., Gamarra, J. M., Krull, J. L., & Chorpita, B. F. (2018). Emergent life events: An in-depth investigation of characteristics and provider responses during youth evidence-based treatment. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication.

Guan, K., Levy, M. C., Kim, R. E., Brown, T. E., Reding, M. E. J., Rith-Najarian, L., … Chorpita, B. F. (2017). Managing in-session "surprises:" Identifying provider responses to emergent life events during evidence-based treatment implementation. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(2), 164-176.

Tsai, K. H., Moskowitz, A. L., Brown, T. E., Park, A. L., & Chorpita, B. F. (2016). Interpreting progress feedback to guide clinical decision-making in children's mental health services. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(2), 199-206.

Brown, T. E., Reding, M. E. J., & Chorpita, B. F. (2015). The uncertain steps on the certain path to progress: Some guesses about the future of cognitive and behavior therapies. *The Behavior Therapist, 38*(6), 144-150.

SELECTED PRESENTATIONS

Chorpita, B.F., Bernstein, A. D., Becker, K. D., Ebesutani, C., Park, A. L., Brown, T. E., … Daleiden, E. L. (2016, October). Measurement feedback and beyond: The place for evidence in clinical decision support. In J. Persons (Chair), *Monitoring progress in psychotherapy: Why and how.* Symposium conducted at the 50th Annual Convention of the Association for Behavioral and Cognitive Therapies, New York, NY.

Brown, T. E., & Chorpita, B.F. (2015, November). *Effects of training and expected treatment progress on clinicians' interpretations of measurement feedback systems.* Poster presented at the 49th Annual Convention of the Association for Behavioral and Cognitive Therapies, Chicago, IL.

Guan, K., Kim, R. E., Brown, T. E., Reding, M. E. J., Sun, M., Rith-Najarian, L., Levy, M., & Chorpita, B.F. (2014, November). *The impact of emergent life events on treatment integrity in EBT implementation.* In C. Arnold (Chair), Furthering treatment integrity measurement science: Exploring challenges and implications for dissemination and implementation. Paper presented at the 48th Annual Convention of the Association for Behavioral and Cognitive Therapies, Philadelphia, PA.

Brown, T. E., Park, A. L., Moore, S. A., & Chorpita, B. F. (2012, November). *How much is enough? Identifying clinically relevant information for decision-making models.* Poster presented at the 46th Annual Convention of the Association for Behavioral and Cognitive Therapies, National Harbor, MD.

Prologue

The use of measurement feedback systems (MFSs), such as clinical dashboards[1], has been found to improve clinical judgments and treatment outcomes (e.g., Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Lambert, Harmon, Slade, Whipple, & Hawkins, 2005). However, despite a growing evidence base highlighting their advantages, dashboards' benefits have remained unrealized in many settings, such as community mental health clinics, due to numerous barriers around implementations (Gleacher et al., 2016). Although these challenges can be at least partially attributable to client factors not encountered in research environments, such as more diverse populations and higher rates of comorbidity (Garland, Kruse, & Aarons, 2003; Southam-Gerow, Weisz, & Kendall, 2003), less has been examined around the interactions between clinicians and the dashboard itself in these contexts. We understand little about which dashboard features improve or interfere with community-based clinicians' use. This dissertation presents research that: (1) examines how clinical dashboard features and clinicians' knowledge affect dashboard interpretation and acceptance, and (2) identifies barriers to ongoing dashboard use in community-based settings.

Shifting the Paradigm of Evidence-Based Treatment

In recent years, clinicians, researchers, agencies, and governments alike have sought to develop and implement mental health treatments with known efficacy and effectiveness (U.S. Department of Health and Human Services, 2007). The push for evidence-based treatments (EBTs) has emerged as a promising solution for integrating science with practice, and numerous initiatives (e.g., Hogan, 2003; National Institute of Mental Health, 2008) have focused on this integration in order to improve outcomes compared to those seen in the less empirically

[1] The terms "clinical dashboards," "dashboards," and "MFSs" are used interchangeably throughout this dissertation.

supported psychological practices of the past. Initial efforts focused on the evidence-based treatment model (e.g., Chambless & Hollon, 1998), which relied on evidence to create and validate empirically-supported procedures that were most often packaged as treatment manuals. These treatments represented a great leap in the evidence-based approach to care but have had mixed results in community settings. For instance, in the child mental health literature, evidence-based treatments have generally outperformed "treatment as usual" in community settings (Weisz, Jensen-Doss, & Hawley, 2006) but have also failed at times to demonstrate meaningful gains over usual care (Barrington, Prior, Richardson, & Allen, 2005; Southam-Gerow et al., 2010; Weisz et al., 2015). These difficulties have been attributed to a variety of challenges found in community settings, including emergent life events, higher levels of comorbidity, and higher levels of symptom severity (Chorpita, Korathu-Larson, Knowles, & Guan, 2014; Garland et al., 2003; Southam-Gerow et al., 2003). When faced with these challenges, providers are significantly less likely to deliver the intended EBT (Guan et al., 2017) or to use one at all (Weersing, Weisz, & Donenberg, 2002). Thus, although the standard EBT approach finds its strength in the empirically supported evidence base, providers faced with manuals are less able to adapt to individual differences than may be possible with individualized care models (e.g., Burchard, Bruns, & Burchard, 2002) or treatment-as-usual.

Chorpita and Daleiden (2014) have identified the issue as one of *design-time/run-time* imbalance. Creating an entity with a *design time* focus results in establishing certain attributes in advance of its use. Traditional EBT manuals often prioritize a design-time approach, which handles potential uncertainty by enforcing strict rules, guidelines, and sequences at the expense of flexibility (e.g., what procedures should be in the manual? how should they be ordered?) In contrast, content created emphasizing a *run-time* focus allows for more flexibility to interact with

the environment but at the possible expense of guidelines around how to best do so, e.g., treatment-as-usual approaches that are individualized to clients' needs but in the absence of research-based guidance. Several approaches have aimed to balance the use of a knowledge base (design-time) with the ability to make informed adaptations (run-time). The Modular Approach to Therapy for Children (MATCH; Chorpita & Weisz, 2009) was developed in a manualized format that allows for treatment flexibility to accommodate differences in client status during treatment, e.g., shifting to address conduct issues that emerge amidst treatment for primary anxiety. Critically, the modular approach provides flexibility within the framework of an overall structure, represented by flow charts, that reflects empirically supported research. Children treated with MATCH improved at rates that surpassed both traditional manuals and usual care (Weisz et al., 2012). A non-manualized approach was also created with this balance in mind. The Managing and Adapting Practice (MAP) system (Chorpita & Daleiden, 2009) provides a structured, searchable database for the identification of evidence-based practices best suited for specific client profiles, e.g., a 7-year-old girl with anxiety. The balanced approach found in both MATCH and MAP enables flexibility via run-time customization while maintaining the design-time foundation and structure of empirical research.

Using Clinical Dashboards for Evidence-Informed Decisions

Given the promise of emerging treatments that balance design-time and run-time control, it is increasingly important to develop procedures and systems to inform the possible run-time adaptations that would be expected to occur. Along those lines, ongoing data measurement during treatment can be greatly beneficial to the quality of care, has been labeled as critical to ethically responsible services (Stuart & Lilienfeld, 2007), and is essential when using systems designed with a run-time/design-time balance since subjective judgments are prone to biases that

may lead to inefficient or ineffective clinical decisions. Confirmation biases and sunk cost fallacies are well-established decision-making errors that can adversely affect clinical judgment (Lilienfeld & Lynn, 2014). Further, decision-making biases are not simply due to lack of education or intelligence as higher levels of intelligence can sometimes make biases even stronger, as seen with "unrealistic optimism," where decision makers have such high confidence in their judgments that they feel that can do whatever they want without worry (Sternberg, 2004). For therapists and medical professionals alike, increased experience is associated with greater levels of confidence – but not better skill or outcomes (e.g., Marteau, Wynne, Kaye, & Evans, 1990; Stein & Lambert, 1984). Meehl (1954; Grove & Meehl, 1996) examined studies of clinicians' judgments that were clinical (i.e., informal, subjective, impressionistic) vs. statistical (i.e., formal, mechanical, algorithmic), and found that the statistical method was equal or superior to informal clinical judgment in every case examined. Another study found that decisions based on algorithms created from clinicians' self-reported decision-making processes were subsequently more valid than the clinicians' actual decisions when applied to new cases (Goldberg, 1970). These findings and biases call for a decision-making approach more in line with Meehl's statistical judgments. A more formal, statistical approach to data collection and examination provides a means for uncovering insights into what has worked and when the evidence base should be consulted for more effective practices.

Considering the benefits of using statistical data to inform treatment decisions, clinical dashboards have emerged as increasingly important tools in organizing and tracking such data around mental health treatment. Dashboards organize critical information, identify problems, monitor progress, and assist in the selection of treatment strategies (Bickman, 2008; Chorpita, Bernstein, Daleiden, & The Research Network on Youth Mental Health, 2008), and their use

improves outcomes in both adult (Lambert et al., 2005; Reese, Norsworthy, & Rowlands, 2009) and youth populations (Bickman et al., 2011). Clinicians have demonstrated superior prediction accuracy when using empirical data rather than clinical judgment (Lutz et al., 2006), and dashboards can enable this process by displaying client demographic information, progress measures, progress data, and practices delivered in each session (Chorpita et al., 2008). Additionally, dashboards can display benchmarks and expected treatment outcomes alongside observed treatment progress to provide clinicians additional context for their decision making (Chorpita, Daleiden, & Bernstein, 2016; Lambert et al., 2005).

Improving Dashboard Implementations for Community-Based Clinicians

Although numerous studies have demonstrated the clinical benefits of tracking and using measurements to improve treatment, the practice remains relatively rare. A recent national survey of providers found that only 13.9% of clinicians monitored treatment progress with standardized measures at least monthly and 61.5% never used them at all (Jensen-Doss et al., 2016). Given that clinical dashboards rely heavily on the use of standardized measures, the use rates of dashboards are likely lower still. In the same survey, 45% of clinicians reported that they would prefer not to gather any progress data. Despite – or perhaps because of – these issues, it remains of paramount importance to bridge the dashboard research-practice gap and provide clients the associated treatment outcome improvements. This multi-chapter dissertation examines factors that encourage or impede effective dashboard use by community-based clinicians.

Chapter 1: Designing and disseminating effective dashboards. The first study of this dissertation examined how dashboard features and user expertise affect community-based clinicians' interpretations of and attitudes toward dashboards. Data were collected from clinicians who had previously been trained in dashboards to find how well they were able to

interpret dashboard information and what features appeared most useful for them in doing so. The findings are presented to guide future dashboard design and education efforts.

Chapter 2: Identifying barriers to community-based dashboard usage. The second study aimed to identify barriers to sustained dashboard use by clinicians and supervisors in community mental health settings. Supervisor interviews and clinician questionnaires were used to find the perceived challenges and benefits of ongoing dashboard use. The identified shared and discrepant barriers between clinical roles point toward considerations for improved dashboard implementation efforts.

References

Barrington, J., Prior, M., Richardson, M., & Allen, K. (2005). Effectiveness of CBT versus standard treatment for childhood anxiety disorders in a community clinic setting. *Behaviour Change, 22*(1), 29–43.

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), 1114–1119.

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423–1429. http://doi.org/10.1176/appi.ps.002052011

Burchard, J. D., Bruns, E. J., & Burchard, S. N. (2002). The wraparound approach. In B. J. Bruns & K. Hoagwood (Eds.), *Community treatment for youth: Evidence-based interventions for severe emotional and behavioral disorders* (pp. 69–90). New York, NY: Oxford University Press.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9489259

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & The Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health, 35*, 114–123. http://doi.org/10.1007/s10488-007-0151-x

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566–579. http://doi.org/10.1037/a0014565

Chorpita, B. F., & Daleiden, E. L. (2014). Structuring the collaboration of science and service in pursuit of a shared vision. *Journal of Clinical Child and Adolescent Psychology, 43*(2), 323–338. http://doi.org/10.1080/15374416.2013.828297

Chorpita, B. F., Daleiden, E. L., & Bernstein, A. D. (2016). At the intersection of health information technology and decision support: Measurement feedback systems... and beyond. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 471–477. http://doi.org/10.1007/s10488-015-0702-5

Chorpita, B. F., Korathu-Larson, P., Knowles, L. M., & Guan, K. (2014). Emergent life events and their impact on service delivery: Should we expect the unexpected? *Professional Psychology: Research & Practice*, *45*(5), 387–393. https://doi.org/10.1037/a0037746

Chorpita, B. F., & Weisz, J. R. (2009). *MATCH-ADTC: Modular approach to treatment for children with anxiety, depression, trauma, and conduct problems*. Satellite Beach, FL: PracticeWise.

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *Journal of Behavioral Health Services & Research, 30*(4), 393–405. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14593663

Gleacher, A. A., Olin, S. S., Nadeem, E., Pollock, M., Ringle, V., Bickman, L., … Hoagwood, K. (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. http://doi.org/10.1007/s10488-015-0642-0

Goldberg, L. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*(6), 422–432. http://doi.org/10.1037/h0029230

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293–323. http://doi.org/10.1037/1076-8971.2.2.293

Guan, K., Levy, M. C., Kim, R. E., Brown, T. E., Reding, M. E. J., Rith-Najarian, L., … Chorpita, B. F. (2017). Managing in-session "surprises:" Provider responses to emergent life events during evidence-based treatment implementation. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(2), 164-176. http://doi.org/10.1007/s10488-015-0692-3

Hogan, M. F. (2003). The President's New Freedom Commission: Recommendations to transform mental health care in America. *Psychiatric Services, 54*(11), 1467–1474.

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*, 1–14. http://doi.org/10.1007/s10488-016-0763-0

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice

suggestions. *Journal of Clinical Psychology*, *61*(2), 165–174. http://doi.org/10.1002/jclp.20113

Lilienfeld, S. O., & Lynn, S. J. (2014). Errors/biases in clinical decision making. In *The Encyclopedia of Clinical Psychology*. doi:10.1002/9781118625392.wbecp567

Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schürch, 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. http://doi.org/10.1002/cpp.496

Marteau, T. M., Wynne, G., Kaye, W., & Evans, T. R. (1990). Resuscitation: Experience without feedback increases confidence but not skill. *BMJ, 300,* 849–850. http://doi.org/10.1136/bmj.300.6728.849

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. http://doi.org/http://dx.doi.org/10.1037/11281-000

National Institute of Mental Health. (2008). *National Institute of Mental Health strategic plan.* Bethesda, MD.

Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy Theory, Research, Practice, Training, 46*(4), 418–431. http://doi.org/10.1037/a0017901

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American*

Academy of Child and Adolescent Psychiatry, *49*(10), 1043–52. http://doi.org/10.1016/j.jaac.2010.06.009

Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2003). Youth with anxiety disorders in research and service clinics: Examining client differences and similarities. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 375–385.

Stein, D. M., & Lambert, M. J. (1984). On the relationship between therapist experience and psychotherapy outcome. *Clinical Psychology Review*, *4*, 127–142.

Sternberg, R. J. (2004). Why smart people can be so foolish. *European Psychologist*, *9*, 145-150. http://dx.doi.org/10.1027/1016-9040.9.3.145

Stuart, R. B., & Lilienfeld, S. O. (2007). The evidence missing from evidence-based practice. *The American Psychologist*, *62*(6), 615–616. http://doi.org/10.1037/0003-066X62.6.615

U.S. Department of Health and Human Services (2007). U.S. Department of Health and Human Services Strategic Plan. Washington, DC.

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child Psychology*, *31*(2), 168–180.

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., … The Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, *69*(3), 274–282. https://doi.org/10.1001/archgenpsychiatry.2011.147

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*(7), 671–689. http://doi.org/10.1037/0003-066X.61.7.671

Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. Y. (2015). Shrinking the gap between research and practice: Tailoring and testing youth psychotherapies in clinical care contexts. *Annual Review of Clinical Psychology, 11*, 139–63. http://doi.org/10.1146/annurev-clinpsy-032814-112820

Chapter 1:

Designing and Disseminating Effective Dashboards

Abstract

Dashboard vignettes were administered to a sample of Minnesota-based mental health service providers participating in a booster training session. Participants were categorized as novice or expert dashboard users and presented with vignettes that varied around the presence or absence of two dashboard components: progress benchmarks and practice panes. Collected data included: perceived adequacy of displayed information for making clinical judgments, confidence in making those judgments, overall attitudes towards dashboard use, and performance on knowledge-based items. Vignette condition did not significantly affect participants performance on knowledge items, confidence in making decisions, or overall attitudes towards dashboards. Expert users were more likely to report that complex dashboards contained the optimal level of data for decision-making purposes. These findings suggest that dashboard design and use may benefit from an approach rooted in semiformality, which would allow an adjustable presentation of data dependent on user preference and expertise.

Keywords: clinical dashboards, implementation, design, semiformality

Introduction

For many areas within the medical field, decisions are based on quantifiable information and decision trees, and for at least 60 years, the medical field has recognized that technology can play a role in that decision making (Ledley & Lusted, 1959, as cited in Shortliffe, Buchanan, & Feigenbaum, 1979). Mental health research on decision-making aids, such as clinical dashboards, began much more recently with the increased prevalence of evidence-based practices and assessment measures, and, in the past decade, research has found that the use of dashboards has a positive impact on mental health outcomes (Bickman et al., 2011; Chorpita et al., 2008, 2016; Lambert et al., 2005; Shimokawa, Lambert, & Smart, 2010). Despite these findings, little is known about which dashboard components or factors contribute to these improvements. Similarly unknown is how clinicians respond to the inclusion or exclusion of dashboard features. A recent survey of 49 measurement feedback systems (MFSs) found that nearly all examined systems tracked standardized outcomes (Lyon, Lewis, Boyd, Hendrix, & Liu, 2016); however, the presence of other tracking capabilities varied widely, reflecting the general lack of research around which features are beneficial or worthy of inclusion in these tools.

One seemingly useful feature identified across nearly 25% of dashboard systems examined by Lyon et al. (2016) was the ability to track interventions delivered by providers. As an example of this, Chorpita et al. (2008) designed and have since refined a dashboard implementation that contains a “Practice Pane” and a “Progress Pane” (see Figure 1). The Practice Pane allows providers to plot and display which therapeutic practices have been delivered in each session. Although this information alone would appear to be useful, its utility is enhanced further via the adjacent display of a Progress Pane, which displays the tracked values

of standardized and idiographic outcome measures collected from the client over time. Providers can integrate the information from each pane to make informed interpretations of client progress. For instance, a slow improvement rate on a progress measure following the introduction of a new practice can signal a provider to identify other evidence-based practices that may better target the area of concern. Although the benefits of this feature seem reasonable to assume, there are no known studies that examine how tracked practices affect clinician usage.

Another feature identified by Lyon et al. across dashboard systems was the ability to track or measure individual treatment targets. The definition and implementation of treatment targets can vary across systems and use cases. For instance, the use of expected values (e.g., Chorpita et al., 2008; Chorpita, Daleiden, & Bernstein, 2016) allows providers to track what *should* happen alongside observed values, i.e., what *has* happened. Expected practice values can be used to reflect what had been planned for each individual session. Expected progress values can be used to display a client's expected outcomes, as informed by one or several evidence bases that may be available to the clinician, e.g., theory, literature, case-specific historical information, or local aggregate evidence (Daleiden & Chorpita, 2005; Regan, Daleiden, & Chorpita, 2013). Similarly, clinical cut-off values can be displayed alongside progress measures to allow providers to track progress against targeted benchmark levels. These expected-observed comparisons can be used to inform clinical decisions via the assessment of treatment integrity, quality, and performance both within individual sessions and across multiple sessions (Regan et al., 2013). However, despite the seeming benefits of these features, no known evidence exists around the positive effects of including expected progress values or benchmarks alongside observed values. This absence of evidence may contribute to their relatively low presence across dashboard products.

Although the potential additive value of these dashboard components could provide direction for future design choices, there exists a risk of overloading providers with too much data to be interpretable. Effective dashboard design becomes paramount to address this potential issue. A simple yet hallmark design principle from the human-computer interaction literature is "know the user" (Hansen, 1971). Human-computer interaction research on user differences has focused on level-of-experience as a primary consideration in design, where experts can deal with greater complexity and need less informative feedback (Aykin & Aykin, 1991). The variable needs of experts and novices can lead to less fulfilling interactions if a user's priorities are neglected. For instance, a dashboard design may include features well-suited to novice users, such as "red light" indicators when a client is "not on track" (e.g., Lambert et al., 2005), but those same features may frustrate expert users if the underlying data that fed the indicator are obfuscated. In building systems, designers often overestimate their abilities to predefine all the ways users will want to use the system (Malone, Lai, & Grant, 2001); instead, *semiformality* is encouraged. As discussed by Malone et al. (2001), a semiformal system does not formally define all the ways a system can be used. Rather, it blurs the boundary between information acted upon by the computer vs. information to be acted upon by the user. A semiformally designed dashboard may present expert users with a more complex display relative to novices, enabling the experts to interpret a fuller set of data. Dashboard customization could allow for these novice/expert differences in design, but customizable dashboards were found in only 10% of examined systems (Lyon et al., 2016). If the principle of semiformality is important to dashboard design, the general lack of customization across dashboards may contribute to less effective or less frequent use due to experiences that are too simplified for experts or too complex for novices.

The current study examined clinicians who are dashboard users to determine how dashboard components and user level of dashboard expertise affect users' interpretations of data and attitudes towards dashboards. The research questions addressed in this study were: (1) are there benefits to tracking practices on a clinical dashboard? (2) are there benefits to displaying progress benchmarks on a clinical dashboard? and (3) do expert and novice users benefit from different levels of dashboard complexity? Given the belief that tracking practices (research question 1) and progress benchmarks (research question 2) provides valuable clinical insights, the first and second hypotheses stated that the presence of each component would lead to clinicians having more confidence in their clinical assessments and holding more positive opinions of dashboard usage. Similarly, the presence of practices and progress benchmarks was also hypothesized to improve clinicians' abilities to read dashboard data from the progress panel by providing greater context to the information displayed. Given the belief that variable levels of dashboard complexity would benefit different levels of user expertise (research question 3), the third hypothesis stated that novice users on less complex dashboards would feel more comfortable with the amount of data present, be more confident in their clinical assessments, and hold more positive opinions of dashboards, whereas expert users would experience the same benefits on more complex dashboards.

Method

Participants

Participants for this study were Minnesota-based providers who were participating in training booster sessions offered by PracticeWise, LLC around the Managing and Adapting Practice (MAP) direct service curriculum (Chorpita & Daleiden, 2009). The providers had each previously been trained on MAP in some capacity previously; however, the time since original

training periods was variable within the group and upwards of 12 years prior to the booster session. Since that originally received training, the MAP curriculum had evolved considerably in scope and content. The potential for discrepancies between providers' knowledge of and skill with MAP and its capabilities spurred the booster training effort, which aimed to familiarize providers with the current system. Prior to the training, 56 providers completed an online assessment to determine the competency of their knowledge of MAP components, including dashboards, practice guides, process guides, and the PracticeWise Evidence-Based Services (PWEBS) database, a searchable repository of randomized clinical trials of treatments for children's mental health problems. Of these 56 providers, 45 providers participated in one-day training booster sessions that were targeting these MAP components and, along with the booster, were given dashboard assessments pre- and post-training. Two booster sessions were conducted – one each at the Metro Minnesota and North Minnesota training sites – with each participant attending one (27 attendees at Metro MN, 18 attendees at North MN).

The 45 participants (35 female, 9 male, 1 unspecified), aged 28 to 74 years (M = 47.1 years, SD = 10.2 years) were primarily Masters-level clinicians (39 Masters-level, 6 Doctoral-level) and were majority Caucasian (38 Caucasian, 5 Asian, 1 Black or African American, 1 American Indian or Alaska Native). They reported primary practice settings of outpatient clinics (n = 31), intensive home- or community-based treatments (n = 7), school settings (n = 4), and out-of-home treatments, e.g., residential inpatient hospital-based or therapeutic foster care (n = 3). The providers reported an average of 16 years experience (SD = 8.1 years) and had all previously been trained on MAP to some degree (M = 6.3 years since training, SD = 3.5 years). Regarding dashboard usage, most reported never or almost never using them (n = 31) while the remainder reported using them once in awhile (n = 13) or "about as often as not" (n = 1). No

significant differences were found between sites on age, gender, ethnicity, degree, primary treatment setting, years of experience, years since MAP training, or reported frequency of dashboard use.

Materials

Prescreening dashboard knowledge assessment. The prescreening dashboard assessment contained 29 questions to gauge competency and six questions to assess dashboard usage and opinions (see Appendix A). The 29 questions were based on four vignettes, which were presented as static images of PracticeWise dashboards that displayed mock treatment information, including progress panes (i.e., measures and collected data); practice panes (i.e., practices delivered in each session); and basic client demographics. The questions were designed to focus on different dashboard skills and domains. For skills, 17 questions assessed the ability to read the dashboard (e.g., identify demographic info, assessment measures and values, etc.); five questions assessed the ability to select measures for the dashboard (e.g., choosing standardized measures that would be most useful in making clinical decisions); and seven questions assessed the ability to integrate all dashboard components and make interpretations. These same 29 questions were also categorized as focusing on certain dashboard domains, including context (e.g., basic demographic info; two items), progress measures (13 items), practices delivered (10 items), and the integration of progress and practice measures (four items).

Participants' answers were scored against consensus correct responses as identified by the PracticeWise development team. Participants were categorized as "novice" or "expert" dashboard users based on how they performed relative to others on the prescreening assessment. Based on the pool's performance, the users were split around the median number of correct responses, with 31 participants falling into the lower performing "novice" range (11-17 answers

correct) and 25 participants falling into the higher performing "expert" range (18-22 answers correct).

Pre- and post-training dashboard assessments.

Pilot. Pre- and post-training measures were piloted by five advanced clinical psychology doctoral students who recorded their responses in a Word document, and their feedback was used to modify knowledge questions for clarity.

Conditions. After participants were split into novice and expert dashboard users based on their prescreening performance, they were randomly assigned to conditions that reflected the type of content they would see within their respective pre- and post-training dashboard assessments. The random assignment was conducted within each expertise level to ensure a balanced spread of conditions, which varied across two independent variables: (a) progress benchmark line present/absent, and (b) practice pane present/absent. Participants in the "Practice/Benchmark" condition received vignettes that displayed practice panes and progress measure benchmarks along with progress panes (see Figure 2). The "Practice/No-Benchmark" condition had vignettes with practice and progress panes but without progress measure benchmarks (see Figure 3). "No-Practice/Benchmark" vignettes displayed progress panes and progress measure benchmarks but no practice panes (see Figure 4). "No-Practice/No-Benchmark" vignettes displayed progress panes only, i.e., neither a practice pane nor progress measure benchmarks were displayed (see Figure 5). All other vignette content and questions were identical across all conditions.

Vignettes. The pre- and post-training measures each had the same structure (see Appendices B & C). Two vignettes were presented, with progress measures always present and with practice panes and benchmarks present or absent per the participant's condition as discussed

above. For each vignette, three questions focused on dashboard content, e.g., what value did the client have on this measure when she began treatment? These knowledge questions were designed to be answerable even if only the progress pane was visible, i.e., information critical to answering the questions was present regardless of a participant's condition. Following each set of knowledge questions, participants were asked to assess treatment progress on a Likert scale (1 – *very poor* to 7 – *very good*), followed by their confidence in that decision on a Likert scale (1 – *not at all confident* to 7 – *very confident*) and how they felt about the amount of information available to them in making that decision (1 – *too little* through 4 – *just right* through 7 – *too much*). Participants were then asked a multiple-choice question of what they would do next in treatment: (a) *Repeat the practice that was most recently delivered;* (b) *Repeat a practice that was previously delivered (but not most recently);* (c) *Change to a new practice;* (d) *Initiate terminating phase of treatment;* or (e) *I do not know what I would do next in treatment*. As with the progress decision, participants were then asked to assess their confidence in their decision along with their comfort regarding the amount of displayed information.

Usage questions. After answering questions regarding the two vignettes, participants were asked their opinions on four clinical dashboard usage items, using Likert scale ratings of 1 (*completely disagree*) to 7 *(completely agree*). These items, also present in the prescreening measure, were adapted from the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, Morris, Davis, & David, 2003). UTAUT was constructed based on conceptual and empirical similarities found from a comparison of eight previously developed models that examined individual acceptance of new information technologies. UTAUT demonstrated an adjusted R^2 value of .70, explaining 70% of the variance in users' intentions to use information technology. The four questions selected for use (see final items in Appendices B

& C) have previously been found by the author to most vary with positive and negative opinions of dashboard usage (Brown, 2014), with question 1 reverse coded to align higher values with more positive opinions.

Procedure

All potential training recipients completed the online prescreening assessment in November or December 2016 to assess their levels of MAP competency with dashboards, practice guides, process guides, and the PWEBS database. Based on assessment results, participants were categorized as novice or expert dashboard users based on their assessment performance relative to the median.

In March 2017, participants who had completed the prescreening assessment were invited to attend one-day, in-person MAP training booster sessions. Participants were first provided feedback on their performance on the pre-training assessment and then immediately completed a pre-training, paper-based dashboard assessment measure containing two dashboard vignettes. Within each level of expertise (based on the prescreening assessment), participants were randomly assigned to conditions that determined the presence or absence of practice and benchmark information displayed on their dashboard vignette. Following administration of the pre-training assessment, experienced PracticeWise trainers delivered the MAP booster session content, with the same trainers used during each training session. Since all participants had had some level of experience with MAP, training content included didactic and rehearsal exercises that generally focused on advanced or more difficult concepts and resources. For dashboards, didactic training focused on observed vs. expected/benchmark values and how they might be represented and interpreted on both the progress and practice panes (see Appendix D). Participants also built a dashboard together as a large group and had an opportunity to build their

own dashboards, using an existing case if possible. At the close of the booster session, participants completed a post-training, paper-based dashboard assessment measure containing two additional vignettes. Participants remained in the same randomly assigned condition as their pre-training assessment with regards to the display of practice panes and progress benchmark values. Trainers observed that all participants appeared to engage thoughtfully with the assessments and completed the tasks on their own.

Results

Independent samples t-tests were conducted to ensure that participants at each training site did not differ significantly from each other. No significant differences between groups were found on prescreening dashboard expertise levels or participant characteristics such as age, education level, and experience.

Data Reconciliation and Inter-Rater Reliability.

During the training sessions, technical challenges arose that necessitated that the pre- and post-training measures be distributed and collected via paper-based questionnaires rather than the originally planned internet-based survey. As such, unforeseen complications arose around participants' responses to the paper-based Likert scale items, which asked participants to rate their confidence in making decisions based on progress and practice information displayed in each vignette and to rate the adequacy of the information displayed for making these decisions. Roughly 36% (16/45) of respondents recorded numeric answers for these items, whereas the remainder of the respondents (64%, 29/45) recorded their answers by making tick marks on the graphical representations of the items' Likert scales.

The tick-mark response style necessitated a data reconciliation phase in which the author and a second coder independently measured each tick to identify the leftmost and rightmost

boundaries of each mark to find a midpoint measurement, which was divided by the length of each Likert scale's paper representation to obtain an estimated whole number value. Interclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated using SPSS statistical package version 22 (SPSS Inc, Chicago, IL) based on a single-rater, absolute-agreement, two-way mixed-effects model comparing coders' estimates of tick-marked Likert values. The single measure ICC was .984, $F(885, 885) = 124.95$, $p < .001$, 95% CI [.982, .986], indicating excellent reliability (Koo & Li, 2016). With this finding, the author's recorded data were used as the primary data source for study analysis purposes.

Questions

Question 1: Are there benefits to tracking practices on a clinical dashboard? The first research question was investigated by comparing participants who received pre- and post-training vignettes with a practice pane present ($n = 22$) vs. those who received vignettes with a practice pane absent ($n = 22$). The hypothesis proposed that the presence of practice panes would be associated with: (a) greater confidence in clinical assessments of dashboard data; (b) more positive opinions of dashboard usage; and (c) improved performance on dashboard knowledge items. For question 1a, conditions were compared based on the total of four items that gauged participants' confidence ratings around choosing how to proceed in each vignette's treatment, e.g., whether to repeat, change, or terminate a treatment practice. Practice-oriented confidence ratings did not significantly differ between the practice-pane-present condition ($M = 16.00$, $SD = 3.35$, range: 10-23) and the practice-pane-absent condition ($M = 15.57$, $SD = 4.05$, range: 7-22), $t(42) = 0.39$, $p = .702$, $d = 0.12$. For question 1b, conditions were compared based on the total of UTAUT-based items that gauged participants' views on using dashboards. Views towards dashboard use did not significantly differ between those with practice panes ($M = 44.38$, $SD =$

7.07, range: 24-56) and those without (M = 45.23, SD = 8.93, range: 17-56), $t(41) = 0.34$, p = .733, d = 0.11. For question 1c, the number of correct answers on the 12 dashboard knowledge items was examined. Performance on dashboard knowledge items with practice panes present (M = 8.27, SD = 1.78, range: 3-12) did not significantly differ from performance with practice panes absent (M = 8.36, SD = 1.81, range: 4-10), $t(42) = 0.17$, p = .867, d = 0.05. Thus, contrary to hypotheses, no significant subjective benefits were observed in the presence of practice panes.

Question 2: Are there benefits to displaying progress benchmarks on a clinical dashboard? The second research question was investigated by comparing participants who received pre- and post-training vignettes with progress benchmarks present (n = 22) vs. those who received vignette with progress benchmarks absent (n = 22). The hypothesis proposed that the presence of progress benchmarks would be associated with: (a) greater confidence in clinical assessments of dashboard data; (b) more positive opinions of dashboard usage; and (c) improved performance on dashboard knowledge items. For question 2a, conditions were compared based on the total of four items that gauged participants' confidence ratings around how well each vignette's client was progressing. Progress-oriented confidence ratings did not significantly differ between the benchmark-present condition (M = 17.00, SD = 2.64, range: 12-22) and the benchmark-absent condition (M = 17.45, SD = 3.44, range: 12-26), $t(42) = 0.49$, p = .626, d = 0.15. Exploratory post hoc tests were conducted on novice pre-training confidence levels around progress judgments, and found that they were lower when progress benchmarks were present (M = 7.70; SD = 1.16, range: 6-9) vs absent (M = 9.25, SD = 2.18, range: 6-13); however, this difference was not found to be significant, $t(18) = 2.07$, p = .061, d = 0.92. Benchmark presence had no significant difference on novices' post-training confidence ratings nor on experts' confidence ratings at either pre- or post-training assessments.

For question 2b, benchmark present/absent conditions were compared based on the total of UTAUT-based items that gauged participants' views on using dashboards. Views towards dashboard use did not significantly differ between those with progress benchmarks (M = 46.38, SD = 4.91, range: 33-56) and those without (M = 43.31, SD = 10.00, range: 17-56), $t(41) = 1.27$, $p = .213$, $d = 0.39$. Exploratory analyses found no significant differences on dashboard opinions based on user expertise or pre/post training timepoints.

For question 2c, the number of correct answers on the 12 dashboard knowledge items was examined. Performance on dashboard knowledge items with progress benchmarks present (M = 7.91, SD = 2.18, range: 3-12) did not significantly differ from performance with progress benchmarks absent (M = 8.73, SD = 1.16, range: 6-10), $t(42) = 1.55$, $p = .130$, $d = 0.47$. Notably, although not significant, performance suffered in the presence of progress benchmarks, contrary to the hypothesis that performance would improve. Exploratory post hoc tests revealed significantly lower knowledge scores for novice participants with benchmarks present (M = 6.9, SD = 2.38, range: 3-9) vs. absent (M = 8.8, SD = 1.14, range: 7-10), $t(18) = 2.28$, $p = .035$, d = 1.02, whereas experts demonstrated no significant differences whether benchmarks were present or absent (respectively, M = 8.75, SD = 1.65, range: 6-12; M = 8.67, SD = 1.23, range: 6-10; $t(22) = 0.14$, $p = .890$, $d = 0.06$). Further exploration revealed that novices' pre-training dashboard knowledge scores were significantly lower when benchmarks were present (M = 3.20, SD = 1.03, range: 1-4) vs. when benchmarks were absent (M = 4.37, SD = 0.81, range: 3-5), $t(19) = 2.89$, $p = .009$, $d = 1.25$. Benchmark presence had no significant difference on novice's post-training knowledge scores nor on expert's knowledge scores at either pre- or post-training assessments. These exploratory findings suggest that displaying progress benchmark to novices without training may impair their ability to interpret the data.

Question 3: Do expert and novice users benefit from different levels of dashboard complexity? The third research question was investigated by separately examining whether novice and expert users varied across levels of dashboard complexity on several measures, including: opinions towards dashboard use; confidence in making clinical decisions; and ratings of how well the displayed data fit their needs for making clinical interpretations. These variables were examined using pre- and post-training measurements as well as an overall combined total. For standardization purposes, the ratings of data display adequacy (measured on a 1-7 Likert scale, with anchors of 1 = *too little*, 4 = *just right*, and 7 = *too much*) were normalized by calculating values relative to the *just right* value, e.g., a rating of 5 was rescaled to +1, a rating of 2 was rescaled to -2, etc. A one-way ANOVA was used to investigate differences between the stated variables across the four display conditions (progress benchmarks present/absent, practice pane present/absent) for each level of user expertise (novice, expert).

For novices, one-way ANOVA revealed no statistically significant differences between conditions on pre-, post-, or total dashboard opinions, confidence ratings, or ratings of display adequacy (see Table 1). For experts, one-way ANOVA revealed a statistically significant difference between groups on post-training ratings of the adequacy of the displayed data, $F(3, 20) = 4.20$, $p = .019$, $d = 1.57$ (see Table 2). No other statistically significant differences were observed. A Tukey post hoc test revealed that data display adequacy ratings were statistically significantly closer to *just right* for experts who saw both progress benchmarks and practice panes (1.17 ± 3.25) compared to those who saw only practice panes (-3.50 ± 1.87, $p = .046$) and those who saw neither (-4.17 ± 2.93, $p = .020$). There was no statistically significant difference with those who saw only progress benchmarks ($p = .339$).

Discussion

This study sought to explore the impact of dashboard features on clinicians' effective use of the tools in clinical decision-making processes. The first research question focused on the practice pane component, which allows clinicians to track practices planned and used in client sessions. The findings reveal that, contrary to the hypothesis, clinicians' confidence in their interpretations and their attitudes towards using dashboards did not improve in the presence of practice panes. Similarly, contrary to the hypothesis, the presence of practices on a dashboard did not improve clinicians' abilities to read dashboard data. However, additional post-hoc consideration of the provided dashboard knowledge items highlighted that the items' content did not draw on information provided by the practice pane, which likely contributed to the null finding. Practice-oriented dashboard knowledge items, such as "does it appear that the client received exposure for anxiety?", would allow for more appropriate exploration of dashboard knowledge as relevant to practice panes. Nonetheless, the overall practice pane findings imply that the display of practices does not subjectively improve clinicians' dashboard experience as compared to using dashboards that track progress measures alone.

The second research question focused on the use of benchmarks or clinical cutoff scores alongside graphed progress measurements. Once again, the overall findings were contrary to the hypotheses that the presence of benchmarks would improve clinicians' ability to read dashboard data, their confidence in doing so, and their general attitudes towards dashboards. Notably, within these results, data showed that the presence of progress benchmarks prior to booster training decreased novice users' dashboard knowledge performance and was associated with a similar (though not statistically significant) negative effect on novice users' confidence levels in making progress-based interpretations. These adverse effects were no longer present following

training and were not present at all for expert dashboard users. This finding indicates that progress benchmarks may be better suited for trained and expert users, who are more able to handle greater complexity (Aykin & Aykin, 1991)

The final research question examined the topic of dashboard complexity more broadly across both novice and expert users. Although an earlier discussed result found that benchmarks adversely affected untrained novice users around progress interpretations, here we found that the level of dashboard complexity had no effects overall on novice users' attitudes towards dashboards, their confidence in making judgments, or their evaluations of whether the amount of data displayed was optimal. The level of dashboard complexity also did not affect expert users' dashboard confidence or attitudes. However, expert users were more likely to report that complex dashboards contained the optimal level of data for decision-making purposes. The seeming contradiction of unaffected confidence levels even when recognizing inadequate data may be attributable to tendencies to rely on clinical judgment over actuarial data (Dawes, Faust, & Meehl, 1989).

Several limitations should be considered regarding the current study. Most notably, as discussed in the method section, data collection on Likert items was hampered due to paper-based measures that provided affordances for tick marks rather than numeric responses. Confusion in completing the Likert scale items, i.e., participants' confidence levels and judgments of information adequacy, contributed to a varied response style across participants, whereby discrete values were specified by only one third of respondents. Although the data reconciliation effort was completed with an excellent level of inter-rater reliability, this reliability reflects agreement on the interpretation of the tick mark values but does not ensure a data set free from random error for the recorded values. As such, the reconstructed data may not

best reflect the values that participants wished to record, raising potential concerns around the validity of the related results and subsequent conclusions. The null findings related to confidence and judgments of information adequacy may also be a consequence of these data collection challenges. Additionally, given the number of conditions and variables that were examined for this study, the sample size does not provide for significant statistical strength, particularly around the post hoc exploratory findings. Nevertheless, those findings alongside the main effects may act as starting points for additional investigation of dashboard components.

Several considerations are warranted around the definition of dashboard users and their various levels of expertise. Although the term "user" is applied to study participants, the reported low levels of pre-training dashboard use by participants highlight that participants may be best not considered active dashboard *users.* Their actual levels of dashboard use may limit the ability to draw conclusions around full-fledged and active dashboard users. The study's findings may instead reflect another construct, such as adeptness around working with novel visual displays of information. Regarding dashboard user expertise, conclusions related to expertise levels may be viewed with caution since the categorization of novice and expert users was conducted relative to the available sample rather than relative to identified benchmarks for dashboard expertise, which do not currently exist. Furthermore, challenges arise in assessing the appropriate level of internal consistency for the dashboard knowledge test items. Measures such as Cronbach's alpha provide a means to assess this factor but may be difficult to apply meaningfully across the given knowledge items. Although the items were all dashboard-related, they assessed a variety of domains that may not be best conceptualized as a single coherent construct, such that internal consistency as measured by alpha may not be desirable (Taber, 2018). Given the range of dashboard domains targeted within these measures and the relatively few items examined, the

dashboard knowledge measurement should be viewed with caution and would benefit from a larger number of items and vignettes to better cover the relevant domains. Nonetheless, future directions may be informed by the differences observed in this study between varying levels of dashboard knowledge skills. Additionally, although the study examined variables related to clinicians' clinical judgments, such as confidence, it did not examine the quality of the judgments themselves. The current study's method originally included clinical progress ratings determined for each vignette via expert consensus, but further discussion highlighted the challenges inherent in identifying objectively "correct" clinical interpretations while using non-trivial vignette examples. As such, clinicians' ability to interpret clinical data on dashboards accurately remains unknown. However, this study's use of dashboard knowledge items may serve as a proxy for reading dashboard data.

The study focused on a MAP dashboard implementation and training protocol, which may limit the generalizability of the findings to other dashboard implementations. However, the dashboard components examined here – progress benchmarks and practices delivered – appear in some form in numerous other dashboard implementations (Lyon et al., 2016), so general conclusions around their utility to clinicians may be used to spark further investigations elsewhere. Nevertheless, a broader sample of dashboard implementations would provide a stronger conclusion around these components' utility to clinicians. Future research to replicate this study's findings may be conducted in the service of a broader examination of the value of dashboard components at a modular level. A modular approach to dashboard research and design would allow for increased flexibility to account for user level of expertise, thus embracing the tenets of semiformality. In considering the findings related to dashboard complexity, it should be noted that complexity was operationalized here based on the number of dashboard components

present or absent. A more direct measurement of participants' perceptions of dashboard complexity would provide clearer insight into preferences around these variable levels of display, and future studies would benefit from assessing and interpreting this information accordingly. A closer examination of level of expertise as a main effect would also be beneficial to more directly explore how confidence levels change with increased dashboard expertise. Additionally, dashboard implementations would benefit from additional user-focused design with an emphasis on the user interface and experience (UI/UX). Improvements to dashboard usability would likely also improve user attitudes towards dashboards and, consequently, user willingness to use them. Collaborative efforts with design departments and professional services would likely provide insights uncommon to research psychologists. Additional opportunities may be available in association with groups such as the Healthcare Information and Management Systems Society (HIMSS), a not-for-profit organization focused on better health through information and technology

Taken together, this study's findings suggest initial considerations into designing more effective dashboards. Most notably, optimal implementations would benefit from careful introduction of new features. Although progress benchmarks and tracked practices may provide valuable information to assist in treatment planning, neither feature was found to provide a subjectively improved experience for users of dashboard implementations. However, if new features such as these are added, more novice users may require training to avoid being overwhelmed. Additionally, as expert users are made aware of the possibility of more detailed information, they may begin to consider starker implementations inadequate for their needs. Dashboard designs grounded in semiformality may best serve this balance as we aim to continue to improve clinicians' dashboard experiences in the service of improving client outcomes.

Tables, Figures, and Appendices

Table 1

One-way ANOVA for Novice Users across Four Vignette Conditions

	Pre-training		Post-training		Total	
Ratings	*F*	*p*	*F*	*p*	*F*	*p*
Dashboard Use Attitudes	(3, 17) = 0.34	.796	(3, 15) = 1.01	.416	(3, 15) = 0.66	.589
Confidence	(3, 16) = 0.57	.645	(3, 16) = 0.26	.855	(3, 16) = 0.49	.694
Adequacy of Displayed Data	(3, 15) = 0.65	.595	(3, 16) = 1.49	.255	(3, 15) = 0.86	.484

Note. Four conditions include progress benchmark present/absent and practice panes present/absent. Examined ratings represent pre-training, post-training, and total measurements of dashboard use attitudes, confidence in clinical judgments, and rated adequacy of displayed data for making clinical decisions.

Table 2

One-way ANOVA for Expert Users across Four Vignette Conditions

	Pre-Training		Post-Training		Total	
Ratings	*F*	*p*	*F*	*p*	*F*	*p*
Dashboard Use Attitudes	(3, 20) = 0.20	.894	(3, 20) = 0.24	.868	(3, 20) = 0.23	.878
Confidence	(3, 20) = 0.26	.854	(3, 20) = 2.49	.089	(3, 20) = 1.13	.363
Adequacy of Displayed Data	(3, 20) = 0.63	.603	(3, 20) = 4.20	.019*	(3, 20) = 2.87	.062

Note. Four conditions include progress benchmark present/absent and practice panes present/absent. Examined ratings represent pre-training, post-training, and total measurements of dashboard use attitudes, confidence in clinical judgments, and rated adequacy of displayed data for making clinical decisions.
*$p < .05$.

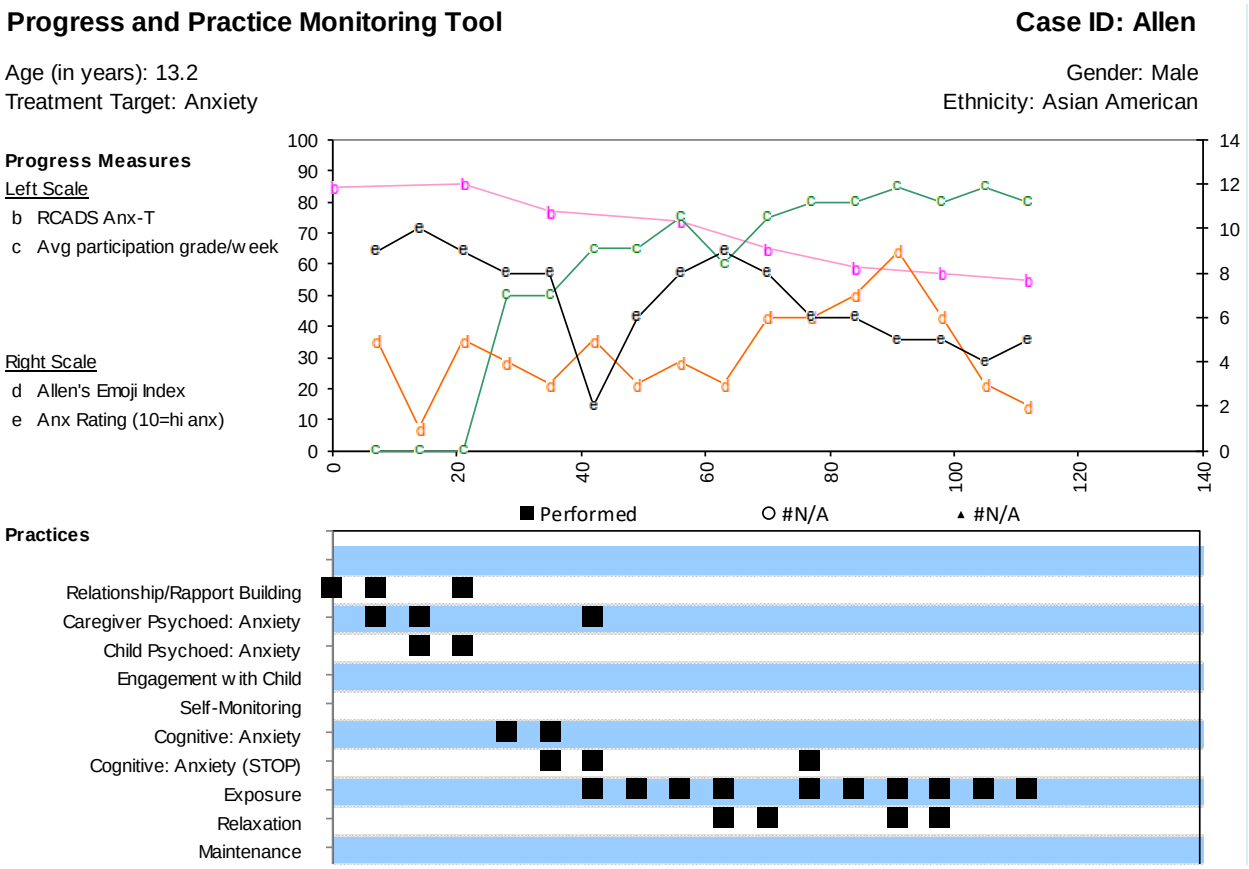


Figure 1. Clinical dashboard with progress and practice panes (template provided by PracticeWise LLC)

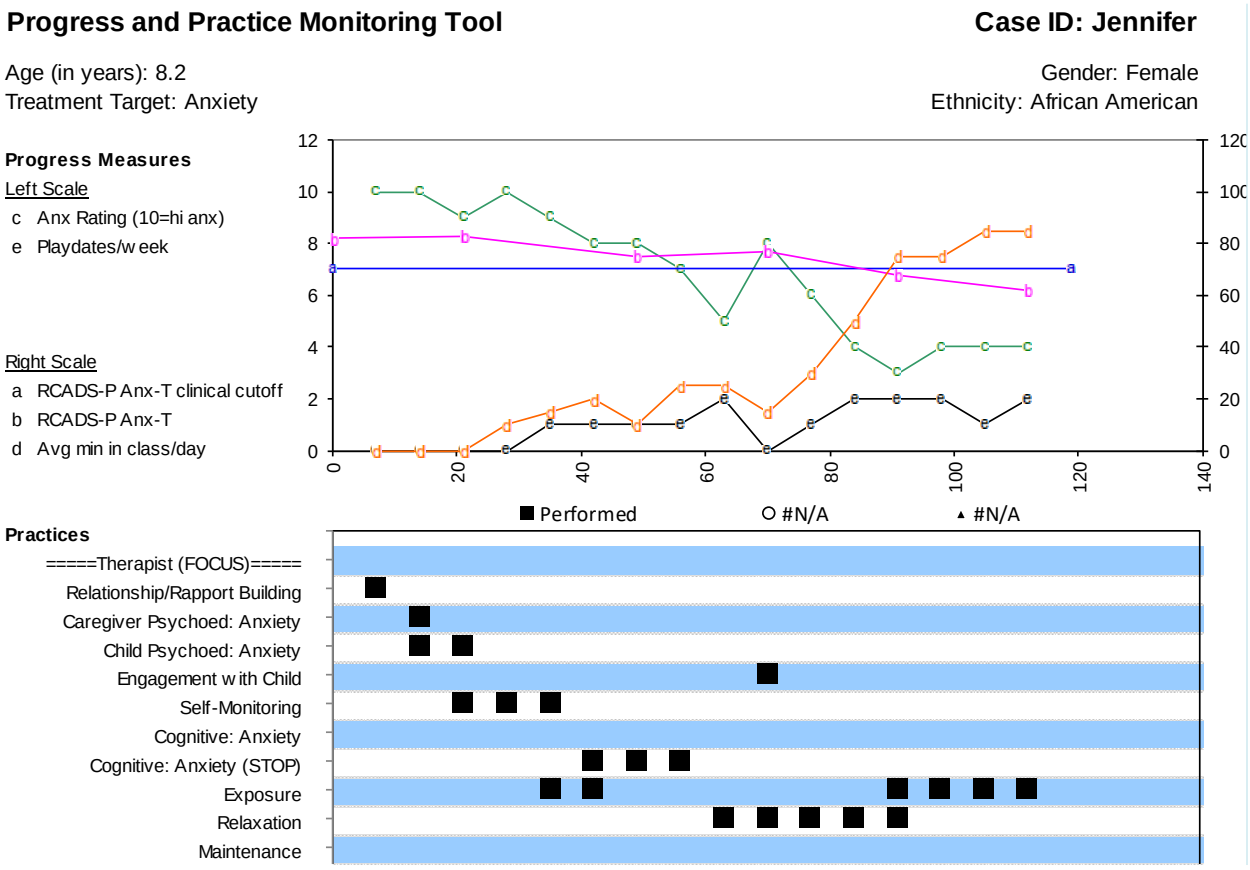


Figure 2. Sample dashboard vignette: Practice/Benchmark condition

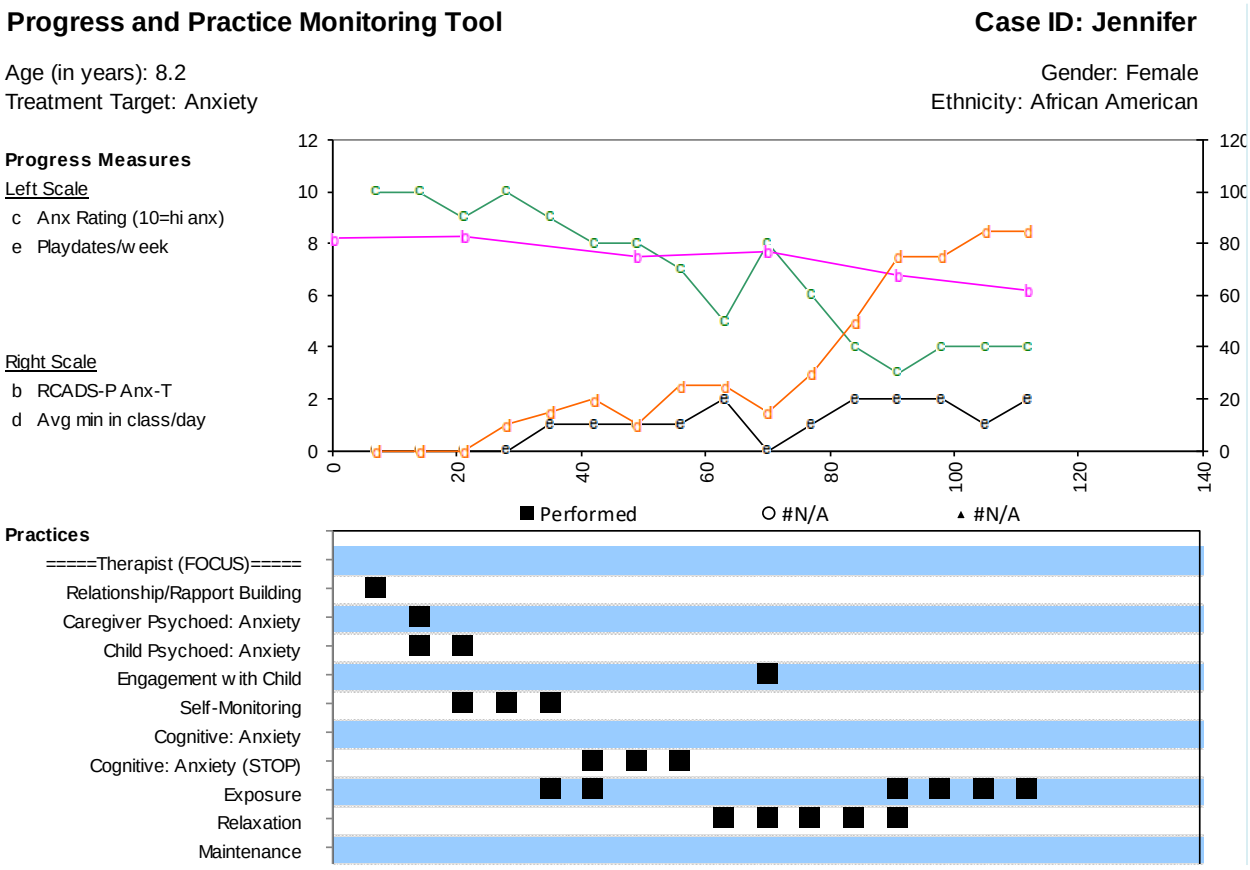


Figure 3. Sample dashboard vignette: Practice/No-Benchmark condition

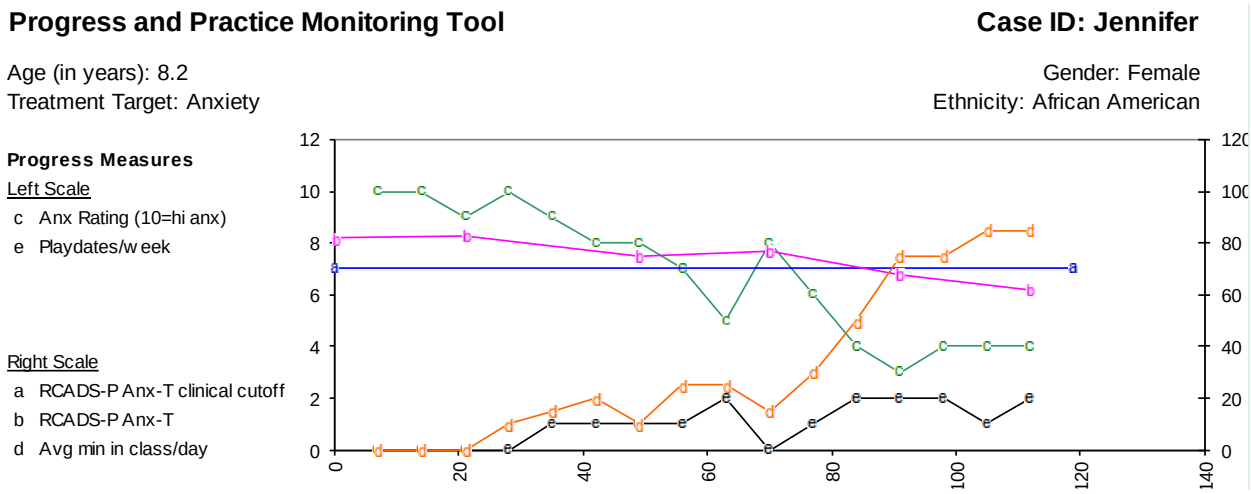


Figure 4. Sample dashboard vignette: No-Practice/Benchmark condition

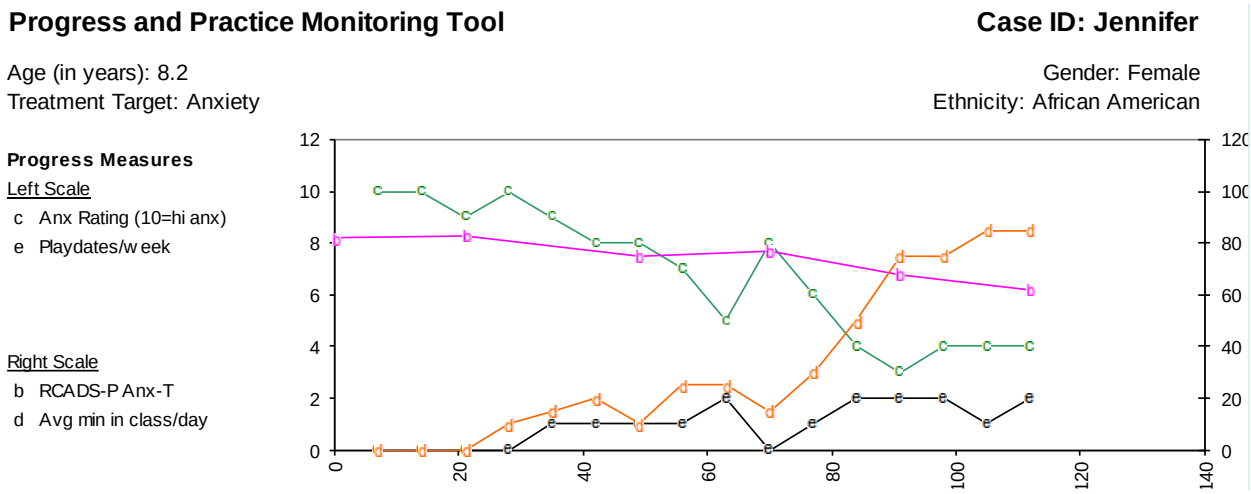
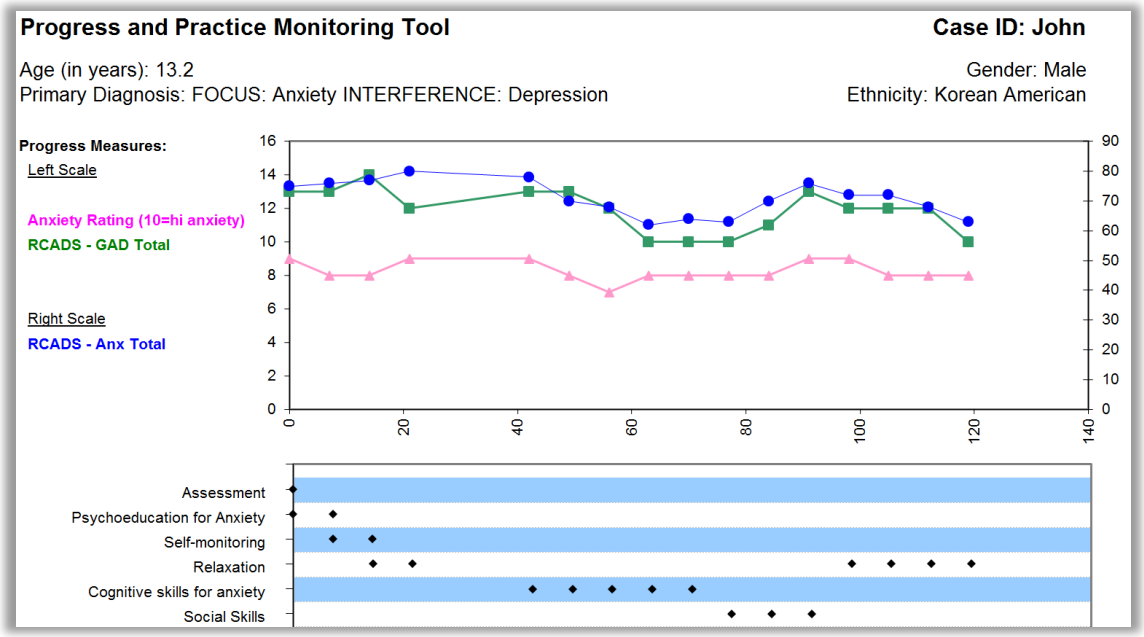


Figure 5. Sample dashboard vignette: No-Practice/No-Benchmark condition

Appendix A: Prescreening Dashboard Knowledge Assessment

Dashboard Reasoning Measure

Vignette 1:



Items for Vignette 1:

1. What is the focus of treatment for John?
   a) Anxiety
   b) Conduct
   c) Depression
   d) Trauma

2. What was John's peak rating on the Total Anxiety scale of the RCADS?
   a) 9
   b) 14
   c) 63
   d) 80

3. Over the last three sessions, John's Anxiety Rating has
    a) Improved
    b) Deteriorated
    c) Stayed the same
    d) Unable to determine based on the information provided

4. Which practice did John receive most often?

    a) Self-monitoring
    b) Relaxation
    c) Cognitive skills for anxiety
    d) Social Skills

5. Which practice did John receive most recently?

    a) Assessment
    b) Psychoeducation for Anxiety
    c) Relaxation
    d) Social Skills

6. John finds it difficult to stay in class due to high levels of general anxiety. These challenges have also led to a lower mood. His primary treatment goals are to reduce anxiety and increase his time spent in class. Which of the following would be the best dashboard modification to improve monitoring of John's progress?

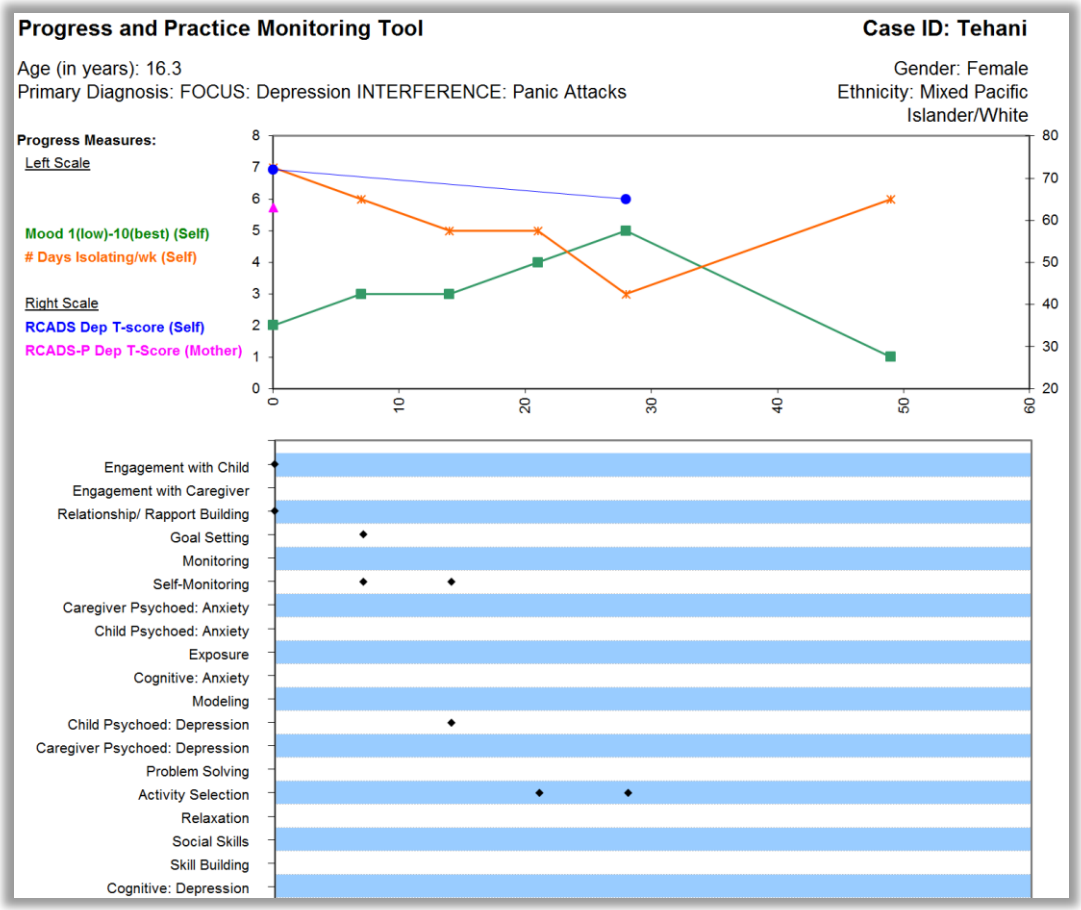
    a) Add "RCADS-Social Phobia Total score self-report" (measured once a month)
    b) Add "RCADS-Major Depression score self-report" (measured once a month)
    c) Add "Average minutes in class/day teacher-report" (measured once a week)
    d) Replace RCADS measures (self-report) with RCADS-P measures (parent-report)

7. Consider how measurement of treatment progress could be improved on John's dashboard. Which of the following suggestions is LEAST likely to help improve the communication of John's treatment progress?
    a. Add who the reporter was for each measure (e.g., self-report)
    b. Note whether the raw score or t-score is plotted for the RCADS scales
    c. Add a benchmark clinical cutoff line
    d. Move the RCADS GAD-Total measure to the right scale with the RCADS Anx Total measure

8. The original PWEBS search results for John's anxiety returns self-monitoring, relaxation, and exposure as some of the most common practices found in research studies. Based on his goals and the practices delivered so far, what would be the best treatment course for John?

    a) Continue with "Relaxation" module
    b) Begin "Exposure" module
    c) Return to "Social Skills" module
    d) Begin depression treatment with "Psychoeducation for Depression" module

Vignette 2:



Items for Vignette 2:

9. How old is Tehani?

    a) 3
    b) 16
    c) 63
    d) Unable to determine

10. What was Tehani's initial T-score on the RCADS Depression scale (self-report)?

    a) 6
    b) 7
    c) 65
    d) 72

11. How many measures are in the progress pane?
    a) 1
    b) 2
    c) 3
    d) 4

12. How many clinical events (sessions) have occurred according to this dashboard?

    a) 1
    b) 4
    c) 5
    d) 8

13. Which progress measure on Tehani's dashboard was collected the fewest number of times?

    a) Mood (self-report)
    b) Number of days isolating per week (self-report)
    c) RCADS (self-report)
    d) RCADS-P (parent-report)

14. Over the first five sessions, Tehani's depression

    a) improved slightly.
    b) got slightly worse.
    c) got significantly worse.
    d) was variable.

15. Which of the following practices has not been delivered during treatment so far?

    a) Engagement with Caregiver
    b) Goal Setting
    c) Child Psychoeducation: Depression
    d) Activity Selection

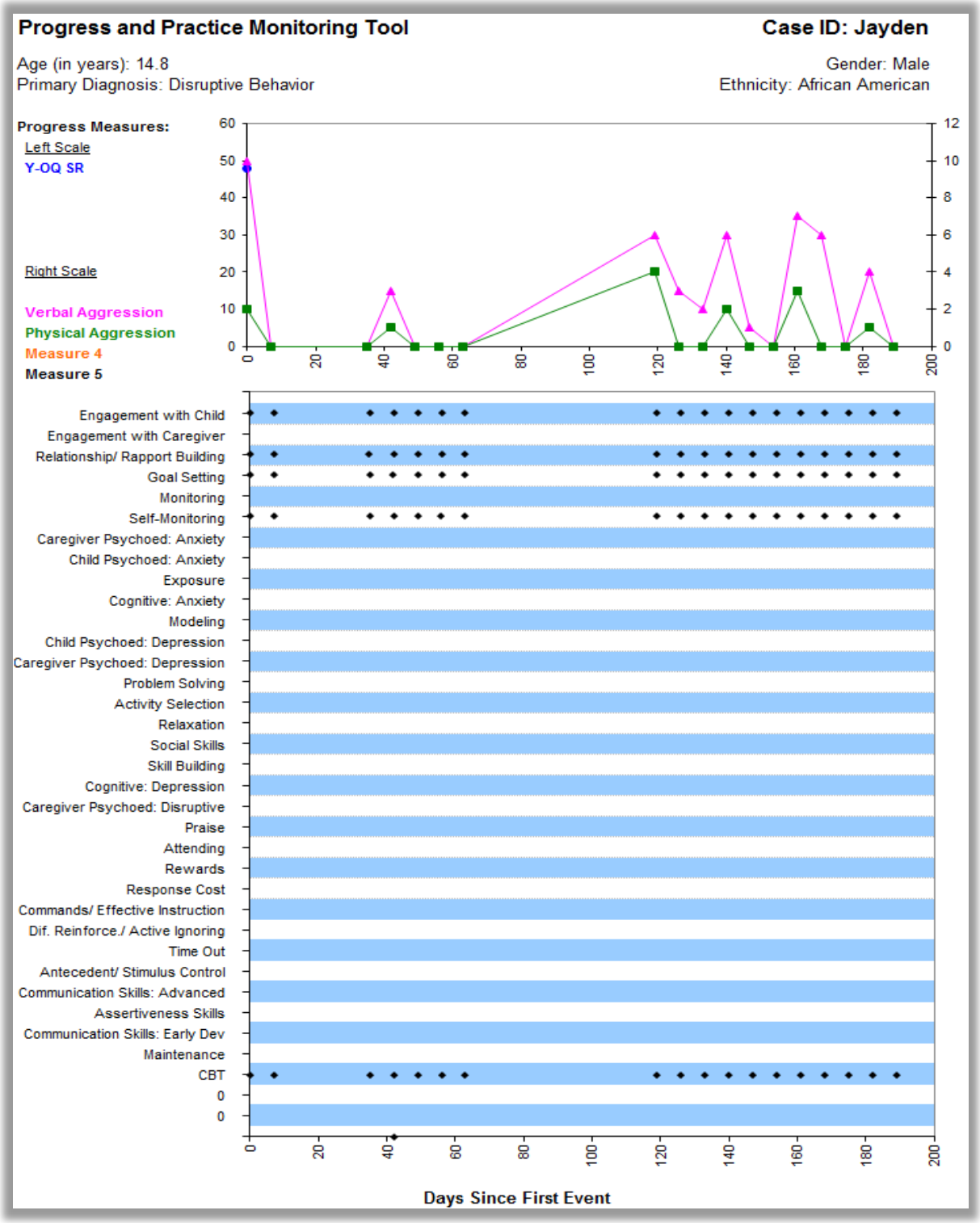
16. Which of the following practices has been repeated in two sessions?

    a) Engagement with Child
    b) Relationship/Rapport Building
    c) Goal Setting
    d) Self-Monitoring

17. After coming in weekly for a month of treatment, Tehani no shows for two appointments in a row. When she comes on the third week, she reports a decrease in mood and an increase in isolating behaviors. Which of the following responses would be the LEAST effective use of the clinical dashboard?

    a) Add an idiographic measure of panic attacks
    b) Return to Engagement with Child to assess barriers to attending treatment
    c) Start treatment for panic (e.g., exposure for panic) without adding an assessment measure
    d) Repeat Goal Setting and determine if it is necessary to switch focus to treatment of anxiety

Vignette 3:


Progress and Practice Monitoring Tool
Case ID: Jayden
Age (in years): 14.8
Primary Diagnosis: Disruptive Behavior
Gender: Male
Ethnicity: African American
Progress Measures:
Left Scale
Y-OQ SR
Right Scale
Verbal Aggression
Physical Aggression
Measure 4
Measure 5
Engagement with Child
Engagement with Caregiver
Relationship/ Rapport Building
Goal Setting
Monitoring
Self-Monitoring
Caregiver Psychoed: Anxiety
Child Psychoed: Anxiety
Exposure
Cognitive: Anxiety
Modeling
Child Psychoed: Depression
Caregiver Psychoed: Depression
Problem Solving
Activity Selection
Relaxation
Social Skills
Skill Building
Cognitive: Depression
Caregiver Psychoed: Disruptive
Praise
Attending
Rewards
Response Cost
Commands/ Effective Instruction
Dif. Reinforce./ Active Ignoring
Time Out
Antecedent/ Stimulus Control
Communication Skills: Advanced
Assertiveness Skills
Communication Skills: Early Dev
Maintenance
CBT
0
0
Days Since First Event

Items for Vignette 3:

18. Jayden had a large gap in treatment sessions, returning around Day ____.

    a) 42
    b) 62
    c) 119
    d) 189

19. Jayden's level of aggression on Day 42 of treatment is _____ what is reported on Day 161.

    a) better than
    b) worse than
    c) equal to
    d) Unable to determine

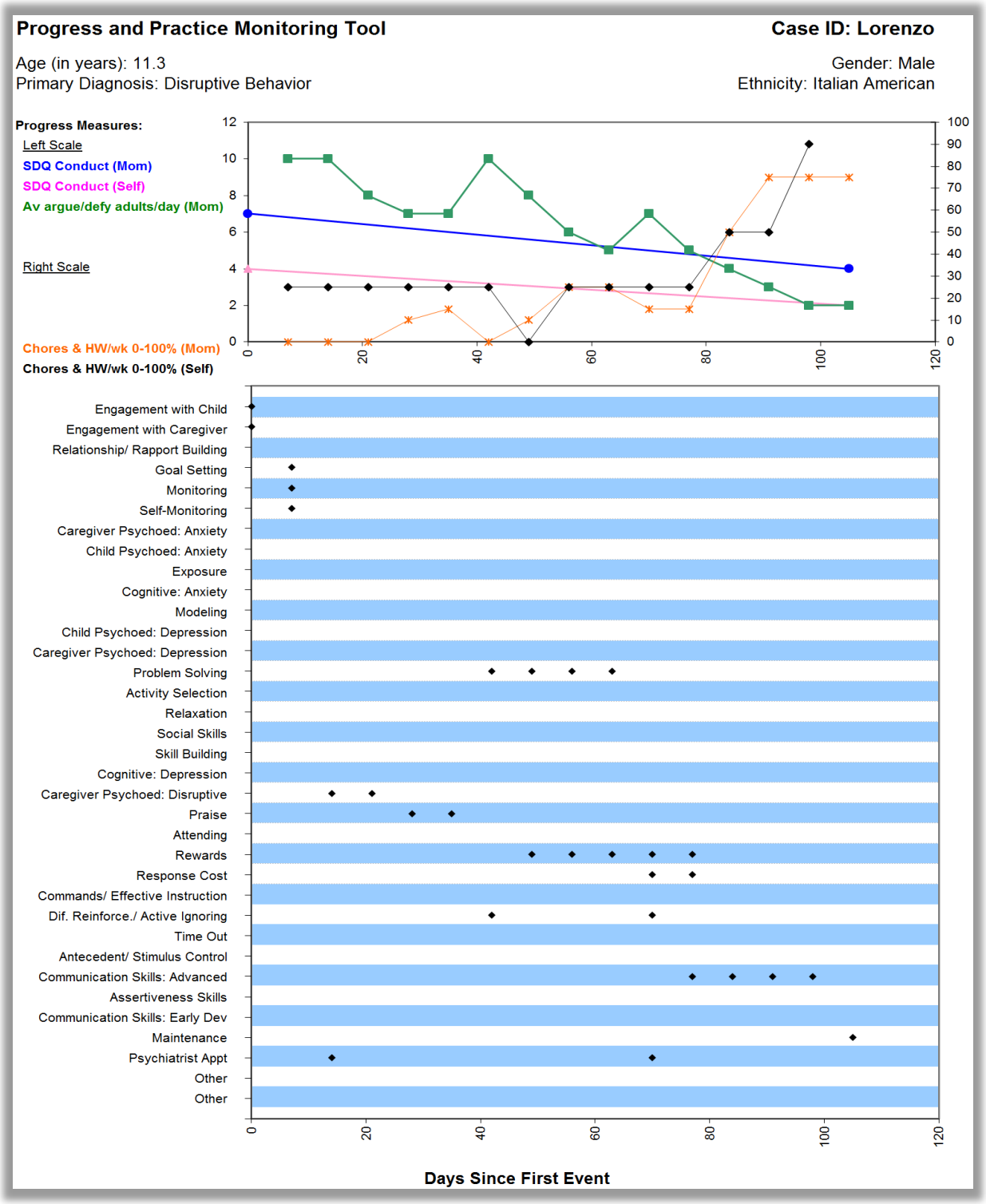
20. Jayden is failing school and has been suspended for getting into fights on three occasions. Which of the following is the BEST adaptation to Jayden's dashboard to improve monitoring on the progress pane?

    a) Add another well-validated self-report measure of Disruptive Behavior in addition to the Y-OQ
    b) Replace Verbal Aggression with a measure of Academic Performance
    c) Move the Verbal Aggression and Physical Aggression measures to the left scale with the Y-OQ
    d) Add teacher report of the number of times Jayden engages in verbal or physical aggressive behavior in the classroom (e.g., shouting, using profanities, kicking chairs, punching walls, etc.)

21. How could the practice pane of Jayden's dashboard be improved?

    a) Re-administer the Y-OQ and add its data point
    b) Cease tracking individual practices and focus on tracking delivery of CBT
    c) Only code practices when they are the primary focus of a session
    d) Add additional practice indicators to eliminate gap

Vignette 4


Progress and Practice Monitoring Tool
Case ID: Lorenzo
Age (in years): 11.3
Primary Diagnosis: Disruptive Behavior
Gender: Male
Ethnicity: Italian American
Progress Measures:
Left Scale
SDQ Conduct (Mom)
SDQ Conduct (Self)
Av argue/defy adults/day (Mom)
Right Scale
Chores & HW/wk 0-100% (Mom)
Chores & HW/wk 0-100% (Self)
Engagement with Child
Engagement with Caregiver
Relationship/ Rapport Building
Goal Setting
Monitoring
Self-Monitoring
Caregiver Psychoed: Anxiety
Child Psychoed: Anxiety
Exposure
Cognitive: Anxiety
Modeling
Child Psychoed: Depression
Caregiver Psychoed: Depression
Problem Solving
Activity Selection
Relaxation
Social Skills
Skill Building
Cognitive: Depression
Caregiver Psychoed: Disruptive
Praise
Attending
Rewards
Response Cost
Commands/ Effective Instruction
Dif. Reinforce./ Active Ignoring
Time Out
Antecedent/ Stimulus Control
Communication Skills: Advanced
Assertiveness Skills
Communication Skills: Early Dev
Maintenance
Psychiatrist Appt
Other
Other
Days Since First Event

Questions for Vignette 4

22. How many times did Lorenzo's mother fill out the SDQ?
    a) 1
    b) 2
    c) At every session
    d) Unable to determine

23. At Day 21, which of the measures below showed that Lorenzo had improved since starting treatment?

    a) Average times arguing/defying adults/day (Mom)
    b) Chores & Hmwk/wk 0-100% (Mom)
    c) Chores & Hmwk/wk 0-100% (Self)
    d) None of the above

24. Did Lorenzo see a psychiatrist during the course of treatment?

    a) Yes, once.
    b) Yes, twice.
    c) No.
    d) Unable to determine

25. According to Lorenzo's report during his most recent session, he completed _____ of his chores and homework for the week.

    a) 9
    b) 11
    c) 75%
    d) 90%

26. Which measure(s) improved while the Communication Skills: Advanced practice was delivered?

    a) Completion of chores and homework/week (Self)
    b) Average times arguing/defying adults/day (Mom)
    c) Both of the above
    d) Neither of the above

27. After the initial "Connect" phase of treatment (between days 14 and 35), who was primarily involved in treatment sessions?

    a) Lorenzo
    b) Lorenzo's mom
    c) Lorenzo's psychiatrist
    d) Unable to determine

Use the following information for the next two questions:

Lorenzo's average times arguing/defying adults per day spiked on Day 42, based on Mom's report.

28. What practice had been delivered during the previous week's session?

    a) Goal Setting
    b) Problem Solving
    c) Caregiver Psychoeducation: Disruptive
    d) Praise

29. What practice(s) was/were delivered in the session when the elevation was reported?

    a) Praise
    b) Praise & Problem Solving
    c) Problem Solving & Differential Reinforcement/Active Ignoring
    d) Problem Solving

Dashboard Usage Questions

For Statements 30-33, reply on a scale of 1 (*completely disagree*) to 7 (*completely agree*)

30. I would feel apprehensive about using clinical dashboards. _____

31. I would like working with clinical dashboards. _____

32. Using clinical dashboards is a good idea. _____

33. I would find clinical dashboards useful in clinical work. _____

34. Since I was trained, I have used clinical dashboards with _____ of my cases.

    a) none
    b) some
    c) most
    d) all
    e) I was never trained in MAP or to use clinical dashboards.

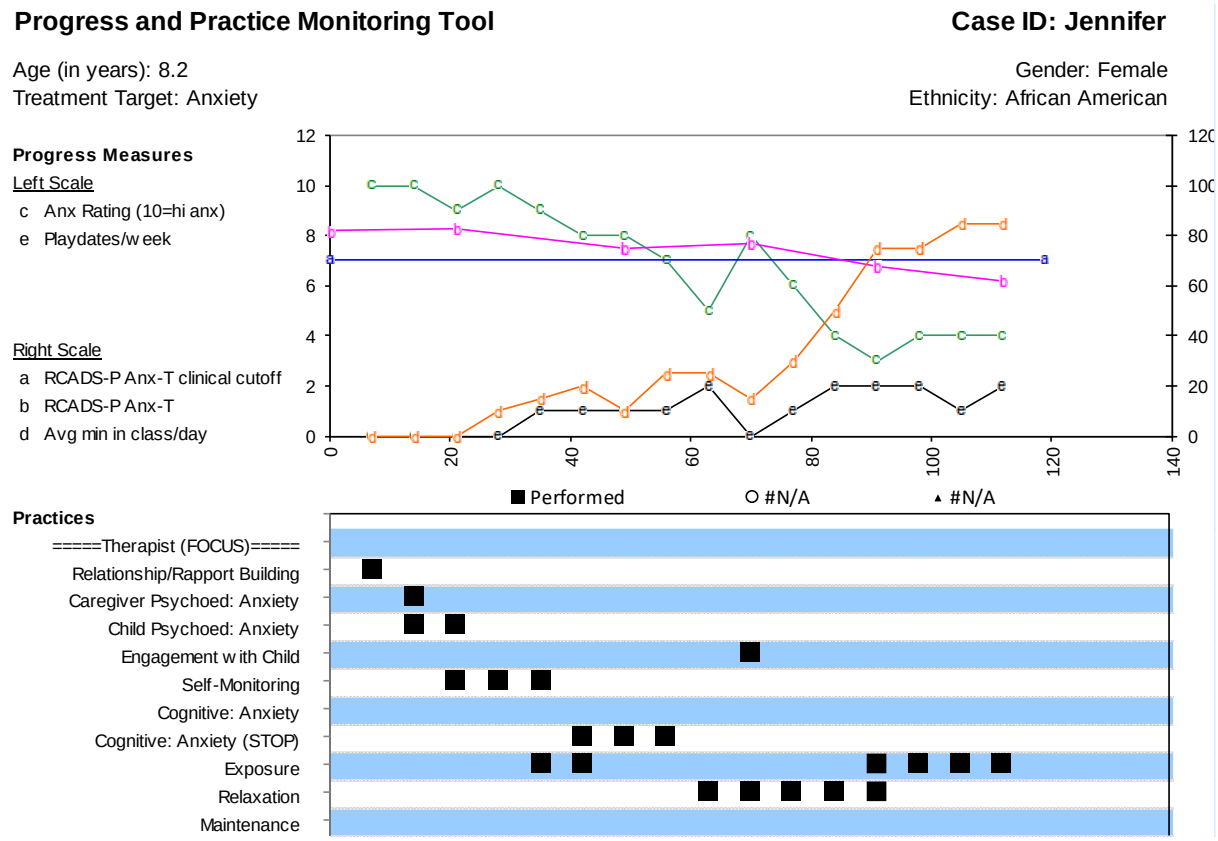
35. The last time I used a clinical dashboard with a case was _____

    a) within the last month
    b) within the last six months
    c) within the last year
    d) over a year ago
    e) over 5 years ago

Appendix B: Pre-Training Dashboard Vignettes (Practice/Benchmark condition is displayed)

Measure

Vignette 1:



Items for Vignette 1:

35. How many times did Jennifer's mother fill out the RCADS-P?
    e) 2
    f) 6
    g) At every session
    h) Unable to determine

36. At Day 21, which of the measures below showed that Jennifer had improved since starting treatment?

    e) Anxiety rating
    f) Average minutes in class/day
    g) Playdates/week
    h) None of the above

37. According to Jennifer's report during her most recent session, she spent an average of _____ minutes in class per day.

    e) 2
    f) 8
    g) 50
    h) 85

4. Given info displayed, how would you rate treatment progress?

    1 (*very poor*) ---------- 7 (*very good*)

    Relative to your response in question #4:

    a. How confident are you in your rating of treatment progress?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    b. While rating treatment progress, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

5. Given info displayed, what would you do next in treatment?
    a. Repeat the practice that was most recently delivered.
    b. Repeat a practice that was previously delivered (but not most recently).
    c. Change to a new practice.
    d. Initiate terminating phase of treatment.
    e. I do not know what I would do next in treatment.

Relative to your response in question #5:

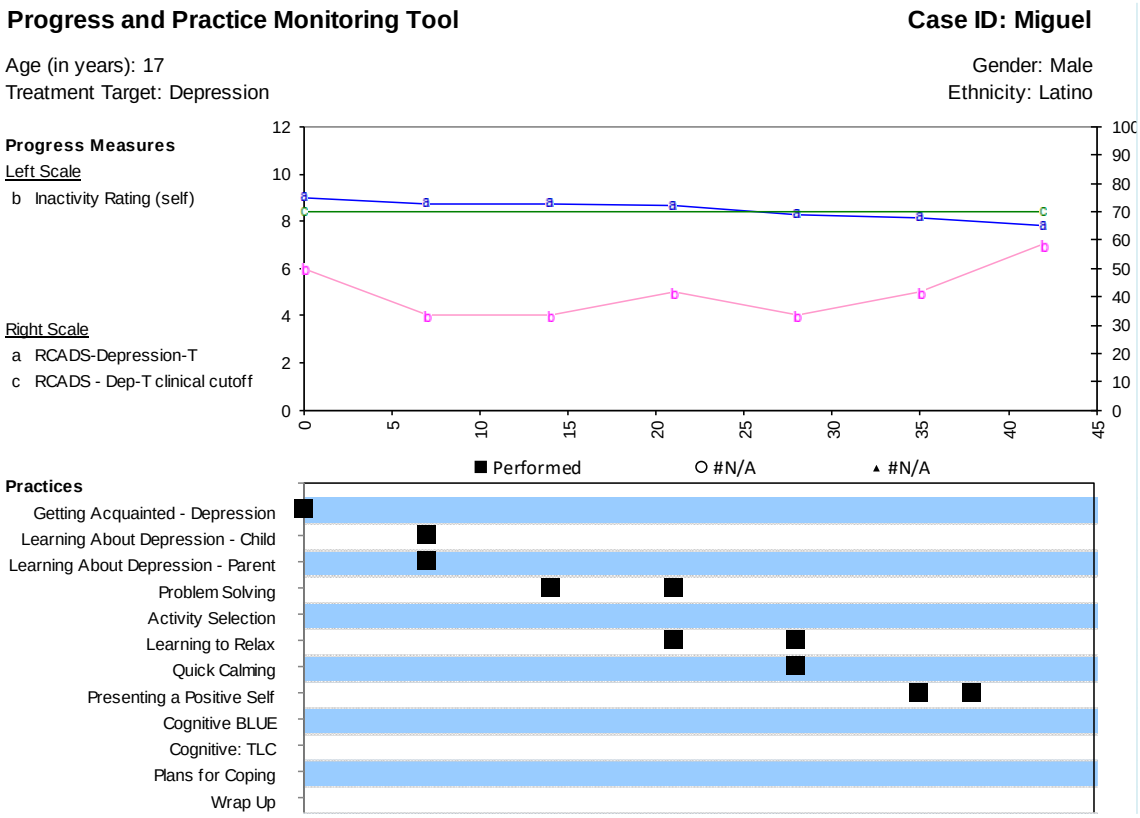
a. How confident are you in choosing the next step in treatment?

1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

b. While choosing the next step in treatment, the amount of information displayed felt like it was:

1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

Vignette 2:



Items for Vignette 2:

1. What was Miguel's peak rating on the RCADS Depression scale?
    e) 6
    f) 9
    g) 75
    h) 100

2. Over the last three sessions, Miguel's Inactivity Rating _____.
    e) has improved.
    f) has deteriorated.
    g) has stayed the same.
    h) cannot be analyzed based on the information provided.

3. How many measures are in the progress pane?
    e) 1
    f) 2
    g) 3
    h) 4

9. Given info displayed, how would you rate treatment progress?

    1 (*very poor*) ---------- 7 (*very good*)

    Relative to your response in question #9:

    a. How confident are you in your rating of treatment progress?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    b. While rating treatment progress, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

10. Given info displayed, what would you do next in treatment?
    a. Repeat the practice that was most recently delivered.
    b. Repeat a practice that was previously delivered (but not most recently).
    c. Change to a new practice.
    d. Initiate terminating phase of treatment.
    e. I do not know what I would do next in treatment.

Relative to your response in question #10:

a. How confident are you in choosing the next step in treatment?

   1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

b. While choosing the next step in treatment, the amount of information displayed felt like it was:

   1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

Dashboard Usage Questions

For Statements 11-14, reply on a scale of 1 (*completely disagree*) to 7 (*completely agree*)

11. Going forward, I would feel apprehensive about using clinical dashboards.
    _____

12. Going forward, I would like working with clinical dashboards.
    _____

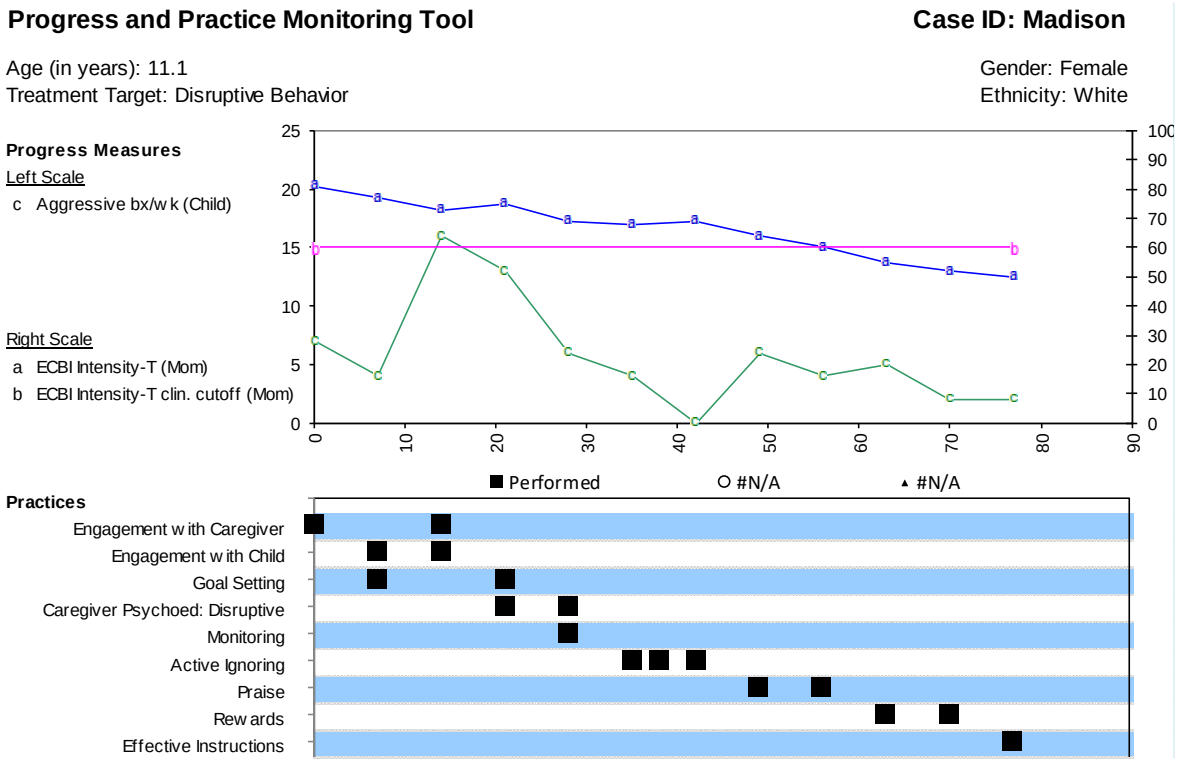
13. Using clinical dashboards is a good idea.
    _____

14. Going forward, I would find clinical dashboards useful in clinical work.
    _____

Appendix C: Post-Training Dashboard Vignettes (Practice/Benchmark condition is displayed)

Measure

Vignette 1:



Items for Vignette 1:

4. Based on Mom's report, what was Madison's lowest rating on the ECBI Intensity scale?
   i) 0
   j) 10
   k) 13
   l) 50

5. Since the start of treatment, Mom's report on the ECBI Intensity scale _____.
   i) shows that Madison's behavior has improved.
   j) shows that Madison's behavior has deteriorated.
   k) shows that Madison's behavior has not changed.
   l) does not provide enough information to assess for change in Madison's behavior.

6. How many measures are in the progress pane?
    i) 1
    j) 2
    k) 3
    l) 9

6. Given info displayed, how would you rate treatment progress?

    1 (*very poor*) ---------- 7 (*very good*)

    Relative to your response in question #4:

    c. How confident are you in your rating of treatment progress?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    d. While rating treatment progress, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

7. Given info displayed, what would you do next in treatment?
    a. Repeat the practice that was most recently delivered.
    b. Repeat a practice that was previously delivered (but not most recently).
    c. Change to a new practice.
    d. Initiate terminating phase of treatment.
    e. I do not know what I would do next in treatment.

    Relative to your response in question #5:

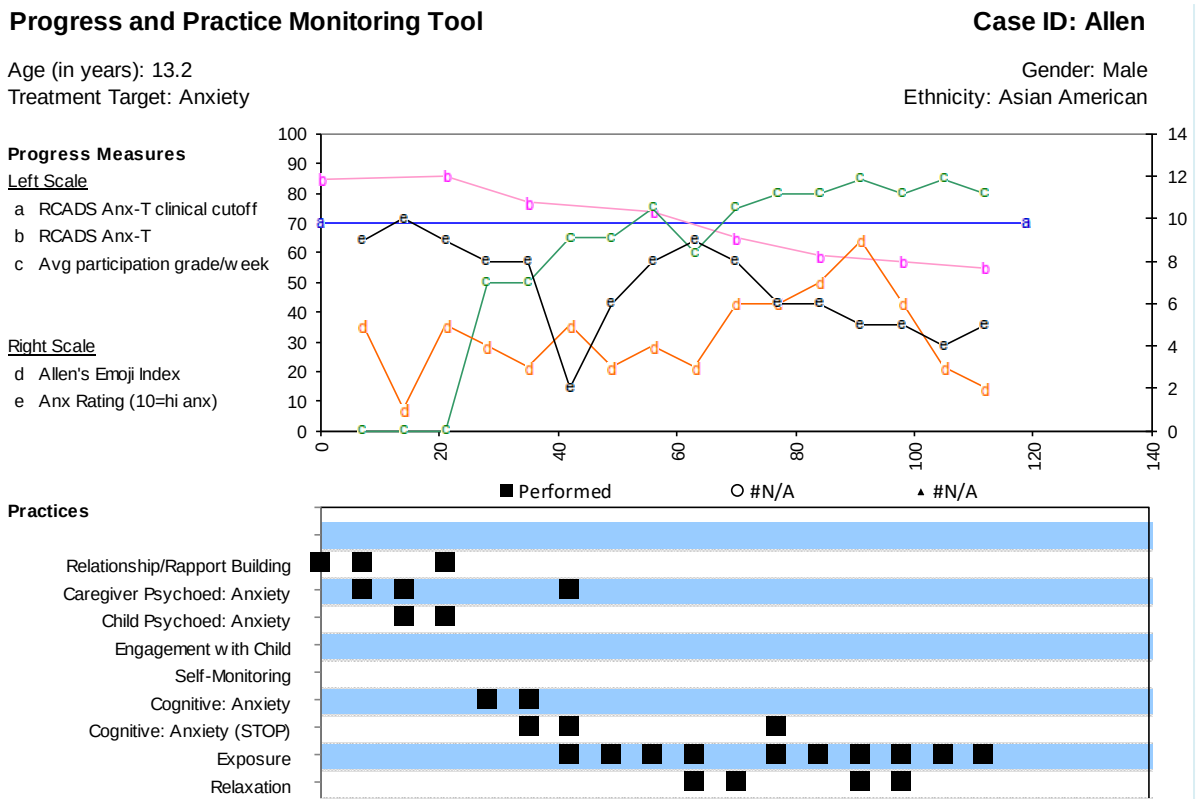
    c. How confident are you in choosing the next step in treatment?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    d. While choosing the next step in treatment, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

Vignette 2:



Items for Vignette 2:

6. What was Allen's lowest rating on his Emoji Index?
   a) 0
   b) 1
   c) 8
   d) 15

7. Over the last three sessions, Allen's Emoji Index _____.
   a) has improved.
   b) has deteriorated.
   c) has stayed the same.
   d) cannot be analyzed based on the information provided.

8. Allen's Anxiety Rating was worst on what day since starting treatment?

    a) 1
    b) 14
    c) 42
    d) 71

15. Given info displayed, how would you rate treatment progress?

    1 (*very poor*) ---------- 7 (*very good*)

    Relative to your response in question #9:

    c. How confident are you in your rating of treatment progress?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    d. While rating treatment progress, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

16. Given info displayed, what would you do next in treatment?
    a. Repeat the practice that was most recently delivered.
    b. Repeat a practice that was previously delivered (but not most recently).
    c. Change to a new practice.
    d. Initiate terminating phase of treatment.
    e. I do not know what I would do next in treatment.

    Relative to your response in question #10:

    c. How confident are you in choosing the next step in treatment?

        1 (*not at all confident*) ---------- 4 (*moderately confident*) ---------- 7 (*very confident*)

    d. While choosing the next step in treatment, the amount of information displayed felt like it was:

        1 (*too little*) ---------- 4 (*just right*) ---------- 7 (*too much*)

Dashboard Usage Questions

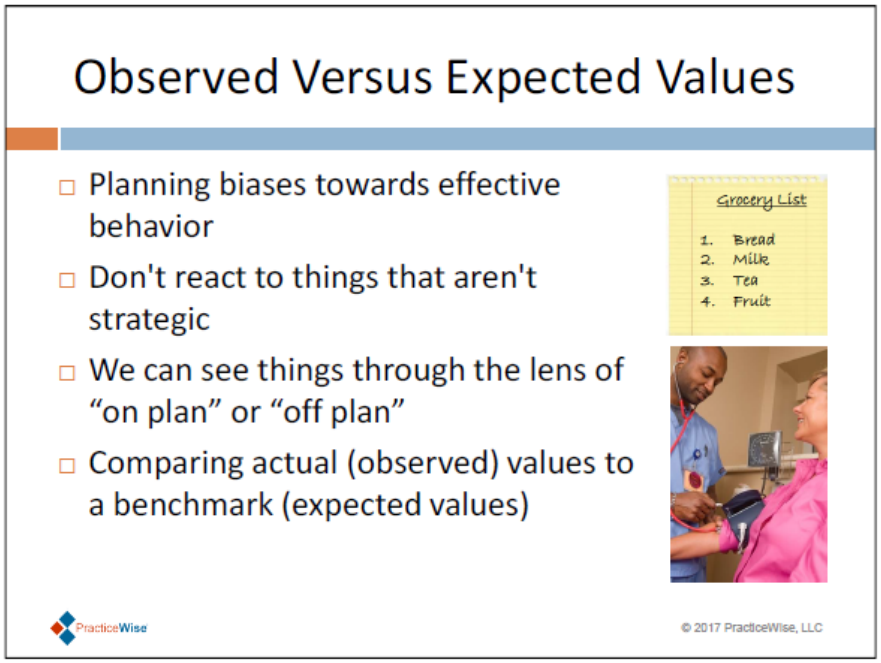
For Statements 11-14, reply on a scale of 1 (*completely disagree*) to 7 (*completely agree*)

17. Going forward, I would feel apprehensive about using clinical dashboards. _____

18. Going forward, I would like working with clinical dashboards. _____

19. Using clinical dashboards is a good idea. _____

20. Going forward, I would find clinical dashboards useful in clinical work. _____

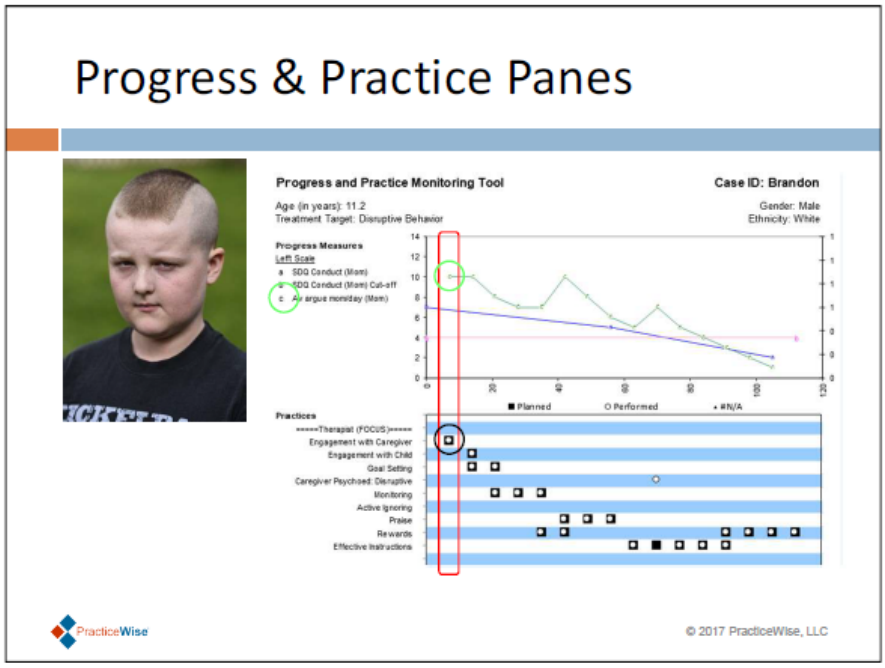


SCRIPT:

In the previous example, Britney's provider had some expectations that Britney should be getting bette

(1) Research suggests that having a plan makes us act more effectively or can bias us towards effectiv
(2) Similarly, planning in the service world can help avoid reacting to things that aren't strategic.
(3) Just as with the grocery list, we are able to see things through the lens of "on plan" or "off plan" to
(4) In this way you are comparing your actual (or observed) blood pressure or body mass to a benchm

Let's take a look at an example of how features of the advanced dashboard allow you to compare expe

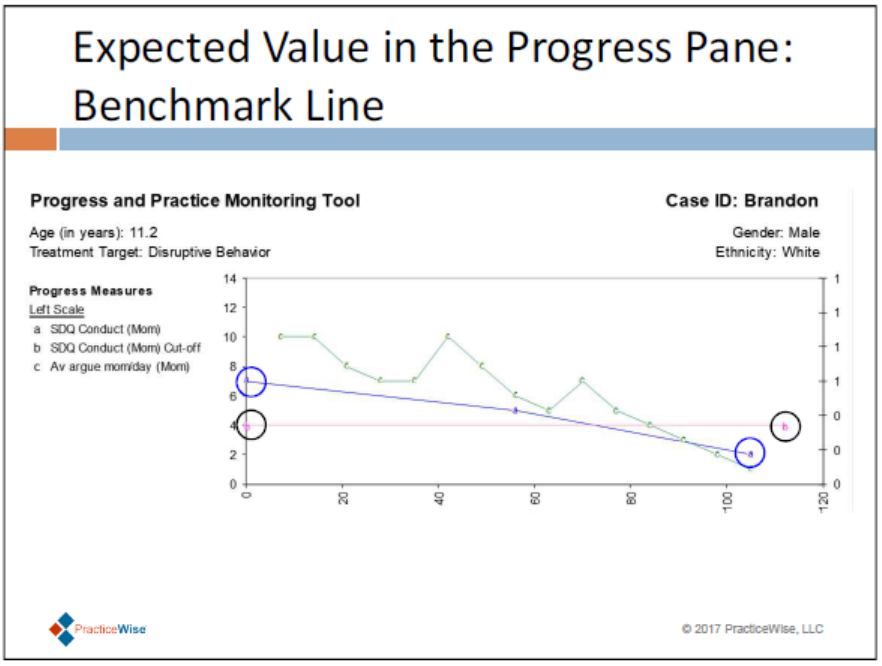


SCRIPT:

Meet Brandon, an 11 year old boy with conduct problems. Here, we see his progress pane together wit

(1) These panes can be interpreted together by examining the information vertically. For instance
(2) On Day 7 of treatment, Brandon's mother reported that he argued with her an average of 10x/day,
(3) which is tracked with a scale labeled 'c' on the progress pane. Scales on advanced dashboards can
(4) Also on Day 7 of treatment, the therapist delivered his planned practice: Engagement with the Care
(5) By examining this information vertically, you can see how treatment practices and treatment progr

We will return to examine the practice pane in more detail in a moment. For now, let's take a closer lo



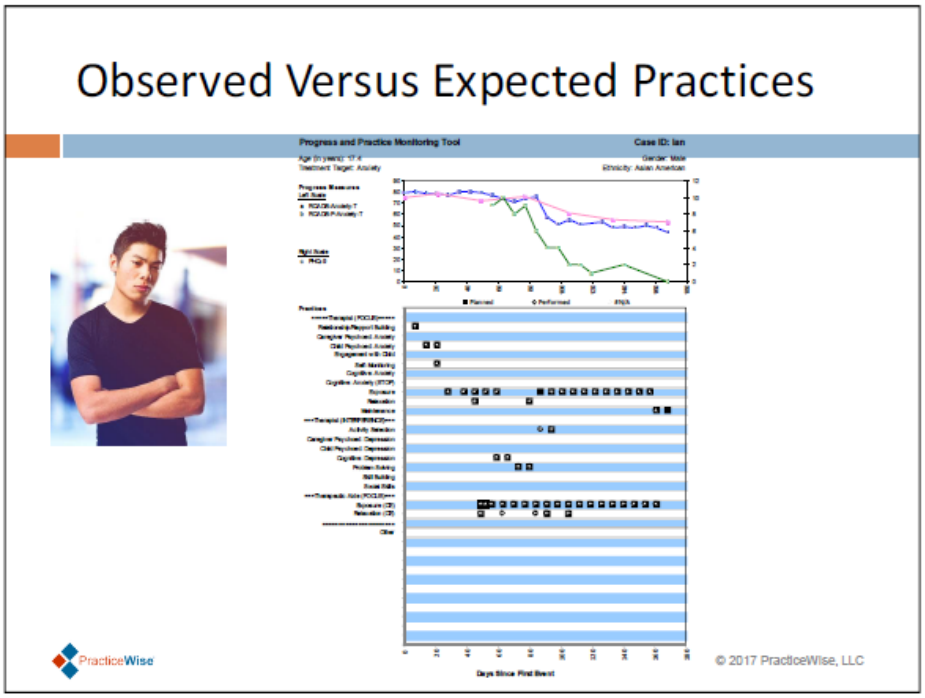
SCRIPT:

You can see that Brandon's provider has plotted two measures – the average number of times Brandon

Tracking Brandon's progress on the SDQ Conduct score over the course of his treatment episode gives

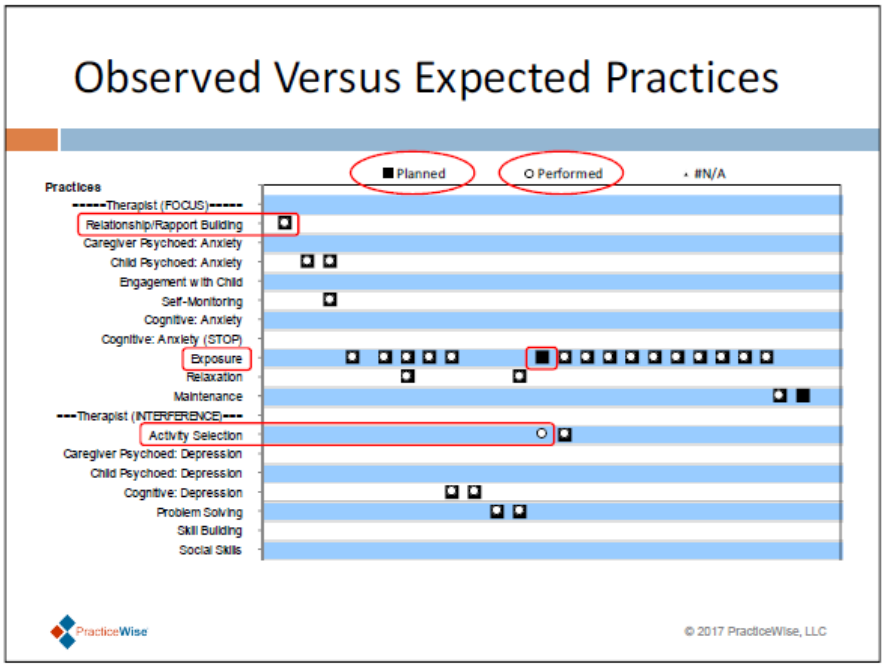
(1) his provider plotted the clinical cut-off score for the SDQ Conduct scale at the start of the clinical
(2) You can see that Brandon was scoring above this level at the start of treatment suggesting that he h
(3) His provider also plotted the clinical cut-off score at the end of treatment in order to produce a clea
(4) You can see that when the SDQ was administered on the 106th day of treatment, Brandon's score o

By entering the clinical cut-off line to establish a benchmark, his provider is able to compare what is e



SCRIPT:

Meet Ian, a 17 year old with anxiety. Let's see how his provider illustrated expected vs observed pract



SCRIPT:

Ian's provider opted to use symbols to indicate which practices were planned and which practices were

(1) the black square to show what was planned before the session occurred (for instance in supervisior
(2) the white circle to indicate what actually happened in session.
(3) For instance, you can see that the first planned session was Relationship/Rapport Building which c
(4) One example is when the therapist planned to engage in exposure after shifting back to treatment f
(5) Instead, on this session, you can see that Ian's provider introduced Activity Selection. This practice

Let's take a look at another example where the provider is struggling a bit more to stick with the plan.

References

Aykin, N. M., & Aykin, T. (1991). Individual differences in human-computer interaction. *Computers & Industrial Engineering, 20*(3), 373–379. http://doi.org/10.1016/0360-8352(91)90009-U

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423–1429. http://doi.org/10.1176/appi.ps.002052011

Brown, T. E. (2014). *Effects of training and expected treatment progress on clinicians' interpretations of measurement feedback systems*. Unpublished manuscript, Department of Psychology, University of California, Los Angeles.

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & The Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health, 35*, 114–123. http://doi.org/10.1007/s10488-007-0151-x

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566–579. http://doi.org/10.1037/a0014565

Chorpita, B. F., Daleiden, E. L., & Bernstein, A. D. (2016). At the intersection of health information technology and decision support: Measurement feedback systems...and beyond. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 471–477. https://doi.org/10.1007/s10488-015-0702-5

Daleiden, E. L., & Chorpita, B. F. (2005). From data to wisdom: Quality improvement strategies supporting large-scale implementation of evidence-based services. *Child and Adolescent Psychiatric Clinics of North America, 14*(2), 329–349. http://doi.org/10.1016/j.chc.2004.11.002

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*.

Hansen, W. J. (1971). User engineering principles for interactive systems. *Fall Joint Computer Conference*, 523–532. http://doi.org/10.1145/1479064.1479159

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology*, *61*(2), 165–174. http://doi.org/10.1002/jclp.20113

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. http://doi.org/10.1007/s10488-016-0719-4

Malone, T. W., Lai, K., & Grant, K. R. (2001). Two design principles for collaboration technology: Examples of semiformal systems and radical tailorability. In G. M. Olson, T. W. Malone, & J. B. Smith (Eds.), *Coordination theory and collaboration technology* (pp. 125-160). New York, NY: Psychology Press.

Regan, J., Daleiden, E. L., & Chorpita, B. F. (2013). Integrity in mental health systems: An expanded framework for managing uncertainty in clinical care. *Clinical Psychology: Science and Practice, 20*(1), 78–98.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298–311. http://doi.org/10.1037/a0019247

Shortliffe, E. H., Buchanan, B. G., & Feigenbaum, E. A. (1979). Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proceedings of the IEEE, 67*(9), 371–375. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1455705

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Chapter 2:

Identifying Barriers to Community-Based Dashboard Use

Abstract

Experiences with clinical dashboards were explored with a sample of New York-based mental health service providers and supervisors. Semi-structured interviews were conducted with supervisors to identify issues related to continued dashboard use. Clinicians completed a survey with items informed by the supervisor-identified themes. Data comparison found supervisor-clinician agreement that time constraints and lack of agency support act as significant barriers for dashboard use. Clinicians were more likely to also endorse barriers related to organizational context, such as agency and supervisor prioritization. These findings suggest that an increased top-down focus on dashboard use within agencies may lead to fewer barriers and greater dashboard utilization by clinicians.

Keywords: clinical dashboards, implementation, barriers, supervisors, clinicians

Introduction

Despite the benefits of using standardized outcome measures to inform treatment, usage rates among community-based clinicians remain low. As reported by Jensen-Doss et al. (2016), only 13.9% of clinicians monitored treatment progress with standardized measures at least monthly and 61.5% never used them at all. Earlier studies on standardized measures reported even more harrowing results, finding that 92% of surveyed providers had never used results from mandated standardized measures (Garland et al., 2003). These low rates may be reflective of dissemination challenges, such as providers having limited access to tools such as dashboards (Chorpita, Bernstein, Daleiden, & The Research Network on Youth Mental Health, 2008). However, low usage rates have also persisted even after dashboard systems are implemented, with half of surveyed clinicians reporting that they do not use the feedback provided to them (e.g., de Jong, 2012). In considering dashboard implementation efforts, de Jong (2016) identified challenges across three themes: (1) design and planning, (2) organizational context, and (3) sustainability

Challenges around designing and planning dashboards can arise from a lack of understanding of clinicians' backgrounds during the design process. Clinicians' conceptual understanding of and attitudes toward dashboards may either enhance or interfere with successful implementations. Across implementations, some clinicians have reported "buying into" dashboards once they saw their value, whereas others reported difficulty incorporating the data into decision making or using computers in general (Borntrager & Lyon, 2015; Gleacher et al., 2016). Drawing from the human-computer interaction (HCI) literature, a user-centered design approach could offer a method to address such barriers prior to an implementation. These models encourage iterative, participatory designs that involve users early in the process as collaborators

(e.g., Agre, 1995; Bannon, 1991), as contrasted with implementations that are thrust upon unknowing clinicians. Along these lines, Lyon, Lewis, Boyd, Hendrix, and Liu (2016) proposed the *contextualized technology adaptation process* (CTAP), which aims to integrate HCI design processes with implementation science models. The CTAP framework itself faces several implementation challenges, including the difficulty of applying iterative design processes to established implementations and the paradigm shift required to move from deliberate, research-minded design approaches towards a rapid, iterative process. However, overall, user-informed implementations appear to be a promising means of addressing usage barriers.

Organizational context also emerges as an area of potential implementation challenges. Although few studies have directly examined the impact of organizational context on dashboard implementation efforts, studies of EBP implementations have found that positive organizational factors, such as on-site champions and more engaged, less stressful climates, are associated with more positive clinician attitudes and higher levels of sustainment (Aarons et al., 2012; Aarons et al., 2016; Langley, Nadeem, Kataoka, Stein, & Jaycox, 2010). Similar suggestions, including ongoing consultation and engendering a sense of ownership in providers, have been proposed to improve implementations of evidence-based assessments (EBAs). Post-hoc analysis of two dashboard implementation case studies have suggested that these patterns may hold for dashboards, where the more successful implementation had leadership with greater day-to-day involvement and oversight (Bickman et al., 2016). Additional research is needed on organizational factors specific to dashboard implementation, but the existing EBP and EBA research provides promising starting points.

Each of these factors – design and organizational context – have a direct role in the sustainability of an implementation as well. A dashboard implementation is unlikely to persist

when clinicians become easily frustrated with its design and receive little organizational support. An emergent consequence of these combined factors is the challenge of post-rollout sustainability. Issues such as the amount of additional time required by clinicians to use dashboards alongside their usual clinical and administrative workload, particularly in high-stress environments, can be especially burdensome when it leads to duplicating work efforts (Borntrager & Lyon, 2016; Gleacher et al., 2016).

In addition to the three overarching challenges areas identified by de Jong (2016), a relative lack of clinician training in EBAs raises challenges in planning, organizational support, and sustainability of dashboard implementations. Graduate training of community clinicians often does not include instruction on EBA principles (Hunsley & Mash, 2005), so implementation efforts may benefit from an increased focus on identified gaps in clinician knowledge around the use and benefits of evidence-based approaches (Borntrager & Lyon, 2016). Additional support may be necessary before clinicians are equipped to incorporate dashboard data into their clinical work (Callaly, Hyland, Coombs, & Trauer, 2006). For instance, untrained clinicians may be less likely to accurately interpret progress measurements, which necessitate the consideration of recent trends rather than single datum points (Tsai, Moskowitz, Brown, Park, & Chorpita, 2016).

Across each of these challenges, organizations' supervisors and leaders play essential roles in developing and promoting climates that address potential barriers to evidence-based strategies (Aarons, Farahnak, Ehrhart, & Sklar, 2014; Langley et al., 2010). As such, a mismatch between leadership-identified barriers and clinician-identified barriers could be especially detrimental to clinicians' ongoing use of dashboards in clinical work. Investigations have been conducted to examine views on evidence-based practices across multiple stakeholders (e.g.,

Aarons, Wells, Zagursky, Fettes, & Palinkas, 2009), but similar analyses are lacking around tools such as dashboards. The current study aims to target both leadership/supervisor and clinician levels to examine the barriers that arise around dashboard use. Discrepant endorsements of barriers will provide additional insight into the implementation challenges that may arise or be exacerbated by these misalignments. Specifically, the following questions were addressed in this study: (1) what factors do leadership and supervisors identify as barriers and benefits to using dashboards? (2) how do clinicians rate their experience with barriers and benefits around dashboard use? (3) in what areas are there agreement and discrepancies between leadership- and clinician-identified barriers and benefits to dashboard use?

Method

Design

This study used an exploratory sequential mixed methods design aimed at understanding and comparing barriers and benefits to dashboard use by supervisors and clinicians. Mixed methods research provides greater insights into phenomena than either qualitative or quantitative approaches alone (Robins et al., 2008). In an exploratory sequential design, qualitative data are first collected and analyzed, and the identified themes are used to inform the creation of a quantitative measure to further explore the research questions. Results of the two phases are then linked to enable a fuller examination of the phenomena (Creswell & Clark, 2018). This triangulation-driven approach seeks convergence and corroboration between the data collected in each phase of the study (Greene, Caracelli, & Graham, 1989). The sequential design format used for this study may also be conceptualized as a QUAL → quan model, where the initial qualitative stage serves as the primary driver for exploration of data (Palinkas et al., 2011).

The primary objective of the qualitative first phase of this study was to gain an in-depth understanding of current barriers and benefits of dashboard use, as identified by supervisors within mental health agencies, and to use that information to construct a quantitative measure. The primary objective of the quantitative second phase of this study was to use the newly created quantitative measure to gather information from clinicians to discover how they rank their experiences with dashboard barriers and benefits. During the third phase of this study, the two sets of findings were examined in concert to identify areas of agreement and discrepancy between supervisors and clinicians regarding dashboard use.

Participants

This study focused on participants with experience with the dashboard implementation present in the Managing and Adapting Practice (MAP) system (Chorpita & Daleiden, 2009). Several mental health agencies were identified as possible participants in this study via collaboration with MAP trainers associated with the NYU Langone Medical Center. In both the qualitative and quantitative phases of the study, participants at each site engaged with the acknowledgement that their participation was optional and declining would not adversely affect them. Site *R* provides child and youth services within community- and school-based environments as part of a county mental health department based in southern New York state. Site *U* provides child and youth services within school-based environments based in New York City. Supervisors and clinicians at both sites had received MAP training from the same set of NYU-based MAP trainers during separate training sessions. These NYU-based trainings and subsequent consultation calls were conducted with materials based on licensed PracticeWise training curricula and content. However, it should be noted that the trainers were not certified MAP training professionals.

Qualitative sample. Convenience sampling recruitment was conducted by emailing three initial points of contact at MAP-trained sites identified by NYU-affiliated MAP trainers. Contacts at sites *R* and *U* agreed to identify potential supervisors to interview for the qualitative phase and clinicians for the quantitative phase. Site *Z* was similarly contacted and then sent a follow-up message one week after an initial non-response; however, no response was received, and they were subsequently removed from study consideration. For interview purposes, Site *R* identified two supervisors, and Site *U* identified four supervisors. All identified supervisors agreed to participate, providing the recommended minimum sample size of six qualitative interview participants (Onwuegbuzie & Collins, 2007). The qualitative sample was composed entirely of Caucasian women supervisors with Masters-level degrees, who had been trained on MAP an average of 10.67 months (*SD* = 1.03) prior to their interviews. Notably, although each supervisor had been MAP-trained and was supervising MAP accordingly, none had been specifically trained to function as a MAP supervisor.

Quantitative sample. The target population for the quantitative phase was MAP-trained clinicians identified by supervisors at the two participating sites. At each site, 19 clinicians were identified as possible participants (38 total). The identified recipients were emailed a description of the study and an estimate of the survey duration (approximately 10 minutes, based on pilot testing). Recipients were offered a $10 Amazon.com e-gift-card for completion of the survey and were informed that their participation or lack thereof would have no adverse effects. A total of 21 participants from 38 in the identified pool (55.3%) participated, with 8/19 (42.1%) for site *R* and 13/19 (68.4%) for site *U*, with no significant difference between site response rates (p = .103). The response rates met or exceeded the range of online response rates identified by Nulty (2008) across studies that used internet-based surveys (20.0% to 47.0%).

The 21 respondents (20 female, 1 male), aged 25-57 (*M* = 34.4 years, *SD* = 10.2 years) were all Masters-level clinicians (3 Masters of Mental Health Counseling, 18 Masters of Social Work) and were majority Caucasian (16 Caucasian, 3 Latinx, 1 mixed Caucasian/Latinx, 1 mixed Caucasian/Native American). They reported primary practice settings of schools (*n* = 17) and community mental health centers (*n* = 4). The respondents reported an average of 7.5 years of experience (*SD* = 5.1 years) and had been recently trained on MAP (*M* = 7.3 months since training, *SD* = 3.2 months). Respondents reported an average of 22.2 face-to-face clinical hours per week (*SD* = 6.6 hours).

Several significant differences were identified between respondents from sites *R* and *U* in addition to the rural vs. urban nature of their respective settings. Site *R* respondents (*M* = 43.9 years, *SD* = 10.2) were significantly older than site *U* (*M* = 28.6 years, *SD* = 4.0), $p = .003$. Similarly, regarding years of clinical experience, site *R* respondents (*M* = 12.8 years, *SD* = 2.8) had significantly more experience than site *U* (*M* = 4.3 years, *SD* = 3.2), $p < .001$. Additionally, site *R* respondents' primary practice settings were split between community mental health centers and schools (50% each) whereas site *U* respondents all worked primarily in schools. The site respondents did not significantly differ on gender, ethnicity, degree, average weekly face-to-face hours, time since MAP training, or how often they use dashboards.

Materials

Qualitative.

***Semi-structured interview*.** A semi-structured interview was used to gather qualitative information on dashboard barriers and benefits from site supervisors. Interview questions were adapted from Gleacher et al. (2016) and informed by difficulties related to evidence-based

assessments (e.g., Hunsley & Mash, 2005) and the challenge areas identified by de Jong (2016): design, organizational context, and sustainability (see Appendix E).

Quantitative.

Online survey. An online quantitative survey was developed by using design- and method-level approaches (Fetters, Curry, and Creswell, 2013) to link the survey to themes identified from analysis of the qualitative data. The language used and themes identified by interviewees in the qualitative phase formed the foundation for the survey's questions. Participants were asked to provide their opinion on 37 random-ordered items related to dashboard barriers and benefits using a 7-point Likert scale (*completely disagree* to *completely agree*). Items were scored or reverse-scored as appropriate to establish higher values as being more positively valenced during data analysis and interpretation. They were also asked to rank their top three benefits and barriers of using dashboards, with an option to identify "other" reasons. Ranked items were scored with weighted values, such that items ranked "1" received a score of 3, items ranked "2" received a 2, and those ranked "3" received a 1. An open-ended final survey question allowed users to enter free text to describe their overall experience with dashboards. The survey was implemented in Qualtrics and piloted by three MAP-trained advanced clinical psychology graduate students to ensure that the questions were well-understood. The full survey can be found in Appendix F.

Usage questions. Four clinical dashboard usage items, using a 7-point Likert scale (*completely disagree* to *completely agree*), were included in the survey based on adaptations from the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, Morris, Davis, & David, 2003). UTAUT was constructed based on conceptual and empirical similarities found from a comparison of eight previously developed models that examined individual

acceptance of new information technologies. UTAUT demonstrated an adjusted R^2 value of .70, explaining 70% of the variance in users' intentions to use information technology. The four questions selected for use have previously been found by the author to most vary with positive and negative opinions of dashboard usage (Brown, 2014) and provided additional attitudinal data around dashboard usage.

Procedure

Qualitative. Prior to interviews, interviewees received emails with information related to informed consent and were notified that they could withdraw at any time without penalty. All interviews were conducted via telephone and were recorded with interviewees' consent. Interviews ranged in length from 35 to 53 minutes (M = 46 minutes) and were conducted individually except for site *R*'s two participants, who were interviewed together due to interviewee time constraints. Transcripts were produced using automated NVivo Transcription services and then edited manually for verbatim accuracy.

Qualitative coding process. After the interviews were transcribed, they were imported into NVivo 12 for coding. An inductive approach based on a constant comparative method (Charmaz, 2006; Glaser, 1965; Glaser & Strauss, 1967) was used to analyze the interview content while accounting for both a priori and emergent themes. An initial review was conducted to identify broad themes of discussion through the process of structural coding, which provides a starting point organized around the research questions (Saldaña, 2015). An in vivo coding phase followed in which participants' verbatim words drove identification of codes. Following the code identification and initial coding phase, a second comprehensive coding effort was conducted to examine intra-rater reliability. Cohen's kappa was run to determine agreement between initial and follow-up coding efforts. Per guidelines outlined by Landis and Koch (1977), there was

almost perfect agreement between the two sets of codes, $\kappa = .973$ (95% CI, .951 to .995), $p < .001$. Upon establishing reliability of the coded material, the initial set of codes were reexamined to organize, combine, and synthesize them into unique categories and linked themes using hierarchal trees within NVivo. The resultant tree hierarchy served as a foundation for qualitative examination and subsequent quantitative survey creation.

Quantitative. MAP-trained clinicians were identified by supervisors following the qualitative phase. All identified clinicians were invited to participate in the study via an email that included a brief description of the purpose and a link to the Qualtrics survey. To encourage participation, clinicians received a $10 Amazon.com e-gift-card for completing the survey. The survey was open for two weeks, with a reminder email sent after one week. Data were downloaded from Qualtrics and analyzed in SPSS (version 22) and Microsoft Excel (Office 365).

Results

Qualitative

Supervisors' views of dashboard barriers and benefits were examined via analysis of qualitative data collected via interviews. Comments from the six interviewees (over five interviews) were distilled into 30 separate codes with 19 identified around barriers to dashboard use and 11 around benefits of dashboard use. Of these 30 codes, the most frequently coded across all comments were: "Additional Task in Workload" (15.13%, barrier); "Helps Treatment Planning" (9.66%, benefit); "Tracking EBAs" (5.46%, benefit); "Lacking Accountability within Agency" (5.46%, barrier); "Consultation Calls" (5.04%, benefit); "Low Supervisor Priority" (4.62%, barrier); and "Data Entry Difficulties" (4.62%, barrier). The 30 codes were assigned a total of 238 times across interview comments (per interview, $M = 47.6$ codes, $SD = 9.1$).

Constant comparison was used to classify the 30 coded themes into four major categories that spanned barriers and benefits: Sustainability, Design, Organizational Context, and Evidence-Based Assessment Knowledge and Use. Subcategories specific to barrier and benefit categorizations were identified within these categories, and the full sets of nodes for barriers and benefits are depicted in Figures 6 and 7, respectively. Frequencies, percentages, and illustrative quotes for barriers and benefits are seen in Tables 3 and 4, respectively. The most frequently endorsed category was around benefits related to Evidence-Based Assessments (25.63% of total comments). The majority of these comments (16.39% of total) were related to dashboard utility around treatments, including as related to treatment planning, evidence-based assessment tracking, and demonstrating alignment between measures and session content. The other subcategory included benefits related to MAP in general, i.e., features more indirectly associated with dashboards (9.24% of total). Categories related to barriers were next most common, with 21.85% of comments related to barriers associated with Sustainability, with the most notable subcategory – Time Commitment – responsible for 20.17% of total comments itself. Barriers around Design also constituted a substantial proportion of comments (15.55%), with 10.92% of comments related to challenges around user interface difficulties. The remaining barrier categories – Organizational Context and Evidence-Based Assessment Knowledge & Use – represented a roughly equal proportion of comments (11.76%, 11.34%, respectively). Organizational Context barrier codes were split between concerns related to low dashboard priorities at the agency (7.14%) and supervisor (4.62%) levels. Evidence-Based Assessment barriers were split between challenges related to the initial MAP training (6.30%) and using evidence-based assessments (5.04%). Benefits related to Design (5.46%), Organizational Context (5.04%), and Sustainability (3.36%) rounded out the remaining comments. Chi-square

tests found no significant differences between sites *R* and *U* based on coding percentages across categories.

Examination of all coded comments revealed that 60.50% were identified as barriers and 39.50% as benefits. The ratio of benefits to barriers varied notably across categories (see Figure 8). Sustainability comments were most likely to be associated with barriers, with 86.7% of coded Sustainability comments falling into that category. Design and Organizational Context were also barrier-heavy, with 74.0% and 70.0% coded as barriers, respectively. Only EBA Knowledge & Use demonstrated a higher likelihood of coded benefits, with 69.3% of EBA comments associated with benefits.

Quantitative

Positively and negatively valenced items. Clinicians' views of dashboard barriers and benefits were examined via analysis of responses to survey items that had been created based on supervisors' interview responses. Responses to the 37 Likert-scale dashboard items were scored and sorted by percentage of respondents who endorsed negatively and positively valenced values relative to promoting dashboard use. The most frequently negative items included two items related to Sustainability ("It is easy to fall behind on keeping dashboards updated.": 95.24% negatively valenced responses; "I do not have enough time to review dashboards with my clients in sessions": 71.43% negatively valenced) as well as three items related to Organizational Context ("Someone at my employer/agency keeps track of how much I use dashboards": 76.19% negative; "It is a priority for my supervisors that I use dashboards": 71.43% negative; "It is a priority for my employer/agency that I use dashboards": 71.43% negative). These five items also received the lowest adjusted Likert values, ranging from 2.05 to 2.95 (see Figure 9).

The most frequently positive items included four items related to Evidence-Based Assessments & Use: “Collecting and tracking data is useful for treating clients” (90.48% positively valenced); “I can provide highest quality treatment based on my clinical judgment alone” (85.71% positive); “Post-training MAP consultation calls were helpful around using dashboards” (80.95% positive); and “It is challenging to interpret dashboard data” (76.19% positive). Three items related to Design were also frequently positively endorsed: “The graphs on dashboards provide useful visuals for treatment progress” (80.95% positive); “I feel comfortable using Excel” (71.43% positive); and “I feel comfortable with computers and technology in general” (71.43% positive). These seven items also received high adjusted Likert values, ranging from 5.19 to 5.81 (see Figure 10).

Site differences on items. Several items were identified to have significantly different response values between sites *R* and *U*. Site *U* was found to have a more negatively valenced response than site *R* on the following items: “Dashboards are not worth the extra time or effort,” $t(17.6) = 3.26$, $p = .004$; “I do not have enough time to review dashboards with my clients in sessions,” $t(19) = 3.06$, $p = .006$; “I wait until clients have completed treatment to review how their dashboards look,” $t(18.5) = 2.95$, $p = .008$; “Dashboards are less useful for family sessions,” $t(19) = 2.76$, $p = .012$; “The post-training MAP consultation calls were helpful around using dashboards,” $t(19) = 2.76$, $p = .012$; “Other clinicians at my employer/agency feel that dashboards are useful,” $t(19) = 2.58$, $p = .018$; “I find it challenging to choose measures to track on a dashboard,” $t(19) = 2.41$, $p = .026$; “Dashboards are useful tools for treatment planning,” $t(19) = 2.40$, $p = .027$; “I felt overwhelmed during MAP training because I had recently started at my employer/agency,” $t(18.5) = 2.36$, $p = .030$; and (marginally) “It is a priority for my employer/agency that I use dashboards,” $t(19) = 2.09$, $p = .051$.

Category valences. Responses to dashboard items were also examined across primary category classifications. The ratio of positively valenced to negatively valenced responses varied notably across categories (see Figure 11). Organizational Context items were most likely to be associated with negatively valenced responses (59.0%) and had an average adjusted Likert value of 2.99. Sustainability items were also more likely to receive negatively valenced responses (50.6%) with an average adjusted Likert value of 3.74. Design items were more likely to receive positively valenced responses (49.5%) with an average adjusted Likert value of 4.47. EBA Knowledge & Use items were most likely to receive positively valenced response (65.3%) with an average adjusted Likert value of 4.97.

Top endorsed challenges and benefits. Weighted rank order values from top-3 rankings were used to identify the most strongly endorsed challenges and benefits to using dashboards (see Table 5). The most highly endorsed challenges for both sites were "time required to enter data" and "no integration with electronic health record (EHR)." The most highly endorsed benefits for both sites were "having graphs/visuals for treatment progress" and "monitoring treatment progress."

Dashboard use attitude items. Dashboard use attitude items, as adapted from the UTAUT (Venkatesh et al., 2003), produced positively valenced responses for all items, ranging from apprehension around using dashboards (M = 4.14, SD = 1.71; reverse coded) to use of clinical dashboards being a good idea (M = 5.14; SD = 1.71). No significant differences were found between sites *R* and *U* on these items.

Mixed

Integration of the qualitative and quantitative data collection strands was conducted at multiple levels of the study: design-level with the selection of an exploratory sequential design;

method-level by using the qualitative data to inform the quantitative survey creation; and interpretation-level by connecting the qualitative data with the quantitative data using a joint display, which allows data to be visually brought together to examine new insights (Creswell and Clark, 2018). In tables 6 and 7 (for barriers and benefits, respectively), sample quotes from qualitative interviews are compared with results from the quantitative survey. Additional sample quotes, where available, are represented from clinician responses to the free-text dashboard opinion question within the survey.

Discussion

Through the preceding investigation, a mixed methods design was employed using an exploratory sequential approach to examine barriers and benefits to dashboard use as identified by supervisors and clinicians. Data gathered from the supervisor-focused qualitative phase drove creation of the quantitative survey for clinicians, and data from each were compared to identify areas of agreement and discrepancy.

Within the qualitative phase, interviews were coded based on emergent themes and a priori research. Four major categories of barriers/benefits were identified and confirmed: sustainability, organizational context, design, and evidence-based assessment and use. Subcategories were identified within each area and informed creation of quantitative survey questions around more targeted areas. Examination of coded qualitative comments revealed a greater representation of barriers than benefits, with sustainability concerns (specifically around the time commitment necessary to use dashboards) as the most commonly mentioned barrier to dashboard use. Design concerns, especially around user interface challenges, were also frequently identified as barriers related to regular dashboard use, with issues related to data entry, underuse of dashboard functionality, setup difficulties, and a general lower level of comfort with

technology. Organizational context, e.g., agency and supervisor support of dashboard use, as well as evidence-based assessment (EBA) knowledge were less commonly endorsed as barriers. Supervisor-endorsed benefits of dashboards were most prominently represented in the EBA knowledge category, where treatment utility and implementation of the MAP system were both noted subcategories. The emergence of a MAP subcategory points to supervisors' integrated mental representations of dashboards and the larger MAP framework. Of particular interest was the frequent endorsement (5.04% of coded comments) of consultation calls as a benefit. Although not specific to dashboards, their perceived importance to dashboard use aligns with research demonstrating the implementation benefits of post-training consultations (e.g., Beidas, Edmunds, Marcus, & Kendall, 2012).

Within the quantitative phase, clinicians similarly ranked time constraints as the top barrier to dashboard use. They also endorsed a lack of integration with electronic health records. Although a lack of EHR integration was categorized primarily as a design barrier, it also has clear implications related to time constraints as the separation of clinical tracking systems inevitably requires additional time for data entry and retrieval. Clinicians ranked their top benefits of dashboards to be the data visualization capabilities and the ability to monitor treatment progress. These endorsements point to a general level of buy-in to EBA use in treatments. The positively valenced UTAUT items, which assessed for attitudes towards dashboards, also suggest a generally sympathetic outlook towards dashboard use to improve treatment plans and outcomes.

At the category level, organizational context (i.e., agency and supervisor support) was most associated with negatively valenced judgments. Specific items contributing to this were related to a perceived lack of agency and supervisor support and prioritization around

dashboards. It is noteworthy that this category was most negatively valenced despite its related items (agency and supervisor support) not being ranked as top barriers by clinicians. This discrepancy suggests that clinicians do not see the lack of support as a significant issue relative to other more immediate concerns, such as time required to use dashboards. However, organizational context has been shown to play a significant role in successful implementations (Gleacher et al., 2016; Langley et al., 2010; Nadeem, Cappella, Holland, Coccaro, & Crisonimo, 2016), especially in school-based environments such as those where the majority of the participating clinicians provide services. A lower level of agency and supervisor prioritization seems likely to contribute to negative views of dashboard use and may also contribute to the sustainability time concerns associated with dashboard use not being prioritized in work requirements. Looking to the other categories, dashboard design presents challenges to clinicians but is broadly viewed as positive, buoyed by the strength of data visualizations. Finally, as reflected in the top ranked benefits of dashboard use, clinicians generally view EBA use as a positive, especially around treatment progress monitoring.

Several differences were found between clinician respondents from sites *R* and *U*, with site *U* being significantly more likely to endorse negative views towards dashboards on several individual items across all categories. Notably, site *U* was also found to be younger and with significantly fewer years of clinical experience. This combination of findings rounds counter to expectations that younger clinicians may be more open to dashboard and technology use. Morris & Venkatesh (2000) note the role of age in technology adoption, where younger workers are more strongly influenced by attitudes towards technology whereas older workers are more strongly influenced by subjective norms and perceptions of how easy it is to use. These findings suggest that younger clinical workforces, such as site *U*, may benefit from implementation

efforts focused on improving attitudes towards dashboards, whereas older clinical workforces, such as site *R*, may receive more benefit from special focus on organizational contexts and "user-friendly" design.

For both sites, clear challenges exist around time constraints and lack of agency and supervisor support. Supervisor/clinician discrepancies appear in the views related to organizational context, i.e., agency and supervisor priorities. Both roles reported a lower sense of agency prioritization of dashboard use (e.g., supervisor: "It's not really enforced agency-wide;" clinician: "I completed them to finish them for my agency mandate, and my agency does not make me use them going forward."), but, relative to clinicians and other categories, supervisors appeared to underestimate the negative impact of these lower priorities. Consequently, a lessened top-down push for dashboard use likely contributes to clinicians viewing dashboards as less critical or useful for their needs. A lower agency/supervisor prioritization may also lead to a lower likelihood that clinicians' dashboard time constraints would be recognized and addressed since the time required for their effective use would also not be prioritized. Supervisors also appeared to overestimate the challenges associated with dashboard design as compared to clinicians' views around their continued use. Although this discrepancy may be due to supervisors overattributing dashboard challenges to extra-agency factors, usability and EHR integration improvements to dashboard design may indeed provide a means to minimize time commitments required to use dashboard effectively.

These findings should also be considered in the broader context of the examined agencies. As stated in the method section, the supervisors at each agency had been trained on MAP but had not been trained as MAP supervisors. One notable distinction to be made between these two roles is the inclusion of an organizational context focus within formal MAP supervisor

trainings. MAP supervisor trainings include an emphasis on recognizing and addressing some of the very organizational contextual difficulties, such as agency support efforts, that were identified as barriers by both clinicians and supervisors at the studied sites. Although supervisors trained in MAP may be adept at providing relevant clinical supervision, their nonstandard training is less likely to include a focus on clinician reinforcers that can contribute to a more successful dashboard implementation. As such, the concerns reported by the participants in this study may be partially reflective of the nonstandard MAP training approach rather than solely due to generalizable views of dashboards. This consideration may be especially relevant given the context of the included agencies, which are contained within a broader New York-based system that had experienced a recent history of failed non-MAP dashboard implementation efforts (see Gleacher et al., 2016). As such, the low levels of agency support reported by both sites as organizational context barriers may be reflective of a particularly challenging environment for dashboard implementations.

Several limitations should be considered regarding the current study. In the qualitative phase of the study, the collection and coding process may be prone to bias due to the single-coder nature of this study. This factor was addressed by using a foundation of a priori research along with multiple cycles of constant comparison to identify and isolate qualitative codes prior to conducting a test for intra-rater reliability between full code cycles. Another limitation relates to the sample size, particularly since the interviews were split between two sites (*R* and *U*) with a single two-interviewee session used to gather information from site *R* supervisors. Although the qualitative sample size met minimum guidelines outlined by Onwuegbuzie and Collins (2007) and no significant differences were identified between themes and codes found for sites *R* and *U*, the lower sample size did not allow for continued data collection to confirm the point of thematic

saturation (see Francis et al., 2010). The lower qualitative sample size was reflective of both low supervisor scheduling availability as well as the limited number of agency supervisors who had supervised MAP cases, especially since both supervisors and clinicians at each agency had received MAP training within a relatively recent timeframe prior to the study's initiation.

Future studies would also benefit from a more robust collection of both qualitative and quantitative data, especially in the examination of different treatment settings and clinicians' views. For treatment settings, a majority of study participants worked in a school environment, so conclusions may not generalize to all community-based services. Regarding clinicians' views, under the exploratory sequential mixed method approach used in this study, supervisors' qualitative input fed the creation of the quantitative clinician-facing survey, which may have limited the ability for clinicians to identify concerns that were otherwise outside the awareness of their supervisors. The brief responses contained within the survey's free-text item did not mention difficulties that were novel from the supervisor-identified categories; however, an increased qualitative focus on clinicians would allow for a more in-depth exploration of factors that may not have been endorsed by supervisors. Another limitation is related to the collection of data. Via the informed consent process, participants were provided the name of the primary investigator as well as the advising faculty member, who may have been recognized as the co-creator of MAP (Chorpita & Daleiden, 2009). As such, responses may have been influenced by the known affiliation with the system under investigation. Statements ensuring confidentiality and an absence of negative consequences were emphasized during recruitment and informed consent phases to minimize these potential effects. Finally, as discussed previously, the MAP trainings that the participants had received prior to engaging with this study were nonstandard implementations that was not delivered directly by PracticeWise trainers. Although the trainings

had used portions of licensed curriculum and content, they did not reflect official implementation protocols, and the findings may be influenced by these training differences.

Future research can expand on these findings by conducting in-depth interviews with clinicians to better understand the barriers that were reported here in quantitative form. Clinician time constraints limit their abilities to participate in extended interviews (much as they limit dashboard use), but additional incentives may be used to encourage participation. These efforts would be especially informative if used to examine barriers across various treatment settings, as this study found site differences at subcategory levels that may influence effective dashboard implementations. Based on the praise given to consultation calls and the identification of less agency support as a significant barrier, additional consideration should also be focused on better enabling agencies to establish ongoing in-house supports and dashboard "champions" who can enable and encourage increased dashboard use. Finally, an increased focus on improving dashboard user interfaces would help address barriers related to design concerns and may consequently target barriers related to time constraints by providing a smoother and more efficient dashboard experience.

Clinician dashboard use continues to face barriers despite evidence of their utility and relatively high levels of buy-in for evidence-based practices. Challenges across sustainability, organizational contexts, design, and evidence-based assessments contribute to a difficult landscape for effective dashboard utilization. However, benefits have also been identified across each of these realms, with opportunities to improve upon dashboard implementations by examining discrepancies between supervisor and clinician positions. Increased levels of dashboard prioritization at agency and supervisor levels as well as design improvements to assist

with clinician time constraints may offer the best opportunities to improve levels of dashboard use and, consequently, clinical outcomes.

Tables, Figures, and Appendices

Table 3

Codes, Frequencies, and Examples of Supervisor Comments Regarding Barriers to Dashboard Use

Category, subcategories, and coded themes	Frequency	% of total comments	Example
	144	60.50%	
Category: Sustainability	52	21.85%	
Time commitment	48	20.17%	
Additional task in workload	36	15.13%	"I guess the barrier is just the time. I don't think people have the time, or I think they probably would do it."
Not reviewed in client sessions	9	3.78%	"I never did it in session with the client."
Ongoing technical difficulties	3	1.26%	"We often have technological difficulties with our server upstate. We had outages... and we can't get onto our own files for one reason or another."
High stress environment	4	1.68%	
Frequent client crises	1	0.42%	"Sometimes, several clients [have constant crises], where it's hard to really focus on one specific thing every week like constantly."
Challenges gathering data	3	1.26%	"It has to be during session time because… in the schools it's just straight through like: go to a class, pick up a client if they're there, have a session, bring the client back to class, go find another client, so there's no waiting room time or anything like that."

Category: Design	37	15.55%	
Lacking EHR integration	5	2.10%	"I think if the dashboard were built into our EHR, it would be great. But it's not."
Inadequate coverage of risks/diagnoses	6	2.52%	"I do a lot of safety planning with my kids, and I have a lot of high risk clients who are high risk because of suicidal thoughts or attempts. There's not a lot of stuff on the dashboard to work with that."
User interface challenges	26	10.92%	
Data entry difficulties	11	4.62%	"I know that it's really fussy. One thing that is challenging is having to not be able to copy and paste repeated data because it will mess it all up, and then you have to scrap the whole dashboard and start from scratch."
Not using features fully/as intended	6	2.52%	"I don't know how much I used the dashboard to keep track of my work I guess, other than maybe the fear hierarchy – keeping track of the scores on that. But the rest of it… like what I did in the sessions with all those dropdowns… if I'm being honest, I don't know how much I really used it."
Challenges with setup	5	2.10%	"Something that was confusing and I think was confusing to a lot of people was the whole left/right thing, so that took a little bit of processing."
Lower technical knowledge/comfort	4	1.68%	"I think some of us folks are a little bit older… and I think it's like learning a new… it's like a different language sometimes."
Category: Organizational context	28	11.76%	
Low agency priority	17	7.14%	
Lacking support within agency	4	1.68%	Q: "Would you say there's anyone within the agency that provides support around dashboards?" A: "Yeah, there's nobody around."
Lacking accountability within agency	13	5.46%	"I bet that it would probably be used more appropriately if agencies chose to make the choice for themselves to really truly incorporate the dashboards into the documentation requirements. Then I'm sure we would be using it."
Low supervisor priority	11	4.62%	"It's never come up in my individual supervision or group supervision."

Category: Evidence-based assessment knowledge & use	27	11.34%	
Evidence-based challenges	12	5.04%	
Clinical judgment preferred	5	2.10%	"I would like to use more standardized instruments that I do now, and, even though I don't, I know some people don't use them at all."
Inexperience with using evidence-base	3	1.26%	"I had one supervisee who… I think it would have been harder for her, just the way that she thought and how she organized herself and everything. I think that would have been more of a problem. She's smart, and she has a lot to offer the kids, but I don't think that she really thought in the way that you have to in order to effectively use it."
Difficult choosing and using measures	4	1.68%	"It's difficult... for some clients coming up with weekly ratings for two separate things. Because I know that's sort of the ideal. I think one is easy enough, but two becomes more challenging, and sometimes if it's not done in a really thoughtful way, it can then become an extra burden on the client to have to answer those."
Training	15	6.30%	
New to MAP & therapy	10	4.20%	"A lot of them are learning how to do therapy in general. You get a brief overview in grad school, but really learning what it means to do CBT – what you say to the kid in front of you or how you explain that to a parent. So I think they're focusing a lot on that process."
Overwhelmed by training	5	2.10%	"I do think that that initial week of training was just so packed with everything. I'm not saying that I would like two weeks of training. I don't know how they would work around it, but it was a little overwhelming at first."

Table 4

Codes, Frequencies, and Examples of Supervisor Comments Regarding Benefits of Dashboard Use

Category, subcategories, and coded themes	Frequency	% of total comments	Example
	94	39.50%	
Category: Evidence-based assessment knowledge & use	61	25.63%	
Treatment utility	39	16.39%	
Helps treatment planning	23	9.66%	"We're coaching to use it to inform the clinical decisions. Just thinking about if you give a client a depression assessment – really looking at what items they score higher on and focusing energy on that, or looking at reasons why the scores may have gone up or down and adjusting treatment from there."
Tracking EBAs	13	5.46%	"I like where you do record... the scores of whatever assessments you're using, like the weekly and the every few months or whatever. That's a nice tool because we as an agency don't have a place in a client's chart where you just see all of the past scores."
Aligns with session context/progress	3	1.26%	"I know that with my own clients, it was accurate. That was cool, and she really liked seeing that and reflecting on that."
MAP	22	9.24%	
Consultation calls	12	5.04%	"The team who trained us, they're really accessible, and very helpful, knowledgeable , and great to work with. I found them to be really helpful through the whole process. Without that, I don't think that it would be a successful tool at all."
Dashboard integration with MAP	6	2.52%	"I think it was the dashboard and the stuff behind the dashboard. Like the practice guides that were informing treatment and the suggestion of certain measures to be using and things like that. That definitely changed the way the case was being handled."
Practice guides	4	1.68%	"Having the practice guides take a lot of anxiety out of it because they're saying, okay, I have to do psychoeducation on depression with this kid today. You have sort of a roadmap for that in the practice guides."

Category: Design	13	5.46%	
Easy to use	9	3.78%	"Once I got past the initial setup of the dashboard, it's a pretty fluid streamlined process."
Data visualizations	4	1.68%	"I'm a visual learner, so being able to visually track the client's progress and look where you need to go back and readjust is the most helpful part of it."
Category: Organizational context	12	5.04%	
Agency support/priority	8	3.36%	"We have the director of analytics now, and he's very hands-on. We're using them to record for compliance and paperwork, so I think more and more our agency is very much moving in that direction [of using dashboards more frequently.]"
Supervisor support/priority	4	1.68%	"Yeah, supervisors are asking about it in supervision each week."
Category: Sustainability	8	3.36%	
In-session use	8	3.36%	"It helped me shift the way I worked with a mom. It helped the mom shift the way that she was engaging with him by looking at the correlation between the more time that they spent just the two of them really focused on something positive and engaging, the fewer arguments they were having throughout the course of the week."

Table 5.

Top Clinician-Identified Challenges and Benefits of Dashboard Use, Weighted by Rank

Challenge	Total	Site *R*	Site *U*
Time required to enter data	37	10	27
No integration with electronic health record (EHR)	28	10	18
Difficult getting measurable data from clients	13	10	3
No time to share with clients in session	9	4	5
Difficult choosing measures to track	8	4	4
Not worth the extra time/effort	8	0	8
Computer/network issues	6	4	2
Not an employer/agency priority	4	2	2
Confusing/complicated to enter data	3	0	3
Lack of employer/agency support	3	3	0
Challenging to interpret data	3	0	3
Difficult using with high risk/crises cases	2	0	2
Not a supervisor priority	1	1	0
Other (“Don’t see the purpose”)	1	0	1
Preference to use clinical judgment alone	0	0	0
Benefit	**Total**	**Site *R***	**Site *U***
Having graphs/visuals for treatment progress	33	10	23
Monitoring treatment progress	32	9	23
Having treatment data in one place	14	3	11
Assisting with treatment planning	11	7	4
Examining treatment progress alongside practices used in sessions	11	7	4
Tracking practices in sessions	9	4	5
Staying compliant with employer/agency expectations	5	1	4
Improving client treatment outcomes	4	2	2
Sharing with clients in sessions	4	3	1
Staying compliant with supervisor expectations	2	2	0
Other (“Having access to practice guides related to specific treatment needs”)	0	0	1

Table 6

Joint Display Comparison of Dashboard Barrier Data from Qualitative and Quantitative Strands

	Supervisors		Clinicians		
Category	Interviews	% negative	Survey response	% negative	Example comment
Sustainability		86.7%		50.6%	
Time commitment	"It's just an extra step I think for clinicians who are already overwhelmed."		"It is easy to fall behind on keeping dashboards updated" (95.24% agree; 2.05 rating) • Top ranked challenge to dashboard use		"The concern is the time is takes to complete dashboard given already extremely hectic schedules that dashboards tend to fall on the back burner."
High stress environment	"I tend to work with high risk, high crises clients, so when they're more in crisis, there's less time for that."		"I do not have enough time to review dashboards with my clients in session" (71.43% agree; 2.95 rating)		"I think they are helpful and useful tools, but I do not think they are practical in our setting, with high caseloads and heavy crisis."
Organizational context		70.0%		59.0%	
Low agency priority	"If it were incorporated into like the other regular practices of the agency, then I could see it being used more properly."		"It is a priority for my employer/agency that I use dashboards" (71.43% disagree; 2.71 rating)		"I have not used dashboards since the training ended. We are swamped with other documentation, and it is not a priority for myself or for my agency."
Low supervisor priority	"We didn't actually look at the dashboard, which probably maybe is bad, but sometimes there's so much that has to be done in an hour."		"It is a priority for my supervisors that I use dashboards" (71.43% disagree; 2.33 rating)		n/a

Design		74.0%		32.9%	
Lacking EHR integration	"I mean, I wish it was kind of just built into our electronic health record because it is an extra step, and paperwork is maddening."		"Dashboards are a burden because they are not part of our electronic health record (EHR) system" (71.43% agree; 3.00 rating) • Second ranked challenge to dashboard use		"Would like to use as part of collaborative documentation but without it being a part of EHR, it's not feasible or realistic."
Inadequate coverage of risk/diagnoses	"Sometimes several clients [have constant crises or other things going on], where it's hard to really focus on one specific thing every week like constantly."		"Dashboards are less useful when working with clients with frequent crises or high-risk concerns" (52.38% agree; 3.90 rating)		n/a
UI challenges	"I think that if it were simpler, it would be more helpful; like less to it."		"It was difficult to set up a new dashboard" (47.62% agree; 3.95 rating)		"Dashboards themselves were a bit too finicky in Excel. There was a constant fear of making a small error and having to create a dashboard from scrap again."
EBA knowledge & use		30.7%		20.7%	
Evidence-based challenges	"I'm thinking about the younger kids maybe, where it's harder to get like accurate self-reporting from them."		"I find it challenges to choose measures to track on a dashboard" (38.10% agree; 4.38 rating)		"I found it difficult to determine a measurement area and sticking with that thing throughout the training period, primarily with new clients."
Training	"Training… was four or five days, which seems like a lot, but I guess that in hindsight it didn't really feel like a lot because it was jam-packed and a lot of information."		"I felt overwhelmed during MAP training because of the amount of content covered." (52.38% agree; 3.76 rating)		"Learning the material was a little overwhelming and fast-paced."

Table 7

Joint Display Comparison of Dashboard Benefit Data from Qualitative and Quantitative Strands

	Supervisors		Clinicians		
Category	Interviews	% positive	Survey response	% positive	Example comment
EBA knowledge & use		69.3%		65.3%	
Treatment utility	"It's really nice being able to see it on paper because, you know, we're human and we get caught up in our interactions with our clients, and sometimes we might think they're making a certain kind of progress or not, but then being able to refer to the dashboard to see the concrete numbers helps to put that into perspective"		"Collecting and tracking data is useful for treating clients" (90.48% agree; 5.81 rating) • Second ranked benefit to dashboard use		"I find dashboard to be very helpful in tracking and monitoring treatment outcomes."
MAP	"I think that many have felt that MAP itself is helpful and really nice way to structure therapy, especially if you're feeling stuck with the client or there's not a clear intervention."		"The post-training MAP consultation calls were helpful around using dashboards" (80.95% agree; 5.48 rating)		"Consultation calls were helpful. Dashboards/MAP did push me to use a strategy I may not have used otherwise with a client."
Design		26.0%		49.5%	
Easy to use	"I literally just plug in the data, and it does the rest."		"Entering data into dashboards is confusing or complicated" (66.67% disagree; 4.76 rating)		"Creating initial dashboards is time-consuming, but once completed, entering data was relatively simple."
Data visualizations	"I personally like how it computes all of the data onto the chart and to be able to see it mapped out like that is very useful."		"The graphs on dashboards provide useful visuals for treatment progress" (80.95% agree; 5.81 rating) • Top ranked benefit to dashboard use		"I think they are a great tool and can provide useful data to share with clients that includes a visual for helping clients see their own progress."

Organizational context		30.0%		26.7%	
Agency support/ priority	"I think everybody feels very much that they would be very helpful. Definitely. I mean, I know from the top-down, they definitely feel that way."		"Someone within my employer/agency provides support to help me use dashboards in clinical work" (52.38% disagree; 3.29 rating)		n/a
Supervisor support/ priority	"In the context of supervision […] obviously it's encouraged."		"It is a priority for my supervisors that I use dashboards" (71.43% disagree; 2.33 rating)		n/a
Sustainability		13.3%		40.5%	
In-session use	"Typically, when I do them with my clients, I'll pull up the last one we did, and we'll talk about what the changes are and what's been different and why."		"It is useful to review dashboards with my clients in session" (71.43% agree; 5.00 rating)		"My clients enjoyed seeing the progress."

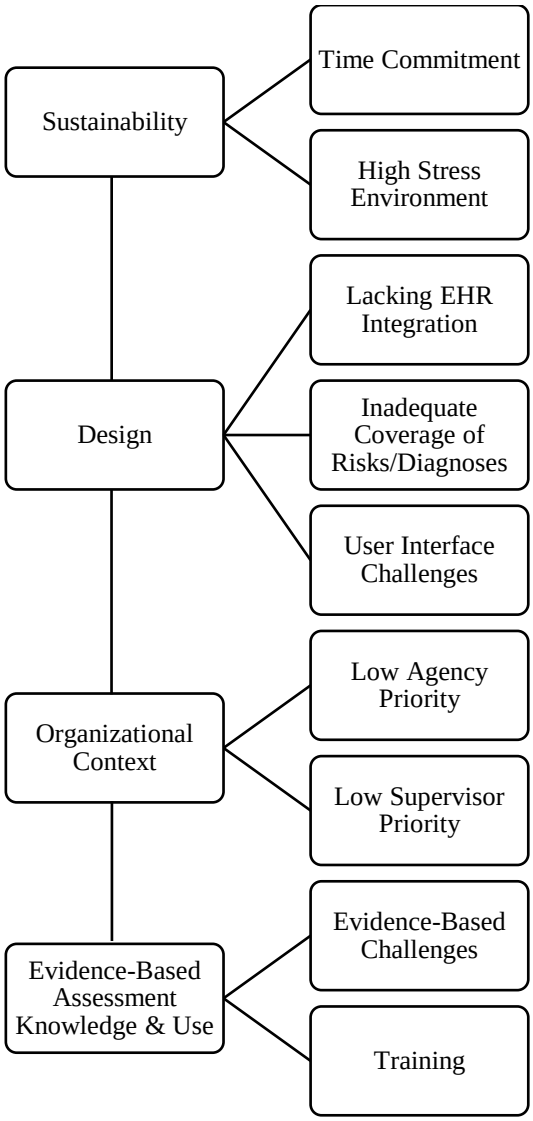


Figure 6. Categories and subcategories of coded dashboard barriers.

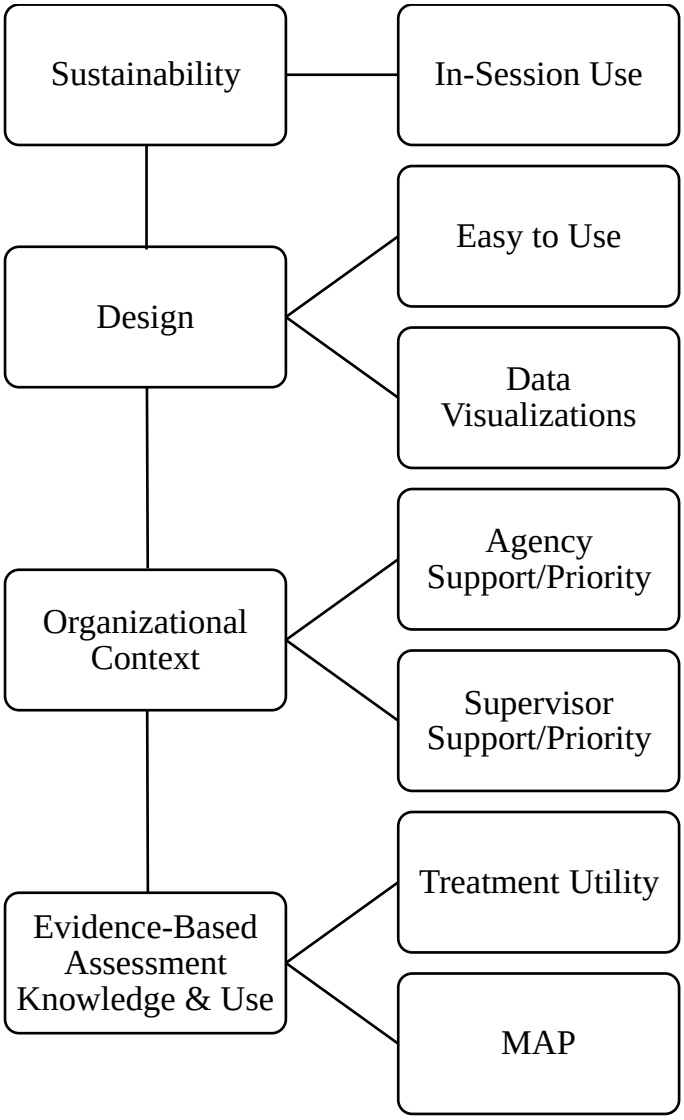


Figure 7. Categories and subcategories of coded dashboard benefits.

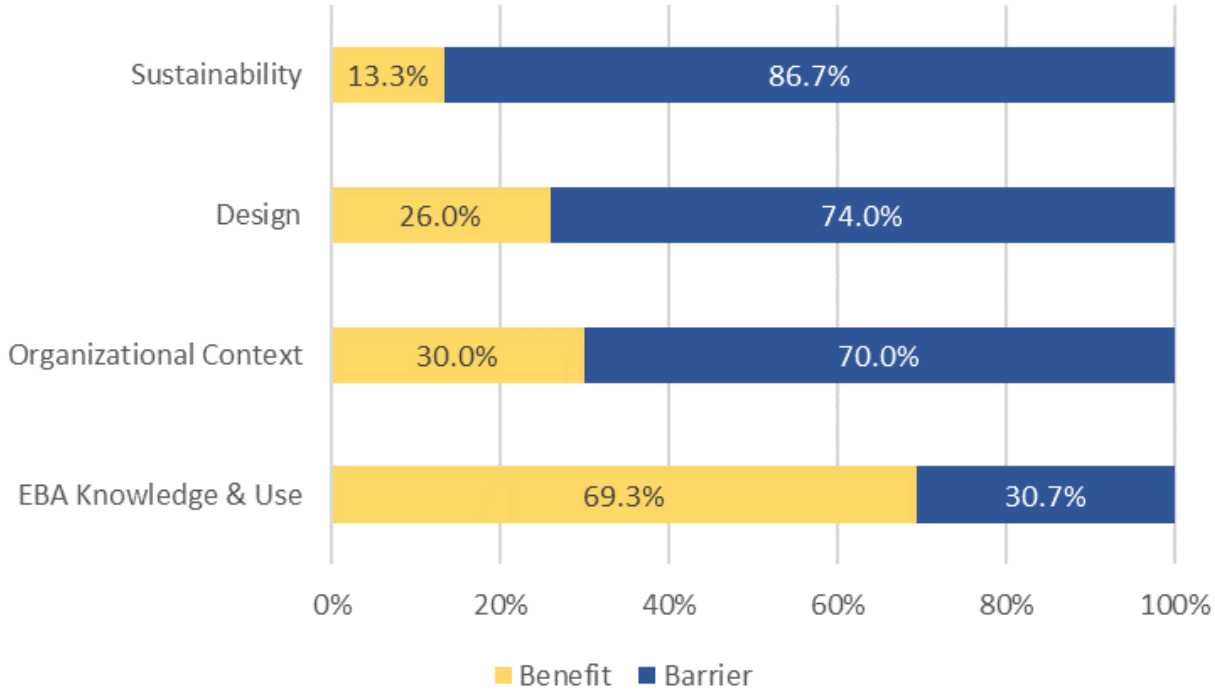


Figure 8. Dashboard barriers and benefits by category based on most frequently endorsed code types in qualitative interviews.

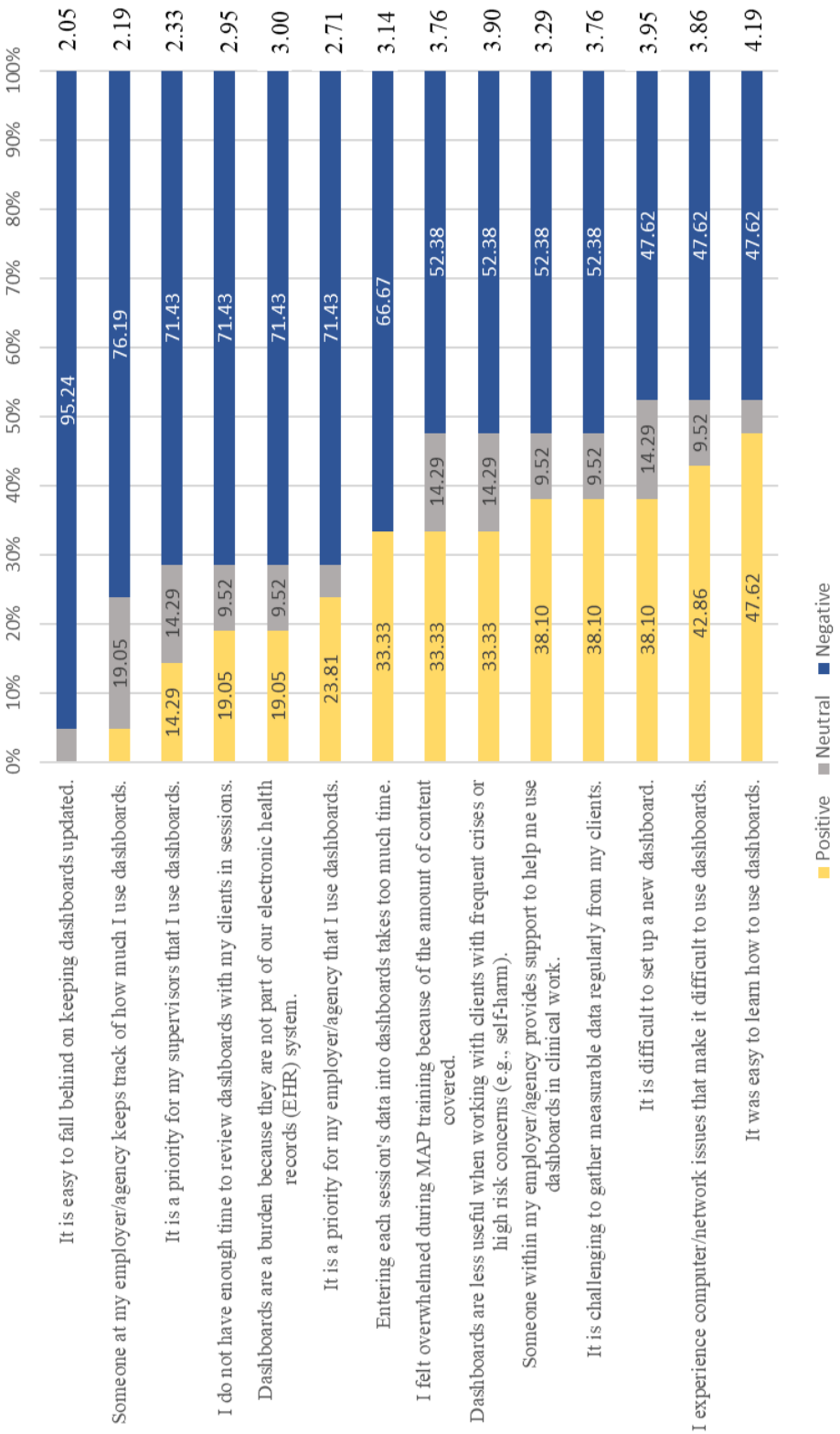


Figure 9. Dashboard items most frequently endorsed with negatively valenced responses. The right axis represents the average adjusted Likert value (1 *negative* – 7 *positive*) for each item.

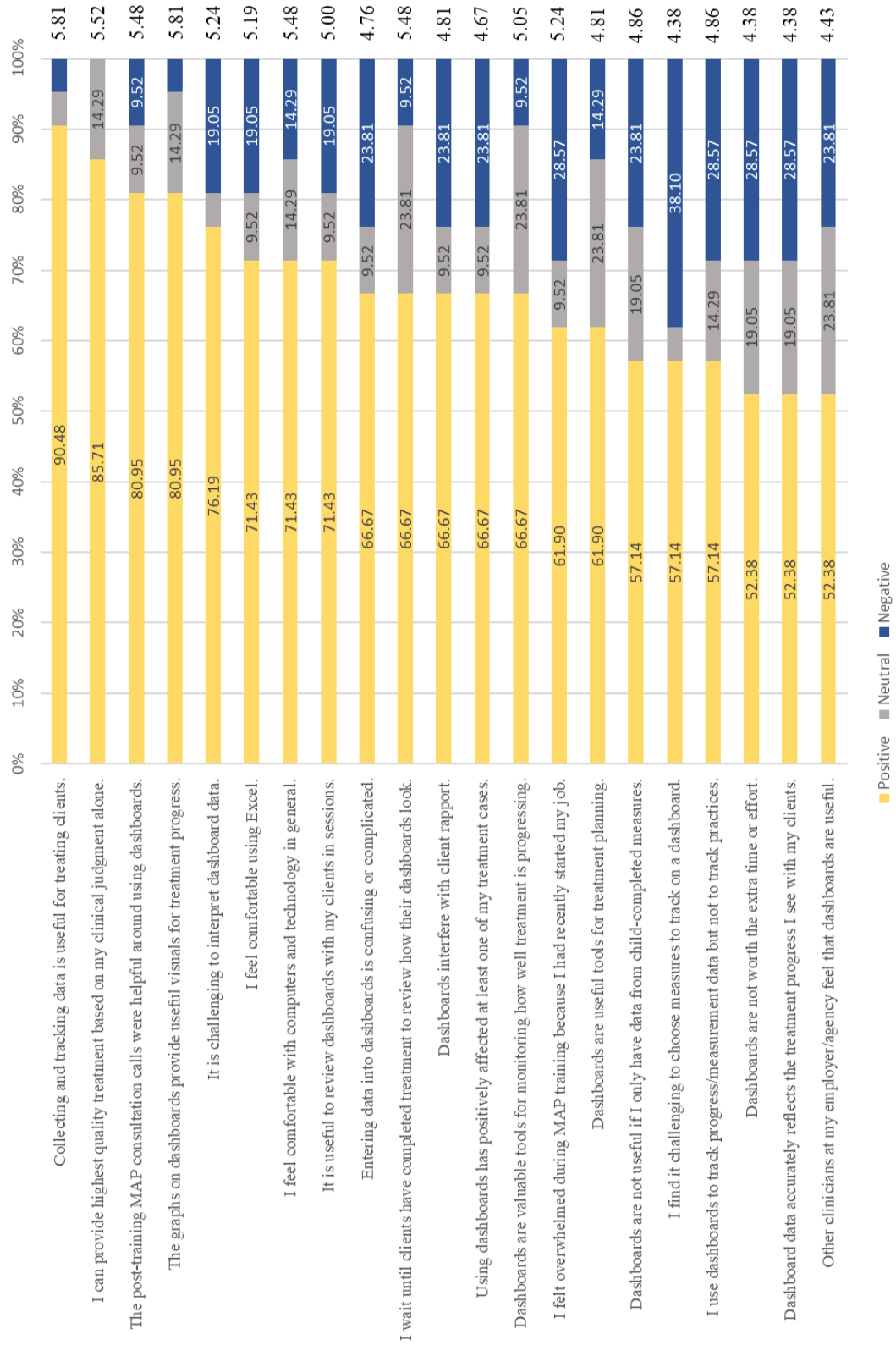


Figure 10. Dashboard items most frequently endorsed with positively valenced responses. The right axis represents the average adjusted Likert value (1 *negative* – 7 *positive*) for each item.

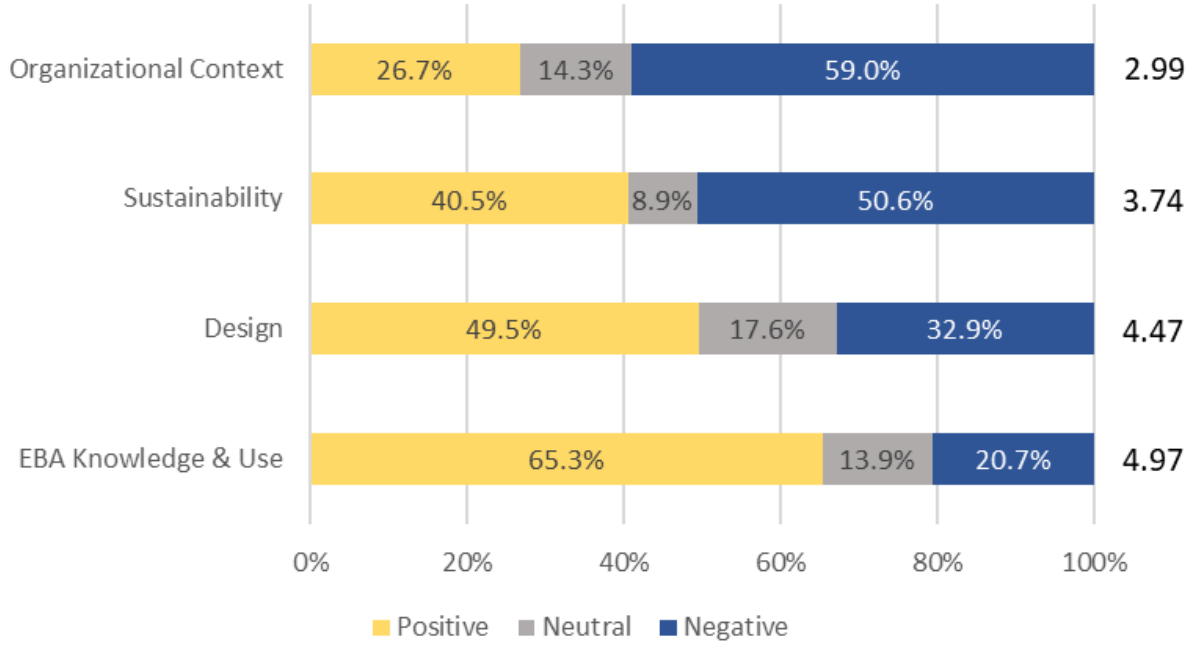


Figure 11. Dashboard barriers and benefits by category based on frequency of positively, negatively, and neutrally valenced responses to quantitative Likert items. The value on the right axis represents the average adjusted Likert value (1 *negative* – 7 *positive*) across each category.

Appendix E

Semi-Structured Interview (Qualitative Phase)

General Overview

1. Overall experience with using ongoing use of dashboards to monitor progress?

2. What are clinicians'/supervisors' feelings around dashboards? Around technology in general?

3. How would you describe clinicians'/supervisors' experience using dashboards? What do they like or dislike?

4. What supports have been helpful for them around using dashboards?

5. What have been some of the barriers to successful implementation?

Management Support

1. How well has your agency supported the implementation of dashboards? (For example: giving time, training, administrative support)

    a. Do you think dashboards are a priority at the supervisory/management level? Why or why not?

    b. How much do you feel like other clinic obligations (limited time in general) or initiatives impact dashboard use?

2. How have clinicians' peers responded to dashboards, i.e., encouraging, dismissive, helpful, etc.?

3. Is there someone at your agency that particularly helps support dashboards? What does he or she do that is helpful?

Technical Issues

1. How do clinicians/supervisors feel about the design of the dashboard? (for example: the layout, interface, graphics, etc.)
    a. What are some of the technical difficulties, if any, with the dashboard?

    b. Which components of dashboard do they feel most comfortable using?

c. Which components of dashboard do they feel least comfortable using?

2. What do you believe could have been done differently to make them feel more comfortable with the technical aspects of using the dashboard?

3. Access to computers?

Clinical Use

1. For clinicians and supervisors, what role do they see data playing? Are they using it to make decisions, just validating decisions they make other ways, or not using it at all?

2. How often are dashboards used in supervision? Are they being used to make decisions/suggestions, validating decisions, check for required compliance, or not at all?

3. Are there situations/cases where dashboard use is more likely? Less likely?

4. Do they think the measures reported by the program accurately reflect their client's current state? (i.e., does it mesh with what they see in session?)

5. Is there sufficient time during a session for clients and caregivers to complete dashboard measures? Is there sufficient time to discuss their responses?

6. Did clinicians/supervisors feel comfortable interpreting dashboard data? If not, what would be helpful to increase their comfort levels?

7. Could clinicians generally identify a time when dashboard data have (or have not) changed the way they handled a case?

8. Overall, do you believe clinicians/supervisors feel that a program like the dashboard could be helpful to them in a clinical setting? Why or why not?

Appendix F

Online Survey (Quantitative Phase)

Dashboard Benefits and Barriers

Start of Block: Consent

Q1
UNIVERSITY OF CALIFORNIA LOS ANGELES STUDY INFORMATION SHEET
Dashboard Benefits and Barriers

Todd Brown, M.A., C.Phil. and Bruce Chorpita, Ph.D. from the Psychology Department at the University of California, Los Angeles (UCLA) are conducting a research study.

You were selected as a possible participant in this study because you have completed training on the use of clinical dashboards. Your participation in this research study is voluntary.

Why is this study being done?
This study is being conducted to examine benefits and challenges that you have experienced around using clinical dashboards in the treatment of your clients. Most notably, the use of clinical dashboards may be impeded by a number of factors that make it difficult to use them on a regular basis. This study aims to identify those factors so that they may be examined and addressed to improve clinician experiences and client outcomes.
What will happen if I take part in this research study?
If you volunteer to participate in this study, the researcher will ask you to do the following: You will be asked to report your views around benefits and challenges to dashboard use. This task will be conducted via this one-time internet-based survey.
How long will I be in the research study?
Participation will take a total of about 15 minutes maximum, and possibly shorter than that.
Are there any potential risks or discomforts that I can expect from this study?
There are no anticipated risks or discomforts.
Are there any potential benefits if I participate?
You may benefit from the study by helping to identify challenges to dashboard use within your clinical practice and agency. The results of the research may lead to efforts to address these challenges and improve your experience.
What other choices do I have if I choose not to participate?
Your participation in this study is optional, and no alternative tasks are required if you choose not to participate.
Will I be paid for participating?
You will receive a $10 Amazon.com digital gift card by completing the survey. The gift card will be delivered to the email address you specify by April 15, 2019.

Will information about me and my participation be kept confidential?

Any information that is obtained in connection with this study and that can identify you will remain confidential. It will be disclosed only with your permission or as required by law. Confidentiality will be maintained by storing your identifying information (name & email – collected for gift card purposes) in a document separate from your responses. Only the research team will have access to your information.

What are my rights if I take part in this study?

You can choose whether or not you want to be in this study, and you may withdraw your consent and discontinue participation at any time. Whatever decision you make, there will be no penalty to you, and no loss of benefits to which you were otherwise entitled.

Who can I contact if I have questions about this study? The research team:

If you have any questions, comments or concerns about the research, you can talk to the one of the researchers. Please contact:

Todd Brown
toddbrown@ucla.edu
xxx-yyy-zzzz

Bruce Chorpita
chorpita@ucla.edu
xxx-yyy-zzzz

UCLA Office of the Human Research Protection Program (OHRPP):

If you have questions about your rights as a research subject, or you have concerns or suggestions and you want to talk to someone other than the researchers, you may contact the UCLA OHRPP by phone: (310) 206-2040; by email: participants@research.ucla.edu; or by mail: Box 951406, Los Angeles, CA 90095-1406.

Q26 By proceeding, I consent to participate based on the information provided above.

- ○ I consent.
- ○ I decline.

Page Break

End of Block: Consent

Start of Block: Demographics/Background

Q25 **Please provide the following background information about yourself and your clinical work.**

Q7 Age:

_____

Q11 Gender:

- ○ Male
- ○ Female
- ○ Non-binary/third gender
- ○ Prefer to self-describe _____
- ○ Prefer not to say

Q6 I identify my ethnicity as: (select all that apply)

- [ ] Asian
- [ ] Black or African American
- [ ] Hispanic or Latina/o/x
- [ ] Middle Eastern or North African
- [ ] Native American
- [ ] Native Hawaiian or Pacific Islander
- [ ] White or Caucasian
- [ ] Other _____

Page Break

Q14 Highest level of degree obtained:

- ○ MSW, ASW, LCSW
- ○ MFT, LMFT
- ○ MD
- ○ PhD
- ○ PsyD
- ○ Other _____

Q15 Degree specialty:

- ○ Psychology (Clinical/Counseling)
- ○ Education/School/Counseling
- ○ Social Work
- ○ Marriage and Family
- ○ Other _____

Q16 Year you completed degree:

_____

Q17 Currently licensed?

- ○ Yes
- ○ No

Page Break

Q12 Agency/employer with whom I primarily work:

- ○ Site *U*
- ○ Site *R*
- ○ Other _____

Q19 Primary setting for clinical practice:

- ○ Academic medical center
- ○ Community mental health center
- ○ Hospital
- ○ Private practice
- ○ School
- ○ Other _____

Q22 Average face-to-face clinical hours per week:

_____

Q20 Average clinical supervision hours per week:

_____

Page Break

Q24 Have you received training on the MAP system? (Managing and Adapting Practice, i.e., PWEBS Database; Practitioner Guides; Clinical Dashboards)

- ○ Yes
- ○ No

Q29 When did you receive MAP training? (best estimate)

	Month	Year
Please Select:	▼ January ... December	▼ 1990 ... 2019

Q30 Overall, how often do you use clinical dashboards?

- ○ Not at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ A great deal

Q28 I work with dashboards in the following role:

- ○ Clinician
- ○ Supervisor
- ○ Both Clinician and Supervisor
- ○ Neither Clinician nor Supervisor

End of Block: Demographics/Background

Start of Block: Dashboard questions

Q31 **The following questions will ask you about your experiences using clinical dashboards. Please provide answers that best reflect your views.**

Q32 Use the scale below to rate how much you agree with the following statements.

(note: Likert shortened here for display purposes only)	Complete Disagree	Mostly Dis.	Slightly Disagree	Neither Agree Nor Disagree	Slightly Agree	Mostly Agree	Completely Agree
Entering each session's data into dashboards takes too much time.	○	○	○	○	○	○	○
Dashboards are not worth the extra time or effort.	○	○	○	○	○	○	○
It is easy to fall behind on keeping dashboards updated.	○	○	○	○	○	○	○
I do not have enough time to review dashboards with my clients in sessions.	○	○	○	○	○	○	○
Entering data into dashboards is confusing or complicated.	○	○	○	○	○	○	○
I feel comfortable using Excel.	○	○	○	○	○	○	○
Someone within my employer/agency provides support to help me use dashboards in clinical work.	○	○	○	○	○	○	○
It is a priority for my supervisors that I use dashboards.	○	○	○	○	○	○	○
It is a priority for my employer/agency that I use dashboards.	○	○	○	○	○	○	○
Dashboards are a burden because they are not part of our electronic health records (EHR) system.	○	○	○	○	○	○	○
It is difficult to set up a new dashboard.	○	○	○	○	○	○	○

Someone at my employer/agency keeps track of how much I use dashboards.	○	○	○	○	○	○	○
It is challenging to gather measurable data regularly from my clients.	○	○	○	○	○	○	○
I can provide highest quality treatment based on my clinical judgment alone.	○	○	○	○	○	○	○
I felt overwhelmed during MAP training because of the amount of content covered.	○	○	○	○	○	○	○
I felt overwhelmed during MAP training because I had recently started at my employer/agency.	○	○	○	○	○	○	○
I experience computer/network issues that make it difficult to use dashboards.	○	○	○	○	○	○	○
Dashboards are less useful when working with clients with frequent crises or high risk concerns (e.g., self-harm).	○	○	○	○	○	○	○
Dashboards are only useful when working with clients with certain diagnoses.	○	○	○	○	○	○	○
Dashboards are not useful if I only have data from child-completed measures.	○	○	○	○	○	○	○
I feel comfortable with computers and technology in general.	○	○	○	○	○	○	○

I wait until clients have completed treatment to review how their dashboards look.	○	○	○	○	○	○	○
I find it challenging to choose measures to track on a dashboard.	○	○	○	○	○	○	○
I use dashboards to track progress/measurement data but not to track practices (i.e., what is done in sessions).	○	○	○	○	○	○	○
It is challenging to interpret dashboard data.	○	○	○	○	○	○	○
Dashboard data accurately reflects the treatment progress I see with my clients.	○	○	○	○	○	○	○
Dashboards are less useful for family sessions.	○	○	○	○	○	○	○
Dashboards interfere with client rapport.	○	○	○	○	○	○	○
Other clinicians at my employer/agency feel that dashboards are useful.	○	○	○	○	○	○	○
Using dashboards has positively affected at least one of my treatment cases.	○	○	○	○	○	○	○
Dashboards are valuable tools for monitoring how well treatment is progressing.	○	○	○	○	○	○	○
The post-training MAP consultation calls were helpful around using dashboards.	○	○	○	○	○	○	○
It is useful to review dashboards with my clients in sessions.	○	○	○	○	○	○	○

Collecting and tracking data is useful for treating clients.	○	○	○	○	○	○	○
The graphs on dashboards provide useful visuals for treatment progress.	○	○	○	○	○	○	○
Dashboards are useful tools for treatment planning.	○	○	○	○	○	○	○
It was easy to learn how to use dashboards.	○	○	○	○	○	○	○

Q33 Rank your top 3 benefits of using dashboards. ('1' for your top benefit, '2' for the next best, '3' for the third best)

_____ Monitoring treatment progress
_____ Tracking practices used in sessions
_____ Examining treatment progress alongside practices used in sessions
_____ Having treatment data in one place
_____ Having graphs/visuals for treatment progress
_____ Assisting with treatment planning
_____ Staying compliant with supervisor expectations
_____ Staying compliant with employer/agency expectations
_____ Sharing with clients in sessions
_____ Improving client treatment outcomes
_____ Other

Page Break

Q34 Rank your top 3 challenges around using dashboards. ('1' for your biggest challenge, '2' for the second biggest challenge, '3' for the third)

_____ Time required to enter data
_____ No time to share with clients in sessions
_____ Confusing/complicated to enter data
_____ Not a supervisor priority
_____ Not an employer/agency priority
_____ Lack of employer/agency support
_____ No integration with electronic health record (EHR)
_____ Challenging to interpret data
_____ Difficult gathering measurable data from clients
_____ Computer/network issues
_____ Difficult using with high risk/crises cases
_____ Preference to use clinical judgment alone
_____ Difficult choosing measures to track
_____ Not worth the extra time/effort
_____ Other

Page Break

Q35 Use the scale below to rate how much you agree with the following statements as related to your clinical work.

	Completely Disagree	Mostly Disagree	Slightly Disagree	Neither Agree Nor Disagree	Slightly Agree	Mostly Agree	Completely Agree
Going forward, I feel apprehensive about using dashboards.	○	○	○	○	○	○	○
Going forward, I would like working with clinical dashboards.	○	○	○	○	○	○	○
Going forward, I would find dashboards useful in clinical work.	○	○	○	○	○	○	○
Using clinical dashboards is a good idea.	○	○	○	○	○	○	○

Page Break

Q37 Finally, how would you describe your overall experience with using dashboards?

_____

_____

Start of Block: Gift Card Info

Q39 **Thank you for completing this survey.**

Please complete the following form to receive your $10 Amazon.com digital gift card, which will be sent by **April 15, 2019**.

As a reminder, your name and email will be used only for the purposes of gift card distribution. Your survey responses will remain confidential and stored separately from your identifying information.

- Name _____
- Email address _____

End of Block: Gift Card Info

References

Aarons, G. A., Farahnak, L. R., Ehrhart, M. G., & Sklar, M. (2014). Aligning leadership across systems and organizations to develop a strategic climate for evidence-based practice implementation. *Annual Review of Public Health, 35,* 255–274. https://doi.org/10.1146/annurev-publhealth-032013-182447

Aarons, G. A., Glisson, C., Green, P. D., Hoagwood, K., Kelleher, K. J., Landsverk, J. A., … Schoenwald, S. (2012). The organizational social context of mental health services and clinician attitudes toward evidence-based practice: A United States national study. *Implementation Science, 7,* 56. https://doi.org/10.1186/1748-5908-7-56

Aarons, G. A., Green, A. E., Trott, E., Willging, C. E., Torres, E. M., Ehrhart, M. G., & Roesch, S. C. (2016). The roles of system and organizational leadership in system-wide evidence-based intervention sustainment: A mixed-method study. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(6), 991–1008. https://doi.org/10.1007/s10488-016-0751-4

Aarons, G. A., Wells, R. S., Zagursky, K., Fettes, D. L., & Palinkas, L. A. (2009). Implementing evidence-based practice in community mental health agencies: A multiple stakeholder analysis. *American Journal of Public Health*, *99*(11), 2087–2095. https://doi.org/10.2105/AJPH.2009.161711

Agre, P. E. (1995). Conceptions of the user in computer systems design. In P. J. Thomas (Ed.), *The social and interactional dimensions of human-computer interfaces*, pp. 67-106. Cambridge: Cambridge University Press

Bannon, L. (1991). From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In J. Greenbaum & M. Kyng (Eds.), *Design at work*, pp. 25-44. Hillsdale, NJ: Lawrence Erlbaum Associates.

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, *63*(7), 660–665. https://doi.org/10.1176/appi.ps.201100401

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. https://doi.org/10.1007/s10488-015-0647-8

Borntrager, C., & Lyon, A. R. (2015). Client progress monitoring and feedback in school-based mental health. *Cognitive and Behavioral Practice*, *22*(1), 74–86. https://doi.org/10.1016/j.cbpra.2014.03.007

Brown, T. E. (2014). *Effects of training and expected treatment progress on clinicians' interpretations of measurement feedback systems*. Unpublished manuscript, Department of Psychology, University of California, Los Angeles.

Callaly, T., Hyland, M., Coombs, T., & Trauer, T. (2006). Routine outcome measurement in public mental health: Results of a clinician survey. *Australian Health Review*, *30*(2), 164–173.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Los Angeles: Sage.

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & The Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to

structure the decision models in service organizations. *Administration and Policy in Mental Health, 35*, 114–123. https://doi.org/10.1007/s10488-007-0151-x

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566–579. https://doi.org/10.1037/a0014565

Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3rd ed.). Los Angeles: Sage.

de Jong, K. (2016). Challenges in the implementation of measurement feedback systems. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 467–470. https://doi.org/10.1007/s10488-015-0697-y

de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research*, *22*(July), 464–474. https://doi.org/10.1080/10503307.2012.673023

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, *48*(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health, 25*(10), 1229-1245. doi:10.1080/08870440903194015

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *Journal of Behavioral Health Services & Research*, *30*(4), 393–405. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14593663

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems, 12*(4), 436-445. doi:10.2307/798843

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. New York: Aldine de Gruyter.

Gleacher, A. A., Olin, S. S., Nadeem, E., Pollock, M., Ringle, V., Bickman, L., … Hoagwood, K. (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. https://doi.org/10.1007/s10488-015-0642-0

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.2307/1163620

Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*, *17*(3), 251–255. https://doi.org/10.1037/1040-3590.17.3.251

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*, 1–14. https://doi.org/10.1007/s10488-016-0763-0

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Langley, A. K., Nadeem, E., Kataoka, S. H., Stein, B. D., & Jaycox, L. H. (2010). Evidence-based mental health programs in schools: Barriers and facilitators of successful implementation. *School Mental Health*, *2*, 105–113. https://doi.org/10.1007/s12310-010-9038-1

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. https://doi.org/10.1007/s10488-016-0719-4

Morris, M. G., & Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing work force. *Personnel Psychology*, *53*(2), 375–403. https://doi.org/10.1111/j.1744-6570.2000.tb00206.x

Nadeem, E., Cappella, E., Holland, S., Coccaro, C., & Crisonimo, G. (2016). Development and piloting of a classroom-focused measurement feedback system. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 379–393. https://doi.org/10.1007/s10488-015-0651-z

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education*, *33*(3), 301–314. https://doi.org/10.1080/02602930701293231

Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281–316.

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health, 38*(1), 44–53. https://doi.org/10.1007/s10488-010-0314-z

Robins, C. S., Ware, N. C., DosReis, S., Willging, C. E., Chung, J. Y., & Lewis-Fernandez, R. (2008). Dialogues on mixed-method and mental health services research: Anticipating challenges, building solutions. *Psychiatric Services, 59*(7), 727–731. Retrieved from http://ps.psychiatryonline.org/doi/pdf/10.1176/ps.2008.59.7.727

Saldaña, J. (2015). *The coding manual for qualitative researchers.* (3rd ed.). Los Angeles: Sage.

Tsai, K. H., Moskowitz, A. L., Brown, T. E., Park, A. L., & Chorpita, B. F. (2016). Interpreting progress feedback to guide clinical decision-making in children's mental health services. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(2), 199–206. https://doi.org/10.1007/s10488-015-0630-4

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Epilogue

In reflection of the studies contained within this manuscript, several challenges emerged at various stages of the research lifecycle. Whereas some of these difficulties may reflect limitations appropriate for acknowledgement within each study's write-up, others stand out as areas particularly worth exploring further in an effort to both recognize their occurrence and to minimize the likelihood of similar missteps in the future. This closing chapter to the dissertation stands to serve that role.

We turn first to Study 1 of this dissertation. Study 1 aimed to examine how dashboard features and user expertise affect community-based clinicians' interpretations of and attitudes toward dashboards. The study contained within this manuscript represented a portion of a larger training and investigation effort conducted in coordination with PracticeWise. As discussed in its chapter, the study focused on Minnesota-based clinicians who had had prior MAP and dashboard experience and were set to receive a one-day training booster session to refresh, consolidate, and standardize their understanding and knowledge of the MAP components. Study 1's role was focused on the dashboard feature, and the effort was directed there accordingly.

Early in the process of conducting this study, the most significant limitation of the effort, if not this full manuscript, arose during data gathering. Participants completing a study measure did so in a manner that led to questionable validity around a significant portion of the gathered data. The measures included vignettes and questions related to dashboard knowledge, clinical judgments, confidence ratings, and dashboard attitudes. Initially, these measures were intended to be administered electronically via websites accessible during the training sessions. An earlier stage of the overall booster session effort had included a successful implementation of a similar measure that was used for prescreening purposes around dashboard knowledge. However,

roughly two days prior to the training days, technical challenges were encountered that made the survey's web-based implementation infeasible in time for the training. Consequently, rather than administering a web-based measure, a paper-based measure was provided to participants. This paper-based version of the measure included an item layout and design that led participants to make tick marks rather than note discrete values. The subsequent fallout of these marks was the necessity for a manual measurement effort to establish reliability between raters; however, despite successfully establishing measurement reliability, the validity and interpretability of the content was impaired due to the random error associated with the unknown "true" values. Study 1's findings were thus weakened accordingly. Additional approaches may have also helped with the interpretation of the compromised data collection. For instance, the affected variables (decision confidence and adequacy of displayed data) could have been analyzed via a method that simplified the values to positively, negatively, or neutrally valenced. Although such an approach would lead to less nuanced statements and analyses due to the loss of detail provided by the Likert values, the data validity would likely improve for those judgments given the clearer assessments of where a tick fell relative to those ranges on the Likert line.

This challenge could have been avoided with additional attention paid to several key factors. Most critically, the study would have benefitted from a comprehensive pilot effort that reflected the study setting and possible contingencies as best possible. The measures had been piloted amongst several advanced graduate students prior to initial implementation efforts in order to receive feedback on timing, clarity, and general content. The piloted content was delivered via an emailed Word document that contained that survey questions because the web-based content continued to be under development by PracticeWise staff as the study approached its implementation. As a result, the pilot recipients were able to successfully provide feedback on

timing, clarity, and general content – but did not have the opportunity to provide feedback on the intended delivery method. Survey questions were answered with typed values within the emailed Word document, which was then returned via email as well. When the decision was made to cease development on the web-based survey, the Word document was delivered to be distributed to trainees in a paper format. The lack of piloting of the paper format, which had been recognized as the ultimate fallback plan in case of technical difficulties, contributed to a missed opportunity to catch and correct the data entry form. This result highlights the critical importance of fully piloting and planning for all foreseeable contingencies, lest one is not considered and suffers accordingly. The prolonged web-based development process could have served as a warning sign that additional care should be spent to plan for these technical difficulties. Additionally, a more hands-on approach to the survey's administration during the training could have caught the issue while there was still an opportunity for corrective efforts. Although travel to Minnesota for the training days was not undertaken due to the time- and cost-prohibitive nature of the trip, taking the effort to be present and as directly involved as possible with the training day would have enabled recognition of the mistake immediately rather than several days after the training had been completed and the data sheets scanned. Alternatively, providing additional study background information to the training team members could have enabled them to notice the concern in my stead, allowing them to contact me when something went awry. Furthermore, additional hands-on efforts and overseeing of the later-abandoned web-based effort would have allowed for either self-driven development efforts or an earlier recognition of the challenges being faced there. An earlier awareness of these challenges would have provided a more in-depth and less last-minute effort to ensure that the measures to be delivered met the

study's needs. Taken together, each of these approaches point to the importance of being as intimately involved as possible throughout each step of the study process.

In addition to improving the measures' means of administration, the measure content could have also benefitted from a more extensive piloting effort via the identification of additional insights or hypotheses that may have emerged through early discussions of additional implications. For example, the vignette measures included questions that were designed to assess participants' accuracy in rating clinical progress and practice decisions. Although these questions were in fact piloted and resulted in consensus agreements on ratings across multiple raters, further consideration during the analysis phase identified increased levels of uncertainty on these judgments based on discussions amongst the primary author and other consulting team members. While not fully foreseeable given the earlier consensus decisions, an increased level of attention and discussion to these factors earlier in the development process may have allowed the concerns to surface with enough time to modify the measure accordingly. The study would have benefitted noticeably if the ratings of decision confidence could have been presented alongside the actual decision performance. Similarly early discussions may have also surfaced additional questions around how participant confidence levels do interact with dashboard expertise. Although this manuscript proposed hypotheses around this interaction, a more in-depth pilot phase could have pointed to an area for further exploration instead.

The advantages of increased direct involvement extend beyond measure development and administration. Additional consideration of and curiosity around the broader training and study effort would have contributed to improved insights into the context in which the training was being held. These insights, whether focused on training history or demographic differences, would have contributed to additional questions for exploration and discovery during the analysis

phase. Although no significant differences were found between the Minnesota sites on collected demographic information, a greater understanding of the diversity between the sites may have pointed to additional factors or aspects around dashboard use worth exploring for variances dependent on "unreported" demographics and treatment settings.

The identification of hypotheses to explore could have benefitted from additional consideration of the study's traits as well. The current study looked at hypotheses across a number of independent and interacting variables, often leading to findings that were underpowered given the sample size in each condition. Since the booster training effort had a limited and known cap on participants from the start, the study's hypotheses could have benefitted from focusing on just those that could be sufficiently powered. Although the interactions and post hoc findings in this study point to interest potential future directions, the overall presentation of findings from the existing study could have been strengthened by a more focused approach give the sample size. Additionally, factors explored within the existing may have benefitted from a less assumed and more direct approach. For instance, the hypotheses around novice/expert preferences for level of complexity would have been made stronger with an item asking each participant to rate the vignettes' complexity, rather than assuming "more/less complex" labels based on component presence alone. Finally, careful use of wording throughout the study and manuscript would assure a clearer set of conclusions. For instance, recognizing earlier that the "users" may not have been appropriately labeled as such. Similarly, noticing the potential for misinterpretation via the use of "expert" and "novice" labels, when those judgments were relative to the study sample only. A more accurate word choice would allow results to be communicated better reflective of the actual findings.

Similar considerations can be explored for Study 2 of this manuscript. Study 2 aimed to identify barriers to sustained dashboard use by clinicians and supervisors in community mental health settings. Supervisor interviews and clinician questionnaires were used to find the perceived challenges and benefits of ongoing dashboard use, and the identified shared and discrepant barriers highlighted potential areas for improvement in implementation efforts.

One of the greater challenges faced with Study 2 was the initiation of the study itself. Prior to identifying and executing the described study with the New York-based agencies, two earlier efforts were made to explore the same themes with California-based community mental health centers. In each case, agency shifts in research prioritization played a role in the moves away from active participation with the study despite initial engagement, and these shifts each occurred after study planning and early execution had commenced with each. Although full shifts away from participation with the study were moderately unexpected, an earlier identification of these shifts may have been possible with an increased level of communication with the agency team around their concerns and potential upcoming barriers. The study status may also have benefitted from a more focused early effort on communicating the benefits the agency would receive through participation the study. Any such partnership requires a balancing act of maintaining regular contact while not becoming a burden, and the shifting financial and managerial landscapes made that assessment difficult at times. Nonetheless, more contact-oriented position on that spectrum would likely have uncovered the issues at an earlier time point, allowing more time to identify alternative study partners and approaches.

Just as awareness of emerging organizational factors could be helpful in foreseeing issues, awareness of past and ongoing organizational factors can do the same. Like the benefits previously discussed of understanding the context of the Minnesota study, a deeper

understanding of the full context of the New York agencies could have provided insights and additional areas for exploration during this study. Most notably, the recent failure of a dashboard implementation within the broader New York system highlights an area that could have better informed questions asked as well as interpretations of findings. Similarly, an earlier consideration of the distinction between trained MAP supervisors and supervisors who had been trained in MAP may have encouraged additional considerations and insights into the current organizational context of the New York agencies. In both cases, there had not been an explicit mention of either factor during the data gathering phase, which may have contributed to these facts remaining unknown at the time. However, once again, an increased immersion in the context of the study could have provided additional opportunities for these discoveries to be made naturally.

The coding of qualitative data in Study 2 presented several areas for potential refinement. To start, a larger sample size would have been especially beneficial to ensure adequate coverage across both agency sites as well as general thematic saturation in uncovering relevant codes. Acquisition of more supervisor participants would have been challenging due to the limited number of agency supervisors associated with MAP as well as their ongoing difficult schedules. However, given additional a longer window of qualitative data collection, each of these factors may have had the possibility to improve. Additionally, the use of incentives, as was done with the survey participants, may have encouraged additional participation at the supervisor level as well. Optimally, these additional participants would be identified more evenly across sites, especially in one-on-one interviews, to ensure the best representation of the participants' independent thoughts. In coding the data, establishing reliability with a second coder would have been beneficial to ensure that the identified codes did not reflect biases that may emerge from a

single-coder approach. Although this study focused its efforts on a priori research to serve as the foundation for barriers to dashboard use, otherwise invisible biases may have influenced the categorization of a statement into one category or another. Finally, the mixed method approach used in this study was designed to build a quantitative clinician-facing survey based on topics identified in the qualitative supervisor interviews. This unidirectional workflow allows for the possibility that clinicians were unable to voice their own concerns if supervisors had not already identified it as an area for exploration. Although the brief short-answer prompt in the survey did not produce unique themes, a more clinician-focused qualitative study would provide clinicians free rein to highlight heretofore unspoken concerns. Although clinicians in community mental health centers typically have very little free time in their workday to participate in these initiatives, the continued use of incentives as well as a larger window for study participation would improve the likelihood of successful engagement.

Taken together, the studies contained within this manuscript included several factors, both inside and outside the author's control, that contributed to challenges in their implementation. However, with an eye towards building from these difficulties, the outlined adjustments and shifts in approach provide a means by which valuable research – on dashboard use and otherwise – may be implemented in a more effective and successful manner.