Challenges and Ideas from a Research Program on High Quality, Evidence-Based Practice in School Mental Health

Mark D. Weist[1], **Eric A. Youngstrom**[2], **Sharon Stephan**[3], **Nancy Lever**[3], **Johnathan Fowler**[1], **Leslie Taylor**[1], **Heather McDaniel**[1], **Lori Chappelle**[4], **Samantha Paggeot**[4], and **Kimberly Hoagwood**[5]

Eric A. Youngstrom: eay@unc.edu; Sharon Stephan: sstephan@psych.umaryland.edu; Nancy Lever: Nlever@psych.umaryland.edu; Johnathan Fowler: jfowler79@gmail.com; Leslie Taylor: taylorlk@mailbox.sc.edu; Heather McDaniel: laskyh@mailbox.sc.edu; Lori Chappelle: LCC10@SCDMH.ORG; Samantha Paggeot: SEP95@SCDMH.ORG; Kimberly Hoagwood: kimberly.hoagwood@nyumc.org

[1]University of South Carolina, Department of Psychology, USC, 1512 Pendleton Street, Columbia, SC 29208

[2]University of North Carolina at Chapel Hill, Chapel Hill, NC

[3]University of Maryland School of Medicine, Baltimore, MD

[4]Waccamaw Center for Mental Health, Conway, SC

[5]New York University Medical Center, New York, NY

Abstract

Objective—Reviews the progression of a research program designed to develop, implement and study the implementation of "achievable" evidence-based practices (EBPs) in schools. Reviews challenges encountered and ideas to overcome them to enhance this avenue of research.

Method—Presents two federally funded randomized controlled trials involving comparison of a four-component targeted intervention (Quality Assessment and Improvement, Family Engagement and Empowerment, Modular Evidence-Based Practice, Implementation Support) versus a comparison intervention focused on Personal Wellness. In both studies primary aims focused on changes in clinician attitudes and behavior, including the delivery of high quality, evidence-based practices and secondary aims focused on student level impacts.

Results—A number of challenges, many not reported in the literature are reviewed, and ideas for overcoming them are presented.

Conclusions—Given the reality that the majority of youth mental health services are delivered in schools and the potential of school mental health (SMH) services to provide a continuum of mental health care from promotion to intervention, it is critical that the field consider and address the logistical and methodological challenges associated with implementing and studying EBP implementation by clinicians.

Correspondence concerning this article should be addressed to Mark Weist, Department of Psychology, USC, 1512 Pendleton Street, Columbia, SC 29208. weist@mailbox.usc.edu.
Author Note: Mark D. Weist, Department of Psychology, University of South Carolina (USC); Eric A.Youngstrom, Department of Psychology, University of North Carolina at Chapel Hill; Sharon H. Stephan and Nancy Lever, Department of Psychiatry, University of Maryland School of Medicine; Johnathan Fowler, Leslie Taylor and Heather McDaniel, Department of Psychology, USC; Lori Chappelle and Samantha Paggeot, Waccamaw Center for Mental Health; Kimberly Hoagwood, Department of Child and Adolescent Psychiatry, New York University Medical Center.

Keywords

school mental health; quality; family engagement and empowerment; evidence-based practices; challenges

This paper describes the evolution of an interconnected program of practice and research beginning in Baltimore in the 1990s that focuses on implementing high quality, evidence-based mental health practices in the challenging, real world environment of schools. We describe (1) the program's early development, (2) two large research grants (R01s from the National Institute of Mental Health, NIMH), (3) challenges encountered and ideas emerging to address them, and (4) key lessons learned in advancing this program of research. Given documented problems in implementing manualized interventions in schools (Evans & Weist, 2004; Schaeffer et al., 2005), the modular approach to evidence-based practice is central in all of our work (EBP; Chorpita, 2006; Chorpita & Daleiden, 2009; Chorpita, Daleiden, & Weisz, 2005), and is a major emphasis of this article.

Early Experiences: Defining Quality in School Mental Health

Expanded School Mental Health

When we began involvement in school mental health (SMH), in the early 1990s there was increasing awareness that most youth needing mental health services were not receiving them, due to a variety of problems with traditional community mental health. The System of Care movement (see Stroul & Friedman, 1986) helped document this gap and found that while schools were the setting where many youth were identified for services, they were woefully ill-equipped to respond effectively to these needs. The *expanded* school mental health construct addressed the gap by having schools joining forces with community mental health (CMH) centers and programs. Within this framework, CMH staff augment the work of school-employed staff to develop and implement a full continuum of effective programs and services for youth in general and special education in schools (Weist, 1997; Weist & Murray, 2007). The approach emphasizes developing a *shared agenda,* with school, community systems, families, and other stakeholders working together to assure the best quality programs to promote student wellness and reduce and remove barriers to their learning (Andis et al., 2002; see Adelman & Taylor, 1999). Please note that this emphasis on a shared agenda overlaps considerably with principles of community-based participatory research (Viswanathan et al., 2004).

The expanded SMH approach became grew progressively in Baltimore, expanding from a few sites in the mid-1990s (Flaherty & Weist, 1999) to more than 100 in 2007 (Weist, Paternite, Wheatley-Rowe, & Gall, 2009). Much of the work in Baltimore involved learning what constitutes high quality SMH through a number of qualitative studies funded by the Agency for Healthcare Research and Quality (see Nabors, Reynolds, & Weist, 2000) including developing principles and indicators for high quality services (Weist et al., 2005). These developments co-occurred with our research team's use of modular evidence-based practices (EBPs) based on the work of the Hawaii Department of Health (see Chorpita & Daleiden, 2009), and strategies for effective family engagement and empowerment (Hoagwood, 2005; McKay & Bannon, 2004). From 1995 on this research was supported through connections to the Center for School Mental Health (CSMH), originally one of two federally funded national centers for SMH awarded to the lead author and colleagues and currently the only federally funded center for SMH, which has always had a major emphasis on bridging research and practice in this emerging field (see Weist & Christodulu, 2000). Please note that the CSMH, funded by the Health Resources and Services Administration operates within the context of a National Community of Practice (COP) on Collaborative

School Behavioral Health funded by the Office of Special Education Programs. Based on the recognition that any systematic agenda builds from the platform of relationships (see Wenger, McDermott, & Snyder, 2002), the COP provides support for convening, communication, collaboration and mutual support among thousands of stakeholders in the U.S. with a vested interest in school mental health (Cashman, Linehan, & Rosser, 2007).

Within this context, two R01s were obtained from the National Institute of Mental Health (NIMH) to explore strategies for achieving high quality, evidence-based practice (EBP) by SMH clinicians in the demanding and fluid environment of schools (see Weist, 1999). The first R01 (2003-2007) explored a three-component framework of systematic quality assessment and improvement (QAI), family engagement and empowerment (FEE), and modular intervention in a randomized controlled trial (RCT) for clinicians operating in three prominent SMH programs. Clinicians were randomly assigned to the target condition involving these components, or a comparison condition focused on Personal Wellness (PW). Three sites were chosen to increase the sample size to over 50 clinicians working in more than 40 schools to increase statistical power, but multiple sites created methodological challenges (reviewed in this paper). Findings and lessons learned from this first R01 (see Weist et al., 2009), set the stage for the second and current R01 (2010-2014), also an RCT involving assignment of clinicians to the target versus PW comparison intervention. The current study operates in one large school district (46 clinicians in 34 schools) to address method problems of multiple sites, adds the key element of implementation support (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), focuses on disruptive behavior disorders (more manageable for clinicians), and integrates method enhancements from the first study. Both studies have been guided by Research Advisory Panels (RAPs), including prominent researchers in child and adolescent and school mental health, and representing family advocacy and state and national policy makers. In spite of having many advantages (e.g., long, progressively developing program of research, support from the CSMH, active guidance by RAPs), a number of key challenges have been encountered in these two studies, which we review in this paper, along with recommendations to begin to overcome them.

First R01: Enhancing Quality in Expanded School Mental Health

As presented, the first R01, *Enhancing Quality in School Mental Health* (NIH#1R01MH71015NIH#1R01MH72003 – 2007, M. Weist, PI) tested a three-component framework (QAI, FEE, modular intervention) on proximal outcomes for clinicians (see below), and sought to assess distal impacts on students treated by them. To maximize sample size the study involved three prominent SMH programs operating in Baltimore, Dallas, and throughout the state of Delaware. The study evaluated the QAI intervention on 1) service quality, reflected in implementation of selected indicators, the use of evidence-based practices, and the impact on family engagement and empowerment, and 2) knowledge and attitudes about evidence-based practice, and perceptions of organizational climate and counseling self-efficacy.

We hypothesized that in order to successfully implement evidence-based mental health practices, SMH clinicians would need to simultaneously meet the contextual demands of effectively working in schools. Thus, the QAI component was built on the foundation of an expanded version of a school mental health report card — the School Mental Health Quality Assessment Questionnaire (SMHQAQ; Weist et al., 2005; Weist, Ambrose, & Lewis, 2006), which included 10 general principles and 40 specific indicators of best practice in SMH. The SMHQAQ covers areas considered crucial for successful SMH practice, including developing and maintaining relationships with school administrators, teachers, health staff and others; understanding relevant education policies; and effectively integrating mental

health promotion and intervention into the school day (see Paternite, Weist, Axelrod, Anderson-Butcher, & Weston, 2006; Stephan, Davis, Burke, & Weist, 2006).

Clinicians in the QAI arm of the study learned EBPs via modularized training about "common elements" of evidence-based practice for four disorder areas: Anxiety, Depression, ADHD, and Disruptive Behavior Disorders. Clinicians received competency training in core techniques and procedures (e.g., exposure for anxiety) that have demonstrated effectiveness in studies of manualized treatment protocols. The method was adopted from the pioneering work of Chorpita, Daleiden and colleagues conducted with the Hawaii Department of Health, laying out the common elements for specific disorder areas (see Chorpita & Daleiden, 2009). At the time, this was an innovative way of packaging interventions, allowing the clinicians flexibility in selecting and arranging modules as appropriate for each case, while still being guided by the "evidence base." The method was consistent with our experience of SMH clinicians' rejection of traditional manualized interventions, often based on concerns about their perceived "one size fits all" approach, and associated concerns about the rigid need for adherence in spite of changing presentations in students and their circumstances (Curry & Reinecke, 2003; Schaeffer et al., 2005).

Finally, clinicians in the targeted condition received training and support in FEE. Despite our knowledge that effectively working with families underpins virtually all effective child and adolescent mental health interventions, school-based clinicians in particular struggled with effective FEE (Hoagwood, 2005; Jensen & Hoagwood, 2008; Lowie, Lever, Ambrose, Tager, & Hill, 2003). Schools have unique challenges for engaging parents including school operating hours, no transportation of parents, no childcare options for younger siblings, and parent comfort level with addressing mental health issues in a school setting, School clinicians may also vary in comfort level in working with families and in their skillsets in engaging and partnering with families. Training and support for FEE was developed based on the work of colleagues McKay and Hoagwood, including first interview engagement strategies (McKay et al., 2004), and Hoagwood's (2005) identification of four key domains of family engagement in children's mental health: (1) *Engagement*: Forming a connection with families in initial sessions and ensuring an open dialogue about any concerns (e.g., poor prior experiences with the mental health system), goals and expectations, and strategies to maximize the keeping and helpfulness of sessions. (2) *Collaboration*: Maintaining a collaborative approach in therapeutic interactions with families, versus operating from a position of expert guidance. (3) *Support*: Assisting families rapidly with making connections to address pressing needs (e.g., tutoring, employment), and supporting efforts to improve family functioning and child and adolescent behavior. (4) *Empowerment*: Pointing out family strengths and successes; promoting the family "taking charge to achieve treatment goals; and instilling hope.

Study findings revealed strong support for our primary hypotheses that the three-component framework would yield improvement in indicators of SMH quality and evidence-based practice implementation by clinicians, but student level differences in outcomes between the two study groups were not discerned, related in part to a measurement problem of very poor return rate of student and parent psychosocial measures and our inability to access school records (Weist et al., 2009).

As a formative evaluation, this study generated ample information about both implementation and methodological barriers. Clinicians in the QAI condition received modular training on four disorder areas over the course of a year, with as many as eight skills being taught per area, in addition to the SMH QAI and family engagement training components. Clinician feedback clearly documented concerns that too much material was covered in too little time, and that skill application was challenging in the context of a

difficult, fluid and stressful school environment. In addition, although the first study incorporated some behavioral rehearsal and bi-weekly training, it lacked some more recently promoted factors of implementation support (IS; see Graczyk, Domitrovich, & Zins, 2003; Fixsen et al., 2005). Also, *on-site* IS was not provided to clinicians in their schools, making it even more difficult for them to generalize related skills into their specific practices in schools. Beyond these concerns, there was a major methodological challenge of variation in study procedures across the three sites, in spite of twice monthly conference calls with research staff and senior trainers. For example, clinicians in the Delaware site did not prioritize family involvement in services and were not responsive to changing in spite of many requests by and support from the study team to do the same. Clinicians in Dallas continued to emphasize therapeutic approaches such as family systems therapy (Jones, 1994) that were not supported by the study, at times, directly interfering with implementing the targeted intervention. A number of these challenges were addressed in the next R01 presented in the following.

Second R01: Strengthening the Quality of School Mental Health Services

In 2010, the team received additional NIMH funding for the study, *Strengthening the Quality of School Mental Health Services* (R01MH0819141; 2010-2014, PI M. Weist) to build upon the first R01 by: (1) further strengthening the targeted intervention via a more narrow training focus on evidence-based practice for one disorder area, youth presenting acting out behaviors consistent with disruptive behavior disorders, (2) utilizing theoretically grounded strategies for implementation support (IS; Fixsen et al., 2005; Fixsen, Blase, Naoom, & Wallace, 2009), (3) assessing ongoing clinician use of practice elements and FEE strategies, and (4) operating in one system--a very large community mental health program operating in 34 schools in Horry County, SC (and subsequently also in Georgetown and Williamsburg counties), to avoid the multi-site error variance experienced in the first R01. Assessment for the current R01 includes the *Family Engagement/Empowerment Observation System* (FEEOS; Weist, 2009), an observational measure for assessing key FEE constructs represented in the literature (Hoagwood, 2005) and the *Practice Elements Behavioral Observation System (PEBOS*; Stephan & Becker, 2010) for tracking clinician usage of ten practice elements targeted for disruptive behavior problems in this project (i.e., praise, commands, rewards, monitoring, time out, response cost, psychoeducation, active ignoring, problem solving, and communication skills). Families targeted for recruitment in the study include those with a child aged 6-17 years, diagnosed with a disruptive behavior disorder or displaying disruptive behaviors of treatment concern, and enrolled for services with a participating school-based clinician. Again, the study is an RCT, with clinicians assigned to the targeted condition (referred to as Clinical Services Support, or CSS) or to a condition focused again on PW. CSS clinicians receive twice monthly two-hour training on QAI, FEE and Modular EBP by skilled, experienced and respected senior trainers, and these same trainers visit staff in their schools at least 6 times per year to provide on-site IS. IS involves observing clinicians during treatment sessions with caregivers, and then providing collegial support and coaching (including behavioral rehearsal) on effective FEE and Modular EBP. IS also includes offering emotional support to clinicians (e.g., empathy, stress management ideas) and administrative support (e.g., replenishing files with handouts; see Fixsen et al., 2009). Discussions among senior trainers and clinicians are bi-directional, with the clinicians providing input on ways to strengthen the components of effective practice (QAI, FEE, Modular EBP), as well as the IS they are receiving.

Implementation support: challenges and solutions—The study is currently entering its fourth and final year. Based on collaborative feedback (see Elias, Zins, Graczyk, & Weissburg, 2003), improvements made to IS within the study include (a) tailoring the amount of IS based on clinician need, with some clinicians clearly needing more IS than

others; (b) pairing clinician coaches with those who are less strong in implementation; (c) conducting more integrative case discussions and demonstrations that involve the clinician demonstrating multiple EBPs; and (d) using visual aids and reminders for key FEE and Modular EBP skills in sessions.

The behavioral rehearsal component of training and IS has been notably more difficult for clinicians than was the case in prior training experiences, requiring much encouragement, reinforcement and innovative approaches to effectively use this key training strategy. Effective role-playing of skill usage in trainings helps both the "actors" and "observers" (fellow clinicians). Although CSS clinicians knew one another before the second R01, bravely practicing skills using case scenarios in front of one's peers was observed to be and reported by them to be challenging and somewhat stressful. Over time, this "performance reluctance" has lessened somewhat, yet continues to warrant monitoring. One successful strategy has been to ask clinicians who have demonstrated a depth of understanding and high fidelity on the implementation of a particular skill to "coach" other clinicians by playing the role of the therapist, or conversely, playing the caregiver role and posing questions that help to guide their colleagues in using skills effectively and thoroughly.

Clinicians also have encountered challenges with engaging families in their children's mental health treatment. Parents have often not completed homework assignments, which are key to the at-home implementation of selected parenting skills. Senior trainers have needed to provide continual and enthusiastic encouragement to some of the clinicians to keep them focused on parental compliance with these requests, as in some cases, a minimal effort resulted in the clinician giving up on engaging the family, and also requiring the senior trainer to challenge trait labels about the caregivers (e.g., "they are an uncooperative family"). Finally, senior trainers strategically utilize training time within the twice-monthly group meetings to review recently observed sessions, with the clinician's approval, to highlight successes with FEE and EBPs, or areas for growth and improvement, to provide an opportunity for fruitful discussion and adjustment of practices.

Another unique challenge has arisen related to senior trainers for both the CSS and PW groups also serving as the clinicians' direct supervisors. Clearly, single-case observations are insufficient for complete supervision experiences. Field professionals must be able to seek guidance and discuss concerns with supervisors outside of the research initiative's framework. At times, more traditional supervision (e.g., review of fee-for-service paperwork) might naturally "bleed into" IS. To maximize the efficiency of IS within the study, some adjustments needed to be made. During year 1, for example, clinicians requested and received additional, "non-specific" supervision time. This met the needs of clinicians doing highly challenging work in schools, and it also allowed observation time to be used strictly for IS related to EBP and FEE skill usage.

Clinicians received further support in the form of handouts, posters, and other physical resources to remind them of the basic EBPs and FEE skills used with the appropriate clients. CSS clinicians got a laminated, colorful, yet simple list of the 10 EBPs for Disruptive Behavior Disorders, for example, for posting in their offices. Visual aids helped when discussions got off-track or when the clinician was seeking additional direction during the day. Clinician feedback has been uniformly positive about these approaches.

Systematic fidelity monitoring: challenges and solutions—Clinician EBP usage is assessed through the PEBOS (Stephan & Becker, 2010). Following session observations in schools, senior trainers rate the degree to which clinicians demonstrate target skills appropriately. Using the PEBOS, trainers also rate general competence in cognitive-behavioral therapy components, including establishing and reviewing goals and progress,

explanation and use of rationales, quality of teaching, modeling/demonstration, rehearsal, reinforcement/feedback, assignment and review of homework, prompting skill use outside of session, and maintaining positive therapeutic rapport (see Garland, Hawley, Brookman-Frazee, & Hurlburt, 2008). Simultaneously, the FEEOS (Weist, 2009) quantifies clinician usage of FEE strategies linked to key constructs (Hoagwood, 2005). These include general strategies such as empathy, sincerity, warmth, and humor, as well as agreement, trust, engagement, collaboration, support, and empowerment.

Senior trainers use the PEBOS and FEEOS as a real-time fidelity assessment for CSS clinicians. They also compare clinician utilization of EBP and FEE strategies within video and audio recordings across the CSS and PW groups. Annual clinical role plays provide additional pre-post assessment of clinicians in both groups, charting Modular EBP and FEE skill development over time. In addition, at 3 indicated time points during each study year, each clinician selects 1-2 participating families to audiotape in a treatment session, which are later rated using the same measures by team members blind to clinician group assignment. Parental adherence to EBPs (e.g., parent completion or non-completion of related homework) is assessed indirectly, such as during session observations and review of session recordings or role plays.

Interpreting video and audio tapes for EBPs and FEE skills has required new coding schemes. There is substantial overlap across FEE components and other key ingredients of therapy, such as therapeutic alliance (see Dearing, Barrick, Dermen, Walitzer, 2005; Yatchmenoff, 2005), and thus it is difficult to clearly separate and define each component. This makes providing clear descriptions of components to coders a challenging task, particularly for audio taped sessions, which lack nonverbal cues. Decisions about selecting observational coders (e.g., whether coders should have therapy training and practice experience or represent a "blank slate" without background knowledge of coding constructs) further compound difficulties in how to optimally codify data.

Other methodological issues involve targeting practice elements, along with definitions of skill usage based on targets. Videotapes show that many clinicians demonstrate one of the EBPs within clinical scenarios (role plays), but not necessarily the one cued for in the behavioral rehearsal scenario. This has led to more explicit instructions in training of the exact skill clinicians are expected to demonstrate, versus assessment of their analysis and choice of appropriate skills. Similarly, when reviewing session audiotapes, it can be hard to tell if a clinician is reviewing a skill previously taught versus teaching it for the first time. What constitutes the presence or absence of skill? How do we interpret implementation of non-targeted skills? How do we interpret low versus high frequency usage of specific practice elements across clinicians? Identifying and addressing these and other questions, and continuing to discuss these issues with national experts (as in writers for this special issue), will be crucial.

Research Design and Analyses: Challenges and Future Directions

Research Design Choices and Challenges

As major analyses for the first R01 have been completed (Weist et al., 2009), and we are more than halfway into the second and beginning interim analyses, our thinking has evolved about the most appropriate analytic strategies and the significance of these two studies for the further development of the field. Issues in a number of realms have been encountered, including Population of Interest, Level of Analysis, Randomization, and Design Complications, and these are reviewed in the following.

Population of interest—What is the population of interest for extrapolating the results (Straus, Richardson, Glasziou, & Haynes, 2005)? Framed broadly, we might generalize to all students receiving mental health services in the schools enrolled, discussing how demographic and clinical factors might vary between these schools and others nationally. Most narrowly, we might limit analyses only to those students with sufficient data, who also received an adequate dose of treatment. The choice of frame reflects the inherent tension between external and internal validity: More generalizable results are likely to create more missing data, variable doses of treatment, and other sources of heterogeneity that reduce statistical power (Kazdin, 1994). Our choice of frame guided efforts to gather data about the "intent to enroll" as well as the "intent to treat" sample, making it possible to use propensity score methods to investigate potential differences, as well as to weight results to increase generalizability.

Level of analysis—Another main decision was the level of analysis -- is the primary "participant" the clinician, or the students? From a policy and public health standpoint, student-level outcomes are imperative. However, the intervention of interest manipulates the training and support for clinicians, and our hypotheses emphasize effects on clinicians' attitudes, knowledge, and behavior. Multilevel regression models often investigate educational and psychological treatment effects on youth outcomes, typically looking at repeated measures of youth variables (level 1) as nested within youths (level 2), nested in turn within classrooms (level 3) (Raudenbush & Bryk, 2002). The situation here is different, for both conceptual and pragmatic reasons. Conceptually, the traditional multilevel model does not neatly incorporate the clinician as a nesting effect; youths are nested within clinician, but may be drawn from multiple different classrooms within the same school, and sometimes even different schools. Many of the putative advantages of treating repeated measures as nested effects only fully materialize when many of the cases have three or four observations (making it possible to estimate within-case variance around the slope) (Raudenbush & Bryk, 2002), whereas here we planned to have only pre- and post measures on the clinicians and some youth level variables.

Randomization—The decision to treat the clinician as the primary participant in the study clarified many other design decisions. Randomization occurred at the level of the clinician, not the school (and some clinicians served more than one school; administrative support was crucial to ensure that any school served by multiple clinicians did not have clinicians from both assignment arms providing services). Youths were treated as nested within clinician, and treatment was a between-clinician factor. We used minimization (Treasure & MacRae, 1998) rather than simple random assignment, gathering baseline information about clinicians, using a principal components analysis to distill the variables into a few summary scores, and then matching and randomizing within paired block to minimize the differences between groups at baseline. With so many clinical and school variables involved, and a moderate number of clinicians, simple randomization would have resulted in lack of balance on multiple variables.

Post launch design complications—After starting the project, several real life issues have arisen. Some clinicians left the school district. Others changed schools. Working closely with our community mental health center partner, efforts were made to assure that clinicians did not change conditions (e.g., CSS to PW) when changing schools, but this could not be prevented in all cases. Given significant challenges to enroll participants in the study (to meet recruitment milestones and maintain funding), do cases get removed from the study if they change conditions? We have decided the answer is "no" but we are carefully tracking all events for all students/families, and placing them in analytic "baskets" reflecting pure cases (no change in school, clinician or condition; no receipt of outside clinical

services; about 87% of the sample) and cases that involve less than pure circumstances. In addition, new clinicians were added in the second year of each of the two studies to maintain or increase statistical power for the clinician-level analyses. Here we are learning of an experience effect, as the quality of services delivered by clinicians in the CSS arm for a year or more would expected to be much better than that of a brand new clinician. Similarly, we are wrestling with issues of dose. Some students may receive a minimal dose of 4 treatment sessions within 90 days, while others will receive much more. Analyses should be sensitive to different doses, as well as differences in implementation among clinicians in the CSS condition. As we enter the final year of the study emphasizing analyses and dissemination, navigating our way through these analytic challenges is a major focus of the work of our team. Additional measurement/analysis challenges are presented in the next section.

Measurement Choices and Complications

Multiple informants—Consistent with best practices (Campbell & Fiske, 1959), we wanted to use multiple informants and multiple methods to gather information about each variable of interest. Theoretically, multiple methods could be used for each informant (e.g., clinician, youth, parent), each construct (e.g., internalizing problems, externalizing problems, knowledge or use of evidence-based practices), or each level of the analytical model (e.g., time, youth, clinician, school, treatment arm). Crossing these facets with each other generates a ponderously large matrix of possibilities. We concentrated on youth self-report about their own emotional and behavioral problems, augmented with parent report (and treating parent report as primary, when the youths were too young to self-report).

Multiple methods—Using parent and youth report about the youth's emotional and behavioral problems provided multiple sources of information, at least when the student was old enough to complete self-report, and when the parent completed and returned the instruments. In the second study, we have obtained permission to access existing records about attendance, grades, behavioral problems and other variables routinely tracked by the schools about each student. Interpreting and consistently coding student records is no small challenge (Powers, Marks, Miller, Newcorn, & Halperin, 2008). Here we are hopeful about the possibilities for meaningful between group analyses but also time series analyses for students in the target CSS condition, as school record variables are recorded quarterly and we have up to a three-year window of school records for each student in the study.

Given the primacy of the clinician level variables in our hypotheses and analyses, we are using additional sources of information about clinician level factors. To supplement clinician self-report, we are also conducting chart reviews of clinical records, along with observational ratings of videotapes of the clinician interacting within a role-played case scenario. Whenever possible, we have used previously developed and validated rating scales and coding schemes. Consistent with best practices (Vacha-Haase, 1998), we always check the reliability within our own samples (more about this below). We picked the shortest instruments that would provide reliable and valid information about the construct of interest, in light of the well-known concerns about burden and feasibility for every informant involved in community-based research.

Complications—Inevitably, measurement has proven less straightforward than hoped. Three sets of challenges occurred in both projects. First, the constructs and goals often required development of new measures, or substantial adaptation of existing measures. This led to extensive psychometric analyses, and sometimes either changes to the coding scheme or modifications to instruments, prior to conducting the final analyses.

Second, reliability analyses of the instruments sometimes revealed substantial issues not discussed in the primary publication for the existing measure. For example, in the first study we used a modified version of the Client Satisfaction Questionnaire (CSQ; Attkisson & Greenfield, 1999) where a small number of additional "dissatisfaction" items were reverse-keyed. In the study sample, across all informants, the reverse-keyed items showed near-zero corrected item total correlations, and including them reduced the alpha coefficient for the scale whether reversed or not—suggesting that some participants were not reading the item or instructions carefully (Streiner & Norman, 1995). An informal follow-up with some participants confirmed that suspicion. We are confident about the problem source, but there is no perfect solution. Using the full scale would include responses with errors, lowering reliability and attenuating observed correlations. Eliminating the reverse-keyed items would improve the reliability and eliminate many errors, but it also would include scores from people who may not have been reading carefully and who were stuck in a response set. Excluding cases that appeared to have an "invalid profile" would give us greater confidence in the remaining responses; but it would reduce the sample size, lowering statistical power; and the excluded cases probably would not be missing at random (Youngstrom, et al., 2011). Given the extent of the problem, and the fact that we have found it across informants, we suspect that other studies may have encountered similar problems. Our short term solution is to run "sensitivity analyses" where we apply all three approaches--using the total score as planned, using scores based on a reduced set of items, and isolating cases with inconsistent responses across items—and then report the extent to which findings change as a result. Future studies might consider whether the benefits of defining a response inconsistency threshold or adding other validity scales might justify the increased burden.

A third, related issue has been when the factor structure of an instrument in our data clearly does not align with the published structure. This could be due to differences in the analytic methods (Fabrigar, Wegener, MacCallum, & Strahan, 1999), or differences in the sample. Either way, it creates a practical problem of selecting which structure to use for the main analyses. Using the published structure may reduce the reliability and validity in our samples; but switching structures would complicate comparison to prior research. We will concentrate on total scores for primary analyses whenever feasible, and decide whether the differences in structure are meaningful enough to justify a secondary analysis and publication of a measurement-oriented paper.

Statistical Analyses

Power—Estimation of power has been complex. The focus on clinician as the primary unit of analysis creates a bottleneck: even working with the largest community mental health facilities or school districts yields relatively small numbers of clinicians. We have borrowed many strategies recommended for "small n" designs, such as using unbalanced pre- and post-tests, or examining whether there would be improved power by aggregating measures (Venter & Maxwell, 1999). The choice of analytical model to accommodate the nesting also complicates power estimation; "off the shelf" software does not handle mixed regressions or generalized estimating equations (cf. Faul, Erdfelder, Buchner, & Lang, 2009), forcing us to set up Monte Carlo simulations that approximate the design, analysis, and anticipated circumstances (Muthén & Curran, 1997; Muthen & Muthen, 2004). Power simulation requires specifying a long list of parameters: in addition to alpha, N, and effect size, we need to estimate number of clinicians, number of youths (and completeness of parent and youth data), correlations between covariates and converging indicators of the construct, attrition, and so forth. Kraemeris skeptical about the emphasis on a priori power estimation: to know all of the parameters with confidence, one would need to have already done the study (Kraemer, Mintz, Noda, Tinklenberg, & Yesavage, 2006).

Reliability—Feasibility and low burden are at odds with psychometric reliability, especially Cronbach's alpha. Longer scales necessarily have higher internal consistency estimates, all else being equal (Cicchetti et al., 2006). Choosing brevity increases the chances that participants will complete measures, but alpha penalizes brevity. Unfortunately, only a few measures have retest reliability reported, and we did not have the resources to systematically gather retest reliability on all measures. We are using a combination of alternate metrics appropriate for short scales (e.g., median corrected item-total correlation; Streiner & Norman, 1995) along with benchmarking against published reliabilities (Feldt, 1969).

Type of analysis—The zeitgeist in psychological and educational research has been to use mixed regression models (also referred to as hierarchical linear models, multi-level models, and random effects regressions) when confronted with nested data (Raudenbush & Bryk, 2002). However, the degree of nesting may not be ideally suited to random effects models. Our number of repeated measures does not capitalize on the potential strengths of the mixed regression approach; we have one or two time points for each youth or clinician, often not enough to estimate the slope, and never enough to estimate variability around the within-person slope. There is a wide range in the number of cases each clinician sees, too. We were not able to find any guidance in the literature about the probable "design effects" of nesting on our youth behavior outcomes or clinician variables.

An alternate analytic approach is to use generalized estimating equations (GEE) to adjust the clinician-level and youth-level standard errors to reflect the nesting, without explicitly attempting to model sources of variance within each clinician (Hanley, 2003; Hardin & Hilbe, 2003). Understanding within-clinician variation is definitely an important goal, but a different study from the one we are doing. If the primary goal were understanding factors within caseloads, and how they changed outcome and moderated the effects of clinician-level interventions, we would want to have 30 or more youth within each clinician at a minimum, and ideally at least 10 youth at each marginal distribution for factors that we thought might affect outcome (i.e., at least 10 males and 10 females per clinician, etc.; Kraemer, 1992). GEE appears to be a much more pragmatic analytic tool for appropriately modeling the nested data that we have at hand, while focusing on the question of how clinician-level interventions change youth outcomes (Hanley, 2003).

For many reasons, we also are putting increasing emphasis on exploratory data analysis techniques (EDA; Behrens, 1997; Tukey, 1977). The constraints of sample size and the size of systems involved, combined with the limits of funding mechanisms available, mean that there is a small zone where the power curve and the feasibility curve overlap. Exploratory approaches to examining the data may help glean policy ideas and second generation hypotheses from expensive data, and they are more likely to be robust. They also have great communication value with less research-oriented stakeholders, and definitely play a complementary role to the more sophisticated inferential methods we are deploying.

Missing data—We knew from the outset that there would be missing data. Depending on the choice of frame for the population, then the successive filters include: consenting to participate, completing at least one evaluation, getting information from collateral sources, completing an adequate course of treatment, and completing all post-test measures without any loss of data to other causes (Moher, Schulz, & Altman, 2001). Pairwise deletion of missing data is widely recognized as flawed (Allison, 2002). Listwise deletion would result in prohibitive shrinkage of functional sample size. However, multiple imputation would face several additional hurdles with these data, including the fact that the sample size is small for imputation purposes at the clinician level, sparse in terms of predictors at the youth or parent level, and imputation algorithms generally are not designed to model nested data (Engels &

Diehr, 2003). Imputing data separately within clinician would be untenable (Graham & Schafer, 1999). Many variables have enough missing data and few enough predictors that the increase in standard errors due to imputation may outweigh the gains from increased sample size. Reliance on school records or clinical charts also introduces missing data problems, as such records are inherently idiosyncratic, and may present unpredictable differences in compliance with specific forms used for assessment, clinical notes, and outcomes.

Propensity scores might offer a more viable approach to modeling a different aspect of the missing data, comparing completers to the original intent to treat or intent to enroll samples (Guo & Fraser, 2010). Unfortunately, gathering the descriptive information about the larger sample, particularly if trying to extrapolate to the intent to enroll group, involves ethical issues with regard to consent (and would require some degree of identification to be able to match information to participants). Collaboratively processing these issues also takes time and involves political and other factors in addition to design and ethical principles. It may not be clear until the end of the project exactly what variables are available for use in propensity scoring.

Future Directions: Addressing Some Fundamental Tensions

Many lessons have been learned going through the process of doing community-based research with two different sets of schools and agencies. As indicated, we have had the benefit of considerable previous experience and wisdom on the investigative team and research advisory boards, and have borrowed extensively from the toolkits developed for clinical trials and small sample studies to maximize the understanding and utility of clinician-level variables. What is striking is that many of the "discoveries" are probably not new, but rather, unreported in the literature. Problems with response set and reverse-keyed items, for example, are unlikely to be unique issues to the two studies we conducted (which involve hundreds of participants in four different states); however, they were not discussed previously in the literature. The gap between efficacy and effectiveness research has been known for decades; however, there is surprisingly little guidance about how to work around these pitfalls when trying to take the research off the well-paved trails of efficacy and extend it into the frontiers of the community setting. Focusing too much on the challenges risks inviting reviewers and readers to reject the findings out of hand. Conversely, ignoring the problems or gliding through them hastily is intellectually disingenuous and does the field a disservice, especially if the results are biased, or if future investigators will hit similar snags without more warning. The solutions that we are using (trying to accurately describe and maximize reliability with brief scales, using GEE to model nesting, using propensity approaches to model missing data and generalizability, and conducting sensitivity analyses to chart the effects of different approaches or assumptions) all have published precedents, but are not the most common or familiar strategies in psychology research at present. We can offer a rationale for each choice, but not data demonstrating superiority. We wish to disclose limitations and inform the research community of potential traps without undermining the value of our work or inviting excess skepticism or nihilism. We are learning some of the quagmires that have bogged down much of the "missing efficacy," and are guardedly optimistic that some research tools are available to help work around them. Some of our contributions will be additional entries on the map, which should lead to easier progress in future expeditions, and ultimately, better outcomes.

Conclusion

The degree to which school mental health services are based on evidence is largely unknown (Langley, Nadeem, Kataoka, Stein, & Jaycox, 2010; Rones & Hoagwood, 2000), and it remains uncertain whether EBPs can be implemented with the same degree of fidelity as in

more controlled settings (Stephan, Westin, Lever, Medoff, Youngstrom, & Weist, 2012). However, given the reality that the majority of youth mental health services are delivered in schools and the potential of SMH to provide a continuum of mental health care from promotion to intervention, it is critical that the field consider and address the logistical and methodological challenges associated with implementing and studying EBP implementation in schools (Short, Weist, Manion, & Evans, 2012). Previous studies (Langley et al., 2010; Weist, Stiegler, Stephan, Cox, & Vaughan, 2010) demonstrate the unique considerations to mental health services implementation in schools including competing responsibilities (of clinicians), lack of support from school administration and teachers, lack of family engagement, student absenteeism and inflexibility of manuals to allow for shorter sessions and briefer interventions. Our line of inquiry into modularized intervention approaches is consistent with our overarching goal of identifying, evaluating and disseminating *achievable* EBPs in schools. We have argued that this approach, embedded within a larger framework of SMH quality and school-family-community partnerships, allows clinicians to utilize the research literature to inform their practice despite competing characteristics of the school settings including complex student case presentation and unique time demands (e.g., summer breaks, brief sessions).

There is a large interconnected agenda for the SMH field to advance. A clear priority is for strong training and ongoing coaching and implementation support to enable clinicians, in the demanding and fluid environment of schools to implement high quality and evidence-based practices. This is not a small task given the challenging nature of school buildings, which lack the formal administrative and institutional supports for clinical service provision that are part of the fabric of clinics and hospitals. This avenue of research reflects both the priority and the difficult reality of working in schools, "where youth are."

Acknowledgments

References

Adelman HS, Taylor L. Mental Health in school and system restructuring. Clinical Psychology Review. 1999; 19:137–163.10.1016/S0272-7358(98)00071-3 [PubMed: 10078417]

Allison, PD. Missing data. Thousand Oaks, CA: Sage; 2002.

Andis P, Cashman J, Praschil R, Oglesby D, Adelman H, Taylor L, Weist MD. A strategic and shared agenda to advance mental health in schools through family and system partnerships. International Journal of Mental Health Promotion. 2002; 4:28–35.10.1080/14623730.2002.9721886

Attkisson, CC.; Greenfield, TK. The UCSF Client Satisfaction Scales: I The Client Satisfaction Questionnaire-8. In: Maruish, ME., editor. The use of psychological testing for treatment planning and outcomes assessment. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1999. p. 1333-1346.

Behrens JT. Principles and procedures of exploratory data analysis. Psychological Methods. 1997; 2:131–160.10.1037//1082-989X.2.2.131

Campbell DT, Fiske DW. Convergent and discriminant validation by multitrait-multimethod matrix. Psychological Bulletin. 1959; 56:81–105.10.1037/h0046016 [PubMed: 13634291]

Cashman, J.; Linehan, P.; Rosser, M. Communities of practice: A new approach to solving complex educational problems. Alexandria, VA: National Association of State Directors of Special Education; 2007.

Chorpita, BF. Modular cognitive-behavioral therapy for childhood anxiety disorders. New York, NY: Guilford Press; 2006.

Chorpita, BF.; Daleiden, EL. The child and adolescent mental health division biennial report: Effective psychosocial intervention for youth with behavioral and emotional needs. 2009. Retrieved from Hawaii Department of Health, Child and Adolescent Mental Health Division website: http://hawaii.gov/health/mentalhealth/camhd/library/pdf/ebs/ebs013.pdf

Chorpita BF, Daleiden EL, Weisz JR. Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. Mental Health Services Research. 2005; 7:5–20.10.1007/s11020-005-1962-6 [PubMed: 15832690]

Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, Tyrer P. Rating Scales, Scales of Measurement, Issues of Reliability. The Journal of Nervous and Mental Disease. 2006; 194:557–564.10.1097/01.nmd.0000230392.83607.c5 [PubMed: 16909062]

Curry, JF.; Reinecke, MA. Modular cognitive behavior therapy for adolescents with major depression. In: Reinecke, M.; Datillo, F.; Freeman, A., editors. Cognitive therapy with children and adolescents. New York, NY: Guilford Press; 2003. p. 95-127.

Dearing RL, Barrick C, Dermen KH, Walitzer KS. Indicators of client engagement: Influences on alcohol treatment satisfaction and outcomes. Psychology of Addictive Behaviors. 2005; 19:71–78.10.1037/0893-164X.19.1.71 [PubMed: 15783280]

Elias MJ, Zins JE, Graczyk PA, Weissburg RP. Implementation, Sustainability, and Scaling Up of Social-Emotional and Academic Innovations in Public Schools. School Psychology Review. 2003; 32:303–319.

Engels JM, Diehr P. Imputation of missing longitudinal data: A comparison of methods. Journal of Clinical Epidemiology. 2003; 56:968–976.10.1016/S0895-4356(03)00170-7 [PubMed: 14568628]

Evans SW, Weist MD. Implementing empirically supported treatments in schools: What are we asking? Child and Family Psychology Review. 2004; 7:263–267.10.1007/s10567-004-6090-0

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. 1999; 4:272–299.10.1037//1082-989X.4.3.272

Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavioral Research Methods. 2009; 41:1149–1160.10.3758/BRM.41.4.1149

Feldt LS. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. Psychometrika. 1969; 34:363–373.10.1007/bf02289364

Fixsen DL, Blase KA, Naoom SF, Wallace F. Core implementation components. Research On Social Work Practice. 2009; 19:531–540.10.1177/1049731509335549

Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, National Implementation Research Network; 2005. FMHI Publication #231

Flaherty LT, Weist MD. School-based mental health services: The Baltimore models. Psychology in the Schools. 1999; 36:379–389.10.1002/(SICI)1520-6807(199909)36:5<379::AID-PITS2>3.3.CO; 2-4

Garland AF, Hawley KM, Brookman-Frazee L, Hurlburt MS. Identifying common elements of evidence-based psychosocial treatment for children's disruptive behavior problems. Journal of the American Academy of Child & Adolescent Psychiatry. 2008; 47:505–514.10.1097/CHI.0b013e31816765c2 [PubMed: 18356768]

Graczyk, PA.; Domitrovich, CE.; Zins, JE. Facilitating the implementation of evidence-based prevention and mental health promotion efforts in schools. In: Weist, MD.; Evans, SW.; Lever, NA., editors. Handbook of school mental health: Advancing practice and research. New York, NY: Springer; 2003. p. 301-318.

Graham, JW.; Schafer, JL. On the performance of multiple imputation for multivariate data with small sample size. In: Hoyle, RH., editor. Statistical Strategies for Small Sample Research. Thousand Oaks, CA: Sage; 1999. p. 1-29.

Guo, S.; Fraser, MW. Propensity score analysis: Statistical methods and applications. Los Angeles, CA: Sage; 2010.

Hanley JA. Statistical analysis of correlated data using generalized estimating equations: An orientation. American Journal of Epidemiology. 2003; 157:364–375.10.1093/aje/kwf215 [PubMed: 12578807]

Hardin, J.; Hilbe, J. Generalized estimating equations. Boca Raton, FL: Chapman Hall; 2003.

Hoagwood KE. Family-based services in children's mental health: A research review and synthesis. Journal of Child Psychology and Psychiatry. 2005; 46:690–713.10.1111/j. 1469-7610.2005.01451.x [PubMed: 15972066]

Jensen, PS.; Hoagwood, KE. Improving children's mental helath through parent empowerment: A guide to assisting families. New York, NY: Oxford University Press; 2008.

Jones, E. Family systems therapy. New York, NY: Wiley; 1994.

Kazdin, AE. Methodology, design, and evaluation in psychotherapy research. In: Garfield, SL.; Bergin, AE., editors. Handbook of psychotherapy and behavior change. 4th. New York, NY: Wiley; 1994. p. 19-71.

Kraemer, HC. Evaluating medical tests: Objective and quantitative guidelines. Newbury Park, CA: Sage; 1992.

Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. Archives of General Psychiatry. 2006; 63:484–489.10.1001/archpsyc.63.5.484 [PubMed: 16651505]

Langley AK, Nadeem E, Kataoka SH, Stein BD, Jaycox LH. Evidence-based mental health programs in schools: Barriers and facilitators of successful implementation. School Mental Health. 2010; 2:105–113.10.1007/s12310-010-9038-1 [PubMed: 20694034]

Lowie, JA.; Lever, NA.; Ambrose, MG.; Tager, SB.; Hill, S. Partnering with families in school mental health. In: Weist, MD.; Evans, S.; Lever, NA., editors. Handbook of school mental health: Advancing practice and research. New York: Kluwer Acadmic/Plenum Publishers; 2003. p. 135-148.

McKay MM, Bannon WR. Engaging families in child mental health services. Child And Adolescent Psychiatric Clinics Of North America. 2004; 13:905–921.10.1016/j.chc.2004.04.001 [PubMed: 15380788]

McKay MM, Hibbert R, Hoagwood K, Rodriguez J, Murray L, Legerski J, Fernandez D. Integrating evidence-based engagement interventions into "real world" child mental health settings. Brief Treatment and Crisis Intervention. 2004; 4:177–186.

Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parellel-group randomised trials. The Lancet. 2001; 357:1191–1194.10.1016/S0140-6736(00)04337-3

Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods. 1997; 2:371–402.10.1037//1082-989X.2.4.371

Muthen, L.; Muthen, B. M-Plus User's Guide. 3rd. Los Angeles, CA: Muthen & Muthen; 2004.

Nabors LA, Reynolds MW, Weist MD. Qualitative evaluation of a high school mental health program. Journal of Youth and Adolescence. 2000; 29:1–14.

Paternite, CE.; Weist, MD.; Axelrod, J.; Anderson-Butcher, D.; Weston, K. School mental health workforce issues. In: Hoge, M.; Morris, J.; Daniels, A.; Adams, N.; Huey, L.; Stuart, G., editors. A thousand voices: The national action plan on behavioral health workforce development. Rockville, MD: U.S Substance Abuse and Mental Health Services Administration; 2006. p. 199-210.

Powers RL, Marks DJ, Miller CJ, Newcorn JH, Halperin JM. Stimulant treatment in children with attention-deficit/hyperactivity disorder moderates adolescent academic outcome. Journal Of Child And Adolescent Psychopharmacology. 2008; 18(5):449–459.10.1089/cap.2008.021 [PubMed: 18928410]

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd. Newbury Park, CA: Sage; 2002.

Rones M, Hoagwood K. School-based mental health services: A research review. Clinical Child And Family Psychology Review. 2000; 3(4):223–241.10.1023/A:1026425104386 [PubMed: 11225738]

Schaeffer CM, Bruns E, Weist MD, Stephan SH, Goldstein J, Simpson Y. Overcoming challenges to evidence-based interventions in schools. Journal of Youth and Adolescence. 2005; 34:15–22.10.1007/s10964-005-1332-0

Short K, Weist MD, Manion I, Evans SW. Tying together research and practice: Using ROPE for successful partnerships in school mental health. Administration and Policy in Mental Health and Mental Health Services Research. 2012; 39:238–247.10.1007/s10488-011-0342-3 [PubMed: 21461778]

Stephan, S.; Davis, E.; Burke, P.; Weist, M. Supervision in school mental health. In: Neil, T., editor. Helping others help children: Clinical supervision of child psychotherapy. Washington D.C.: American Psychological Association; 2006.

Stephan, S.; Becker, K. Development and initial findings of the Practice Elements Behavioral Observation System (PEBOS). In: Poduska (Chair), J., editor. Measuring fidentliy of program implementation in schools Symposium presented at the annual. National Institutes of Health Science of Dissemination and Implementation; Bethesda, MD: 2010.

Stephan S, Westin A, Lever N, Medoff D, Youngstrom E, Weist M. Do school-based clinicians' knowledge and use of common elements correlate with better treatment quality? School Mental Health. 2012; 4:170–180.10.1007/s12310-012-9079-8

Straus, SE.; Richardson, WS.; Glasziou, P.; Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 3rd. New York: Churchill Livingstone; 2005.

Streiner, DL.; Norman, GR. Health Measurement Scales: A Practical Guide to their Development and Use. 2nd ed. New York, NY: Oxford University Press; 1995.

Stroul, BA.; Friedman, RM. A System of Care for Severely Emotionally Disturbred Children and Youth. CASSP Technical Assistance Center, Georgetown University Child Development Center; Washington, DC: 1986.

Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. BMJ. 1998; 317:362–363.10.1136/bmj. 317.7155.362 [PubMed: 9694748]

Tukey, JW. Exploratory data analysis. Reading, MA: Addison-Wesley; 1977.

Vacha-Haase T. Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. Educational & Psychological Measurement. 1998; 58:6–20.10.1177/0013164498058001002

Venter, A.; Maxwell, SE. Maximizing power in randomized designs when N is small. In: Hoyle, RH., editor. Statistical Strategies for Small Sample Research. Thousand Oaks, CA: Sage; 1999. p. 31-58.

Viswanathan, M.; Ammerman, A.; Eng, E.; Garlehner, G.; Lohr, KN.; Griffith, D.; et al. Whitener, L. Community-Based Participatory Research: Assessing the Evidence. 2004. Retrieved from the Agency for Healthcare Research and Quality website: http://www.ncbi.nlm.nih.gov/books/NBK37280/

Weist, MD. Expanded school mental health services: A national movement in progress. In: Ollendick, TH.; Prinz, RJ., editors. Advances in clinical child psychology. Vol. 19. New York: Plenum Press; 1997. p. 319-352.

Weist MD. Challenges and opportunities in expanded school mental health. Clinical Psychology Review. 1999; 19:131–135. [PubMed: 10078416]

Weist, MD. Family Engagement/Empowerment Observational System Center for School Mental. Health, University of Maryland School of Medicine; Baltimore, MD: 2009.

Weist MD, Ambrose M, Lewis C. Expanded school mental health: A collaborative community/school example. Children & Schools. 2006; 28:45–50.10.1093/cs/28.1.45

Weist MD, Christodulu KV. Expanded school mental health programs: Advancing reform and closing the gap between research and practice. Journal of School Health. 2000; 70:195–200.10.1111/j.1746-1561.2000.tb06472.x [PubMed: 10900597]

Weist MD, Lever N, Stephan S, Youngstrom E, Moore E, Harrison B, et al. Stiegler K. Formative evaluation of a framework for high quality, evidence-based services in school mental health. School Mental Health. 2009; 1:196–211.10.1007/s12310-009-9018-5

Weist MD, Murray M. Advancing school mental health promotion globally. Advances in School Mental Health Promotion, Inaugural Issue. 2007:2–12.10.1080/1754730X.2008.9715740

Weist MD, Paternite CE, Wheatley-Rowe D, Gall G. From thought to action in school mental health promotion. International Journal of Mental Health Promotion. 2009; 11:32–41.10.1080/14623730.2009.9721790

Weist MD, Sander MA, Walrath C, Link B, Nabors L, Adelsheim S, Moore E, Jennings J, Carrillo K. Developing principles for best practice in expanded school mental health. Journal of Youth and Adolescence. 2005; 34:7–13.10.1007/s10964-005-1331-1

Weist MD, Stiegler K, Stephan S, Cox J, Vaughan C. School mental health and prevention science in the Baltimore City schools. Psychology in The Schools. 2010; 47:89–100.10.1002/pits.20453

Wenger, E.; McDermott, R.; Snyder, WM. Cultivating communities of practice: A guide to managing knowledge. Boston: Harvard Business School Press; 2002.

Yatchmenoff DK. Measuring Client Engagement From the Client's Perspective in Nonvoluntary Child Protective Services. Research On Social Work Practice. 2005; 15(2):84–96.10.1177/1049731504271605

Youngstrom EA, Youngstrom JK, Freeman AJ, De Los Reyes A, Feeny NC, Findling RL. Informants are not all equal: predictors and correlates of clinician judgments about caregiver and youth credibility. Journal of Child and Adolescent Psychopharmacology. 2011; 21:407–415.10.1089/cap.2011.0032 [PubMed: 22040186]