

Formative Evaluation of a Framework for High Quality, Evidence-Based Services in School Mental Health

Mark Weist · Nancy Lever · Sharon Stephan · Eric Youngstrom · Elizabeth Moore · Bryan Harrison · Laura Anthony · Kenneth Rogers · Kimberly Hoagwood · Aya Ghunney · Krystal Lewis · Kerri Stiegler

Published online: 25 August 2009
© Springer Science+Business Media, LLC 2009

Abstract The purpose of this study was to test a three component framework for enhancing quality in school mental health (SMH), focusing on quality assessment and improvement (QAI), family engagement/empowerment, and modular evidence-based practice (EBP) implementation in three established SMH programs. The study involved a 2-year, multisite (Delaware, Maryland, Texas) formative evaluation with clinicians randomly assigned to participate in either the QAI (target) intervention or a Wellness Plus Information (WPI, comparison) intervention. As hypothesized, clinicians who participated in the QAI condition demonstrated significantly greater implementation of quality indicators and greater implementation of EBP as compared to clinicians in the WPI condition. However, contrary to original hypotheses, findings did not

reveal differences between the conditions in knowledge or attitudes toward EBP, clinician self-efficacy, or student psychosocial outcomes. Implications for future research on quality improvement in SMH are discussed, with an emphasis on the need to examine the impact of increased implementation and resource support to SMH clinicians.

Keywords Children · Adolescents · School mental health · Quality · Evidence-based practice · Implementation support

Introduction

In spite of the significant progress in the school mental health (SMH) field (Evans, Weist, & Serpell 2007; Robinson, 2004; Weist, Evans & Lever 2003), the research base remains limited (Hoagwood et al., 2007), and programs struggle to provide the necessary support (administrative and resource) to deliver high quality services and implement evidence-based practices (Evans & Weist, 2004). The use of formal evidence-based interventions, particularly involving manualized strategies, is likely to encounter many obstacles, with clinicians typically showing poor or limited adherence (Schaeffer et al., 2005). In this paper, we report on a study that sought to bridge this gap between research-supported intervention and common practice in school mental health through the development and assessment of the first systematic framework for Quality Assessment and Improvement (QAI), with an emphasis on research supports that could be replicated in the real-world environment of schools. The QAI framework involved three dimensions: Effectively Working in Schools, Effectively Working with Families, and Implementing Evidence-Based Practices.

Funded by the National Institute of Mental Health, US Department of Health and Human Services, #1R01MH71015-01A1; 2003-2007. Additional support for this project was provided by cooperative agreement U45 MC 00174-10-0 from the Office of Adolescent Health, Maternal and Child Health Bureau (Title V, Social Security Act), Health Resources and Services Administration.

M. Weist (✉) · N. Lever · S. Stephan · E. Moore · B. Harrison · L. Anthony · K. Rogers · A. Ghunney · K. Lewis · K. Stiegler
Division of Child and Adolescent Psychiatry, Center for School Mental Health, University of Maryland School of Medicine, 737 West Lombard Street, 4th Floor, Baltimore, MD 21201, USA
e-mail: mweist@psych.umaryland.edu
URL: <http://csmh.umaryland.edu>

E. Youngstrom
University of North Carolina, Chapel Hill, NC, USA

K. Hoagwood
Columbia University, New York, NY, USA

Effectively Working in Schools

Mental health clinicians working in schools include those from disciplines traditionally connected to education, including school counseling, psychology and social work; and increasingly from other disciplines including clinical and counseling psychology, clinical social work, and child and adolescent psychiatry (Rappaport, Osher, Garrison, Anderson-Ketchmark, & Dwyer 2003; Weist, Ambrose, & Lewis 2006). Both school-employed and collaborating mental health staff from the community often need training on how to work effectively in schools, including developing and maintaining relationships with school administrators, teachers, health staff and others; understanding relevant educational regulations and policies; and effectively integrating mental health promotion and intervention into the school day (Paternite, Weist, Axelrod, Anderson-Butcher, & Weston 2006; Stephan, Davis, Callan Burke, & Weist 2006; Weist, 1997). In addition, people who have not worked in schools should be prepared for environmental differences compared to traditional child and adolescent mental health settings. For example, work in schools typically involves less administrative support and a greater emphasis on prevention and environmental change (Power, 2003; Power, Manz, & Leff 2003; Robinson, 2004; Weist et al., 2003).

To assist mental health clinicians to work effectively in schools, in this study, we used an expanded version of a school mental health report card—the School Mental Health Quality Assessment Questionnaire (SMHQAQ; Weist et al., 2005, 2006), which includes 10 principles and 40 indicators of best practice in SMH.

Effectively Working with Families

Effectively working with families underpins virtually all effective intervention in child and adolescent and school mental health (Hoagwood, 2005; Jensen & Hoagwood, 2008). Family involvement in schools has been linked to improvement in child academic outcomes across development (Barnard, 2004; Hill et al., 2004; Hoagwood, 2005; Jeynes, 2005; Marcon, 1999; Vanderbleek, 2004). Despite the widely recognized importance of family involvement in SMH programs, there remains a gap between best and actual practice (Lowie, Lever, Ambrose, Tager, & Hill 2003). A number of programmatic barriers to family involvement exist, including limited resources (e.g., funding, staffing) to provide evening/weekend appointments, child care, and transportation. Further, schools can be very unwelcoming places to family members (Bickham, Pizarro, Warner, Rosenthal, & Weist 1998).

Studies have demonstrated that actively engaging families in the treatment process from the beginning results in

better attendance and follow-through with mental health services. In a review of empirically supported approaches to working with families, Hoagwood (2005) identified four key domains: (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 truly collaborative approach in therapeutic interactions with families, versus operating from a position of solely providing expert guidance. (3) *Support*: Assisting families rapidly with making connections to address pressing needs (e.g., tutoring, employment), and in general, playing a supportive role in efforts to improve family functioning and child and adolescent behavior. (4) *Empowerment*: Promoting family involvement at the highest level, with clinicians reducing perceived barriers, equipping families with the means to contribute to and guide services, and instilling hope (also see McKay, 2004).

Implementing Evidence-Based Practices

A key dimension of training, prior to and following professional education, is the implementation of evidence-based practices (EBPs). Studies indicate that there is little administrative support for the implementation of EBPs in most child and adolescent and SMH settings; and without this support, these interventions are likely to be implemented poorly (Evans & Weist, 2004; Graczyk, Domitrovich, & Zins 2003; Kutash, Duchnowski, & Lynn 2006). Major limiting factors on EBPs in school mental health are the lack of time, resources and support to effectively implement them (Shernoff, Kratochwill, & Stoiber 2003), along with staff resistance to implementation due to these limitations and to the perception that manualized interventions are not well suited for schools (Schaeffer et al., 2005). To address these and other concerns, “modular” strategies to child and adolescent evidence-based practice are being advanced (Chorpita, Becker, & Daleiden 2007). A modular approach provides clinicians with competency training in core techniques and procedures (“practice elements”) that have demonstrated effectiveness as part of evidence-based protocols for particular problem areas (e.g., exposure, cognitive restructuring for anxiety disorders). For example, Chorpita’s *Modular Cognitive-Behavioral Therapy for Childhood Anxiety Disorders* (2006) targets the common features of anxiety and the most prominent evidence-based skill training approaches to address them (e.g., exposure). Although providers are guided by evidence-based fundamentals, they also have the flexibility in selecting and arranging modules as appropriate for each case. An inherent advantage to this training style is that it

allows for flexibility and individualization, as opposed to manuals, which are often criticized for their “one size fits all” approach (Curry & Reinecke, 2003). Not only does the modular approach build upon the common elements of protocols with a demonstrated evidence base, there is growing support for the efficacy of modularized treatment for children’s mental health problems (Chorpita et al., 2004).

The purpose of this study was to test this three-component framework focusing on QAI, family engagement/empowerment, and modular EBP implementation in three established SMH programs. The study represented a formative evaluation with refinement of the QAI intervention from Year 1 to Year 2. Throughout the course of the study, ongoing information pertaining to the feasibility and effectiveness of intervention procedures was collected through regular interviews with senior clinicians in training roles from each of the three study sites (elaborated on in the following), who regularly sought feedback from clinician participants regarding aspects of the study, including challenges and ideas for overcoming challenges. In addition, significant feedback was sought from senior trainers and participating clinicians at the end of the first study year, in preparation for the second year. Space constraints prohibit a detailed examination in this paper of qualitative feedback provided by trainers and clinicians and analyzed and processed by research investigators toward an achievable QAI strategy in SMH. However, themes related to lessons learned and intervention improvement are presented throughout the paper.

Method

Study Overview

The study involved a 2-year, multisite, stratified random assignment, formative evaluation study. Within each site (Delaware, Maryland, Texas), schools and the clinicians serving them were randomly assigned to be involved in either the Quality Assessment and Improvement (QAI, target) intervention or the Wellness Plus Information (WPI, comparison) intervention. Information provided to clinicians in the WPI condition was an overview of 10 principles for best practice in school mental health (Weist et al., 2005). The two respective interventions were implemented in Years 1 (2004–2005) and 2 (2005–2006), with improvements and refinement in the interventions in Year 2 based on lessons learned, feedback, and new developments in the field. Primary Aim 1 of the study was to evaluate the impact of the QAI intervention on the quality of services used in treating specific disorders. We hypothesized that clinicians randomly assigned to the QAI

condition would have higher knowledge and demonstrated skills in implementing evidence-based practices for specific child and adolescent emotional/behavioral disorders. Primary Aim 2 of the study explored impacts on clinician’s attitudes toward EBP and clinical self-efficacy. In addition, we explored a Secondary Aim on satisfaction and outcomes for students treated by clinicians in the two conditions.

Research Sites and Assignment of Schools to Targeted and Comparison Conditions

Three SMH programs participated in the study, University of Maryland School Mental Health Program (SMHP), Christiana Care Visiting Nurses Association (CCVNA), and the Dallas Youth and Family Centers (DYFC). The SMHP was operating in 22 public schools in Baltimore City. This included ten elementary schools, one elementary/middle school, six middle schools, and five high schools. Program schools are characterized by high levels of poverty and community problems, such as violence, substance abuse, and crime. About 85% of students in the schools were African American, with the majority of remaining students being Caucasian. Across schools in the program, roughly 75% of students received reduced/free lunches and 19% were in special education.

The CCVNA provides health and mental health services through 16 wellness centers located in public high schools (urban, rural, suburban) throughout the state of Delaware. Roughly 60% of students served by the program were Caucasian, 35% African American, and around 5% were Hispanic. Around 20% of students received reduced/free lunches, and around 11% were in special education.

The DYFC provides expanded school mental health services to all 220 of the schools in the Dallas Independent School District. They use a cluster model with each cluster serving up to 25 schools. Two clusters participated in the study—North Oak Cliff and Woodrow, which were selected based on similarity to each other in sociodemographic characteristics of students and families and school factors. The North Oak Cliff cluster served 20 schools (11 elementary, 7 middle, 2 high); about 85% of students were Hispanic, 13% were African American, and 3% were Caucasian, with 76% receiving reduced/free lunches, and around 9% were in special education. The Woodrow Cluster was served 22 schools (17 elementary, 3 middle, 2 high); about 80% of students were Hispanic, 13% were African American, and 8% were Caucasian, with 70% of students receiving reduced/free lunches, and 8% in special education. The Dallas program had 21 targeted and 21 comparison schools participating in the study.

Table 1 Summary of research sites

Name of program	Services provided to	Number of target schools	Number of comparison schools	School levels served	Geographic	Race/ethnicity	Free/reduced lunch	Special education
The Christiana Care Visiting Nurses Association	Wellness Centers in Delaware Public High Schools	8	8	High School (16)	Rural, Suburban, Urban	60% Caucasian, 35% African American, 5% Hispanic	20%	11%
The Dallas Youth and Family Centers: North Oak Cliff and Woodrow Clusters	The Dallas Independent School District	21	21	North Oak: Elementary (11), Middle (7), High School (2) Woodrow: Elementary (17), Middle (3), High School (2)	Urban	North Oak: 85% Hispanic, 13% African American, 3% Caucasian Woodrow: 80% Hispanic, 13% African American, 8% Caucasian	North Oak: 76% Woodrow: 70%	North Oak: 9% Woodrow: 8%
University of Maryland School Mental Health Program	Baltimore City Public Schools	11	11	Elementary (7), Elementary/Middle (4), Middle (5), High School (6)	Urban	Caucasian 85% African American, 14% Caucasian, 1% Other	75%	19%
Total		40	40					

Within each of the three research sites, all schools being served by the respective SMH programs agreed to participate in the project. Prior to project start-up, within each program, schools were matched according to level (i.e., elementary, middle, and high school), clinician characteristics (full time equivalent, ethnicity and years of clinical experience), and major school and student sociodemographic characteristics (e.g., school size, racial background of students, percentage who receive reduced/free lunch) and based on these factors a stratified random assignment of school/clinician to condition was used. Once pairs of schools were matched, a coin was tossed to determine which school was assigned to the QAI group and which to the WPI group. After randomization, analyses indicated that the groups were statistically equivalent along the dimensions. Table 1 highlights school participation and demographics of the three sites participating in the study.

Study Participants and Recruitment Strategies

The primary subjects in the study were clinicians employed by one of three established SMH programs in targeted and comparison schools. Clinicians were able to provide a full continuum of prevention and intervention services to youth in school, primarily those in general education. Baltimore and Delaware clinicians were located on-site, whereas Dallas clinicians provided school-based or linked services, with services located on one school campus, and strong connections including transportation support to other schools within a feeder pattern of schools. For all sites, the providers of mental health services were community-based agencies with an established history of providing school-based care working in close collaboration with schools and families. There were a total of 91 participants across the 2 years, with 64 in Year 1 and 66 Year 2. Of the 64 participants in Year 1 (35 QAI and 29 WPI), 24 participants did not continue into Year 2 (13 QAI and 11 WPI). The dropout between Year 1 and Year 2 did not demonstrate a significant difference between the two groups, 37% Quality versus 38% Wellness. Of the 24 participants who did not continue, one clinician from the WPI condition in Baltimore dropped out due to workload demands, two Dallas clinicians (one from each study condition) discontinued due to increased administrative responsibilities, with the remaining 21 participants leaving the programs for other reasons (e.g., graduate school, maternity leave). No particular patterns were noted related to non-participation across clinicians (e.g., study condition, school level, etc.). Eighty-two percent of the clinician participants were female, with 52% Caucasian, 26% African American, 20% Hispanic, and 2% Asian American.

Although clinicians served as the primary participants in the study, additional data were collected from: (1) students

aged 11 and older who received services from the SMH clinicians, (2) parents of students who received services, (3) school staff who referred students, and (4) school principals and/or assistant principals. School staff and principals received a letter requesting their anonymous participation, with participation reflecting their consent. Informed consent was obtained from clinicians to participate in the study, and informed parental consent and student assent was obtained for student participants. After students were seen for five sessions (chosen as an indicator of a meaningful level of treatment services; Weist & Albus, 2004), clinicians participating in both the targeted and comparison schools described the project to students and families during the intake process and asked permission to provide their contact information to the research team. Parents/guardians were provided a packet of information (either via mail or directly from the clinician) including a letter describing the study, informed consent and assent forms, and child and parent satisfaction measures. Two weeks after families received the forms, a telephone call was made to the family asking if they had any questions, if they had consented (assented) to be in the study, and if so, returned the forms. Families were compensated \$20 for participating in the study. In Year 1, a total of 192 parents/guardians completed and returned project measures, and in Year 2, 187 parents/guardians completed and returned the measures (for a total of 379). Clinicians provided the research staff with contact information for 532 families. Therefore, we attained a response rate of 71.2% for parent/guardian questionnaires. However, because information was not gathered on the total number of students seen five or more times, nor on the number of families who declined to provide contact information, we were unable to calculate the response rate among all eligible participants. In addition, there was variability in clinician recruitment of families into the study ranging from 1 case to 20, and in spite of training and regular contact with research staff on parameters for recruitment, biases likely operated. This factor, and the fact that cases were seen for individual intervention (augmented by family intervention) may help to explain why males comprised only 39% of the student sample. The disproportional number of male versus female students involved is consistent with literature showing that boys and men are less likely than girls and women to seek social support and counseling (Hunter, Boyle, & Warden 2004; Rickwood & Braithwaite, 1994), and other findings, e.g. showing higher rates of depression among teenage girls than boys (2 to 1 ratio; Hazler & Mellin, 2004).

In Year 1 of the study, 104 students (age range 11–19, mean 14.1) completed the student questionnaires. Male students represented 38.8% of those returning project measures, and 16.5% of those returning measures were in special education. In terms of race/ethnicity, 42.3% of the

students completing the survey were African-American, 30.8% were Caucasian, 18.3% were Hispanic/Latino, and 8.7% were Other. In Year 2 of the study, 105 students (age range 11–19, mean 14.5) completed and returned project measures, with 32% males. (We did not collect race/ethnicity of student participants in Year 2).

Project Measures

The following measures were completed as part of the study. In Dallas, measures were translated into Spanish as needed. Please note that since the study was a formative evaluation over 2 years, the measurement plan was revised for some measurement domains. For example, in Year 2, we expanded knowledge and attitudes measures.

Primary Aim 1: Exploring Impacts on Use of Evidence-Based Practices and Implementation of Quality Indicators

Depression and ADHD Interviews The interview was designed to review cases systematically by the American Academy of Child and Adolescent Psychiatry. During the interview, clinicians had in front of them a case record for a client treated for any depressive disorder to review.

To assess the quality of clinical services delivered by clinicians, case review interviews were conducted on the quality and implementation of treatment for an individual case involving one of two common disorders, Depression and Attention-Deficit Hyperactivity Disorder (ADHD). In Year 1 of the study, all clinicians in elementary schools were interviewed about an ADHD case, while clinicians in middle and high schools were interviewed about a Depression case. Based on lower numbers of staff completing the ADHD interview in Year 1, and elementary clinicians consistently indicating they worked with youth presenting depression, it was decided that all clinicians would complete the interview using a Depression case in Year 2. Clinicians were asked to select cases that were most representative of their work with students presenting this particular disorder. A case review specialist used a semistructured interview process developed to assess clinician implementation of evidence-based practices in relation to Depression or ADHD. The case review specialist was blind to condition for all participants. The two versions of the tool were developed based on professional association recommendations for high quality and effective practice, including recommendations of the National Assembly on School-Based Health Care, and the American Academy of Child and Adolescent Psychiatry. Literature on evidence-based practice for treating these disorders was also reviewed, and key informant interviews were conducted with acknowledged experts in conducting research and implementing EBPs for these disorders. Interviews of

SMH clinicians were conducted by phone for both disorders in May or June at the end of each study year, and assessed 12 specific dimensions of EBP for these disorders (reviewed in Results). The interviews also included a global impression of the extent to which the clinician followed recommended practice parameters. Two quantitative scores were derived from the interview: The interviewer's global impression of adherence, scaled from 0 ("no or very little adherence; 0–20% of standards met") to 4 ("excellent adherence; 80–100% of standards met"), and a summary score derived from the individual ratings of adherence to each of the component items assessed in the interview. This summary score contained twelve items for the depression and ADHD scales (each originally scaled 0–4). In Year 1 the alpha was .82, and in Year 2 the alpha was .88. To determine interrater agreement on the scale, reliability between raters was assessed on 14% of the interviews in Year 1 (7 interviews). The two raters demonstrated 91% agreement. The most frequent errors were forgetting to code an answer. Where the raters disagreed, they typically disagreed by only one point.

A total of 50 clinicians in the study participated in the case reviews in Year 1 and 40 clinicians participated in Year 2. Interviews took approximately 20–30 min to complete. Although all clinicians were invited to participate, for a variety of reasons (e.g., illness, leaving position before June, no case with relevant diagnosis) not every clinician completed the case review. To protect privacy, student names were not included during case reviews.

Quality of School Mental Health Services The School Mental Health Quality Assessment Questionnaire is a research-based measure developed by our team, which in Year 1 assessed 45 indicators connected to 10 principles for best practice in SMH (SMHQAQ; Weist et al., 2005, 2006a, b). Principles are listed below: (1) All youth and families are able to access appropriate care regardless of their ability to pay; (2) Programs are implemented to address needs and strengthen assets for students, families, schools, and communities; (3) Programs and services focus on reducing barriers to development and learning, are student and family friendly, and are based on evidence of positive impact; (4) Students, families, teachers and other important groups are actively involved in the program's development, oversight, evaluation, and continuous improvement; (5) Quality assessment and improvement activities continually guide and provide feedback to the program; (6) A continuum of care is provided, including school-wide mental health promotion, early intervention, and treatment; (7) Staff holds to high ethical standards, is committed to children, adolescents, and families, and displays an energetic, flexible, responsive and proactive style in delivering services; (8) Staff is respectful of, and competently addresses

developmental, cultural, and personal differences among students, families and staff; (9) Staff builds and maintains strong relationships with other mental health and health providers and educators in the school, and a theme of interdisciplinary collaboration characterizes all efforts; (10) Mental health programs in the school are coordinated with related programs in other community settings.

Clinicians were asked to rate the degree that each indicator was developed and/or implemented in their practice. Staff rated each indicator on Likert scales from "not at all in place" to "fully in place" Based on feedback from Year 1, the Likert scale on the measure was expanded to 6 from 4 points and redundant items were eliminated resulting in 40 items in Year 2. The SMHQAQ was administered to the QAI group at the start and end of each year, and to the WPI group at the end of Year 2. While the administration of the SMHQAQ to the WPI group only at postintervention presented the limitation of not being able to assess for preintervention differences between groups, it was determined that provision of the instrument prior to the end of the study would increase the risk of contamination of the comparison group with content from the QAI group.

Scale level principal component analyses indicated that ten principles cohered into a single strong component in three out of four assessments (with the exception being the posttest at the end of Year 1, where a first component still explained 41% of the variance, and the second two components both had few indicators and many cross-loadings) (Glorfeld, 1995). Because of this, subsequent analyses relied primarily on total scores on the SMHQAQ. The alphas for the SMHQAQ total were .94 (Year 1, Fall), .93 (Year 1, Spring), .94 (Year 2, Fall) and .95 (Year 2, Spring).

Primary Aim 2: Exploring Impacts on Knowledge and Attitudes Toward Evidence-Based Practices, and Clinical Self-Efficacy

Knowledge and Attitudes Toward Evidence-Based Practices Based on the work of the Hawaii Department of Health (Chorpita & Daleiden, 2007) and its comprehensive summary of top evidence-based modular practice elements, we developed the *Practice Elements Checklist (PEC)*. The PEC asks clinicians to provide ratings of the top eight skills for each of the four disorder areas (ADHD, Disruptive Behavior Disorders, Depression and Anxiety). Respondents used a six-point Likert scale to rate current knowledge of the practice element (1 = "none" and 6 = "significant") and frequency of use of the practice element (1 = "never" and 6 = "frequently").

Clinicians completed the PEC at the start and end of Year 2. Internal consistencies were excellent for each of the subscales, ranging from .84 to .92.

To assess knowledge of evidence-based practice, we used the *Knowledge of Evidence-Based Services Questionnaire* (KEBSQ; Stumpf et al., 2009), which asks participants to indicate whether different strategies are or are not supported for use with four disorder areas (anxious/avoidant, depressed/withdrawn, disruptive behavior, and hyperactivity/inattention). Items were generated for the KEBSQ to represent component skills that constitute evidence-based intervention (e.g., relaxation, exposure) as well as techniques that are frequently utilized yet are not part of any evidence-based treatments for mental health disorders in youth (e.g., “insight building”). The KEBSQ has 40 strategies (items), with an item score of 0–4 (with “4” meaning that the respondent correctly identified the strategy as being supported or not for each of the four disorder areas), and a possible total score of 160. Findings from an investigation conducted by Stumpf and colleagues suggest that test–retest reliability of the KEBSQ is acceptable ($r = .56$). In addition, results indicated a significant group by time interaction ($F [1,118] = 55.55$, $P < .001$), demonstrating that the KEBSQ is sensitive to participation in training on EBPs. Internal consistency was relatively poor ($\alpha = .72$) in our sample.

To assess attitudes toward evidence-based practices, we used the *Evidence-Based Practice Attitude Scale* (EBPAS; Aarons, 2004), which assesses general clinician attitudes toward the adoption of evidence-based practice (EBP) with four subscales: Requirements (would an EBP be used if required), Appeal (is EBP intuitively appealing), Openness (would an EBP be adopted if it required following a treatment manual), and Divergence (extent of divergence of EBP from usual care). Clinicians were asked to indicate the extent to which they agreed with each of the 15 items using a four-point Likert scale (0 = “not at all” and 4 = “to a very great extent”). This measure has also been reported to have adequate psychometric qualities in prior research but showed variable internal consistency in our sample with alphas for the respective scales of .91, .75, .78, and .48. These two measures were collected at the end of Year 2 for all clinicians (they were not in the original research plan, and related to delays in completing an IRB amendment could only be collected at the end of Year 2).

Clinical Self-Efficacy We collected measures from clinicians on counseling self-efficacy using the *Counselor Self-Efficacy Scale* (CSS; Sutton & Fall, 1995). The CSS contains 33 items reflecting school counselors’ views on their effectiveness in various counseling roles in the school, and their expectancies for positive outcomes resulting from their work. Each item is rated on a 1–6 Likert scale, with 1 representing “strongly disagree,” and 6 representing “strongly agree.” The internal consistency of the measure has been reported to be adequate (.65–.75) (Sutton & Fall,

1995). While this measure was developed for school counselors, the language was modified slightly to make it more relevant to the work of SMH clinicians (e.g., changing “guidance program” to “counseling program”). This measure was collected from clinicians in both conditions at the beginning and end of study Years 1 and 2.

Secondary Aim: Exploring Impacts on Student Emotional/Behavioral Functioning, and on Satisfaction with Services by Parents and School Staff

Emotional/Behavioral Functioning of Students Treated by Clinicians The *Strengths and Difficulties Questionnaire* (SDQ; Goodman, Meltzer, & Bailey 1998) is a brief screening questionnaire for children and adolescents aged 3–17 that assesses emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. There are several versions of the SDQ, with adequate psychometric properties (Cronbach alpha informant—.73, student—.82; test stability informant—.62). In this study, we used a parent form and a self-report form for youth aged 11–17. Each form is comprised of 25 items that assess the following five domains: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behavior.

Student and Family Satisfaction The Client Satisfaction Questionnaire, Version 8 (CSQ-8), one of the most widely used measures of client satisfaction, was used to measure student satisfaction. The CSQ-8 includes eight items, with each item rated on a four-point scale (Attkisson & Zwick, 1982; Roberts et al., 1984). In addition, four items pertaining to reasons for dissatisfaction were added to the CSQ-8, including (1) The sessions with the counselor were too short, (2) The counselor was not available enough, (3) My counselor did not keep my “business” private, and (4) My counselor gave me unhelpful advice. These items were derived from our prior research in this area, reflecting the most common reasons for student dissatisfaction with mental health services. These items were added to the scale to address possible ceiling effects in satisfaction ratings. The internal consistency for the CSQ-8 was .90 when completed by students in the present sample.

Parent, Referring Staff, and Principal Satisfaction The revised CSQ-8 was then adapted to measure parent, referring staff, and principal satisfaction with services, respectively, with appropriate minor changes in language for these audiences (e.g., for the parent version, instead of “How satisfied are you with the amount of help you received?” the item read, “How satisfied are you with the amount of help your child received?”). As above, four items related to dissatisfaction were added. Based on input

from clinicians and referring staff related to insufficient knowledge to respond to all items (e.g. the sessions with the counselor were too short), the instrument was modified for referring staff to only include five items on which they could more reasonably report. The parent CSQ-8 had an alpha of .91 in the present sample. The referring school staff and school principals completed the CSQ-8 in Year 1 only with alphas of .82 and .89, respectively.

Training of Clinicians in Targeted and Comparison Schools Common training approaches for staff and supervisors in both conditions. Four training events were conducted for participating staff in both conditions, a 2-day more intensive training in the summers of 2004 (Year 1) and 2005 (Year 2), and 1 day refresher training and processing in the springs of 2005 (Year 1) and 2006 (Year 2). All four trainers had significant experience in making professional presentations. In both conditions, training included a mix of didactics, questions and comments, interactive exercises, and open discussion based on social learning theory to maximize training impact, and continuing education credit was provided.

Unique training provided for staff in the QAI condition. As part of the four training events, the QAI condition participants received training related to the following: (1) Understanding Quality Assessment and Improvement, (2) Evidence-Based Practice, (3) Practice Elements for Specific Disorders. In addition, the QAI clinicians used the trainings at the beginning of each year to prioritize and select quality indicators to focus on during the year. At each of the study sites, “senior clinicians” were recruited to serve as project supervisors for staff in the QAI condition. These clinicians were masters level or higher and had at least 3 years experience in SMH and were licensed as mental health professionals. Each worked 1 day per week on the study, and was assigned a group of about ten clinicians from targeted schools with whom to work. They held weekly group meetings with quality groups to review QAI processes and activities in their schools, and strategies for using the evidence base. They also served as liaisons between on-site leaders and CSMH staff in conveying information and offering resources to project staff, and ensuring that study measures were being completed appropriately and in a timely manner. The senior clinicians participated in the twice annual trainings and also participated in an intensive train-the-trainer session at the CSMH focused on teaching modular intervention. Related to number of participants in the study, there were two senior clinicians in Dallas (both male psychologists), two senior clinicians in Baltimore (both female, one psychologist one social worker), and one senior clinician in Delaware (female social worker). Senior clinicians were not participants in the study, and there was no turnover among them

during the duration of the study (although one from Baltimore was out for 3 months related to maternity leave).

As part of the QAI process, quality indicators to be targeted during the school year were selected based on summer SMHQAQ ratings. Each month one to two indicators were emphasized in the quality group meetings. Clinicians reviewed PowerPoints, summaries of QAI strategies, newsletters, and relevant materials developed by the CSMH staff, and worked with senior clinicians to develop activity steps for each indicator. Demands to focus on indicators were reduced from around 15 indicators in Year 1 to around 9 indicators in Year 2 based on clinician feedback regarding manageability. The second dimension trained during weekly meetings was family engagement and empowerment strategies, based on the previously mentioned framework (Hoagwood, 2005), and including materials as described earlier. The third dimension trained was modular cognitive-behavioral intervention strategies (Anxiety, Depression, ADHD, and Disruptive Behavior Disorders), based on the work of the Hawaii Department of Health (Chorpita & Daleiden, 2007) and their ranking of the most common cognitive behavioral practice elements to address these disorders.

Assessing the integrity of training in the QAI condition. To promote fidelity, in Years 1 and 2, quality group sessions for each of the senior clinicians were audiotaped and then reviewed by a senior project staff member with significant experience in SMH and evidence-based practice. Written and verbal feedback was then provided to each of the senior clinicians. In Year 2, a senior project staff member held weekly meetings, by phone or in person, with the senior clinicians as a way to record weekly progress and provide feedback and recommendations to senior clinicians. The weekly fidelity meetings provided an opportunity to give guidance and support to the senior clinicians and ensure that project staff were developing and sharing resources that would support the required content.

Unique Training and Support for Staff in the WPI Condition In summer and spring trainings for both years of the project, staff in the WPI condition received unique training and resources on stress management, relaxation, coping, exercise, nutrition, and preventing burnout, with the training program based on the best practices in the wellness field. Over the course of the school year, staff in comparison schools expressed some interest in informal smaller group wellness meetings with the other clinicians. Although these meetings were encouraged, we did not structure the process or content of these encounters. Similar to the QAI condition, staff and supervisors in the WPI condition received updates on staff wellness through a separate CSMH list serve, and had password protected access to additional materials on the topic through the center’s website.

Results

Preliminary Analyses

All variables were screened for out of range scores and univariate outliers, using box and whisker plots, as well as examining skew and kurtosis. All primary outcome measures had distributions that were approximately normal, well within the range for robust analyses. All tests of statistical significance were reported with alpha of .05, two-tailed. Due to the limited sample size and exploratory nature of many analyses, no posthoc error correction was used; but effect sizes are reported as a way of gauging the size of associations, and statistical significance testing was not used when fewer than 10 cases were available per subgroup. Reliability statistics (alpha and principal component structures) were reported for descriptive purposes. Principal components analysis is appropriate for purposes of data summarization (e.g., identifying major dimensions for description, rather than reporting multiple correlated findings separately), as opposed to theory testing or identifying psychometric latent variables. Principal components analyses can also provide stable and robust estimates with smaller sample sizes when the number of indicators is large or indicator validity is good (Guadagnoli & Velicer, 1988; MacCallum, Widaman, Preacher, & Hong 2001).

Another set of preliminary analyses evaluated potential effects of missing data. Data were not available to compare the participating youths to all youths attending the schools on study measures or demographics. Within the sample of participating youths and clinicians, we constructed dummy codes to examine patterns of missingness and to test whether responders differed from non-responders in terms of demographics or other study variables. There was evidence of clustering within informant (e.g., if one clinician measure was missing, then other clinician measures also were more likely to be omitted), but there were no other statistically significant patterns of missingness.

Scaling: Percentage of Maximum Possible

In order to facilitate comparisons between scores, including visual presentations, we used a scaling method called “Percentage of Maximum Possible” (POMP) scoring developed and recommended by Cohen and colleagues (Cohen, Cohen, Aiken, & West 1999). It is a simple transformation where raw scores are adjusted so that they can range from zero to 100%. POMP scoring makes no assumptions about the shape of the distribution (unlike *z*-scores, which assume a normal distribution), and the anchors (zero and 100%) are tied to the full possible range of the measure, not observed parameters in the data that could change from sample to sample. Using POMP scoring

allows consistent use of scales on the axes of graphs, and makes interpretation more intuitive to readers. POMP scoring is also helpful in evaluating “material effects” such as interventions to improve quality and adoption of evidence-based principles (Cohen et al., 1999).

Findings for Primary Aim 1 Interview on evidence-based practices for depression. At the end of Year 1, 19 clinicians in the Quality condition and 15 in the Wellness condition completed interviews about their use of evidence-based practices for the treatment of depression. The clinicians in the Quality condition used markedly higher levels of evidence-based practices, based on the summary rating of the components used (81 vs. 62% of maximum possible for the Wellness group), $t(23.9) = 3.69, P < .005$, and also on the interviewers’ global impression (84 vs. 62%), $t(32.0) = 4.10, P < .0005$. This represented a dramatic difference in the amount of evidence-based content employed, with corresponding effect sizes of $d = 1.35$ and 1.50 —where .8 is conventionally considered a “large” effect.

Also of note, the clinicians in the Wellness condition tended to show greater variability in their amount of evidence-based practices (SD of 17 vs. 12% in the Quality condition), $F(12, 18) = 3.72, P = .063$. The findings suggest that the intervention actually reduced the variability in implementation in the Quality group as well as raising the average amount of implementation. Table 2 presents the breakdown in terms of specific skills and content provided in the treatment of depression. In Year 1, the Quality group performed significantly better across most skills and content areas, with large effect sizes for everything except for involving caregivers in treatment or providing psychoeducation about depression to caregivers.

A similar pattern held in the second year, when 13 counselors in the Quality and 11 in the Wellness condition completed interviews (see Table 3). The effect size was $d = 1.02$ for the summary score and 1.17 for the global impression, both favoring the Quality group. Again, the intervention also tended to reduce the variability in implementation of evidence-based approaches in the Quality condition, $F(10, 12) = 3.69, P = .068$. The effect sizes in the second year tended to be smaller. The only comparisons achieving conventional thresholds for statistical significance were increases in providing psychoeducation to the caregiver about depression, and also using more rating scales. Large (but not statistically significant) effects were also found for conducting more systematic interviews about the symptoms of depression, and also promoting generalization of skills across settings. The small sample size and thus low power can also help explain the lack of significant findings. Upon comparison of site data, there were no significant site differences in either study year (Table 4).

Table 2 Comparison of specific skills or content areas covered in quality ($n = 19$) versus wellness ($n = 15$) based on depression interviews, year 1

Specific skill or content area	Quality <i>M</i> (SD)	Wellness <i>M</i> (SD)	Cohen's <i>d</i>
Promoting generalization of skills across settings	3.4 (0.8)*	2.2 (1.2)	1.18
Diagnostic interview with caregiver	3.5 (0.8)*	2.5 (1.1)	1.11
Maintenance planning	3.2 (1.0)*	1.9 (1.4)	1.07
Assessing for presence of comorbid symptoms	3.2 (1.0)*	2.3 (1.0)	.89
Homework assignments	3.5 (1.1)*	2.4 (1.5)	.88
Rating scale usage (interviewer judgment)	2.9 (0.8)*	2.1 (1.5)*	.88
Psychoeducation about depression to student	3.3 (1.0)*	2.5 (1.0)	.87
Interviewer rating of skills taught	4.0 (0.0)*	3.9 (0.4)*	.78
Treatment plan considers seriousness	3.3 (1.0)*	2.5 (1.3)	.76
Interviewed about specific depression symptoms	3.1 (0.3)	2.9 (0.3)	.60
Psychoeducation about depression to caregiver	2.6 (1.4)	2.1 (1.3)	.39
Were caregivers involved in the treatment?	2.8 (1.3)	2.7 (1.6)	.13

* Denotes the larger mean or standard deviation, $P < .05$, two tailed

Clinicians at the elementary level completed a similar interview about evidence-based practices for ADHD at the end of Year 1. However, only nine clinicians in the Quality condition and five in the Wellness condition had relevant cases and completed the interview. The small sample size precluded investigating the psychometric properties of the items. Please note that the small sample size for ADHD cases was likely related to two factors, (1) the requirement that student participants be 11 or older, which restricted the range of eligible students for the study for clinicians at the elementary level, and (2) school clinicians were primarily providing individually focused services augmented by working with families, more appropriate for the treatment of internalizing versus externalizing disorders. Based on the interviewer global impressions, there was a similar pattern of findings: The Quality group obtained an average score of 86 versus 75% for the Wellness group, $d = .96$. Again, the Quality group also had a smaller standard deviation, 18 versus 13%. However, findings could not

meet standards for significance despite the large effects due to small sample size.

School Mental Health Quality Assessment Questionnaire (SMHQAQ). A group of 27 counselors in the Quality condition completed the SMHQAQ at the beginning and end of Year 1. The average total score on the SMHQAQ moved from a 47 to a 66% over the course of the year, $t(26) = 7.91, P < .0005$. This was a large effect size for the intervention leading to increased implementation of indicators, $d = 1.4$. Counselors with higher scores at baseline tended to still have higher scores at the end of Year 1, $r = .57, P < .005$. Looking at the 10 principles subsumed within the SMHQAQ, the clinicians showed statistically significant improvement in each area except the first principle, all $P < .05$, with effect sizes ranging from $d = .49$ (principle #1) to $d = 1.26$ (principle #5).

A subset of 20 clinicians continued in the Quality condition from Year 1 to Year 2. Comparing their SMHQAQ scores from spring of Year 1 to Fall of Year 2 suggests that

Table 3 Differences between conditions on interview of evidence-based practices for depression and ADHD

Scale	Quality Mean (SD)	Wellness Mean (SD)	Cohen's <i>d</i>
Depression—rating of specific treatment components year 1	81% (12%)	62% (17%)	1.35*
Global depression rating year 1	84% (19%)	57% (20%)	1.50***
<i>n</i>	19	15	
Depression—rating of specific treatment components year 2	83% (9%)	72% (15%)	1.02*
Global depression rating year 2	87% (13%)	66% (23%)	1.17*
<i>n</i>	13	11	
Global ADHD year 1	86% (13%)	75% (18%)	0.96
<i>n</i>	9	5	

All scales using POMP scoring. Cohen's *d* measures effect size of advantage for quality over wellness

* $P < .05$, ** $P < .005$, *** $P < .0005$, two-tailed. Follow-up ANOVAs indicated that there were no significant site differences or site condition interactions

Table 4 Comparison of specific skills or content areas covered in quality ($n = 14$) versus wellness ($n = 11$) based on depression interviews, year 2

Specific skill or content area	Quality <i>M</i> (SD)	Wellness <i>M</i> (SD)	Cohen's <i>d</i>
Interviewed about specific depression symptoms	3.0 (0.0)	2.6 (0.7)*	1.13
Psychoeducation about depression to caregiver	3.5 (0.8)*	2.5 (1.1)	1.00
Rating scale usage (interviewer judgment)	3.4 (0.8)*	2.5 (1.4)	.95
Promoting generalization of skills across settings	3.4 (0.7)	2.7 (0.9)	.90
Maintenance planning	3.2 (1.1)	2.5 (1.1)	.56
Treatment plan considers seriousness	2.9 (0.9)	2.5 (1.3)	.45
Psychoeducation about depression to student	3.4 (1.0)*	3.1 (0.5)	.43
Assessing for presence of comorbid symptoms	3.0 (1.0)	2.6 (0.9)	.40
Were caregivers involved in the treatment?	3.3 (0.9)	2.9 (1.3)*	.40
Diagnostic interview with caregiver	3.3 (1.0)	3.2 (1.2)	.10
Interviewer rating of skills taught	4.0 (0.0)	4.0 (0.0)	n/a
Homework assignments	3.2 (0.9)	3.5 (0.7)	-.28

* Denotes the larger mean or standard deviation, $P < .05$, two tailed

gains were maintained, with SMHQAQ Total scores averaging 67% at the end of Year 1 and 66% at the beginning of Year 2, $t(19) = .56$, ns. The stability of clinician implementation of indicators was also high, $r = .50$, $P < .05$. Overall, findings suggest that the skills gained during Year 1 were maintained into the outset of Year 2.

Twenty-eight clinicians completed the SMHQAQ at the beginning and end of Year 2. They demonstrated a significant increase in implementation of indicators over the course of the year, with total scores rising from a mean of 62 to 73%, $t(27) = 4.61$, $P < .0005$, $r = .54$, $P < .005$. Significant gains were made in adherence to eight out of ten principles, with the exceptions being principles #7 and #8. The effect sizes of the gains ranged from $d = .36$ to $d = .97$.

In addition, a group of clinicians in the Wellness condition completed the SMHQAQ at the end of Year 2 ($n = 15$). Their scores were significantly lower on average than the clinicians in the Quality condition at the end of Year 2, with a mean of 65% versus 77%, $t(32) = 2.77$, $P = .009$, $d = .98$. The effect sizes for the specific principles ranged from $d = -.04$ (Principle #1) to 1.07 (Principle #10), with statistically significant advantages for the Quality group on Principles #3, #5, #8, and #10. Overall, the findings indicate that the Quality group produced large gains in implementation of quality indicators over both years, with the level of implementation substantially exceeding the level in the Wellness comparison group, and the implementation gains were successfully maintained over the summer into the next school year.

Findings for Primary Aim 2 Knowledge of Evidence-Based Services Questionnaire (KEBSQ; End of Year 2 Only). There were no significant differences in performance on the KEBSQ at the end of Year 2, $t(49) = 0.38$,

$P > .05$. The effect size was very small, suggesting that the null results were not due to low statistical power. However, the psychometric properties of the KEBSQ may have made it difficult to detect effects. The internal consistency of the total score was quite low for a scale of its length, and the median corrected item-total correlation of .25 lay near the .20 threshold where authorities recommend discarding the item, and 17 out of 40 items fell below that threshold. A principal components analysis found two well-identified components (not a single score), and yet roughly a third of the items did not load on either major component. The two treatment conditions did not differ significantly on either component score. Informal feedback from participants was that the KEBSQ was difficult to complete, due to length and the complexity of judgments required for completing each item.

Findings for Secondary Aim Evidence-Based Practice Attitudes Survey (EBPAS). The EBPAS was completed at the end of Year 2 by clinicians in both the Quality and Wellness groups. There were no significant differences between the two groups in terms of average scores (or differences in standard deviations) on any of the four scales or total score on the EBPAS. Part of the reason for a lack of differences may have been due to a high level of positive regard for evidence-based practices in the Wellness condition, where POMP scores averaged from 69 to 84% across scales.

Counselor Self-Efficacy Scale (CSS). There was no difference between QAI and WPI groups in perceived efficacy, with main effects for condition, time, and time \times condition interaction effects insignificant in study Year 1. The only significant effect in Year 2 was for time with a significant decrease in CSS score over the course of the year, $F(1, 51) = 6.36$, $P = .017$, partial eta-squared = .17. This effect was driven by a 3% drop in efficacy in the

Wellness group and a 1.6% drop in the Quality group, neither of which would be clinically meaningful.

Strengths and Difficulties Questionnaire (SDQ) and the Client Satisfaction Questionnaire (CSQ). For measures of student emotional/behavioral functioning (Parent and Student SDQ) and satisfaction with services (Parent, Referring Staff, and School Principal CSQ-8), there were no main effects for intervention; nor were there significant site \times intervention interactions. Follow-up analyses considering nesting effects of student and parent data within clinicians using Hierarchical Linear Modeling (HLM) and Generalized Estimating Equations (GEE) found no significant intervention effects.

Discussion

This study involved the first effort to develop and implement a systematic approach to improve the quality of SMH services, through a three-component framework including QAI, evidence-based intervention through an enhanced modular strategy, and emphasis on family engagement and empowerment. A primary goal of the study was to bridge the gap between research-supported intervention and the real-world environment of the schools toward a realistic approach to integrate evidence-based practices in the schools. The study was a formative evaluation involving refinement of the three-component intervention in Year 1, and application of a refined approach in Year 2. Three well-established school mental health programs from Baltimore, Dallas, and the state of Delaware participated in the study.

With clinicians randomized into conditions of systematic QAI as mentioned earlier, as compared to one focusing on Wellness Plus Information (on principles for best practice in SMH), the study pursued two primary aims, involving evaluation of the QAI intervention on (1) service quality as reflected in implementation of quality indicators and use of evidence-based practices, and (2) knowledge and attitudes about evidence-based practice, and perceptions of organizational climate and counseling self-efficacy. In addition, a secondary aim was pursued, evaluating impact of QAI on parent, student, and school staff satisfaction with services.

Findings strongly confirmed Primary Aim 1: on an interview measure (with interviewer blind to condition), clinicians in the QAI condition were determined to be more likely to use established evidence-based practices. In Year 1, the Quality group performed significantly better across most skills and content areas, with similar findings in Year 2. In addition, clinicians in the QAI condition showed significant improvement in the implementation of quality indicators from summer to spring assessment in Year 1,

with these improvements maintained in the summer of Year 2, and a similar level of improvement again from fall to spring assessment in Year 2. At the spring assessment in Year 2, Wellness clinicians also completed the SMHQAI and findings confirmed a significant difference ($P < .01$) favoring clinicians in the QAI condition. In addition, the intervention appeared to reduce variability in implementing quality indicators for QAI clinicians as well as raising the average amount of implementation. The fact that these gains were accomplished in spite of the formative nature of the study and its implementation across multiple sites is notable, and underscores the potential for increasing high quality and evidence-based practices in school mental health, given the achievable supports and coaching strategies as used here (see Evans & Weist, 2004; Fixsen et al., 2005; Schaeffer et al., 2005).

However, we did not confirm Primary Aim 2, with no statistically significant differences across conditions in knowledge or attitudes about evidence-based practices, or in counseling self-efficacy. The failure to find an effect for knowledge of EBP is inconsistent with interview findings and may reflect problems (including reports that it was confusing, poor internal consistency) with the particular measure used (the KEBSQ). Other measures demonstrated moderate effect sizes favoring the QAI condition (e.g., the PEC Usage scores, producing a median effect size of $d = .4$), but the number of clinicians completing the measures was too small to generate sufficient statistical power for these effects to achieve conventional levels of significance. The failure to find differences in attitudes and self-efficacy may reflect the nature of the work involved. Although clinicians in the QAI condition were exposed to significant training on EBP, they were trained in modular skills for four different disorder areas (depression, anxiety, disruptive behavior disorders, and ADHD), with training in up to five skill areas for each of these disorders in weekly sessions. In addition, these sessions were very dense with training also occurring on family engagement/empowerment, and on pursuit of quality indicators. Written feedback by clinicians in the QAI condition repeatedly documented their perception that training sessions covered too much material in too little time, and that it was difficult to apply skills learned in the busy environment of their schools. Relatedly, the skill training demands on these clinicians may have served to compromise their sense of counseling self-efficacy when compared with staff in the WPI condition, who were presented with essentially no demands to translate learning from training workshops to their actual work in the schools.

In addition, our failure to find differences between conditions in this second primary aim, and for the secondary aim pursuing more distal impacts was likely related to three factors: (1) the lack of on-site implementation

support (see Fixsen et al., 2005) for clinicians in the QAI condition, (2) significant differences in implementation of the QAI intervention across the three sites, (3) the fact that the study represented a formative evaluation, with the QAI intervention under development in Year 1, with a revised intervention and somewhat revised measurement procedures in Year 2. Please note that the formative nature of this study was necessary since there were no examples in the literature of studies that have sought to combine systematic attention to services quality, family engagement and support, and evidence-based practice in SMH. Thus, the study sought to both further develop the framework and pursue feasibility dimensions, while provisionally testing impacts. It could be argued that the outcome assessment through a randomized controlled design was premature, but our goal in integrating these processes was to accelerate the pace of this avenue of research (which was endorsed by NIH grant reviewers). Nonetheless, the mixing of feasibility and impact assessment results in a lack of precision that may have contributed to our failure to document child level outcomes. However, the study provides clear guidance about next steps in this research avenue.

Fixsen et al. (2005) and the work of the National Implementation Research Network (NIRN) are documenting that *implementation support*, moving beyond traditional limited models of supervision, to provide on and off site *coaching*, involving interactive and lively teaching, modeling, behavioral rehearsal and feedback, peer-to-peer support, and administrative support, is crucial to the acquisition and maintenance of trained evidence-based practices. Although weekly meetings with senior trainers did involve modeling, rehearsal and feedback of modular skills, as mentioned, this training agenda was likely too broad, and provided inadequate time for all staff to rehearse and receive feedback on skills (since some groups included as many as 10 people). In addition, we did not build into the research intervention *on-site* implementation support for clinicians at their schools. The combination of the high density and comprehensiveness of training, and the lack of implementation support no doubt compromised clinicians' application of learned skills in their schools, confirmed by qualitative feedback provided by them about the study (as above).

Second, there was significant variability in implementation of the QAI intervention across the three sites. Sessions with senior clinicians were audiotaped and reviewed by the research team, with written reviews and recommendations for improvement sent to them. In addition, in Year 2, weekly phone calls with the senior clinicians were added to promote comparability in approach across the three sites. In spite of these efforts, there appeared to be continuing differences across the sites. For example, one of the sites included two senior clinicians, who were found on

a few occasions to interject their own material into training sessions (e.g., on family systems), and one of these clinicians had relatively poor compliance with staff showing up for weekly meetings. In another site, there were periods of inconsistent holding of weekly meetings by the senior clinician (which was corrected after feedback by the research team), and the program there maintained a different philosophy about family involvement (with a focus on high school students, this was less of a priority). In the third site, one senior clinician went out on maternity leave, resulting in a supervisory burden on the other.

Third, as mentioned, the study involved a formative evaluation, involving improvement and refinement of the materials in Year 1 for application in Year 2. This method variance may have contributed error variance, and in addition, added some complexity and some confusion for clinicians in the second year. For example, related to feedback from Year 1, the School Mental Health Quality Assessment Questionnaire (SMHQAQ, Weist et al., 2005) was improved, eliminating redundant items (resulting in a reduction from 45 to 40 items) and improving Likert ratings (from a four-point to six-point scale). Although the measure was improved, staff did report some difficulty transitioning from the first to the second measure. Also in Year 2, based on feedback from Year 1, we reduced the focus on quality indicators from around 15 targeted in a year to around 7–9 targeted in a year, while increasing modular skills training. This change in focus may have been confusing, and overall, training demands were likely too high.

Another concern that may impact student level outcomes relates to the recruitment process. Students were only eligible for recruitment after participating in five sessions. Many students receiving services do not meet the five session requirement. Thus, the sample may not be representative of all students served by a clinician and may underrepresent students with particular problems (e.g., externalizing problems, poor attendance, frequent suspensions). It would be helpful in future studies to evaluate the differences, if any, between students seen five or more times and students seen less than five times. An additional concern, which likely constrained the study's ability to attain differential impacts on outcome variables of interest was compensatory rivalry (Cook & Campbell, 1983; Cook, Campbell, & Peracchio 1990). For example, in one of the sites, clinicians from both conditions participated in their program's weekly staff meetings together at the program's headquarters. At this site, clinicians participated in the WPI meeting immediately before the regular staff meeting. Although supervisory assignments were separated so that supervisors and staff were matched to the WPI or QAI condition (as in all three sites), it was not possible to completely separate the two groups administratively. On one occasion, one of the coinvestigators observed in a staff

meeting a participant in the QAI condition handing another QAI participant a resource for EBP. A participant in the WPI condition asked what it was and if she could see it, and the QAI participant answered “You can’t have that; you’re in Wellness.” Following this incident, there were other reports of dissatisfaction by staff in the WPI condition of not receiving resources of those in QAI and also desiring weekly meetings, which were then held ad-lib, led by some senior clinicians from WPI, without support of the research team. These kinds of dynamics are difficult to anticipate in implementing studies with clinicians who are actively engaged in SMH services, and point to the need for more sophistication in handling administrative issues. Please note that end of study feedback by staff in the WPI condition indicated that they were for the most part pleased to participate and learned much relevant material that they were applying in their lives.

Although there were limitations (as mentioned earlier) to the study being a formative evaluation, there were many benefits. A user-friendly approach to high quality and evidence-based school mental health services has been developed, and there is now a wealth of resources available broadly. All developed resources including the SMHQAQ, a 140 page guide including information on all 40 indicators, and a comprehensive resources regarding modular skill training and on family engagement empowerment are available and downloadable from the website, www.schoolmentalhealth.org. Resources in all three areas are available in a variety of user-friendly formats including briefs, PowerPoint presentations, newsletters, tools, decision making guides, handouts, and resource and web resource lists and links.

Based on lessons learned from the study, the QAI intervention has been refined and improved to include a more reasonable focus on quality indicators, emphasis on modular intervention for a single disorder category at a time, and enhancement of family engagement/empowerment materials and training to reflect a cross cutting theme to be integrated with other elements. In addition, the study has documented the critical need for implementation support, with a submitted R01 proposal to follow-up on this study in one large SMH program (to avoid the error variance associated with multisite intervention) with the three component intervention enhanced a fourth component of Implementation Support. Further, the materials developed for the WPI condition received strong ratings for their value, and address a critical need for the field related to the stress on SMH staff and the relative lack of resources focusing on self-care and wellness. These materials can be used to include in future randomized controlled designs, as a face valid and acceptable comparison intervention.

There is now a national charge to “improve and expand school mental health programs” (Recommendation 4.2,

PNFC, 2003). Related to this charge, the Substance Abuse and Mental Health Services Administration (SAMHSA) recently released the report of the first national survey of SMH in the US for the 2002–2003 school year (Foster, Rollefson, Doksum, Noonan, & Robinson 2005). The report indicated that over 80% of US schools provided assessment for mental health problems, consultation for behavioral issues, crisis intervention, and referrals for more intensive mental health services. Approximately two-thirds of schools reported providing individual counseling, case management, and group counseling. However, district and state leaders expressed that mental health needs of students were increasing, yet funding was inadequate to meet these needs and predicted to worsen rather than improve. In this context, there is a real need for research that demonstrates achievable strategies for high quality, evidence-based practice in school mental health findings and lessons learned from this study help to build this critically important research avenue and provide directions for next steps, including focusing interventions on a manageable level of learned skills for clinicians, being cautious about multisite implementation, and significantly increasing implementation support, including in-school support for learned evidence-based skills.

Acknowledgments We extend our sincere appreciation to the following people who made critical contributions to the successful conduct of the study—*Senior Clinicians*: Patricia Cotton, Harold Burnett, and Gerald Grubbs; *Site Managers*: Ellie Davis, Rachel Grier, Phyllis Hazel, Jenni Jennings, and Glenn Pearson; *Data Consultant*: Mark Anderson; *Site Coordinators*: Skylar Woods-Nunley, and Nereyda Pederson, *Research Associates*: Linda Kinney, and Matthew Page; *Advisory Panel*: Alan Bellack, Maureen Black, Nick Ialongo, Karen Friedman, Philip Leaf, Richard Sarles, Susan Tager, Stuart Tiegel, Beth Warner; *Consultants*: Bruno Anthony, Bruce Chorpita, and David Pruitt.

References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61–74.
- Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 5*(3), 233–237.
- Barnard, W. M. (2004). Parent involvement in elementary school and educational attainment. *Child and Youth Services Review, 26*, 39–62.
- Bickham, N., Pizarro, J., Warner, B., Rosenthal, B., & Weist, M. (1998). Family involvement in expanded school mental health. *Journal of School Health, 68*(10), 425–428.
- Chorpita, B. (2006). *Modular cognitive-behavioral therapy for childhood anxiety disorders*. New York: Guilford Press.
- Chorpita, B. F., Becker, K. D., & Daleiden, E. L. (2007). Understanding the common elements of evidence based practice:

- Misconceptions and clinical examples. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 647–652.
- Chorpita, B. F., & Daleiden, E. L. (2007). *2007 Biennial report: Effective psychosocial intervention for youth with behavioral and emotional needs*. Hawaii: Child and Mental Health Division, Hawaii Department of Health.
- Chorpita, B. F., Taylor, A. A., Francis, S. E., Moiffitt, C. E., & Austin, A. A. (2004). Efficacy of modular cognitive behavior therapy for childhood anxiety disorders. *Behavior Therapy*, 35, 263–287.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstances for POMP. *Multivariate Behavioral Research*, 34, 315–346.
- Cook, T., & Campbell, D. (1983). The design and conduct of quasi-experiments and true experiments in field settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223–326). Chicago: Rand McNally.
- Cook, T., Campbell, D., & Peracchio, L. (1990). Quasi experimentation. In M. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 491–576). Palo Alto, CA: Consulting Psychologists Press.
- Curry, J. F., & Reinecke, M. A. (2003). Modular cognitive behavior therapy for adolescents with major depression. In M. Reinecke, F. Datillo, & A. Freeman (Eds.), *Cognitive therapy with children and adolescents: A casebook for clinical practice* (pp. 95–127). New York: Guilford Press.
- Evans, S. W., & Weist, M. D. (2004). Implementing empirically supported treatments in the schools: What are we asking? *Clinical Child and Family Psychology Review*, 7, 263–267.
- Evans, S. W., Weist, M. D., & Serpell, Z. (2007). *Advances in school-based mental health interventions: Best practices and program models* (Vol. II). New York: Civic Research Institute.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature (FMHI publication #231)*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Foster, S., Rollefson, M., Doksum, T., Noonan, D., & Robinson, G. (2005). *School mental health services in the United States, 2002–2003*. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3), 377–393.
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 17(3), 125–130.
- Graczyk, P. A., Domitrovich, C. E., & Zins, J. E. (2003). Facilitating the implementation of evidence-based prevention and mental health promotion efforts in schools. In M. Weist, S. Evans, & N. Lever (Eds.), *Handbook of school mental health: Advancing practice and research* (pp. 301–318). New York, NY: Kluwer Academic.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265–275.
- Hazler, R., & Mellin, E. (2004). The developmental origins and treatment needs of female adolescents with depression. *Journal of Counseling and Development*, 82(1), 18–24.
- Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., et al. (2004). Parent academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development*, 75(5), 1491–1509.
- Hoagwood, K. E. (2005). Family-based services in children's mental health: A research review and synthesis. *Journal of Child Psychology and Psychiatry*, 46(7), 690–713.
- Hoagwood, K., Olin, S. S., Kerker, B., Kratochwill, T. R., Crowe, M., & Saka, N. (2007). Empirically-based school interventions targeted at academic and mental health functioning. *Journal of Emotional and Behavioral Disorder*, 15(2), 66–92.
- Hunter, S. C., Boyle, J. M., & Warden, D. (2004). Help seeking amongst child and adolescent victims of peer-aggression and bullying: The influence of school-stage, gender, victimisation, appraisal, and emotion. *British Journal of Educational Psychology*, 74(3), 375–390.
- Jensen, P. S., & Hoagwood, K. E. (Eds.). (2008). *Improving children's mental health through parent empowerment: A guide to assisting families*. New York: Oxford University Press.
- Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban Education*, 40, 237–269.
- Kutash, K., Duchnowski, A. J., & Lynn, N. (2006). *School-based mental health: An empirical guide for decision makers*. Tampa, FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Department of Child & Family Studies, Research and Training Center for Children's Mental Health.
- Lowie, J. A., Lever, N. A., Ambrose, M. G., Tager, S. B., & Hill, S. (2003). Partnering with families in expanded school mental health programs. In M. Weist, S. Evans, & N. Lever (Eds.), *Handbook of school mental health programs: Advancing practice and research*. New York: Kluwer Academic.
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36(4), 611–637.
- Marcon, R. A. (1999). Positive relationships between parent school involvement and public school inner-city preschoolers' development and academic performance. *School Psychology Review*, 28(3), 395–412.
- McKay, M. (2004, March). Engagement of families in evidence-based trauma treatment: Preliminary findings. In K. Hoagwood (Eds.), *Organizational efficacy: Impact of culture, climate, and family engagement on clinical care*. Symposium conducted at the 17th annual Research Conference. A system of care for children's mental health: Expanding the research base, Tampa, FL.
- Paternite, C. E., Weist, M. D., Axelrod, J., Anderson-Butcher, D., & Weston, K. (2006). School mental health workforce issues. In M. Hoge, J. Morris, A. Daniels, N. Adams, L. Huey, & G. Stuart (Eds.), *A thousand voices: The national action plan on behavioral health workforce development* (pp. 199–210). Rockville, MD: U.S. Substance Abuse and Mental Health Services Administration.
- Power, T. J. (2003). Promoting children's mental health: Reform through interdisciplinary and community partnerships. *School Psychology Review*, 32(1), 3–16.
- Power, T. J., Manz, P. H., & Leff, S. S. (2003). Training for effective practice in the schools. In M. Weist, S. Evans, & N. Lever (Eds.), *Handbook of school mental health: Advancing practice and research* (pp. 257–273). New York: Kluwer Academic.
- President's New Freedom Commission on Mental Health. (2003). *Achieving the Promise: Transforming Mental Health Care in America. Final Report for the President's New Freedom Commission on Mental Health (SMA Publication No. 03-3832)*. Rockville, MD: President's New Freedom Commission on Mental Health.
- Rappaport, N., Osher, D., Garrison, E. G., Anderson-Ketchmark, C., & Dwyer, K. (2003). Enhancing collaboration within and across disciplines to advance mental health programs in schools. In M. Weist, S. Evans, & N. Lever (Eds.), *Handbook of school mental*

- health programs: *Advancing practice and research* (pp. 107–118). New York: Kluwer Academic.
- Rickwood, D. J., & Braithwaite, V. A. (1994). Social-psychological factors affecting help-seeking for emotional problems. *Social Science and Medicine*, 39, 563–572.
- Roberts, R. E., Atkisson, C. C., & Mendia, R. M. (1984). Assessing the Client Satisfaction Questionnaire in English and Spanish. *Hispanic Journal of Behavioral Sciences*, 6(4), 385–396.
- Robinson, K. E. (2004). *Advances in school-based mental health interventions: Best practices and program models*. New York: Civic Research Institute.
- Schaeffer, C. M., Bruns, E., Weist, M., Stephan, S. H., Goldstien, J., & Simpson, Y. (2005). Overcoming challenges to using evidence-based interventions in schools. *Journal of Youth and Adolescence*, 34(1), 15–22.
- Shernoff, E. S., Kratochwill, T. R., & Stoiber, K. C. (2003). Training in evidence-based interventions (EBI's): What are school psychology programs teaching? *Journal of School Psychology*, 41, 467–483.
- Stephan, S. H., Davis, E. T., Callan Burke, P., & Weist, M. D. (2006). Supervision in school mental health. In T. Neill (Ed.), *Helping others help children: Clinical supervision of child psychotherapy*. Washington, DC: American Psychological Association.
- Stumpf, R. E., Higa-McMillan, C. K., & Chorpita, B. F. (2009). Implementation of evidence based services for youth: Assessing provider knowledge. *Behavior Modification*, 33(1), 48–65.
- Sutton, J. M., & Fall, M. (1995). The relationship of school climate factors to counselor self-efficacy. *Journal of Counseling & Development*, 73(3), 331–336.
- Vanderbleek, L. M. (2004). Engaging families in school-based mental health treatment. *Journal of Mental Health Counseling*, 26, 211–224.
- Weist, M. D. (1997). Expanded school mental health services: A national movement on progress. In T. Ollendick & R. A. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 19, pp. 319–352). New York, NY: Plenum Press.
- Weist, M. D., & Albus, K. E. (2004). Expanded school mental health: Exploring program details and developing the research base. *Behavior Modification*, 28, 463–471.
- Weist, M. D., Ambrose, M., & Lewis, C. (2006a). Expanded school mental health: A collaborative community/school example. *Children & Schools*, 28(1), 45–50.
- Weist, M. D., Evans, S. W., & Lever, N. A. (Eds.). (2003). *Handbook of school mental health: Advancing practice and research*. New York: Kluwer Academic.
- Weist, M. D., Lever, N. A., Stephan, S. H., Anthony, L. G., Moore, E. A., & Harrison, B. R. (2006, February). *School mental health quality assessment and improvement: Preliminary findings from an experimental study*. Paper presentation at the annual conference. A system of care for children's mental health: Expanding the research base, Tampa, FL.
- Weist, M. D., Sander, M. A., Walrath, C., Link, B., Nabors, L., Adelsheim, S., et al. (2005). Developing principles for best practice in expanded school mental health. *Journal of Youth and Adolescence*, 34(1), 7–13.