Published in final edited form as:

Adm Policy Ment Health. 2017 January ; 44(1): 16–28. doi:10.1007/s10488-015-0650-0.

Community-Sourced Intervention Programs: Review of Submissions in Response to a Statewide Call for “Promising Practices”

Aaron R. Lyon, Ph.D., Michael D. Pullmann, Ph.D., Sarah Cusworth Walker, Ph.D., and Gabrielle D’Angelo, M.S.W.

University of Washington

Abstract

This study was initiated to add to the nascent literature on locally-grown intervention programs in the youth mental health, child welfare, and juvenile justice service sectors, many of which demonstrate practice-based or community-defined evidence, but may not have been subjected to empirical evaluation. Characteristics of applications submitted in response to three public calls for additions to an inventory of research-supported intervention programs were reviewed on evidence for effectiveness, the use of key quality assurance (QA) elements (e.g., clearly specified training or integrity monitoring procedures), and cultural specificity. Findings indicate that four QA processes were identified in approximately half of all submissions: a specific initial training process, the existence of intervention integrity measures, routine outcome monitoring, and ongoing support post-training. An initial training process and integrity measurement were more commonly described among programs determined to have greater research evidence for their effectiveness. Overall, cultural elements were described relatively infrequently and most often reflected surface-level program delivery characteristics (e.g., offering services in languages other than English). Discussion is focused on the alignment of submitted programs with the larger literatures focused on implementation science and cultural competence.

Keywords

Evidence-based practice; policy; implementation science; quality assurance; culture

Introduction

As increasing attention is paid to quality and accountability in the public sector, critical gaps have been identified between typical and optimal practice across a variety of disciplines, including mental health (e.g., McHugh & Barlow, 2010), child welfare (Aarons & Palinkas, 2007), and juvenile justice (e.g., Lipsey, Howell, Kelly, Chapman, & Carver, 2010). Such “research-to-practice gaps” are often highlighted because of their potential to limit the effectiveness of public sector services for populations most at risk for negative outcomes (e.g., low-income and ethnic/cultural minority youth and adults). Interest in these gaps is

further fueled by the realization that the “usual care” interventions delivered in community settings are frequently ineffective (Garland et al., 2013; Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010; Weisz, Jensen-Doss, & Hawley, 2006; Weisz et al., 2013), despite the high costs of maintaining these programs.

As a consequence, an entire field of implementation science has emerged with the purpose of bringing research and practice into better alignment, often via training, consultation, and support strategies to facilitate the uptake and sustained use of innovative programs in real world service settings (Eccles & Mittman, 2006; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Multiple solutions have been described to address quality gaps, ranging from top-down mandates surrounding the use of evidence-based practices (EBP) (e.g., Jensen-Doss, Hawley, Lopez, & Osterberg, 2009) to bottom-up grassroots intervention development and evaluation in the context of specific communities (e.g., Marsiglia & Kulis, 2009). Although much of the focus on EBP implementation has been on top-down approaches, the incorporation of bottom-up strategies holds great promise for ensuring local relevance and buy-in. Nevertheless, a variety of knowledge gaps inhibit the viability of bottom-up approaches. Little information is available about the characteristics of locally-developed programs, such as the extent to which they make use of quality assurance (QA) mechanisms (e.g., integrity assessments) known to be essential for achieving positive clinical outcomes and facilitating larger-scale implementation. Furthermore, even though bottom-up program development holds promise as a strategy for improving the cultural relevance of interventions (Martinez, Callejas, & Hernandez, 2010), no research has examined the extent to which “community-sourced” interventions demonstrate culturally-specific program content. To help fill these knowledge gaps, this paper presents an analysis of “promising practice” applications submitted for inclusion on the Washington State Inventory of Research and Evidence-Based Practices with the goal of evaluating the extent to which submissions are aligned with the broader literature on program effectiveness, implementation, and cultural relevance. Many submitted programs were locally-developed and provide a unique perspective on how direct service providers and administrators are currently addressing these elements.

“Top-Down” Implementation Approaches

Policy initiatives that support the uptake and sustained use of high-quality prevention and intervention programs represent an increasingly common strategy to support widespread EBP implementation and scale-up. This orientation is consistent with mounting evidence that variables operating at larger systemic levels influence the success of implementation initiatives (Aaron et al., 2012; Beidas, et al., in press; Beidas & Kendall, 2010). Most often, this type of implementation research tends to describe policies enacted by states and municipalities to support EBP scale-up across service sectors (e.g., Beidas et al., 2013; Bruns & Hoagwood, 2008; Hoagwood et al., in press; Isett et al., 2007; Rhoades, Bumbarger, & Moore, 2012). Nevertheless, while “top-down” policy initiatives have utility in effecting significant system change, they have also been criticized as attending inadequately to the local context and insufficiently promoting collaboration with front-line service providers or other community stakeholders (Jensen-Doss et al., 2009). It is for many of these reasons that top-down policies have been identified as “necessary but not sufficient”

implementation drivers (Beidas et al., 2013; p.1). Alignment among strategies used at different levels is likely to be important if efficient and effective implementation is to be realized (Aarons, Hurlburt, & Horwitz, 2011). A complementary approach in which collaborative, bottom-up strategies to quality improvement are situated within larger policy changes may be optimal.

Culturally-Grounded and Practice-Based Evidence to Support “Bottom-Up” Quality Improvement

In response to perceptions that many EBP implementation efforts reflect “top-down” knowledge transfer processes that fail to incorporate the needs and backgrounds of service settings and providers, increasing attention has focused on methods of facilitating community engagement and bidirectional knowledge exchange. Indeed, to be successful, large-scale quality improvement efforts should incorporate clear *evaluation of existing services* to identify potentially effective practices already in use and drive implementation decisions (Garland, Bickman, & Chorpita, 2010). This approach is more likely to actively engage community stakeholders than a full *replacement paradigm* in which practitioners are directed to abandon their existing practices in favor of new interventions. It may also increase the probability of implementation success due to stakeholder buy-in achieved through validation and respect. Furthermore, the approach is practical, in that the identification of existing practices with demonstrated utility may reduce the need for the introduction of some new interventions, thereby reducing implementation costs. Below, we review related, overlapping approaches to bottom-up quality improvement, including locally grown interventions and practice based / community defined evidence.

Locally grown approaches to intervention development – which involve local practitioners specifically designing interventions for use with a particular community or population – represent one method for gathering and codifying knowledge about what works in specific practice contexts. For agencies serving significant ethnic and cultural minority populations, this might include documentation of “culturally-grounded” intervention approaches, which place culturally-specific values, beliefs, practices, and socio-historical perspectives at the center of treatment design (e.g., Marsiglia & Kulis, 2009; Resnicow, Baranowski, Ahluwalia, & Braithwaite, 1999). Although culturally-grounded interventions can be considered “local technologies” that have developed in response to specific, unmet community needs, there are currently few mechanisms through which they can be efficiently identified, evaluated, and (if found to be effective) incorporated into a larger set of evidence-based services (Holleran Steiker et al., 2008; Lyon, Lau, McCauley, Vander Stoep, & Chorpita, 2014).

Related to locally-grown and culturally-grounded approaches, practice-based and community-defined evidence represent methods of determining whether existing programs (potentially developed using culturally-grounded methods) demonstrate local validity. In these approaches, local validity may emerge either from the collection of empirical data or from an implicit or explicit endorsement of an intervention approach from stakeholders. Practice-based evidence (PBE) has therefore been defined multiple ways, sometimes focusing on the *information or data* used to evaluate a program or practice (e.g., Jensen et

al., 2012) and other times emphasizing the perceived local relevance and acceptability of *treatment innovations* produced in particular contexts. The PBE definition offered by Isaacs, Huang, Herdandez and Echo-Hawk (2005), for instance, focuses primarily on the extent to which treatment innovations are congruent with a given culture: “a range of treatment approaches and supports that are derived from, and supportive of, the positive cultural attributes of the local society and traditions” (p.16). In this conceptualization, the pursuit of PBE involves a codification of the services being delivered in a community, provided that they are “accepted as effective by the local community.” Community-defined evidence is a refinement of the PBE concept, which Martinez (2008) described as “a set of practices that communities have used and determined by community consensus over time and which may or may not have been measured empirically.” Embedded in this definition is the assumption that individuals within communities have the knowledge and perspective necessary to determine the value of practices, but that such knowledge is rarely incorporated into traditional avenues of intervention development or evaluation (Martinez et al., 2010).

Identification of interventions with established practice-based or community-defined evidence has the potential to support essential bidirectional knowledge flow in a manner that first accepts that there may be expertise and experience present in practice settings that is not reflected in the research literature and then seeks to augment that knowledge with more traditional empirical methods. One step in this process is often the identification of existing practices that may ultimately demonstrate utility in a controlled research trial, but few researchers have undertaken this task. In one notable example, the Pennsylvania Commission on Crime and Delinquency worked with the National Center for Juvenile Justice to solicit “best practices” that would be most likely to stand up to an empirical test (Bumbarger & Campbell, 2011). The authors explained that such an approach recognizes the local expertise of stakeholders and “bring[s] science to the field while also facilitating the organic emergence of this ‘wisdom guided by knowledge’” (Bumbarger & Campbell, 2011; p.274).

At their core, the approaches described above all involve careful community “sourcing” of intervention content to augment the existing scientific literature. Furthermore, if the ultimate goal of the identification of practice-based and community-defined evidence/programs is their dissemination and implementation, it becomes important to gather information about program characteristics that are likely to predict whether they can be effectively taken to scale or can provide appreciable benefits to service recipients. Beyond simply demonstrating evidence for program effectiveness, such efforts should evaluate how new programs provide key implementation support activities and whether they do so in a manner aligned with the burgeoning implementation science literature. Feasible and effective processes for QA – including initial training activities, intervention tracking, and ongoing supports for providers – are among the most critical, as they provide opportunities to create and maintain standards for effective program delivery (Sanders & Kirby, 2014). Indeed, intervention integrity (a.k.a., “fidelity”) monitoring and other forms of ongoing QA are frequently identified among program characteristics essential for widespread implementation and scale up (Southam-Gerow & McLeod, 2013). Nevertheless, the “real world” feasibility of QA processes originally developed in the context of well-resourced clinical trials has been repeatedly questioned (Kendall & Beidas, 2007; Schoenwald, Hoagwood, Atkins, Evans, &

Ringeisen, 2010) and no research has examined the extent to which community-derived programs incorporate commonly accepted QA elements. Below, we describe the results of a project in which information about QA procedures and other key aspects of community practices was collected to augment a policy-driven, statewide EBP implementation effort.

Policy Background

Washington State House Bill 2536 was passed by the 2012 Legislature with the intent that “prevention and intervention services delivered to children and juveniles in the areas of mental health, child welfare, and juvenile justice be primarily evidence-based and research-based” and that “such services will be provided in a manner that is culturally competent.” Specifically, the bill required the Washington State Department of Social and Health Services (DSHS) child-serving divisions (mental health, child welfare and juvenile justice) to produce a baseline report indicating the percent of total expenditures being spent on research and evidence-based practices. The bill further directed the divisions to report, annually, to the legislature their progress in increasing this allocation (see Kerns & Trupin, this issue). This process included the development of an inventory of structured interventions currently in use in the State (henceforth referred to as the “Washington State Inventory”) as well as existing evidence for their effectiveness (Walker, Lyon, Aos, & Trupin, under review). Washington State Inventory programs were identified as evidence-based (i.e., [a] tested in heterogeneous or intended populations with multiple randomized and/or statistically-controlled evaluations, or one large multiple-site randomized and/or statistically-controlled evaluation, where the weight of the evidence from a systematic review demonstrates sustained improvements; [b] can be implemented with a set of procedures to allow successful replication in Washington; and, [c] when possible, has been determined to be cost-beneficial), research-based (i.e., tested with a single randomized and/or statistically-controlled evaluation demonstrating sustained desirable outcomes; or where the weight of the evidence from a systematic review supports sustained outcomes as identified in the term “evidence-based,” but does not meet the full criteria for “evidence-based.”), and promising (based on statistical analyses or a well-established theory of change, shows potential for meeting the “evidence-based” or “research-based” criteria”) (see Walker et al., under review). Promising practices were included in particular to facilitate the inclusion of programs which may have been locally grown and/or which may demonstrate practice-based or community-defined evidence, but which would not be identifiable via a search of the empirical services literature. Additionally, driven by the realization that the initial Inventory (a) would likely represent an incomplete account of interventions currently in use and (b) may not fully address the diverse populations and problems represented among the state’s residents, it was determined that an open call for additional programs should be issued. This call had two primary goals: (a) identify programs that may have already met the Inventory standards for the evidence- or research-based categories, but had been overlooked; and (b) identify additional locally-grown “promising practices” that demonstrate PBE but had not been subject to rigorous evaluation, a subset of which might be eligible for technical assistance surrounding evidence accumulation and eventual dissemination and implementation. From its initiation, the Washington State initiative also included an explicit focus on engaging a diverse and representative group of stakeholders (D’Angelo et al., this issue). Related, the solicitation of promising practices was conducted with explicit

awareness of the limited research available related to the appropriateness and efficacy of many intervention programs for culturally diverse youth (Bernal & Scharrón-del-Río, 2001; Huey & Polo, 2008) and the goal of identifying new programs to improve accessibility and coverage for historically underserved and culturally-diverse residents of the state.

Study Aims

The current study was initiated to examine the characteristics of applications submitted in response to the three public calls for Inventory additions in Washington State. The study aims are to add to the limited research literature on the array of PBE and “locally-grown” approaches to youth service delivery, and to examine these approaches through the lenses of implementation and cultural specificity. Specifically, alignment of the information contained in the applications with the initiative’s emphases on outcome evaluation, methods of evaluating implementation quality, and relevance to culturally diverse groups was evaluated. In doing so, the following research questions (RQ) were assessed: (RQ1) *To what extent do applications reflect intervention programs with demonstrated evidence for effectiveness?* (RQ2) *How often, and in what ways, do applications describe procedures for integrity measurement and quality assurance?* and (RQ3) *How often, and in what ways, do submissions describe culturally-specific program elements?* Programs were submitted in the areas of youth mental health, juvenile justice, and child welfare, and evaluated by a team of researchers at the University of Washington (UW) and the Washington State Institute for Public Policy (WSIPP).

Method

Procedures

Following initial Washington State EBP Inventory generation (Walker et al., under review), researchers developed a Promising Practice Application and review process to gather information about additional programs currently in use in the State. Three separate public calls were issued by DSHS in November of 2012, April of 2013, and November of 2013. Each call was publicized before it was opened and remained open for a period of approximately two months. Although the review process was focused on Washington State, programs could be submitted by individuals or organizations outside of the state as well. Program submission occurred through a web-based survey, the details of which are described below.

Survey submissions underwent a multi-component review process. The first component involved an official review to determine the appropriateness of the program for inclusion in the Washington State Inventory and, if it was determined to be appropriate, its categorization as evidence-based, research-based, or promising. This component was comprised of multiple stages including expert review surrounding the program’s level of evidence for effectiveness and rigorous review of any completed evaluations. A more detailed account of the Inventory review procedures can be found in Walker et al. (under review). The current study also added a second review component focused on identifying the quality assurance procedures and culturally-specific elements referenced in survey responses (see Analysis).

Data Collected

Data were qualitative and derived from the promising practice survey submissions. Among other information, respondents were asked for descriptions of their programs that included target population and outcomes, specific culturally-based aspects, an explicit theory of change, whether the program had a written manual or set of procedures, any evaluations of the program (and the results of those evaluations), and quality assurance procedures (e.g., provision of practice/program specific supervision or consultation, methods of determining competence to deliver the program, and integrity/fidelity measures). Due to the relative infrequency with which relevant content was reported and the fact that cultural content was not required for inclusion in the Washington State Inventory, the question about culturally-specific program elements was only presented to the first round of applicants.

Analysis

Program review for Inventory inclusion involved a two-stage process in which UW expert reviewers assessed program features (e.g., evidence for effectiveness; a well-established theory of change) and provided a recommendation of whether the program could be categorized as a promising, research-based, or evidence-based practice using the definitions provided earlier. Specifically, promising practices required: (a) a pre/post test demonstrating improvement on direct outcomes of interest to DSHS; (b) a pre/post test demonstrating improvement on variables known in the literature to be linked to outcomes of interest; or (c) a well-specified theory of change that is empirically-based. Furthermore, programs that include elements that appeared to be iatrogenic based on known literature were excluded.

In addition to formal Inventory review, promising practice survey responses were also coded using directed content analysis to identify the presence of quality assurance and culturally-specific elements (Hsieh & Shannon, 2005). Directed content analysis makes explicit use of existing theoretical or empirical frameworks to identify initial coding categories, which may then be revised during the coding process (Potter & Levine-Donnerstein, 1999). Coding began with a three UW faculty members with expertise in EBP implementation and cultural competence in juvenile justice, child welfare, and youth mental health reviewing a randomly selected subset of survey responses. Reviewers used a preliminary set of codes and then met to discuss the codes and revise the codebook. The codebook was then trialed independently through multiple iterations in which all team members coded additional survey responses and met for discussion. During this process, new codes relevant to the research questions were added and others removed or consolidated.

Final codes for quality assurance procedures were developed for different components of the training and implementation processes described: (1) initial training integrity, (2) intervention delivery integrity, and (3) ongoing support integrity. Within each category, 3–5 subcodes were identified related to specific aspects of the integrity process. *Initial training integrity* included items assessing whether the following were described: (a) a specific training process for program providers, (b) a method of ensuring the quality of the trainers themselves, (c) direct assessment of the quality of training processes. *Intervention integrity* was comprised of items capturing whether the application mentioned (a) a method of ensuring the quality of the individuals hired/selected to be providers (e.g., via hiring

practices, “certification,” etc.), (b) the existence of a measure of program integrity/process (typically adherence, but could include competence), (c) whether the integrity measure was completed by practitioners, (d) whether the integrity measure was completed by an independent, direct observer, and (e) routine outcome monitoring. *Ongoing support integrity* included items assessing whether the application mentioned the following consultation, coaching, or supervision elements: (a) any variety of practice/program-specific ongoing support (e.g., supervision, consultation, boosters), (b) a method of ensuring the quality of the consultants/supervisors themselves (e.g., via hiring practices, “certification,” etc.), (c) direct assessment of post-training support procedures (i.e., a measure of ongoing support integrity), and (d) if support integrity ratings were done by an independent coder/observer. Some of these items (e.g., direct assessment of post-training support integrity) are not necessarily common practice among programs with substantial evidence for efficacy, and were intentionally designed to be a “high bar” for applications.

Initial codes for cultural aspects of programs were also developed a priori, based on an existing framework for culturally-specific intervention adaptation (Cardemil, 2010), and then revised in the context of the coding process. To be coded a “culturally-specific” component of a program, an application needed to express clear intentionality surrounding specific responsiveness to a group of service recipients. “Group” was defined based on gender, race/ethnicity, nationality, socioeconomic status, sexual orientation, religion, age, or geography. Major categories for cultural codes included: (1) *structural elements* (i.e., how the intervention is organized so as to provide the desired therapeutic effect), (2) specific *program content* (i.e., curriculum or intervention designed to make culturally relevant material central), (3) culturally-specific aspects of *program delivery* (e.g., services or materials are provided in languages other than English; the program is delivered in a setting intended to increase accessibility), (4) *provider behavior* that was not specific to the program (i.e., providers were trained to be generally “culturally competent”), and (5) *cultural match* between service providers and service recipients. Cultural codes were applied whether applications indicated that a component was initially included for a culturally-relevant reason during program development or if they were added in the context of adapting an existing program for a new population. We coded survey responses for culturally-specific aspects for the first round of submissions (81%), as they were the only round that were specifically asked about these issues.

In addition, a “degree of information scale” was developed to assess the extent to which an application generally referenced culturally-specific components or cultural responsiveness. Using this scale, applications were rated “0” (no information provided or only referenced diverse service recipients), “1” (general statement[s] made about the program’s cultural flexibility, relevance, or competence without specifics), and “2” (specific information was provided about the ways the program is culturally responsive). Only applications that received a “2” were reviewed using the culture codes listed above. For both QA procedures and culturally-specific aspects, coders rated each code on a separate 0 to 2 scale that reflected both the quality of information presented as well as the overall quality of the integrity process or culturally-specific component described (0 = absent; 1 = mentioned briefly/low to moderate quality; 2 = described in detail/moderate to high quality). The final codebook is available from the authors upon request.

Following the establishment of a stable codebook and adequate initial inter-rater reliability (absolute consistency $ICC_{(2,2)} \geq .80$), all three coders coded or recoded each survey response for quality assurance and culturally-specific content using the categories described above. Although accompanying materials (e.g., web pages, manuals) were reviewed as part of the official determination of program status (i.e., evidence-based, research-based, and promising), they were not included in the content analysis of applicants' responses to survey questions because they were supplied inconsistently, and only for a minority of the programs.

Codes for each program were compared across coders for each item and final codes were arrived at by consensus. The total intra-class correlation ($ICC_{(2,2)}$; absolute consistency two-way random models with average measures) for QA and cultural coding across all three coders was 0.86, indicating excellent agreement (Cicchetti, 1994). Separate ICCs for QA and culture codes were 0.87 and 0.82, respectively. In the instances when coders all disagreed (i.e., when they scored 0, 1, and 2, respectively), items were assigned the average code of 1. This occurred for 51 of the 1782 codes assigned (2.9%) and was most common for items assessing aspects of intervention integrity (program integrity measure, $n = 8$; integrity measures completed by practitioners, $n = 6$; integrity measures completed by independent observer, $n = 6$; routine outcome monitoring, $n = 6$).

Results

Eighty-two submissions were received in response to the three open calls for promising practices (66, 11, and 5 from the first, second, and third calls, respectively). One submission was excluded because it was a proposal for a program that the applicant was interested in developing and not an existing program or service. Of the 81 remaining submissions, 51 (63%) were submitted by community-based organizations, 11 (14%) by governmental organizations, 8 (10%) by university-based researchers, 5 (6%) by independent service providers, and 5 (6%) by private companies, and 1 was unable to be determined. Seventy-four (91%) of the applications originated in Washington State. Of the 81 submissions, there were six sets of duplicate program submissions. Because all duplicate submissions were submitted by different individuals or agencies, they were retained in the analyses.

Although 68 programs (84%) stated in their application that they had a manual or written set of procedures for program delivery, primary reviewers only obtained a digital or paper copy of the manual from the submitting organization or individual for 18 programs (22%). Fifty-five (68%) indicated that a formal evaluation had been conducted for their program, 27 of which (49%) had been released in a peer-reviewed outlet. Although some programs were relevant to multiple service sectors, the largest number of applications were most relevant to mental health and substance abuse ($n = 34$; 42%), followed by child welfare ($n = 29$; 36%), and juvenile justice ($n = 7$; 9%). The remaining programs focused on education, general prevention, or system change that was not explicitly tied to one service sector ($n = 11$; 14%).

Based on initial application review, 12 programs were categorized as "promising." Twenty-nine submissions were identified for additional review of evidence for effectiveness because formal program evaluations had been conducted. Through this subsequent review process, an

additional 13 programs were identified as promising (25 total; 31%), 5 programs (6%) were identified as “research-based,” and 3 (4%) as “evidence-based” using the definitions described in Walker et al. (under review). The remaining 48 programs (59%) were excluded from the Inventory for a variety of reasons, most commonly an emphasis on outcomes that were not of interest to DSHS (e.g., general wellness) or a lack of well-defined program parameters (e.g., application was for an agency, but not a specific intervention).

Quality Assurance/Integrity

Within the domain of QA, submissions were most likely to mention (i.e., were assigned a “1” or “2” for) having a specific training process (54.3% of all submissions), a measure of intervention integrity (50.6%), routine outcome monitoring (46.9%), or some form of ongoing support (58%). Of these QA aspects, 21%, 24.7%, 25.9%, and 32.1% of submissions were determined to provide a high-quality description and were assigned a “2.” Least commonly mentioned QA aspects included a method of ensuring the quality of trainers (13.6%; 7.4% high quality), direct assessment of *training* integrity (6.2%; 2.5% high quality), direct assessment of *ongoing support* integrity (2.5%; 1.2% high quality), or support integrity ratings that were completed by an independent observer (1.2%; 1.2% high quality). Table 1 displays full results for QA procedure coding, separated by determination status (i.e., whether programs were ultimately categorized as evidence-based, research-based, promising, or excluded). Due to their relatively low frequencies, evidence-based and research-based programs were combined for all analyses.

Because it was hypothesized that evidence-/research-based applications would be more likely to mention all QA components than promising or excluded programs, one-tailed chi-square tests were conducted to determine whether there were cross-group differences in the likelihood that applications referenced different QA elements. Results for dichotomous codes for any mention of specific elements (i.e., 0 versus 1 or 2) indicated significant differences by determination status for description of a specific training process ($\chi^2 = 5.426$, $df = 2$, $p = .033$, $\eta^2 = 0.259$), whether a measure of program integrity exists ($\chi^2 = 8.229$, $df = 2$, $p = .008$, $\eta^2 = 0.319$), and assessment of ongoing support integrity ($\chi^2 = 4.801$, $df = 2$, $p = .046$, $\eta^2 = 0.243$). For each of these variables, promising and research/evidence-based programs referenced relevant information more often than excluded programs. A similar analysis examining differences by determination status in high-quality descriptions of QA elements revealed significant findings for whether a measure of program integrity exists ($\chi^2 = 7.960$, $df = 2$, $p = .001$, $\eta^2 = .313$), and the presence of ongoing support procedures for providers ($\chi^2 = 16.701$, $df = 2$, $p < .001$, $\eta^2 = 0.454$). For program integrity measurement, programs in the research/evidence-based category were more likely to include high-quality information than programs in the promising or excluded categories. For ongoing support procedures, promising and research/evidence-based programs were both more likely to include high-quality information than excluded programs.

Cultural Aspects

Of the 65 applications that were explicitly prompted for culturally-specific content during the first open program call, 16 (24.6%) made no mention of cultural elements, 23 (35.4%) made brief mention (i.e., stating that the program was “flexible” or able to be individualized

to service recipients of any background), and 26 (40%) provided sufficient detail to allow for further cultural coding. A two-tailed chi-square test was performed to determine if the overall amount of information provided about cultural elements (i.e., no mention, brief mention, in-depth discussion) varied by final program status (i.e., combined evidence/research-based, promising, excluded), but found no significant differences ($\chi^2 = 3.96$; $df = 4$; n.s.). Chi-square tests similar to those conducted for QA codes revealed no differences by submission final determination status for any cultural items.

Table 2 provides comprehensive information about the 26 applications that supplied sufficient information to allow for in-depth cultural coding. As displayed, applications were most likely to mention culturally-specific aspects of program delivery ($n = 13$; 50%) and service provider-service recipient match ($n = 8$; 30.8%). Within the program delivery category, applications were most likely to reference providing services in languages other than English ($n = 7$; 26.9%) and least likely to mention service delivery in settings that increased service accessibility or acceptability ($n = 1$; 3.8%). With the exception of the overall program delivery code, relatively few applications referenced cultural aspects with high-quality information (ranging from 3.8% to 15.4%).

Discussion

The current study was designed to evaluate the extent to which submissions to a statewide call for “promising practices” in youth mental health, child welfare, and juvenile justice aligned with the broader literature on program effectiveness, implementation, and cultural relevance. The majority of submissions came from community-based service providing organizations, likely reflecting the relatively high number of those agencies as well as their vested interest in having their services prioritized by DSHS under HB 2536. Of the 81 submissions reviewed, only 8 (10%) were ultimately placed into the evidence-based or research-based categories, indicating that most submissions came from outside of a traditional peer-reviewed, empirically-based pathway for program recognition. This is encouraging, because identifying such programs was a central goal of the promising practice initiative. In contrast, evidence-based and research-based submissions often reflected known interventions with empirical support which had simply not been included in the initial version of the Washington State Inventory (e.g., Homebuilders; Evans, Boothroyd, & Armstrong, 1997) or research-supported adaptations of existing EBPs (e.g., Multisystemic Therapy for Child Abuse and Neglect; Swensen, Schaeffer, Henggeler, Faldowski, & Mayhew, 2010). Newly identified evidence-based and research-based programs spanned all three primary areas of interest (child welfare, youth mental health, juvenile justice). Thirty percent of submissions were ultimately identified as promising practices and added to that category of the Inventory. Similar to evidence- and research-based programs, promising practices spanned all three areas of interest, indicating that the process had relevance across multiple service systems. The full Washington State Inventory, which is updated periodically, can be found at the WSIPP website (<http://www.wsipp.wa.gov/>).

Quality Assurance

For QA, the coding system was devised to test the full range of submission breadth using comprehensive codes that reflected rigorous approaches to QA and provided an opportunity to document some information about the extent and quality of information included (i.e., a “1” versus a “2”). A specific training process, measurement of the integrity of intervention delivery, routine outcome monitoring, and some type of ongoing support were all mentioned (assigned a “1” or “2”) in the majority – or nearly the majority (47% for outcome monitoring) – of submissions. Furthermore, there was clear separation between these elements and the remaining QA codes, with the next most frequently mentioned element included in only 32% of applications (ensuring the quality of providers via hiring or selection practices). In many ways, the four most common QA elements represent a core set of essential quality improvement activities that are reflective of the current state of implementation science. Indeed, substantial bodies of research have focused on the importance of training (Herschell, Kolko, Baumann, & Davis, 2010; Lyon, Stirman, Kerns, & Bruns, 2011), intervention integrity assessment (Southam-Gerow & McLeod, 2013), outcome monitoring (Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Lambert et al., 2003), and ongoing support (Fixsen et al., 2005; Lyon et al., 2011). The relative frequency with which they were identified in the current project is encouraging, as it suggests some important alignment between the QA activities reported by community stakeholders and the broader implementation literature.

Despite this encouragement, findings also indicated significant differences by final determination status in the presence of some of these core QA elements, with evidence- and research-based programs more likely to mention a specific *training process* and *intervention integrity assessment* than promising or excluded programs. However, no differences were found across categories for *outcome monitoring* or for *ongoing support*, suggesting that even applications reflecting programs that had not been subjected to rigorous empirical evaluations (i.e., promising) may acknowledge the value of these activities for the delivery of high-quality services. The greater likelihood of research and evidence-based programs to have direct assessments of training quality may have been due to the need to monitor quality across the multiple trainers and service sites that are typical of randomized trials. In contrast, trainers and program developers were often one and the same for many locally-grown, “promising” programs.

Interestingly, high-quality information descriptions were relatively rare across all QA elements and final determination categories, and this was especially true for excluded programs. Even among evidence-/research-based programs, only *measures of program integrity*, *routine outcome monitoring*, and *descriptions of ongoing support procedures* were detailed with high quality information in half or more of the applications. For promising practices, only *descriptions of ongoing support procedures* included high quality information a majority of the time. It is possible that this finding reflects the application prompts – which were vague as to the specific elements of QA that were being requested – rather than the actual quality with which programs made use of each QA element. This assumption is supported by anecdotal knowledge that some programs with known, high quality QA procedures did not provide responses that could be coded at the highest level. Although

some programs provided detailed accounts of consultation or supervision structures to ensure skill uptake and intervention use (e.g., via mandatory consultation calls for a one-year period with an expert trainer; coded “2”), others simply stated that trainees received “ongoing supervision” surrounding program delivery (coded “1”). Despite this, recent findings from the mental health literature that suggest consultation procedures may be even more important than initial training experiences (Beidas, Edmunds, Marcus, & Kendall, 2012) provide some reason for optimism that identified promising practices may already have some of the necessary structures in place to allow them to be implemented successfully.

Finally, many of the QA procedures coded in the current study reflected “gold standard” approaches to program integrity assessment, some of which may not be feasible in low-resource community service contexts (Kendall & Beidas, 2007; Shoenwald et al., 2011). Indeed, some of the lower-frequency elements – such as intervention integrity ratings completed by an independent observer – may be more of a vestige of the field’s emphasis on randomized controlled trials than a realistic approach to QA in community agencies. Additional QA elements that were identified infrequently are generally uncommon within the empirical literature. Specifically, direct measurements of the integrity of initial training processes and/or post-training supports (regardless of whether they are completed by an independent observer) are rarely reported even in high-quality implementation studies. Nevertheless, the importance of attending to *implementation* integrity (e.g., integrity of training or consultation processes) in addition to traditional *intervention* integrity will likely receive increased attention as awareness of the importance of implementation processes grows (Proctor, Powell, & McMillen, 2013). As implementation integrity procedures continue to be developed, it will be important to design them for real world feasibility from the outset if they are ever to have relevance to community practitioners and organizations.

Cultural Aspects

Considering that the empirical intervention literature has been repeatedly criticized for inadequate attention to ethnically and culturally diverse populations (e.g., Bernal & Scharrón-del-Río, 2001; Hwang, 2009; Miranda, Nakamura, & Bernal, 2003; Nicolas, Arntz, Hirsch, & Schmiedigen, 2009), the current project was committed to identifying programs that could fit the needs of specific cultural groups. Although “cultural group” was operationalized broadly to include gender, race/ethnicity, nationality, socioeconomic status, sexual orientation, religion, age, or geography, nearly all of the culturally specific aspects mentioned in the applications were in reference to youth or family race/ethnicity. Forty percent of the eligible submissions (i.e., those asked explicitly about cultural content) made sufficient mention of cultural elements to allow for further coding. Even among this group, however, cultural aspects were generally identified infrequently, and even less frequently with high-quality information. Interestingly, there were no significant differences by final determination status, regardless of whether applications were examined for any mention of a cultural aspect or for the provision of high-quality information. This suggests that applications for programs that were more closely aligned with a traditional empirical epistemology and its evaluation processes were generally no more or less likely to describe attending explicitly to cultural elements than other programs.

Among applications referencing cultural aspects, mention of *program delivery* components was most frequent. Within the program delivery category, the most common cultural component was the provision of services in languages other than English, a component that was mentioned more often than the provision of program-specific materials in other languages. When multi-language materials were referenced, however, submissions were invariably judged to have provided high quality information. Both of these components are typically included among the “surface-level” efforts to achieve cultural relevance that are frequently discussed in the literature on psychotherapy adaptation (Lau, 2006). Cardemill (2010) has defined surface (or superficial) modifications as “those that consist of small changes to the intervention so as to match the delivery of the intervention to observable characteristics of the target population, but that leave the vast majority of the original intervention intact” (p.12). In contrast, “deep modifications” take into consideration central “cultural aspects of the relevant ethnic group” (Cardemill, 2010, p.12). Although the current project was focused on applications from programs that were *originally developed* for a specific cultural group as well as those that were *culturally adapted*, this distinction remains relevant. In our study, “surface-level” characteristics were coded under the *program delivery* domain whereas “deep” characteristics were included under *program content*. As indicated above, program delivery components were identified considerably more frequently than program content aspects. Some researchers have suggested that “deep” program content elements are more likely to resonate with diverse service recipients and lead to more effective interventions. For instance, in a meta-analysis of culturally-adapted psychotherapy interventions, Benish, Quintana, & Wampold (2011) found that only one component of culturally-specific program content – adoption of the illness myth (i.e., framing psychological suffering within culturally consonant, adaptive interpretations) – predicted more favorable outcomes. In the current study, these program content (i.e., “deep”) elements appeared less frequently than program delivery (i.e., “surface”) elements.

Service provider-service recipient matching was the second most frequently-referenced cultural category. For decades, this has been a topic of discussion in the literatures focused on the delivery of psychosocial services to minority individuals, especially in the psychotherapy literature (Maramba & Nagayama Hall, 2002). Within that domain, some research has suggested the underlying *cognitive match* between service provider and recipient may be more important than surface-level *cultural or ethnic match* (Zane et al., 2005). Interestingly, cultural match coding also demonstrated the largest discrepancy between the number of applications mentioning service provider-service recipient cultural match (i.e., stating that there was some attempt to pair along a culturally-relevant dimension) and the number that provided high-quality information (i.e., explicitly stated how providers were selected and along what dimensions they were matched). This suggests that, although many applicants may have considered cultural match to be a component of culturally relevant service, far fewer chose to elaborate extensively on this component as a central method of achieving cultural specificity. It is unknown, however, the extent to which this may have been a function of the level of attention or priority given to this element in the submitted programs or the method by which this was evaluated in the current project, which did not ask explicitly about this (or any other) cultural element.

Implications

The statewide call for promising practices yielded a sizeable number of applications reflecting a diverse range of applicants and program types. Although the number of applications was highest for the first call (81% of submissions were received at that time) and decreased each cycle, our efforts to promote stakeholder engagement and support the implementation of House Bill 2536 are ongoing (Kerns & Trupin, this issue). The promising practices initiative has transitioned from specified submission windows to a standing call and rolling review of programs. These reviews will continue to include assessment of evidence for program effectiveness and placement of programs into evidence-based, research-based, and promising categories, but will not attend explicitly to QA or cultural elements.

As it continues, the promising practices review process will need to reconcile practice-based and community-defined evidence with more traditional empirical evidence on an ongoing basis. Unlike the CDE approach, which places highest importance on communities' evaluations of practices, the current initiative sought to identify a subset of programs to potentially move into the broader sphere of interventions available for wider dissemination and implementation. In this way, the current project was not only concerned with supporting the use of effective programs within the settings that they originated, but also identifying programs that may have generalizable utility in similar communities statewide. The presence of many core QA elements within identified promising practices is expected to increase the likelihood that this can be done effectively. Nevertheless, it may be also worthwhile to pay additional attention to supporting programs that do not yet make use of the four most common QA elements (i.e., a specific training process, intervention integrity measurement, routine outcome monitoring, ongoing support) to develop or adopt those procedures.

Limitations

This paper has the standard limitations related to self-report data in that it represents a review of *proposals*, not necessarily actual program activities or elements. The proposals may have exaggerated or failed to report certain elements. The review of proposals may also have inadvertently left out programs, organizations, or providers who chose not to respond to such calls for information. Qualitative coding for QA and cultural aspects was based only on the content provided in submitted applications and did not include reviews of external documentation (e.g., webpages, manuals or publications). This was done to standardize the review process across applications (the majority of which did not provide additional materials) and increase the feasibility of review. Nonetheless, for these reasons, the results of QA and cultural assessments should be interpreted with caution as they are likely an underrepresentation of the actual presence of these elements. Furthermore, because the amount of information provided by each application and the quality of the practices/procedures described were interdependent, it was necessary to develop a coding structure that accounted for both aspects of responses simultaneously (i.e., a "2" on a QA element typically reflected a more detailed explanation as well as description of a more rigorous approach).

Summary and Conclusion

In sum, results from the current project suggest that it was possible to assess the alignment of promising practice submissions with broader standards for empirical evidence for effectiveness, QA, and cultural relevance. Although a small number of submitted applications reflected programs that met the criteria for evidence-based or research-based categories, substantially more promising programs were identified. Across all submissions, four core QA processes were identified as most common: a specific training process, intervention integrity measurement, routine outcome monitoring, and ongoing support. Some of these (training process descriptions; integrity measurement) were more common among programs determined to have greater research evidence for their effectiveness. Description of cultural aspects did not vary by submission final determination status. Overall, cultural elements were described relatively infrequently and most often reflected surface-level program delivery characteristics. The information collected for this project provides an important snapshot of components of usual care practice in Washington State. Although far from comprehensive, documentation of submission characteristics is intended to facilitate better integration of evidence-based practice and practice-based evidence with the goal of improving the quality and relevance of services statewide.

Acknowledgments

Dr. Lyon is an investigator with the Implementation Research Institute (IRI), at the George Warren Brown School of Social Work, Washington University in St. Louis; through an award from the National Institute of Mental Health (R25 MH080916) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative (QUERI).

This publication was made possible in part by funding from grant number K08 MH095939, awarded to the first author from the National Institute of Mental Health (NIMH).

References

- Aarons GA, Glisson C, Green PD, Hoagwood K, Kelleher KJ, Landsverk JA. The Research Network on Youth Mental Health. The organizational social context of mental health services and clinician attitudes toward evidence-based practice: a United States national study. *Implementation Science*. 2012; 7(56)
- Aarons G, Hurlburt M, Horwitz S. Advancing a conceptual model of evidence-based practice implementation in child welfare. *Administration and Policy in Mental Health*. 2011; 38(1):4–23. [PubMed: 21197565]
- Aarons GA, Palinkas L. Implementation of evidence-based practice in child welfare: Service provider perspectives. *Administration and Policy in Mental Health*. 2007; 34(4):411–419. [PubMed: 17410420]
- Beidas RS, Aarons G, Barg F, Evans A, Hadley T, Hoagwood K, Mandell DS. Policy to implementation: evidence-based practice in community mental health—study protocol. *Implementation Science*. 2013; 8(38)
- Beidas RS, Edmunds JM, Ditty M, Watkins J, Walsh L, ...Kendall P. Are inner context factors related to implementation outcomes in cognitive-behavioral therapy for youth anxiety? *Administration and Policy in Mental Health*. Advance online publication. in press
- Beidas RS, Edmunds JM, Marcus SC, Kendall PC. Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*. 2012; 63(7):660–665. [PubMed: 22549401]

- Beidas RS, Kendall PC. Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*. 2010; 17(1):1–30. [PubMed: 20877441]
- Benish SG, Quintana S, Wampold BE. Culturally adapted psychotherapy and the legitimacy of myth: A direct comparison meta-analysis. *Journal of Counseling Psychology*. 2011; 58(3):279–289. [PubMed: 21604860]
- Bernal G, Scharrón-del-Río MR. Are empirically supported treatments valid for ethnic minorities? Toward an alternative approach for treatment research. *Cultural Diversity and Ethnic Minority Psychology*. 2001; 7(4):328–342. [PubMed: 11759270]
- Bickman, L.; Kelley, SD.; Breda, C.; de Andrade, AR.; Riemer, M. *Psychiatry Services*. Vol. 62. Washington D.C.: 2011. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial; p. 1423-1429.
- Bruns EJ, Hoagwood EK. State implementation of evidence-based practice for youths, Pt. I: Responses to the state of the evidence. *Journal of Child and Adolescent Psychiatry*. 2008; 47(4):369–373.
- Bumbarger BK, Campbell EM. A state agency-university partnership for translational research and the dissemination of evidence-based prevention and intervention. *Administration and Policy in Mental Health*. 2011; 39(4):268–277. [PubMed: 21901440]
- Cardemil EV. Cultural adaptations to empirically supported treatments: a research agenda. *Scientific Review of Mental Health Practice*. 2010; 7(1):8–21.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6(4):284–290.
- Eccles MP, Mittman BS. Welcome to implementation science. *Implementation Science*. 2006; 1(1)
- Evans ME, Boothroyd RA, Armstrong MI. Development and implementation of an experimental study of the effectiveness of intensive in-home crisis services for children and their families. *Journal of Emotional and Behavioral Disorders*. 1997; 5:93–105.
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. *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; 2005.
- Garland AF, Bickman L, Chorpita BF. Change what? Indentifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health*. 2010; 37(1–2): 15–26. [PubMed: 20177769]
- Garland AF, Haine-Schlagel R, Brookman-Frazee L, Baker-Ericzen M, Trask E, Fawley-King K. Improving community-based mental health care for children: translating knowledge into action. *Administration and Policy in Mental Health*. 2013; 40(1):6–22. [PubMed: 23212902]
- Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: a review and critique with recommendations. *Clinical Psychology Review*. 2010; 30(4):448–466. [PubMed: 20304542]
- Hoagwood KE, Olin SS, Horwitz S, McKay M, Cleek A, Gleacher A, Hogan M. Scaling up evidence-based practices for children and families in New York state: toward evidence-based policies on implementation for state mental health systems. *Journal of Clinical Child and Adolescent Psychology*. in press
- Holleran Steiker LK, Castro FG, Kumpfer K, Marsiglia FF, Coard S, Hopson LM. A dialogue regarding cultural adaption of interventions. *Journal of Social Work Practice in the Addictions*. 2008; 8(1):154–162.
- Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qualitative Health Research*. 2005; 15(9):1277–1288. [PubMed: 16204405]
- Huey SJ, Polo AJ. Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child & Adolescent Psychology*. 2008; 37(1):262–301. [PubMed: 18444061]
- Hwang WC. The Formative Method for Adapting Psychotherapy (FMAP): A community-based developmental approach to culturally adapted therapy. *Professional Psychology: Research and Practice*. 2009; 40(4):369–377.
- Isaacs, LN.; Huang, M.; Hernandez, H.; Echo-Hawk. *The road to evidence: The intersection of evidence-based practices and cultural competence in children's mental health*. Washington D.C: National Alliance of Multi-ethnic Behavioral Health Associations; 2005.

- Isett, KR.; Burnam, MA.; Coleman-Beattie, B.; Hyde, PS.; Morrissey, JP.; Magnabosco, J.; ... Goldman, HH. Psychiatry Services. Vol. 58. Washington D.C.: 2007. The state policy context of implementation issues for evidence-based practices in mental health; p. 914-921.
- Jensen DR, Abbott MK, Beecher ME, Griner D, Golightly TR, Cannon JAN. Professional Psychology: Research and Practice. 2012; 43(4):388–394.
- Jensen-Doss A, Hawley KH, Lopez M, Osterberg LD. Using evidence-based treatments: The experiences of youth providers working under a mandate. Professional Psychology: Research and Practice. 2009; 40(4):417–424.
- Kendall P, Beidas R. Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. Professional Psychology and Research Practice. 2007; 38(1):13–20.
- Lambert MJ, Whipple JL, Hawkins EJ, Vermeersch DA, Nielson SL, Smart DW. Is it time for clinicians to routinely track patient outcome? A meta-analysis. Clinical Psychology: Science and Practice. 2003; 10(3):288–301.
- Lau AS. Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. Clinical Psychology: Science and Practice. 2006; 13(4):295–310.
- Lipsey, MW.; Howell, JC.; Kelly, MR.; Chapman, G.; Carver, D. Improving the effectiveness of juvenile justice programs: A new perspective on evidence-based practice. Washington D.C: Center for Juvenile Justice Reform; 2010.
- Lyon AR, Lau AS, McCauley E, Vander Stoep A, Chorpita BF. A case for modular design: implications for implementing evidence-based interventions with culturally diverse youth. Professional Psychology: Research and Practice. 2014; 45(1):57–66.
- Lyon AR, Stirman SW, Kerns SE, Bruns EJ. Developing the mental health workforce: review and application of training approaches from multiple disciplines. Administration and Policy in Mental Health. 2011; 38(4):238–253. [PubMed: 21190075]
- Maramba G, Nagayama Hall GC. Meta-analyses of ethnic match as a predictor of dropout, utilization, and level of functioning. Cultural Diversity And Ethnic Minority Psychology. 2002; 8(3):290–297. [PubMed: 12143105]
- Marsiglia, FF.; Kulis, S. Diversity, oppression and change: Culturally grounded social work. Chicago, IL: Lyccum; 2009.
- Martinez K. Culturally defined evidence: what is it? And what can it do for Latinos/as? El Boletín: Newsletter of the National Latino/a Psychological Association, Fall/Winter. 2008
- Martinez K, Callejas L, Hernandez M. Community-defined evidence: A ottom-up behavioral health approach to measure what works in communities of color. Report on Emotional and Behavioral Disorders in Youth. 2010; 10(1):11–16.
- Miranda J, Nakamura R, Bernal G. Including ethnic minorities in mental health intervention research: A practical approach to a long-standing problem. Culture, Medicine, and Psychiatry. 2003; 27(4): 467–486.
- McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. The American Psychologist. 2010; 65(2):73–84. [PubMed: 20141263]
- Nicolas G, Arntz DL, Hirsch B, Schmiedigen A. Cultural adaptation of a group treatment for Haitian American adolescents. Professional Psychology: Research and Practice. 2009; 40(4):378–384.
- Potter JW, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. Journal of Applied Communication Research. 1999; 27(3):258–284.
- Proctor EK, Powell BJ, McMillen JC. Implementation strategies: Recommendations for specifying and reporting. Implementation Science. 2013; 8:139. [PubMed: 24289295]
- Renisow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. Ethnicity and Disease. 1999; 9(1):10–21. [PubMed: 10355471]
- Rhoades BL, Bumbarger BK, Moore JE. The role of state-level prevention support system in promoting high-quality implementation and sustainability of evidence-based programs. American Journal of Community Psychology. 2012; 50(3–4):386–401. [PubMed: 22441729]
- Sanders MR, Kirby JN. Surviving or thriving: quality assurance mechanisms to promote innovation in the development of evidence-based parenting interventions. Implementation Science. 2014:1–11. [PubMed: 24398253]

- Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. 2011; 38(1):32–43. [PubMed: 20957425]
- Schoenwald SK, Hoagwood KE, Atkins MS, Evans ME, Ringeisen H. Workforce development and the organization of work: the science we need. *Administration and Policy in Mental Health*. 2010; 37(1–2):71–80. [PubMed: 20145990]
- Southam-Gerow MA, McLeod BD. Advances in applying treatment integrity research for dissemination and implementation science: introduction to special issue. *Clinical Psychology: Science and Practice*. 2013; 20(1):1–13. [PubMed: 23970819]
- Swenson CC, Schaeffer CM, Henggeler SW, Faldowski R, Mayhew AM. Multisystemic Therapy for Child Abuse and Neglect: a randomized effectiveness trial. *Journal of Family Psychology*. 2010; 24(4):497–507. [PubMed: 20731496]
- Walker SC, Lyon AR, Aos S, Trupin E. (under review). The consistencies and vagaries of the Washington State Inventory of Evidence-Based, Research-Based and Promising Practices. The definition of “evidence-based” in a policy context.
- Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*. 2010; 78(2):144–155. [PubMed: 20350026]
- Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: a meta-analysis of direct comparisons. *The American Psychologist*. 2006; 61(7):671–689. [PubMed: 17032068]
- Weisz JR, Kuppens S, Eckshtain D, Ugueto AM, Hawley KM, Jensen-Doss A. Performance of evidence-based youth psychotherapies compared with usual clinical care: a multilevel meta-analysis. *JAMA Psychiatry*. 2013; 70(7):750–761. [PubMed: 23754332]
- Zane N, Sue S, Chang J, Huang L, Huang J, Lowe S, Lee E. Beyond ethnic match: Effects of client-therapist cognitive match in problem perception, coping orientation, and therapy goals on treatment outcomes. *Journal of Community Psychology*. 2005; 33(5):569–585.

Table 1
Percentage of Applications Referencing Quality Assurance Processes by Application Determination Status

	Integrity Processes	Evidence- or Research-Based (n = 8)		Promising (n = 25)		Excluded (n = 48)	
		%mentioning	%providing high-quality information	%mentioning	%providing high-quality information	%mentioning	%providing high-quality information
Initial training	A specific training process is described	75.0	25.0	68.0	20.0	43.8	20.8
	Ensures quality of the trainers themselves	25.0	0.0	16.0	8.0	10.4	8.3
	Training processes are directly assessed*	12.5	0.0	4.0	0.0	6.3	4.2
Intervention Delivery	Ensures quality of the providers (via hiring / selection practices)	25.0	25.0	48.0	24.0	25.0	20.8
	Measure of program integrity exists*†	75.0	62.5	68.0	28.0	37.5	16.7
	Integrity ratings are completed by the practitioners	25.0	0.0	36.0	24.0	16.7	10.4
	Integrity ratings are completed by an independent observer	37.5	25.0	44.0	20.0	20.8	8.3
Ongoing Support	Routine outcome monitoring is standard practice	50.0	50.0	40.0	20.0	50.0	25.0
	Program-specific ongoing support procedures described	75.0	62.5	68.0	56.0	50.0	14.6
	Ensures quality of the consultants/supervisors (via hiring practices or "certification")	37.5	12.5	28.0	16.0	18.8	10.4
Ongoing Support	Post-training support procedures are directly assessed (ongoing support integrity)*†	12.5	0.0	4.0	4.0	0.0	0.0
	Support integrity ratings are completed by an independent observer	0.0	0.0	4.0	0.0	0.0	0.0

* $p < .05$ testing proportion mentioning the element

information on the element
proportion providing high quality
testing $d_i < 0.5$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Percentage of Applications Providing Adequate Information for Cultural Coding (n = 26) and Referencing Specific Cultural Elements

Culturally-Specific Elements	% mentioning	% providing high-quality information
Structural Elements (e.g., modality, dosage, sequencing, or culturally-specific assessments)	23.1	15.4
Program Content (e.g., curriculum or intervention designed to make culturally relevant material central)	23.1	7.7
Program Delivery	50.0	30.8
A. Program delivered in a way explicitly congruent with the target community/culture (e.g., provides food, explicitly highlights collaborative nature of intervention)	15.4	3.8
B. Services provided in languages other than English	26.9	15.4
C. Materials are provided in languages other than English	15.4	15.4
D. Program is delivered in a <u>setting</u> intended to increase accessibility or be more culturally congruent	3.8	3.8
Providers trained to be generally “culturally competent”	15.4	7.7
Service provider-service recipient cultural match	30.8	7.7